Construcción de una base de datos de imágenes de mamografía para la identificación de microcalcificaciones

Claudia Joana Santamaría Moya

Director: Álvaro Ángel Orozco Gutiérrez

Universidad Tecnológica de Pereira
Facultad de Ingenierías Eléctrica, Electrónica, Física y Ciencias de la Computación
Maestría en Ingeniería Eléctrica
Pereira-Risaralda
2014

Índice general

Agradecimientos

Quiero agradecer a mi director Ph.D Álvaro Ángel Orozco Gutiérrez, por su gran apoyo en la realización de este proyecto.

A MSc Hernán Felipe García Arias, por guiarme con su experiencia y por compartir todo su conocimiento, sin duda él fue una persona muy importante en el desarrollo de este trabajo.

A mis padres y hermanos, quienes siempre me han brindado un hogar lleno de amor y me han acompañado a lo largo de este proceso.

A Andrés, por su amor, paciencia y comprensión en los momentos más difíciles de esta etapa y por estar siempre presente en los mejores momentos.

A la Fundación Alejandro Londoño por su colaboración y acompañamiento en el proceso de la construcción de la base de datos.

Resumen

La mamografía es el tipo de imagen más utilizado en la deteccción de cáncer de mama, sin embargo, se caracteriza por tener un bajo contraste y un alto contenido de información no deseada. La detección de microcalcificaciones (mcals) en una mamografía es una tarea díficil aún para un especialista experimentado. Actualmente, la comunidad académica no cuenta con un banco de imágenes mamográficas que cuantifiquen la población local (Latinoamérica, Colombia) y además que permitan la validación del diagnóstico médico confirmado con biopsia, mediante algoritmos de procesamiento de imágenes. En este documento se presenta una metodología para la creación de una base de datos de registros mamográficos de una población local para la detección de mcals. El proceso se divide en dos etapas principales: en la primera etapa, se realiza la construcción de la base de datos, se analizan las fases de almacenamiento, validación médica y etiquetado de las imágenes. En la segunda etapa se realiza el procesamiento de las mamografías para la validación del reporte médico con biopsia confirmada. El proceso inicia con una etapa de preprocesamiento para eliminar artefactos derivados de la adquisición de las imágenes, seguido de un análisis de textura mediante técnicas de análisis de textura fractal (SFTA) y patrones locales binarios (LBP). Finalmente se realiza una clasificación basado en máquinas de aprendizaje para la identificación de los hallazgos reportados por el personal médico. Los resultados evidencian que la base de datos construida, cumple con los parámetros epidemiológicos para representar una población local, y la metodología de identificación de micro-calcificaciones basada en descriptores de textura evidencia porcentajes de acierto del $93{,}2\,\%$ lo cual permite la correcta validación de los hallazgos patológicos de la base de datos recopilada.

Abstract

Breast cancer is a disease of great global impact and its the most common type of cancer affecting womens. This disease its the second death cause in the world. The incidence of this pathology has grown in recent years. Moreover, this is reflected in Colombia, which has passed from fifth to second in rank frequency [1] , [2] , [3] , [4].

For early detection of this cancer, the only effective method is mammography. In this exam is possible to detect microcalcifications (mcals), which are the earliest manifestation of breast cancer [3]. However, the diffeerence between malignant and benign lesions represents a very complex problem, even for an experienced radiologist. Because the characteristics of such images (low contrast and sharpness) features make sometimes necessary to apperal to a second diagnostic or invasive tests. Besides this pathology, resulting in increased costs of analysis and episodes of unnecessary stress in patients [4] , [5]. For these reasons, is important to build mamographic databases from patient records in a local area, which allow us, a medical validation of biopsy diagnosis using imaging processing techniques.

In this work we proposed a methodology for mamographic imaging database building, which allow us quantify, the variability of microcalcification in a local poblationa and brings to the reasearch community an open records to the medical image analysis. The construction of the database was conducted in four stages: First authorization usage of medical records was collect. Second, the acquisition and storage of this data was described. Here, the analysis and validation of the information contained in the medical report was accomplish labeling all types or microcalcification.

Finally, we introduce a microcalcification recognition method based on texture descriptors for medical image processing, to automatically detect an abnormality in a breast image given. The results shows that the proposed model can efficiently detect a microcalcification with $93{,}2\,\%$ accuracy. This brings to the medical community an open source to robustly perform a microcalcification recognition.

1. Introducción

El cáncer de mama es una enfermedad de gran impacto mundial pues tiene alta incidencia en mujeres mayores de 50 años, además, es el tipo de cáncer con mayor tasa de mortalidad en mujeres en casi todos los países [2]. La tasa de incidencia de este tipo de patología ha crecido más en los últimos 30 años, esto se refleja en Colombia, donde pasó del quinto al segundo lugar en frecuencia [1]. Para disminuir las tasas de mortalidad de dicha enfermedad se ha enfatizado en la detección temprana mediante el uso de pruebas de tamizaje como la mamografía y el autoexamen.

Actualmente, la mamografía continúa siendo la principal técnica para el diagnóstico de cáncer de mama pues se ha demostrado que como método de tamizaje reduce significativamente la tasa de mortalidad por cáncer de mama en 20 a 35 % [2], en ella es posible detectar microcalcificaciones (mcals) agrupadas, que son la manifestación más temprana de cáncer de seno y se presentan en la imagen como pequeños puntos de alta intensidad [10]. El hallazgo de microcalcificaciones se considera de gran interés, ya que corresponde aproximadamente a la mitad de los tipos de cáncer detectados.

A pesar de esto , la tarea de diagnóstico se dificulta cuando hay que diferenciar entre lesiones malignas y benignas; tal es el caso de la detección de mcals, cuyas características (tamaño, forma, distribución) varían según el tipo de severidad, esto representa un problema aún para un radiólogo experimentado, pues las características propias de las mamografías digitales en cuanto al bajo contraste y nitidez, limitan su sensibilidad, haciendo que en ocasiones se recurra a técnicas invasivas o a un análisis adicional por parte de otro especialista.

Por estas razones, es importante contar con bases de datos de registros mamográficos para la validación de algoritmos. En los últimos años se han utilizado con frecuencia bases de datos internacionales como la MIAS *(The Mammographic Image Analysis Society)* [11], [12] y la DDSM *(The Digital Database for Screening Mammography)* [13], [14], que sirven para la prueba y validación de metodologías de procesamiento de mamografías a nivel mundial. Sin embargo, actualmente no hay muchos proyectos enfocados en la generación de nuevas bases de datos [15], que contengan registros de mamografías adquiridos con sistemas tecnológicos actualizados, además, a nivel local y nacional no existen bases de datos de registros de pacientes con la taxonomía propia de las mujeres colombianas y que adicional a esto, permitan la validación del diagnóstico médico mediante técnicas de procesamiento de imágenes encaminadas a la detección de lesiones que puedan representar malignidad como las mcals.

La metodología para la creación de la base de datos inicia con la solicitud del consentimiento informado de las pacientes, luego, se describe la adquisición y almacenamiento de los registros mamográficos obtenidos en cada estudio, en la siguiente etapa se realizó un análisis y validación de la información contenida en el reporte médico con el acompañamiento de un especialista y la última etapa corresponde al etiquetado de las imágenes acorde con la información de la etapa anterior y la confirmación del resultado de las biopsias en

los casos que lo necesitaron, se construye un archivo en formato digital con la siguiente información: código de identificación del estudio y de cada imagen, clasificación Bi-Rads, tipo de tejido, clase de anormalidad, severidad, coordenadas del centro y radio que encierra la anormalidad detectada.

Se ha demostrado que en el proceso de adquisición y digitalización de las mamografías es introducido un ruido que puede interferir con las características de interés en dichas imágenes [5], por lo tanto, el procesamiento de este tipo de imágenes debe iniciar con una etapa de filtrado para disminuir el ruido que puede entorpecer etapas posteriores. Para ello, se utilizaron técnicas de filtrado holístico con máscaras de media, mediana, gaussiana y unsharp y las métricas Q [6] y ssim *(structural-similarity Index)* [8] para determinar la efectividad de los mismos.

Aprovechando las propiedades del análisis de textura y que el tejido que conforma una lesión en una mamografía suele cambiar con la presencia de cáncer, se realizó la extracción de características mediante la técnica sfta que descompone una imagen de entrada en un conjunto de imágenes binarias, generando un vector con el tamaño, niveles de gris y contorno de las imágenes binarias resultantes, estas medidas conocidas como fractales son empleadas para describir la complejidad entre los límites del objeto y las estructuras segmentadas de la imagen de entrada [9].

Por último, en la metodología propuesta se realiza la clasificación entre imágenes con calcificaciones benignas, malignas y con ausencia de ellas mediante el aprendizaje de una máquina de soporte vectorial (SMV). Este documento está compuesto de la siguiente forma. La justificación del trabajo se presenta en el capítulo 2. La descripción y formulación del problema se presentan en el Capítulo 3. Los objetivos del desarrollo de este trabajo se presentan en 4. En el Capítulo 5 se encuentra la revisión de la literatura de las técnicas de construcción de bases de datos y la detección de mcals. En el Capítulo 6 se presenta el marco conceptual de la temática a analizar. Los capítulos 7 y 8, describen los materiales utilizados y la metodología a seguir para el desarrollo del trabajo. Finalmente, los capítulos 9 y 10, describen los resultados obtenidos, su análisis y la discusión pertinente; además de las conclusiones y la descripción de posibles trabajos futuros derivados del trabajo desarrollado.

2. Justificación

2.1. Pertinencia

El cáncer de mama es una enfermedad de gran impacto mundial pues tiene alta incidencia en mujeres mayores de 50 años, además, es el tipo de cáncer con mayor tasa de mortalidad en mujeres en casi todos los países [2]. La tasa de incidencia de este tipo de patología ha crecido más en los últimos 30 años, esto se refleja en Colombia, donde pasó del quinto al segundo lugar en frecuencia [1]. Para disminuir las tasas de mortalidad de dicha enfermedad se ha enfatizado en la detección temprana mediante el uso de pruebas de tamizaje como la mamografía y el autoexamen.

Actualmente, la mamografía continúa siendo la principal técnica para el diagnóstico de cáncer de mama pues se ha demostrado que como método de tamizaje reduce significativamente la tasa de mortalidad por cáncer de mama en 20 a 35 % [2]; en ella pueden observarse hallazgos como: microcalcificaciones (mcals), masas, áreas de densidades asimétricas, distorsiones arquitectónicas, entre otros; las mcals que se presentan en la imagen como pequeños puntos de alta intensidad, son lesiones consideradas de gran interés ya que son una manifestación temprana de cáncer [10].

A pesar de esto, la tarea de diagnóstico hace que en ocasiones se dificulte la diferenciación entre lesiones malignas y benignas; tal es el caso de la detección de mcals, cuyas características (tamaño, forma, distribución) varían según el tipo de severidad, esto representa un problema muy complejo aún para un radiólogo experimentado, pues el bajo contraste y nitidez de las mamografías digitales, causan una limitada sensibilidad, haciéndose necesario un doble diagnóstico para alcanzar mejoras o la realización de pruebas invasivas adicionales para determinar con certeza la presencia de cáncer, lo que incrementa los niveles de estrés y ansiedad en las pacientes [5].

El hecho de que una imagen requiera ser revisada por dos radiólogos hace que se incrementen los costos y tiempo de estudio, reduciendo la productividad individual del especialista.

Adicionalmente, y pese a la alta incidencia de cáncer de mama a nivel mundial y del incremento de casos de esta enfermedad en Colombia, en la actualidad no se desarrollan muchos proyectos para la construcción de bases de datos de mamografías [15] que permitan la validación de técnicas de procesamiento de imágenes aplicadas propiamente a la detección de mcals en pacientes de la región, con lo que se obtendría un diagnóstico más certero por parte del especialista, disminuyendo las biopsias innecesarias, los costos de análisis y los episodios de estrés que puedan vivir las pacientes debido a falsos positivos (FP).

Este estudio se lleva a cabo desde un punto de vista interdisciplinario que comprende trabajos en el campo de las ciencias computacionales, ingeniería, ciencias cognitivas, ciencias afectivas, sicología, ciencias del aprendizaje y diseño de equipos, con el fin de comprender y describir de forma clara el protocolo para la

creación de una base de datos regional que permita la validación de técnicas para la detección de microcalcificaciones [16].

2.2. Viabilidad

En los últimos años la tasa de incidencia de pacientes con cáncer de mama en Colombia se ha incrementado, a pesar de esto y de las campañas de lucha contra esta enfermedad no se ha logrado disminuir significativamente el número de muertes [2], [1].

Actualmente se han detectado diversos factores de riesgo que incrementan la incidencia del cáncer de seno, pero no ha sido posible determinar un factor específico que sea causa-efecto en su presentación por lo que la prevención aún no ha sido posible [5].

Por lo tanto, las medidas actuales dirigidas a tratar de disminuir la mortalidad por esta enfermedad se enfocan hacia la prevención secundaria, esto es, la detección temprana de la enfermedad. Se estima que cuando el cáncer de seno se presenta como una masa palpable, hasta un 80 % de los casos será clasificado como en un estadio avanzado del cáncer [5].

En los últimos años, se ha enfatizado mucho en la detección temprana de cáncer mediante el uso de pruebas de tamizaje como la mamografía y el auto examen.

Por las razones expuestas, la creación de bases de datos de mamografías, se convierte en una importante herramienta que permite a los investigadores en este campo, estudiar y desarrollar sistemas de reconocimiento automático de mcals y otros hallazgos que puedan representar malignidad para apoyar el diagnóstico de los especialistas [17], [5], [18], [19].

La mayoría de bases de datos existentes en el estado del arte disponibles para la comunidad académica, incluyen registros de mamografía de pacientes de otros países cuyas características físicas y factores de riesgo de adquirir la enfermedad son diferentes a las mujeres de la población local [15]. Sumado a esto, actualmente, no se desarrollan mucho trabajos de creación de bases de datos y a nivel regional no se cuenta con un banco de datos de este tipo de imágenes. La creación de una base de datos regional, incrementaría la robustez de los sistemas de detección de mcals pues permitiría la validación de algoritmos con imágenes de pacientes locales y de otros países.

Dada la importancia de esta enfermedad, diferentes grupos de investigación alrededor del mundo han enfocado grandes esfuerzos en el diseño de sistemas de apoyo al diagnóstico médico, al igual que se han generado espacios en los últimos años para la publicación de trabajos específicamente en este campo, mediante congresos y revistas especializadas dada su importancia y los grandes alcances que se perciben a partir de su desarrollo.

2.3. Impacto

Con el desarrollo de este proyecto, se busca avanzar en la creación de bases de datos locales y en análisis de los hallazgos presentes en un examen de mamografía, con el fin de darle más robustez al desarrollo de sistemas de apoyo al diagnóstico médico para el reconocimiento automático de mcals o cualquier signo que

pueda representar malignidad en la mama [17], [18], [19], ya que este tipo de estudios son de gran impacto en diversas aplicaciones relacionadas con los campos de desarrollo tecnológico, sicológico, social, entre otros.

Adicional a esto, este proyecto puede tener una favorable relación costo beneficio, evitando un número importante de procedimientos innecesarios, acortando el tiempo de diagnóstico y disminuyendo la ansiedad de las pacientes [5].

Además, el principal aporte metodológico derivado del desarrollo de esta propuesta de investigación, es la creación de la base de datos que servirá para la aplicación de metodologías encaminadas a la detección de mcals en mamografías. Con lo cual se propone una metodología para la creación de la base de datos e implementación de técnicas de preprocesamiento, extracción de características y clasificación de patrones mediante algoritmos de entrenamiento de máquina, que incrementarán la robustez en los sistemas de apoyo al diagnóstico médico [20], [21].

3. Planteamiento del problema

En el presente capítulo se hace una descripción completa del problema de investigación a resolver, desde la medicina, al igual que desde el punto de vista de ingeniería de la forma en la que podría llevarse a cabo la construcción de una base de datos de registros de mamografía de pacientes locales y la validación del diagnóstico médico mediante la aplicación de técnicas de procesamiento de imágenes para la detección de las microcalcificaciones.

3.1. Diagnóstico médico

El cáncer de mama es una enfermedad de gran impacto mundial [2], en Colombia la tasa de incidencia de este tipo de patología ha crecido más en los últimos 30 años, donde pasó del quinto al segundo lugar en frecuencia [1].

Para el diagnóstico de cáncer de mama la mamografía continúa siendo la principal técnica pues se ha demostrado que como método de tamizaje se reduce significativamente la tasa de mortalidad por cáncer de mama en 20 a 35 % [2], pues es posible detectar microcalcificaciones (mcals) agrupadas, que son la manifestación más temprana de cáncer de seno [10] y corresponde aproximadamente a la mitad de los tipos de cáncer detectados.

Sin embargo, el diagnóstico de cáncer de mama es una tarea que se dificulta cuando hay que diferenciar los hallazgos entre lesiones malignas y benignas [10], pues el bajo contraste y nitidez presente en este tipo de imágenes representa un problema aún para un especialista experimentado [5].

Una buena alternativa para ayudar al profesional en la detección de mcals, es crear sistemas de apoyo al diagnóstico de imágenes mamográficas con los cuales se mejore notablemente la sensibilidad y se reduzca la posibilidad de incurrir en un diagnóstico incorrecto, que perjudica directamente al paciente [18], [16], [20].

A pesar de esto, en Colombia y a nivel mundial no hay muchos proyectos de desarrollo de bases de datos que sirvan para la prueba de metodologías computarizadas encaminadas a la detección de mcals y que a su vez validen el diagnóstico médico.

3.2. Bases de datos y detección de microcalcificaciones

Cuando se crean sistemas de apoyo al diagnóstico de cáncer de mama es fundamental contar con diversos bancos de registros mamográficos. Para mejorar la robustez de dichos sistemas a nivel nacional, es importante tener bases de datos que contengan registros mamográficos de pacientes de la región, pues las características físicas y epidemiológicas pueden cambiar de una zona a otra.

Asociado a esto, la detección de mcals que son la manifestación más temprana de cáncer de seno, es muy importante pues corresponde aproximadamente a la mitad de los casos de este tipo de cáncer. La mayoría de los médicos cree que la detección temprana del cáncer de seno salva miles de vidas cada año y que muchas más pudieran salvarse, si las pacientes y sus doctores aprovecharan dichas pruebas.

Para hallar las mcals mediante algoritmos de procesamiento en un mamografía, es necesario analizar características en cuanto a textura, forma y tamaño [17], [18], [22], [11], [14], [13], [23] para su identificación se han desarrollado varias metodologías las cuales pueden ser agrupadas en tres fases: 1) métodos de preprocesamiento, 2) métodos de extracción de características y 3) Sistemas de clasificación.

Para los métodos de preprocesamiento se han utilizado técnicas como filtrado holístico las cuales utilizan métodos de análisis globales, también se han realizado transformaciones basadas en características locales [24], [11], que analizan los niveles de los píxeles, su contraste y el fondo, estas técnicas a pesar de ser sencillas en cuanto a implementación no se adaptan a las propiedades de la imagen, lo que se refleja en la imagen como una pérdida de detalles especialmente en la frontera entre los objetos de interés y el fondo.

Para realizar la segmentación y extracción de características de las microcalcificaciones se debe conocer información a priori acerca de su longitud o tamaño, medidas empleadas por los especialistas en el diagnóstico. Para la extracción de características se han utilizado técnicas basadas en mediciones estadísticas las cuales no han alcanzado porcentajes de detección lo suficientemente altos; por lo tanto, es necesario considerar otras técnicas que permitan porcentajes de acierto mayores, tal es el caso de los descriptores de textura y los modelos activos [22], [23], [21], [25].

La textura de una imagen se puede definir como la variación entre píxeles en una pequeña vecindad, es decir, la textura es un atributo que representa la distribución espacial de los niveles de intensidad en una región dada.

Por esta razón, los descriptores de textura y los modelos activos pueden ser apropiados para el análisis y detección de hallazgos que representen malignidad en la mama, ya que el tejido que la conforma suele cambiar con la presencia de cáncer, permitiendo extraer información cuantitativa de este tipo de imágenes, ya que no se basa en el estudio del valor de cada píxel, sino en la variación entre píxeles contiguos; los modelos estadísticos de orden superior basados en los correspondientes histogramas de niveles de intensidad se ajustan también con dicha caracterización [22].

En la tercera fase se han utilizado sistemas de aprendizaje de máquina que han demostrado en el apoyo al diagnóstico una disminución de los índices de falsos positivos y negativos, algunas técnicas implementadas han sido: Clasificador bayesiano, redes neuronales, máquinas de soporte vectorial.

Por eso, un sistema de detección de mcals debe tener las siguientes ventajas:

- Las técnicas de preprocesamiento deben garantizar la mejora de la imagen y no la pérdida de detalles fundamentales para el diagnóstico.
- Extracción de características que permitan la mejor detección de hallazgos sutiles en las mamografías.
- Correcta clasificación de mcals proporcionando disminución de diagnósticos erróneos.

3.3. Formulación del problema de investigación

A pesar de los altos índices de incidencia y de muerte de cáncer de mama [2], [1], actualmente no se desarrollan muchos proyectos de creación de nuevas bases de datos [15]. En Colombia, puntualmente en la región del eje cafetero no se cuenta con una base de datos etiquetada de registros mamográficos, que permita validar mediante algoritmos de procesamiento en imágenes de pacientes con la taxonomía de las mujeres Colombianas, el diagnóstico médico emitido por los especialistas y su comprobación con biopsia en los casos que representan sospecha de malignidad.

Además, pese a que las metodologías para el realce de contraste y reducción de ruido mejoran elementos en las imágenes y son de bajo costo computacional, éstas pueden presentar pérdida de información fundamental en el análisis de imágenes de mamografía, debido a que las microcalcificaciones son pequeños hallazgos que en imágenes de baja resolución puede estar representadas mediante el tamaño de un píxel [5].

Debido a esto, es importante utilizar técnicas que permitan analizar la distribución espacial de un pixel y los píxeles de su entorno, con el fin de incrementar la precisión en la detección de microcalcificaciones y su severidad (benigna o maligna) [18], [19], [11], [14], [13], [23], con lo cual disminuirían los índices de mortalidad por cáncer de mama y los episodios de estrés innecesarios en las pacientes.

Considerando, primero la importancia de construir nuevas bases de datos y que a nivel local no se cuenta con un banco de imágenes mamográficas, segundo la relevancia de los sistemas de detección de mcals para el diagnóstico de cáncer en una etapa temprana; el problema a resolver consiste en la creación de una base de datos para validar el diagnóstico médico mediante técnicas de preprocesamiento y la extracción de características automáticas que permitan la detección de mcals y la clasificación del grado de severidad de las mismas.

¿Será posible validar el diagnóstico médico mediante las técnicas de preprocesamiento, extracción de características y clasificación aplicadas a la base de datos local creada en este proyecto?

4. Objetivos

4.1. General

Construir una base de datos de imágenes de mamografía para la identificación de microcalcificaciones.

4.2. Específicos

- Construir una metodología para el diseño de una base de datos de registros de mamografía de pacientes de una población local.
- Validar la base de datos mediante la aplicación de técnicas de extracción de características como los descriptores de textura y algoritmos de aprendizaje de máquina para la detección de mcals.

5. Antecedentes bibliográficos

La mamografía es una de las mayores herramientas para la detección temprana de cáncer de mama, pues en esta etapa el cáncer no es palpable. Luego del examen, las imágenes son interpretadas por radiólogos para encontrar algún signo de enfermedad, sin embargo, dicho análisis no es simple pues las características propias de la imagen hacen que en ocasiones sea confusa la categorización de los hallazgos la cual debe responder al estándar de radiología Bi-Rads (*Breast Imaging Report and Data System*). [20]

Por lo expuesto anteriormente, es necesario contar con bases de datos de imágenes mamográficas para realizar la implementación de técnicas de preprocesamiento que permitan mejorar elementos en las mamografías, incrementando a su vez la eficiencia de metodologías en etapas posteriores como la extracción de características y la clasificación de los hallazgos en la mama.

5.1. Protocolo para generar una base de datos

Para la construcción de una base de datos [26] propone una metodología compuesta por por las siguientes tres etapas:

5.1.1. Adquisición de los datos y validación

En esta etapa se inicia el proceso de adquisición y se termina con la validación de los datos según lo analizado por los especialistas. A continuación se describen los pasos realizados en el proceso:

- Se recoge de forma anónima la información y se realiza la eliminación de todo tipo de datos confidenciales del paciente.
- Se almacena en la base de datos información que incluye los datos de imagen en formato DICOM, el informe radiológico, el informe de histopatología y otros informes disponibles, según sea el caso del paciente.
- Tecnólogos en radiología validan y mejoran la información revisando el reporte clínico.
- La localización precisa de las patologías es delimitada como regiones de interés (ROI).
- Los casos más difíciles son revisados por neuro-radiológos y las ROIs para estos casos son delimitadas bajo la dirección de dichos especialistas.
- Como validación adicional se revisan aleatoriamente los datos introducidos a la base de datos.

5.1.2. Normalización y registro

El sistema combina algoritmos de preprocesamiento de imágenes con acceso directo a una base de datos, en esta etapa, se realiza la normalización de las imágenes corrigiendo la rotación y escala de éstas.

5.1.3. Extracción de características

Se analizan características de la imagen como: color, forma, textura o alguna información derivada de la imagen, que pemitan detectar hallazgos benignos y malignos.

5.2. Técnicas de preprocesamiento

A lo largo de la revisión literaria, se encuentra que para la reducción de ruido y mejora de contraste se han utilizado técnicas lineales, no lineales y de análisis multiresolución.

En [11] se proponen técnicas convencionales como CLAHE (*Contrast-Limited Adaptive Histogram Equalization*) y LRM (*Local Range Modification*) que son ampliamente utilizadas para mejora de contraste en imágenes médicas. El primer algoritmo divide la imagen en regiones y aplica ecualización de histograma a cada una de ellas, modificando la intensidad de los valores de la imagen mediante una metodología no lineal maximizando el valor de todos los pixeles de la imagen. La técnica LRM amplía el histograma utilizando la fórmula $y = ax + b$, donde y es la imagen mejorada, x la imagen en escala de grises y los parámetros a y b dependen del contraste local que se calcula mediante interpolación.

Además, utilizan técnicas de análisis multiresolución basadas en wavelets. La técnica 2-D RDWT (*Redundant Dyadic Wavelet Transform*) donde la imagen resultante se reconstruye usando el segundo y tercer nivel en un esquema de descomposición de cuatro niveles. WSRK (*wavelet shrinkage*) la cual se basa en la eliminación del primer y cuarto nivel de descomposición y la *wavelet back-ground* (WBGK) en la cual el preprocesamiento de la mamografía resulta del segundo y tercer nivel con una aproximación de los niveles del fondo los cuales son extraídos del cuarto nivel de descomposición. El mejor rendimiento se obtuvo con las metodologías LRM y con la wavelet basada en expansión lineal (WSRK), con un AZ de 0.932 y 0.926 respectivamente.

5.3. Técnicas de extracción de características

Para la extracción de características que permitan detectar las calcificaciones en una mamografía se han utilizado diversos algoritmos. En [13] se proponen técnicas como PCA (*Principal Component Analysis*), LDA (*Linear Discriminant Analysis*), TSA (*Tensor Subspace Analysis*) y GTDA (*General Tensor Discriminant Analysis*), en [14] no utilizan las dos primeras técnicas (PCA y LDA) porque consideran que no siempre trabajan bien, por lo tanto, proponen TSA y GTDA.

5.3.1. PCA

PCA es un algoritmo de reducción dimensional, que proyecta los datos en las direcciones de las máximas varianzas de tal manera que el error es minimizado. Dado una serie de puntos $x1, x2, \ldots.., xn$, sea w el

vector de transformación y $y_i = w^T x_i$. La función objetivo de PCA es:

$$w_{opt} = arg\left\{max_w \sum_{i=1}^{n} (y_i - \overline{y})^2\right\} = arg\left\{max_w w^T C w\right\} \tag{5.1}$$

Donde $y = \frac{1}{n}\sum yi$ y C es la matriz de covarianza de los datos. Las funciones base de PCA son eigenvectores de los datos de la matriz de covarianza asociados con los más grandes eigenvalores.

5.3.2. LDA

La técnica LDA a diferencia de PCA que busca direcciones que sean eficientes para la representación, busca direcciones que sean eficientes para discriminación. Si se tienen una serie de n muestras $x1, x2, \ldots .., xn$, pertenecientes a k clases. La función objetivo de LDA es como se muestra a continuación:

$$w_{opt} = arg\left\{max_w \frac{w^T S_B w}{w^T S_w w}\right\} = arg\left\{max_w \frac{tr(w^T S_B w))}{tr(w^T S_w w))}\right\} \tag{5.2}$$

$$S_B = \frac{1}{n}\sum_{i=1}^{k} n_i (m^{(i)} - m)(m^{(i)} - m)^T \tag{5.3}$$

$$S_w = \frac{1}{n}\sum_{i=1}^{k}\left(\sum_{j=1}^{n_i} (x_j^{(i)} - m^{(i)})(x_j^{(i)} - m^{(i)})^T\right) \tag{5.4}$$

Donde m es el vector de media de las muestras totales, ni es el número de muestras de las iésimas clases, m(i) es vector de media de las iésimas clases, SW la matriz de dispersión dentro de las clases y SB la matriz de dispersión entre las clases.

5.3.3. TSA y GTDA

TSA es una nueva técnica que aprende de un subespacio tensor con estructuras geométricas y discriminativas del espacio de datos original. Se tiene $X \in R^{n_1 x n_2}$ que denota una imagen de tamaño $n_1 x n_2$, donde X es un tensor de orden 2 en el espacio $R^{n_1} \otimes R^{n_2}$. Además, $(u_1, u_2, \ldots u_n)$ son funciones base ortonormales de R^{n_1}, y $(v_1, v_2, \ldots v_n)$ son funciones base ortonormales de R^{n_2}. Entonces el 2-tensor X puede escribirse como:

$$X = \sum_{ij} (u_i^T x v_j) u_i v_j^T \tag{5.5}$$

GTDA es una técnica modificada, es una extensión del tensor de DSDC *(differential scatter discriminant criterion).*

5.3.4. Descriptores de textura

En [23] se utiliza la técnica de descriptores de textura para extraer características de una calcificación que pueda representar malignidad, estas características son extraídas de un una región de interés (ROI) derivadas la matriz de co-ocurrencia, los elementos de dicha matriz son el conjunto de probabilidades de la ocurrencia de los niveles de gris i y j para los pares de pixeles los cuales estas separados por una distancia d en una dirección θ. Tiedeu utiliza una dirección $\theta = 0°$, distancia d = 4 para obtener finas características de textura y selecciona 13 características, algunas de ellas son: correlación, entropía, energía, inercia, momento diferencial inverso, entre otras, para analizar malignidad en microcalcificaciones

En [21] se utilizan diversas características de textura en imágenes de mamografías (discretas, markovianas, no markovianas, de longitud de ejecución y de fractal. Las características discretas de textura se basan en la segmentación de objetos en regiones de baja, media y alta densidad óptica; la densidad de área y sus radios sonunas de estas. Propone además, la extracción de características markovianas y no markovianas, las primeras se basan en caracterizar la variación de niveles de gris entre pixeles adyacentes en la imagen, comúnmente se calcula la función de probabilidad sobre los pixeles del objeto y valores estadísticos de su distribución; las segundas describen la textura en términos de una estimación global de los diferentes niveles de gris en un objeto.Las características de longitud de ejecución describen la textura según los niveles de gris ejecutados, representa los pixeles que tienen el mismo nivel de gris con el valor p de longitud q en cierta dirección θ. Por último, las de fractal, las cuales consideran el gráficode un objeto en tres dimensiones (la densidad óptica del objeto versus las coordenadas espaciales x y y).

5.4. Técnicas de clasificación de patrones

Para diagnosticar adecuadamente a las pacientes se deben clasificar los hallazgos (si existen) de sus mamografías en benignos o malignos, para esto, luego de extraer características determinantes de las imágenes de sus estudios se aplican algoritmos de aprendizaje de máquina, analizando la literatura los más utilizados para este tipo de imágenes son las redes neuronales (ANN) y las máquinas de soporte vectorial (SVM).

En [23] utilizan un clasificador basado en redes neuronales ya que la clasificación de mcals mediante los métodos del vecino más cercano (KNN) y el Bayesiano no arrojaron los mejores resultados. La ANN utilizada es la BPNN (*Back Propagation Neural Network*), la cual está constituida por una MLP (*Multilayer Perceptron*) y el algoritmo de entrenamiento backpropagation. Para seleccionar las características de entrenamiento de la ANN, realizan un método iterativo el cual inicia con 13 características y la evaluación del GCR (*Good Classification Rate*), luego se extrae una característica a la vez y se recalcula el GCR, el mejor resultado se obtuvo con nueve características, logrando una sensibilidad del sistema de clasificación del tipo de hallazgo (benigno o maligno) de un 100 %, una especificidad del 87.7 % y un AZ de 0.968.

En [13] utilizan una máquina de soporte vectorial para clasificar las mcals detectadas, el proceso utilizado se basa en cuatro pasos: preprocesamiento de la mamografía, extracción de características por ventanas de análisis, luego se aplican algoritmos de aprendizaje para obtener el subespacio de características y luego se aplica la SVM para decidir si el subespacio corresponde o no, a una mcal.

En [14] se utiliza el clasificador TWSMV (*Twin Support Vector Machine*) en el cual la detección es un problema de clasificación binaria, de esta forma se analiza si una mcal está presente o no, en cada parte o bloque de la mamografía. Este clasificador obtiene planos no paralelos alrededor de los puntos que corresponden a

los datos de la clase de cúmulos de mcals. Cada uno de los problemas de programación cuadrática de un par de TWSVM tiene la formulación de una SVM típica.

Con la aplicación de las técnicas GTDA + SMV se obtuvo una alta sensibilidad en la detección y clasificación de mcals de 0.9431 con un error de 0.074; sin embargo, el clasificador TWSVM demuestra unos mejores resultados en cuanto a clasificación con un 0.9677 $\pm 0{,}0574$.

Autor	**Año**	**Técnica**	**Mejora**
Papadopoulos *et.al.* [11]	2008	CLAHE, LRM, RDWT, WSRK, WBGK	Optimización del CAD (computer-aided design), la cual fue dirigida al porcentaje de los píxeles con más contraste y al tamaño mínimo del objeto detectable que podría representar satisfactoriamente mcals. Los mejores resultados se obtuvieron ampliando el histograma (LRM) y con aplicación de la wavelet RDWT.
Sameti *et.al.* [21]	2009	*Texture Features* discrete, markovian, non-markovian, run-length, fractal	Detección de signos de desarrollo de cáncer en las marcaciones realizadas por un especialista en una mamografía (antes de adquirir cáncer), lo que aumenta la investigación de la región sospechosa y reduce el intervalo de tiempo para el siguiente examen.
Zhang and Hua Xie [14]	2012	PCA, LDA, TSA, GTDA, SVM	SVM basada en subespacios de aprendizaje (TSA y GTDA) tiene mejor rendimiento que métodos tradicionales como ANN. Esta metodología es cuatro veces más rápida que la convencional SVM.
Zhang and Xinbo Gao [13]	2012	TSA, GTDA, TWSVM	Mejor detección de mcals y es más rápida que SVM.
Tiedeu *et.al.* [23]	2012	*Textural features*, BPNN	La sensibilidad del sistema es del 100 %, se logra la detección de todos los casos de cáncer.

Tabla 5.1: Mejoras propuestas de los sistemas para la detección de mcals.

6. Marco Conceptual

6.1. Conceptos médicos

El cáncer de seno es la segunda causa principal de muerte por cáncer en las mujeres. La probabilidad de que el cáncer de seno sea responsable de la muerte de una mujer es de aproximadamente 1 en 36 (alrededor de tres por ciento) [27]. La mamografía como prueba de tamizaje es muy utilizada, en ella pueden detectarse lesiones benignas o malignas, no obstante, las características propias del tejido de la mama hace que en ocasiones su observación se dificulte, a continuación se describen conceptos relacionados con un estudio de mamografía: tipos de tejido, de proyecciones, de lesiones, hallazgos detectados y la clasificación de éstos según su severidad.

6.1.1. Tipo de tejido

Existe una gran variación en la densidad del parénquima mamario que se observa en las mamografías. La densidad de la mamografía varía inversamente con la edad. Por lo tanto, las mujeres más jóvenes tienden a tener mamas más densas en cuanto a tejido glandular que las mujeres de más edad, pero muchas mujeres mayores también tienen mamas densas. [5]

El aspecto general de la imagen proporciona información acerca del tejido predominante en la mamografía, estos pueden ser tejido graso, tejido graso glandular, tejido mixto (graso y glandular) y tejido denso. La densidad del parénquima depende de la cantidad de tejido conectivo y de tejido glandular en la mama. Aquellas dominadas por el tejido adiposo, que parecen menos densas, son fáciles de analizar con la mamografía. [5]

6.1.2. Tipos de proyecciones

Para adquirir las imágenes en un examen de mamografía, las proyecciones más utilizadas son la medio lateral oblicua y la cráneo caudal.

Proyección medio lateral oblicua (MLO): en este tipo de proyección se debe observar [28]:

- El ángulo intramamario
- El pezón de perfil
- El pezón al mismo nivel que el borde inferior del músculo pectoral
- El músculo pectoral cruzando la placa con un ángulo adecuado para cada mujer (entre 20^o y 35^o)

Proyección Cráneo Caudal (CC): Junto a la MLO a 45^o se considera como la exploración radiológica habitual de la mama [28]. Debe mostrar:

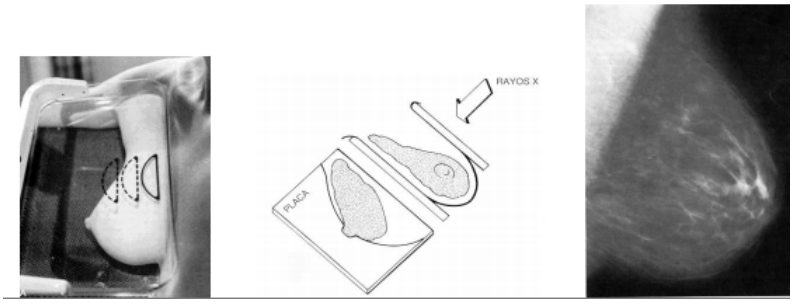

Figura 6.1: Proyección medio lateral oblicua (MLO)

- El pezón de perfil apuntando ligeramente hacia la línea media
- La mayor parte del tejido lateral y medial con la excepción de la cola axilar
- En algunas pacientes puede observarse el músculo pectoral

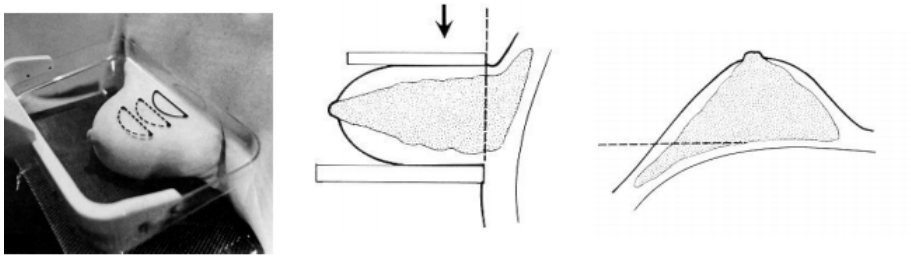

Figura 6.2: Proyección Cráneo Caudal (CC)

6.1.3. Tipos de lesiones en la mama

Para determinar si una mamografía es normal o si presenta lesiones benignas o malignas, el radiólogo debe tener conocimiento acerca de los diversos procesos fisiológicos, patrones de crecimiento, adenosis o fibroadenomas, entre otros, que presente la paciente. [5]

Lesiones benignas de la mama: son de gran interés para el radiólogo ya que, en ocasiones, es difícil diferenciarlas de las malignas. Incluso más difícil aún es la distinción entre los procesos benignos y las mamas normales. Los grupos de las diferentes enfermedades de mama no son absolutos, pero sirven para identificar la probabilidad de un proceso simple o más localizado, como opuesto a procesos múltiples o difusos. Las lesiones benignas se pueden dividir en cuatro grandes grupos: intraglandulares discretas, extraglandulares discretas, subareolares discretas y difusas. [5]

Lesiones malignas de la mama: más del 90 % de los cánceres de mama proceden de las células ductales epiteliales. Así, es interesante determinar qué actividad celular es importante y cuál es premaligna. Los tumores malignos típicos no sólo producen cambios localizados en el sitio en que aparece la lesión, sino que también provocan alteraciones difusas en alguna o todas las estructuras de la mama distantes de la lesión. Básicamente la clasificación de las lesiones malignas considera carcinomas no invasivos e invasivos, tipos histológicos especiales de carcinomas, sarcomas, y tumores metastáticos. [5]

6.1.4. Hallazgos detectados en una mamografía

Un hallazgo corresponde a una región sospechosa que es de interés para el especialista. Los hallazgos a detectar en una mamografía son:

Normal: es difícil afirmar qué es una mamografía normal, dado que la mama está conformada por tejido conjuntivo, glandular y graso que con el paso de los años, por la paridad y otros factores puede adquirir apariencia densa. Es más práctico definir los casos anormales. [5]

Masas: son concentraciones que se muestran en la mamografía como formas relativamente bien definidas. Existen diferentes tipos de masas encontrados en una mamografía, las masas son categorizadas según su localización, forma, tamaño, densidad y sus bordes. La localización puede indicar si la lesión es probablemente maligna. La forma puede ser redonda, ovalada, lobular e irregular, tal como se ilustra en la parte izquierda de la figura 6.3. El tamaño de una masa en una mamografía no determina si la lesión es maligna o benigna, no obstante si indica el grado de progreso. Entre los bordes se incluyen circunscrito, obscurecida, microlobulada, indefinida y espiculada, como se puede ver en la parte derecha de la figura 6.3. [5]

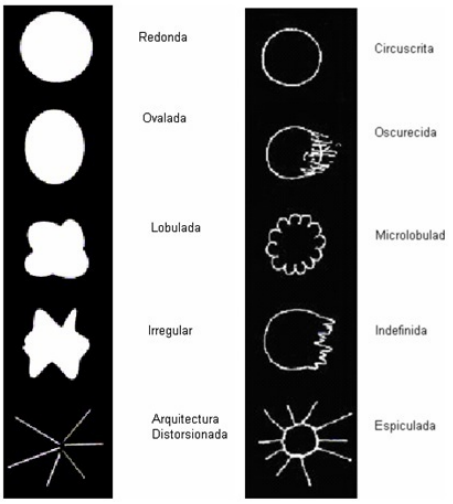

Figura 6.3: Tipos de masas

La apariencia real de las masas se muestra en la figura 6.4, según la forma: a) redonda, b)Ovalada, c) Lobular, d)irregular, según los bordes: e) circunscrita, f) Obscurecida, g)Microlobulada, h) Indefinida.

Microcalcificaciones: Son minúsculos depósitos de calcio dentro de los tejidos de la mama y se muestran en la imagen como puntos de alta intensidad y marcado contraste. Su tamaño puede variar entre 0.05 mm y 2 mm. Pueden aparecer individuales o en grupos. De la forma, tamaño, número y disposición puede deducirse su naturaleza (benigna o maligna). [5]

Por lo general las microcalcificaciones de forma alargada, redonda u ovalada de gran tamaño y uniformes pueden estar asociadas a un proceso benigno. [5]

Las microcalcificaciones benignas suelen tener un tamaño mayor a 2 mm, por el contrario las microcalcificaciones malignas suelen medir menos de 0.5 mm. [5]

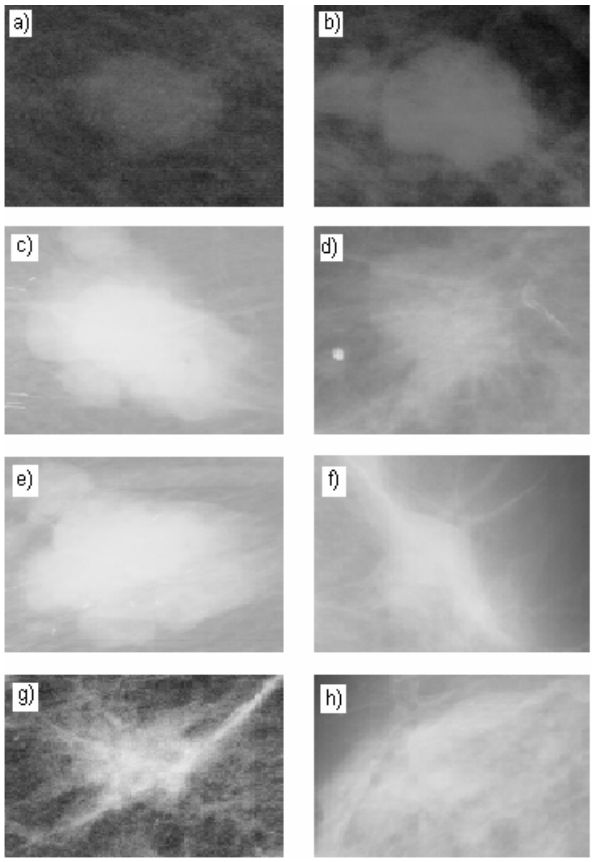

Figura 6.4: Tipos de masas observados en las mamografías

El número de microcalcificaciones que constituyen un arreglo es usado como un indicador de malignidad o benignidad. Mientras que el número de microcalcificaciones es arbitrario, los radiólogos coinciden en que el número mínimo de microcalcificaciones agrupadas deben ser de 4, 5 ó 6, para ser significativas. Un número menor que estas raramente guían a la detección de cáncer de seno. [5]

Existen clasificaciones dadas por asociaciones de radiólogos divididas en tres categorías: típicamente benignas, intermedias, con alta probabilidad de ser malignas. [5]

Microcalcificaciones típicamente benignas.

- Calcificaciones de piel: típicamente se presentan con un centro radiolúcido y tienen forma poligonal. Se ubican en la piel, pero en la proyección pueden aparecer ubicadas en el parénquima. Las formas atípicas pueden ser conformadas mediante vistas tangenciales que las muestren a nivel de la piel. [5]

- Calcificaciones vasculares: forman líneas paralelas o tubulares, están claramente asociadas con vasos sanguíneos. [5]

- Calcificaciones groseras: se pueden observar en fibroadenomas en involución. [5]

- Calcificaciones en forma de barra: poseen forma tubular, pueden ramificarse ocasionalmente, usualmente miden más de un milímetro de diámetro, pueden tener un centro radiolúcido si el calcio rodea en lugar de rellenar los conductos ectásicos. Pueden encontrarse en la ectasia ductal. [5]

- Calcificaciones redondas: son de tamaño variable, miden más de 1 mm, cuando miden menos de 0.5 mm puede usarse el término puntiforme. [5]

- Calcificaciones puntiformes: miden menos de 0.5 mm, son redondas u ovaladas, aparecen bien definidas como depósitos puntuados, raramente se asocian con cáncer, pero si se asocian con otras calcificaciones de forma irregular incrementan la sospecha de ser malignas. [5]

- Calcificaciones esféricas: se pueden extender desde 1 mm hasta 1cm. Pueden encontrarse como restos en un conducto mamario, en áreas de necrosis de tejido graso y a veces en fibroadenomas. [5]

- Calcificaciones bordeadas o tipo cáscara de huevo: son muy delgadas (menos de 1 mm. de espesor) y semejan a depósitos cálcicos sobre la superficie de una esfera. Aunque la necrosis grasa puede producir este tipo de depósitos, la calcificación de la pared de un quiste es la causa más común. El cáncer de mama raramente produce este tipo de calcificaciones. [5]

- Calcificaciones de calcio de leche: sedimento cálcico intraquístico. En la proyección cráneo-caudal es menos evidente y tiene apariencia poco definida como un depósito amorfo, mientras que en la proyección lateral es de bordes definidos, semilunares, de forma creciente, cóncava hacia arriba o lineal dependiendo la porción del quiste. [5]

- Calcificaciones suturadas: se forman alrededor del material de sutura. Son relativamente comunes en las mamas irradiadas, típicamente son lineales o tubulares y en ocasiones pueden identificarse los nudos. [5]

- Calcificaciones distróficas: usualmente aparecen en mamas irradiadas o traumatizadas.Aunque de contornos irregulares usualmente miden más de 0.5 mm. De diámetro y a menudo tienen centro radiolúcido. [5]

La apariencia real de este tipo de lesiones se muestra en la figura 6.5 , a) Tipo piel, b) Tipo puntiformes, redondeadas dispersas c) En anillo, cáscara de huevo, d) Tipo calcio de leche, e) Tipo heterogéneas, groseras en forma de palomita de maíz, f) En forma de barra, g) Lineales gruesas en vara, h) distróficas. [5]

Microcalcificaciones intermedias.

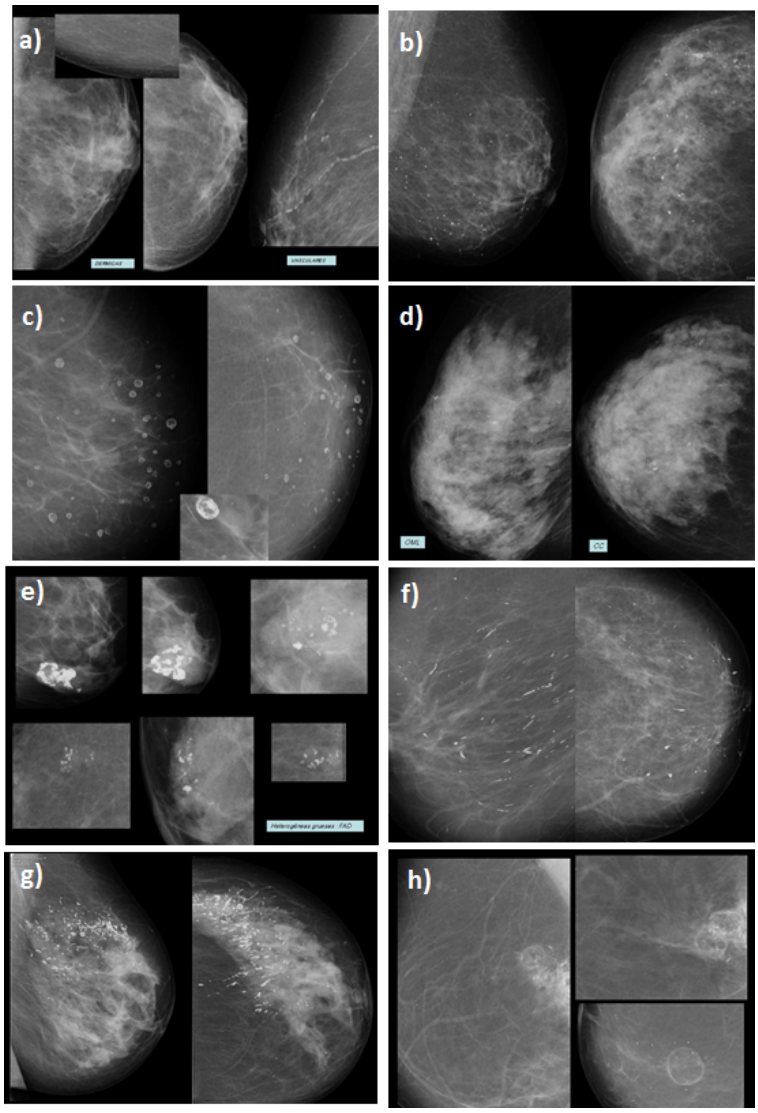

Figura 6.5: Microcalcificaciones benignas observadas en mamografías

Son microcalcificaciones indistintas o amorfas. Pueden parecer redondas o en formas de hojuelas, pero son lo suficientemente pequeñas para que su forma no pueda ser apreciable en forma clara. [5]

En la figura 6.6 se muestra este tipo de calcificaciones en una mamografía, a) heterogéneas, b) amorfas.

Microcalcificaciones con alta probabilidad de ser malignas.

- Calcificaciones heterogéneas: son granulares, no se asocian a procesos benignos o malignos a priori, pero agrupamientos de calcificaciones irregulares en tamaño y forma y además de tamaños menores a 0.5 mm incrementan la sospecha de ser malignos. [5]

- Calcificaciones de líneas finas: son delgadas, irregulares, diseminadas y aparecen formando líneas

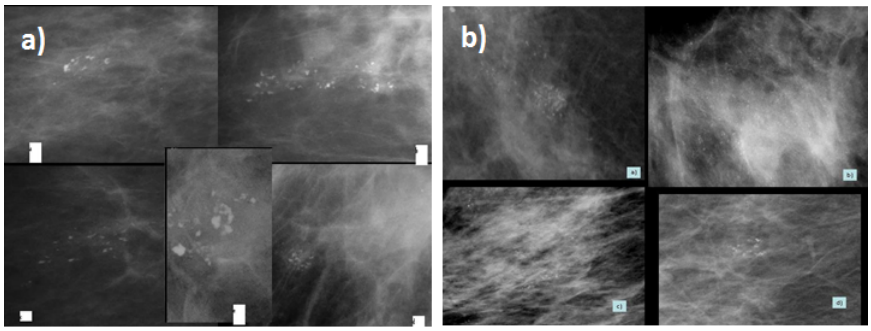

Figura 6.6: Microcalcificaciones intermedias observadas en mamografías

discontinuas, tienen menos de 0.5 mm de ancho. Su apariencia sugiere el llenado de la luz de los conductos afectados irregularmente por cáncer de mama. [5]

La apariencia real de este tipo de lesiones se muestra en la figura 6.7, a) microcalcificación de tipo pleofórmica fina, b) microcalcificaciones lineales finas/ramificadas. [5]

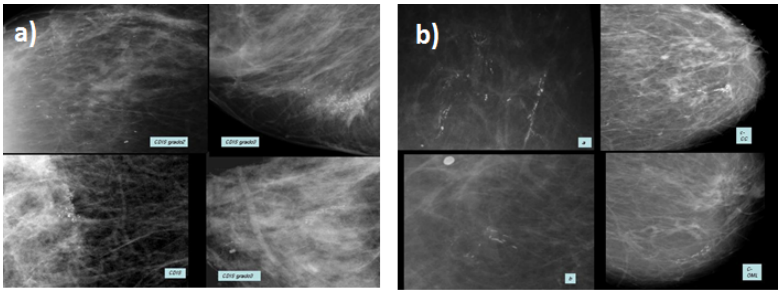

Figura 6.7: Microcalcificaciones malignas observadas en mamografías

Distorsiones arquitectónicas: corresponden a una reacción dermoplásica en la cual existe interrupción focal del patrón del tejido normal, se manifiesta cuando el tejido circundante a la lesión se distorsiona levemente hacia un eje focal. Se trata de identificar una lesión no visible directamente por sus efectos en el tejido circundante. [5]

Densidades asimétricas: corresponden a variaciones en la densidad de algunas regiones de la mama, se observan comparando ambas mamas (derecha e izquierda). No se puede esperar que una sea el espejo de la otra. Este tipo de análisis permite analizar estructuras que pueden pasar como tejido denso si se analiza una sola mamografía en lugar de realizar un análisis conjunto de ambas mamas [5].

6.1.5. Clasificación BI-RADS

Existe una clasificación de los hallazgos de las mamografías en categorías establecida por el colegio estadounidense de radiólogos *(American College of Radiology)*, denominada BI-RADS (*Breast Imaging Reporting and Data System)*, este define los términos para describir anormalidades en las mamografías en categorías

que son predictivas de la probabilidad de malignidad así [29]:

BI-RADS O: evaluación incompleta. Es necesario realizar estudios por imágenes adicionales o comparar con mamografías anteriores. Esto significa que es posible que haya una anomalía que no sea visible o no esté definida con claridad y se necesiten exámenes adicionales, como el uso de una compresión puntual (se aplica compresión a un área menor cuando se hace el mamograma), vistas agrandadas, vistas especiales en el mamograma o ultrasonido.

Esto sugiere también que se debe comparar el mamograma con exámenes anteriores para determinar si con el tiempo han ocurrido cambios en el área.

BI-RADS 1: hallazgos negativos (mamografía normal). No hay ninguna anomalía importante que reportar. Los senos lucen igual (son simétricos), no hay bultos (protuberancias), estructuras distorsionadas, o calcificaciones sospechosas. En este caso, negativo significa que no se encontró algo malo.

BI-RADS 2: hallazgos benignos. También se trata de un resultado negativo de la mamografía (no hay signos de cáncer), pero el médico que realiza el informe prefiere describir el hallazgo como benigno, tal como calcificaciones benignas, ganglios linfáticos en el seno o fibroadenomas calcificados. Esto asegura que otras personas que vean la mamografía no interpretarán equivocadamente este hallazgo benigno como sospechoso. Este hallazgo se incluye en el informe de la mamografía para ayudar en la comparación con futuras mamografías.

BI-RADS 3: probablemente benignos, seguimiento a los 6 meses. Los hallazgos en esta categoría tienen una muy buena posibilidad (más de 98 %) de ser benignos (no cancerosos). No se espera que estos hallazgos cambien con el tiempo. Pero ya que no se ha probado que sea benigno, es útil ver si han ocurrido cambios a lo largo del tiempo en el área de interés.

Por lo general, se hace seguimiento a los 6 meses cuando se repite la evaluación con imágenes y luego regularmente, hasta que se determine que el hallazgo está estable (usualmente un mínimo de 2 años). Este enfoque evita biopsias innecesarias, pero si el área cambia a lo largo del tiempo, permite hacer un diagnóstico en sus inicios.

BI-RADS 4: anormalidad sospechosa, se recomienda una biopsia. Los hallazgos no parecen indicar de manera definitiva que sean cancerosos, pero pudiera ser cáncer. El radiólogo está lo suficientemente preocupado como para recomendar una biopsia. Los hallazgos en esta categoría tienen un rango amplio de niveles de sospecha. Por este motivo, algunos médicos pueden dividir esta categoría aún más:

4A: hallazgo con una sospecha baja de que sea cáncer.
4B: hallazgo con una sospecha mediana de que sea cáncer.
4C: hallazgo de preocupación moderada de que sea cáncer, pero no tan alta como la Categoría 5.

No todos los médicos usan estas subcategorías.

BI-RADS 5: anormalidad que sugiere firmemente la presencia de cáncer. Los hallazgos tienen la apariencia de cáncer y hay una alta probabilidad (al menos 95 %) de que se sea cáncer. Se recomienda firmemente la realización de una biopsia.

La categoría de BIRADS ayuda al médico a decidir cuáles deberían ser las opciones de tratamiento posteriores. En los casos de clasificación BIRADS 3 y 4, cabe considerar la posibilidad de una biopsia que permita con un buen índice de certeza confirmar o descartar malignidad en la muestra tomada.

BI-RADS 6: resultados de biopsia conocidos con malignidad demostrada, se deben tomar las acciones adecuadas. Esta categoría se utiliza únicamente para hallazgos en un mamograma que ya han demostrado ser cancerosos según una biopsia realizada con anterioridad. Los mamogramas se usan de esta forma para ver cómo el cáncer está respondiendo al tratamiento.

6.2. Técnicas de procesamiento de imágenes

6.2.1. Técnicas de segmentación

En ocasiones en el procesamiento de imágenes es necesario aislar o separar objetos de interés, generalmente las imágenes estan compuestas por regiones o zonas que tienen características iguales (nivel de gris, textura, momentos, etc.), las cuales corresponde a los objetos presentes en la imagen. La segmentación de una imagen consiste en dividir la imagen en varias regiones homogéneas y disjuntas a partir de su contorno, su conectividad, o en términos de un conjunto de características de los píxeles de la imagen que permitan discriminar unas regiones de otras. [30]

Los algoritmos de segmentación de imágenes se basan en alguna de las siguientes tres propiedades:

- Discontinuidad en los tonos de gris de los píxeles de un entorno, que permite detectar puntos aislados, líneas y aristas (bordes).
- Similaridad en los tonos de gris de los píxeles de un entorno, que permite construir regiones por división, por crecimiento o por umbralización.
- Conectividad de los píxeles. Se dice que una región D es conexa o conectada si para cada par de píxeles de la región existe un camino formado por píxeles de D que los conecta. Un camino de píxeles es una secuencia de píxeles adyacentes (que pertenecen a su entorno inmediato).

Los métodos de segmentación se pueden agrupar en cuatro clases diferentes [30]:

1. Métodos basados en píxeles: están compuestos por operaciones locales y globales, las primeras se basan en las propiedades de los píxeles y los de su entorno y las segundas se basan en la información global obtenida de la imagen.

1. Métodos basados en bordes.
2. Métodos basados en regiones: utilizan las nociones de homogeneidad y proximidad geométrica, como las técnicas de crecimiento, fusión o división.
3. Métodos basados en modelos.

6.2.2. Técnicas para el filtrado de imágenes

Con las técnicas de filtrado se busca disminuir el ruido que es introducido durante la adquisición y digitalización de las imágenes el cual puede interferir con las características de interés e información propia de una mamografía [5].

Una técnica utilizada para la reducción de ruido es el filtrado espacial, el cual se realiza desplazando una matriz rectangular de dos dimensiones conocida como: ventana, kernel, máscara o núcleo que contiene unos "pesos.º ponderaciones por cada píxel de la imagen. Cuando la ventana se ubica sobre los píxeles, se realiza la evaluación del pixel central de la ventana de acuerdo con los pixeles de alrededor y sus valores de ponderación, este proceso se realiza con cada pixel de la imagen desplazando la ventana. Al proceso de evaluar la vecindad del pixel con los pesos se le denomina çonvoluciónz a la matriz del filtro se le conoce como "kernel de convolución" [31].
La convolución de una imagen f de tamaño M x N con una máscara h de m x n está dada por la siguiente expresión:

$$g(x,y) = \sum_{i=-a}^{a} \sum_{j=-b}^{b} f(x+i, y+j).h(i,j) \tag{6.1}$$

Donde

$$a = \frac{m-1}{2} b = \frac{n-1}{2}$$

En la siguiente figura se muestra el proceso de convolución aplicado a un solo píxel de la imagen.

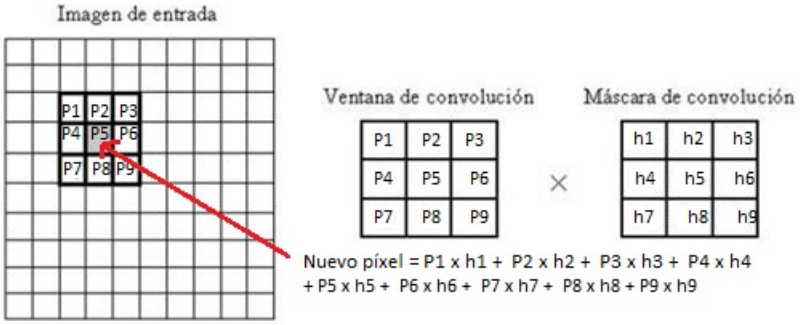

Figura 6.8: Proceso de convolución

Uno de los tipos de filtros más utilizados es el pasa bajo o de suavizado, se utiliza para eliminar ruido o pequeños detalles en una imagen, existen diveros filtros pasabajo, su diferencia radica en el tipo de operación realizada sobre los píxeles de la imagen, a continuación se describen algunos de ellos:

Filtro de media o promedio: este tipo de filtro realiza el promedio sobre el entorno de vecindad n x n de la ventana [32]. El más básico de ellos realiza la media aritmética, como sigue:

$$Ma = \frac{1}{nm} \sum_{(x,y)\in W} f(x,y) \tag{6.2}$$

donde nm es el número de píxeles en la venta W de dimensión n x m.

Este filtro sirve para suavizar las variaciones locales dentro de una imagen.

Filtro de Mediana: es un filtro estadístico no lineal cuya respuesta se basa en el ordenamiento de los pixeles abarcados por la máscara y luego el valor del píxel central es reemplazado por el valor de la mediana determinado del ordenamiento. Por ejemplo, en un entorno de 3 x 3 la mediana es el quinto valor más grande [32].

Filtro Gaussiano: este filtro simula una distribución gaussiana, el valor máximo se ubica en el pixel central y disminuye hacia los extremos según el parámetro de desviación típica s. El resultado de este filtro es un conjunto de valores entre 0 y 1. Si la matriz a transformar está compuesta por números enteros, ésta se divide por el menor de los valores obtenidos. La ecuación para calcularla es:

$$g(x,y) = e^{-\frac{x^2+y^2}{2s^2}} \tag{6.3}$$

$$G(x,y) = \frac{g(x,y)}{\text{mín}_{x,y}(g(x,y))} \tag{6.4}$$

- Filtro Unsharp: este tipo de filtro se utiliza para agudizar las formas en una imagen restando a la imagen original una versión promediada de ésta.

$$g(x,y) = f(x,y) - \overline{f}(x,y) \tag{6.5}$$

En la figura 6.9 se muestra la aplicación de filtros a una imagen:

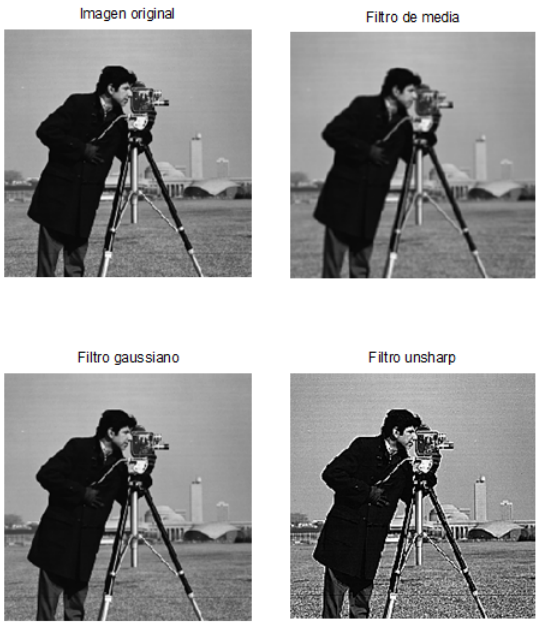

Figura 6.9: Aplicación de técnicas de filtrado

6.2.3. Técnicas para extracción de características

La técnica de descriptores de textura es ampliamente utilizada para extraer características relevantes de un objeto en una imagen. La textura en una imagen puede proveer información invaluable al momento de identificar objetos presentes en ésta. La principal característica de la textura es la repetición de un patrón o de varios patrones contenidos en una región [25]. La distribución de las instancias de un patrón de textura básico que se repite en el espacio puede no ser idéntico [32].

Existen varias técnicas para el tratamiento de las texturas en una imagen: 1) primitivas de textura, 2) modelos estructurales y 3) modelos estadísticos.

1. **Primitivas de textura:** El calificativo texel (*texture element*) hace referencia a una primitiva visual con propiedades que se repiten en diferentes posiciones pero son invariantes, además de las deformaciones y orientaciones en una region dada. Dichos elementos pueden cumplir con la propiedad de que todos los píxeles que lo conforman tienen el mismo nivel de gris. Un texel debe repetirse varias veces dentro de un área, para entender cuántas veces debe hacerlo suponga que tiene una ventana de análisis, a medida que ésta se va haciendo más pequeña, hay menos texels contenidos en ella. Para alguna distancia, la imagen en la ventana ya no aparece texturada, si llegara a hacerlo, la sola translación de la ventana cambiaría la percepción de la textura. De manera silimar ocurre si la ventana es muy grande, pues el campo de vista se alejaría. Por lo tanto, el número adecuado de texels está relacionado con la resolución, si ésta es apropiada la textura es evidente a medida que el campo visual se desplaza por el área de textura. [32]

2. **Modelos estructurales:** Una de las formas de describir los modelos estructurales es a partir de gramáticas, desarrolladas por Ballard y Brown (1982), una de ellas puede ser la gramática de forma que se define como una 4-trupla (vt, vm, R, S), donde:
   - Vt es un conjunto finito de formas.
   - Vm es un conjunto finito de formas, tales que $vt \bigcap vm = \oslash$.
   - R es un conjunto finito de pares ordenado (u,v) tales que u es la forma que consta de un elemento de vt^+ combinada con un elemento vm^+ y v es la forma de un elemento vt^* combinada con un elemento vm^*.
   - S es una forma que consta de un elemento vt^* combinada con un elemento vm^*.
3. **Modelos estadísticos:** Este modelo aplica cuando una textura no tienen una distribución geométrica regular, por lo que es conveniente describirlas mediante modelos estadísticos.

Local Binary Pattern

El operador LBP *(Local Binary Pattern)* es un acercamiento que unifica los modelos estadísticos y los modelos estructurales del análisis de texturas. En el modelo LBP la textura se describe en términos de microprimitivas (textones) y sus reglas estadísticas; opcionalmente, las primitivas pueden ser utilizadas como una medida complementaria del contraste local de la imagen, lo cual mide la fuerza de las primitivas [33].

Definiendo la textura T en un vecindario local de una imágenes en niveles de gris, como la distribución conjunta de los niveles de gris de $P + 1(\mathrm{P} \succ 0) pixelesenunaimagen$.

$$T = t(gc, go, ...gp - 1) \tag{6.6}$$

Donde gc corresponde al valor de gris del píxel central de un vecindario local y gp(p = 0, 1, . .. p-1) corresponde a los valores de gris de P igualmente espaciados sobre un circulo de radio R $(R \succ 0)$ que forma un conjunto de vecinos simétrico circular. Este conjunto de P + 1 píxeles es después denotado por Gp.

En el dominio digital de una imagen, las coordenadas de los vecinos gp están dadas por:

$$xc = R\cos\frac{2\pi p}{P} \tag{6.7}$$

$$yc = R\sin\frac{2\pi p}{P} \tag{6.8}$$

Donde (xc, yc) son las coordenadas del píxel central.

La siguiente figura muestra tres conjuntos de vecinos circulares simétricos con diferentes valores de P y R. Los valores de los vecinos que no caen exactamente en los pixeles son estimados por interpolación lineal. Puesto que la correlación entre pixeles decrece con la distancia, la información textural de una imagen puede ser obtenida de vecindarios locales.

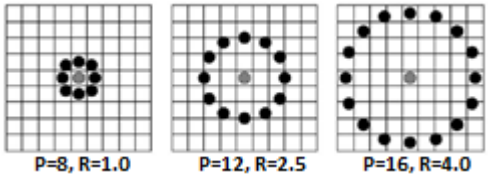

Figura 6.10: Conjuntos de vecinos circulares simétricos con valores de P y R diferentes.

Si el valor del píxel central es restado a los valores de los vecinos, la textura local puede ser representada (sin perder información) como una distribución conjunta del valor del píxel central y las diferencias:

$$T = t(gc, go - gc, ...gp - 1 - gc) \tag{6.9}$$

Suponiendo que las diferencias son independientes de gc, la distribución puede ser factorizada:

$$T \approx t(gc)(go - gc, ...gp - 1 - gc) \tag{6.10}$$

Debido a que t(gc) describe el total de la luminosidad de una imagen, la cual no tiene relación con la textura local de la imagen, t(gc) no provee información útil para el análisis de textura.

Por lo tanto, mucha de la información acerca de las características de textura en la distribución conjunta original es preservada en la distribución conjunta de las diferencias.

$$T \approx t(go - gc, ...gp - 1 - gc) \tag{6.11}$$

Para conseguir la invariancia respecto a cualquier transformación monotónica en niveles de grises, únicamente se consideran los signos de las diferencias.

$$T \approx t[s(go - gc, ...s(gp - 1 - gc)] \tag{6.12}$$

Donde

$$S(x) = \begin{cases} 1, \text{ si } x \geqslant 0 \\ 0, \text{ en caso contrario} \end{cases} \tag{6.13}$$

Ahora, se asignan pesos binomiales 2P a cada signo s(gp-gc) transformando las diferencias de un vecindario en un código LBP único.

Este código caracteriza la textura local de la imagen alrededor de (xc,yc):

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(gp - gc)2^P \tag{6.14}$$

El proceso del operador LBP aplicado a un píxel de la imagen se muestra en la siguiente figura:

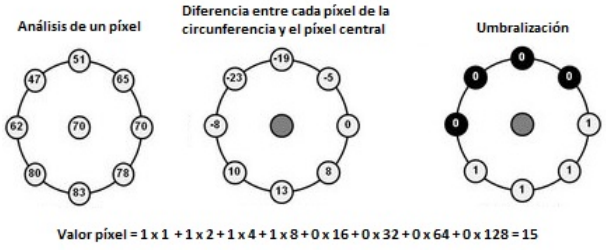

Figura 6.11: Aplicación del operador LBP a un píxel.

Segmentation-based Fractal Texture Analysis

El algoritmo de extracción de características mediante SFTA (*Segmentation-based Fractal Texture Analysis*) se realiza en dos pasos: el primero consiste en descomponer la imagen de entrada en un conjunto de imágenes binarias y el segundo en formar un vector de características de dicho conjunto [9].

Para descomponer la imagen original en un grupo de imágenes binarias puede utilizarse la técnica TTBD *(Two-Threshold Binary Decomposition)*, la cual se describe a continuación:

Esté método toma como entrada una imagen en escala de grises I(x,y) y retorna una serie de imágenes binarias. El proceso inicia computando un conjunto T de valores umbral que son obtenidos seleccionando los niveles de gris igualmente espaciados.

TTBD utiliza la información de la distribución de los niveles de gris de la imagen de entrada para calcular el conjunto de umbrales, esto se consigue empleando el algoritmo multi-nivel de Otsu.

El algoritmo de otsu, encuentra el umbral que minimiza la varianza interclase de la imagen de entrada. Este algoritmo es aplicado a cada región de la imagen hasta obtener el número deseado de umbrales nt, definidos por el usuario [9].

El siguiente paso de TTBD consiste en descomponer la imagen de entrada I(x,y) en un conjunto de imágenes binarias, para esto se seleccionan pares de umbrales de T y se aplica a la segmentación de dos umbrales, así:

$$I_b(x,y) = \begin{cases} 1, tl < I(x,y) < tu \\ 0, otrocaso \end{cases} \tag{6.15}$$

Donde tl y tu denotan respectivamente, los valores inferior y superior de umbral.

El conjunto de imágenes binarias es obtenido aplicando la ecuación anterior a la imagen de entrada usando todos los pares de umbrales contiguos de $T \in nl$ y todos los pares de umbrales $t, nl, t \in T$, donde nl corresponde al máximo nivel de gris de la imagen I(x,y), por lo tanto, el número resultante de imágenes binarias es 2nt.

Todas las imágenes binarias obtenidas mediante TTBD son un superconjunto de todas las imágenes binarias que se obtendría por aplicación de una segmentación de un solo umbral.

La justificación del uso de pares de umbrales para calcular el conjunto de imágenes binarias es que pueden segmentarse objetos que no se logran con la segmentacion de una umbralización regular, además, permite extraer información de regiones cuyos niveles de gris se encuentra en los rangos medios del histograma de la imagen de entrada. En la figura 6.12 se muestra la descomposición de una imagen según tres niveles de brillo: iluminado, medio y oscuro mediante el uso de pares de umbralización.

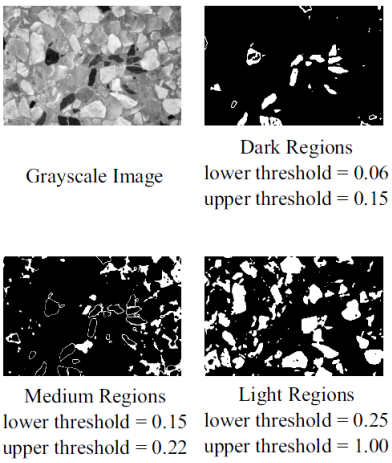

Figura 6.12: Textura y sus correspondientes imágenes binarias obtenidas por segmentación de pares de umbralización.

El siguiente paso del algoritmo SFTA consiste en generar un vector de características de las imágenes binarias resultantes con el tamaño (conteo de píxeles), media de los niveles de gris y contorno de éstas, llamado dimensión fractal. Estas medidas fractal sirven para describir la complejidad entre los límites de objetos y las estructuras segmentadas de la imagen de entrada.

Los límites de las regiones de una imagen binaria Ib(x,y) son representadas como una imagen de bordes denotada por $\Delta(x, y)$ y calculados como sigue:

$$\Delta(x,y) = \begin{cases} 1, \exists(x^{'},y^{'}) \in N_8[(x,y)]: \\ \quad I_b(x^{'},y^{'}) = 0 \wedge \\ I_b(x,y) = 1 \\ 0, otrocaso \end{cases}$$

(6.16)

donde N8[(x,y)] es el conjunto de pixeles que están 8-conectados a (x,y). $\Delta(x, y)$ toma el valor de 1 si el pixel en la posición (x,y) en la correspondiente imagen binaria Ib(x,y) tiene el valor de 1 y tiene por lo menos un vecino con valor de cero. En el caso contrario $\Delta(x, y)$ toma el valor de cero. Los límites resultantes son de un píxel de ancho.
El nivel de gris y tamaño (número de píxeles) complementa la información extraída de cada imagen binaria sin aumentar significativamente el tiempo computacional. Por lo tanto, la dimensión del vector de características de SFTA corresponde al número de imágenes binarias obtenidas por TTBD multiplicado por tres (dimensión fractal, nivel de gris medio y área).

En la figura 6.13 se muestran las dos etapas del método de extracción de características sfta, primero se descompone en imágenes binarias y luego se genera el vector de características.

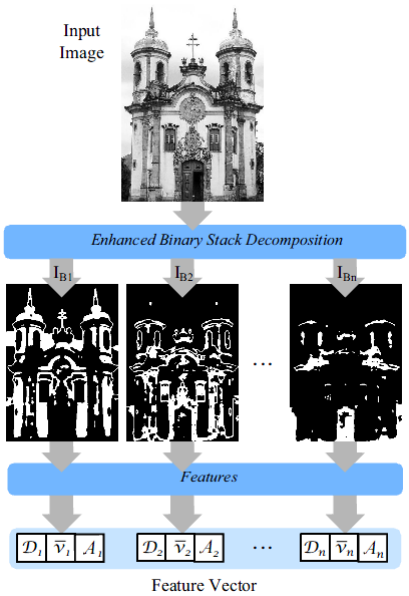

Figura 6.13: Etapas de la técnica sfta de una imagen en escala de grises

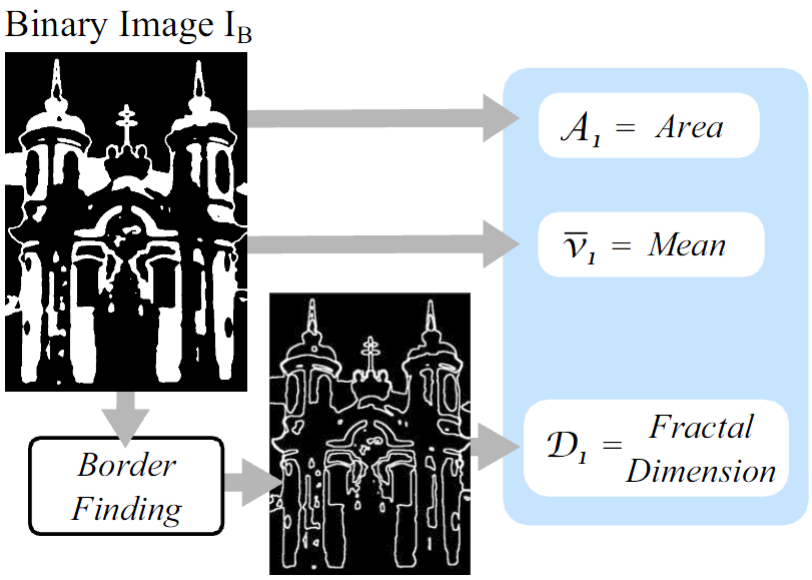

Figura 6.14: Extracción de características de imagen binaria

6.2.4. Métodos de clasificación

Actualmente las técnicas de aprendizaje de máquina pueden aplicarse para resolver problemas en diversos campos del conocimiento. En el ámbito médico la toma de decisiones es un tema crítico ya que un diagnóstico erróneo puede provocar complicaciones para el paciente o episodios de estrés innecesarios. En el caso de las patologías relacionadas con cáncer de mama, el aprendizaje de máquina y el reconocimiento de patrones son herramientas fundamentales para el análisis de las características de los hallazgos presentes en las mamografías.

Uno de los clasificadores más utilizado para el reconocimiento de patrones es el SVM, capaz de predecir la clase de una nueva muestra, a partir de un conjunto de puntos pertenecientes a dos clases. La SVM busca un hiperplano que separe de ”forma óptima” una clase de otra, buscando que éste tenga la máxima distancia con los puntos que están cerca de el mismo. De esta forma los puntos etiquetados estarán a un lado del hiperplano y los otros al otro lado del mismo.

Una SVM [34] primero mapea los puntos de entrada a un espacio de características de una dimensión mayor (i.e.: si los puntos de entrada están en $\Re^2$ entonces son mapeados por la SVM a $\Re^3$ y encuentra un hyperplano que los separe y maximice el margen entre las clases en este espacio como se aprecia en la figura 6.15.

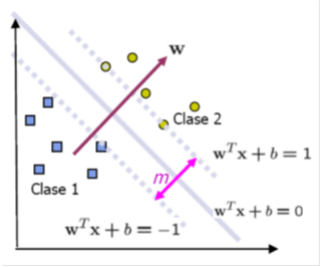

Figura 6.15: Frontera de decisión para la separación de dos clases utilizando SVM

Maximizar el margen m es un problema de programación cuadrática (QP) y puede ser resuelto por su problema dual introduciendo multiplicadores de Lagrange. Sin ningún conocimiento del mapeo, la SVM encuentra el hyperplano óptimo utilizando el producto punto con funciones en el espacio de características que son llamadas *kernels*. La solución del hyperplano óptimo puede ser escrita como la combinación de unos pocos puntos de entrada que son llamados vectores de soporte.

Caso Linealmente Separable

Si se tiene un conjunto S de puntos etiquetados para entrenamiento como se aprecia en la figura 6.16.

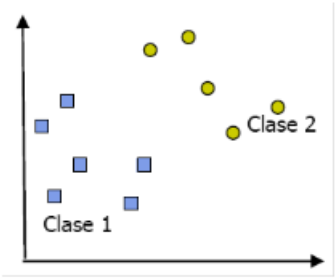

Figura 6.16: Caso linealmente separable

Cada punto de entrenamiento $x_i \in \Re^N$ pertenece a alguna de dos clases y se le ha dado una etiqueta $y_i \in \{1, 1\}$ para $i = 1, ...l$. En la mayoría de los casos, la búsqueda de un hyperplano adecuado en un espacio de entrada es demasiado restrictiva para ser de uso práctico. Una solución a esta situación es mapear el espacio de entrada en un espacio de características de una dimensión mayor y buscar el hyperplano óptimo allí. Sea $z = \phi(x)$ la notación del correspondiente vector en el espacio de características con un mapeo ϕ de $\Re^N$ a un espacio de características Z. Se desea encontrar el hyperplano notado en 6.17:

$$w \bullet z + b = 0 \tag{6.17}$$

Definido por el par (w,b), tal que se pueda separar el punto xi de acuerdo a la función mostrada en 6.18:

$$f(x_i) = sign(w \bullet z_i + b) = \begin{cases} 1\ y_i = 1 \\ -1\ y_i = 1 \end{cases} \tag{6.18}$$

Donde $w \in Z$ y $b \in \Re$. Más precisamente, se dice que el conjunto S es linealmente separable si existe (w,b) tal que las inecuaciones en 6.19, sean válidas para todos los elementos del conjunto S.

$$\begin{cases} (w \bullet z_i + b) \geqslant 1,\ y_i = 1 \\ (w \bullet z_i + b) \leqslant -1,\ y_i = 1 \end{cases} \text{ i = 1, ...l} \tag{6.19}$$

Para el caso linealmente separable de S, se puede encontrar un único hyperplano óptimo, para el cual, el margen entre las proyecciones de los puntos de entrenamiento de dos diferentes clases es maximizado.

Caso No Linealmente Separable

Si el conjunto S no es linealmente separable, se deben ser permitir violaciones a la clasificación en la formulación de la SVM.

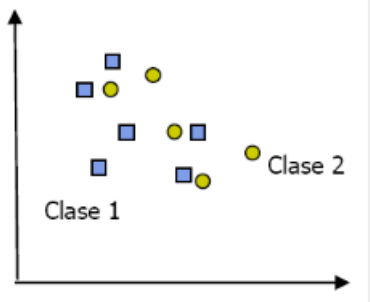

Figura 6.17: Caso linealmente separable

Para tratar con datos que no son linealmente separables, el análisis previo puede ser generalizado introduciendo algunas variables no-negativas $\xi \succ 0$ de tal modo que 6.19 es modificado a:

$$y_i(w \bullet z_i + b) \geqslant 1 - \xi \tag{6.20}$$

Los $\xi_i \neq 0$ en 6.20 son aquellos para los cuales el punto xi no satisface 6.19. Entonces el término $\sum_{i=1}^{l} \xi_i$ puede ser tomado como algún tipo de medida del error en la clasificación. El problema del hyperplano óptimo es entonces redefinido como la solución al problema:

$$\begin{aligned} &\text{mín} \left\{ \frac{1}{2} w \bullet w + c \sum_{i=1}^{l} \xi_i \right\} \\ &y_i(w \bullet z_i + b) \geqslant 1 - \xi, \text{ i = 1, ...l} \\ &\xi_i \geqslant 0, \text{ i = 1, ...l} \end{aligned} \tag{6.21}$$

Donde C es una constante. El parámetro C puede ser definido como un parámetro de regularización. Este es el único parámetro libre de ser ajustado en la formulación de la SVM. El ajuste de éste parámetro puede hacer un balance entre la maximización del margen y la violación a la clasificación.

Uso del Kernel para el caso no linealmente separable

Como no se tiene ningún conocimiento de ϕ, existe una buena propiedad de la SVM, en la cual no es necesario conocer la función ϕ. Para ello solo se necesita una función K(• , •) llamada *Kernel* que calcule el producto punto de los puntos de entrada en el espacio de características Z como se muestra en 6.22.

$$\mathrm{z}_i \bullet z_i = \varphi(x_i) \bullet \varphi(x_j) = K(x_i, x_j) \tag{6.22}$$

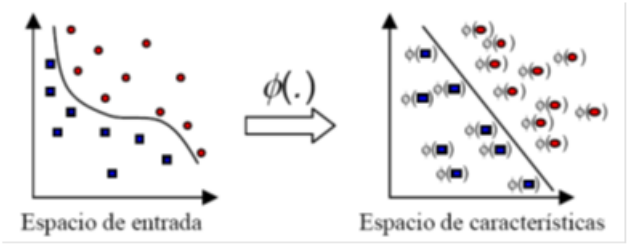

Figura 6.18: Idea del uso de un Kernel para la transformación del espacio de los datos

Idea del uso de un Kernel para la transformación del espacio de los datos. Las Funciones que satisfacen el teorema de Mercer pueden ser usadas como productos punto y por ende pueden ser usadas como kernels. Para construir un clasificador SVM se puede usar el Kernel polinomial de grado d como se muestra en 6.23.

$$K(x_i, x_j) = (1 + x_i \bullet x_j)^d \tag{6.23}$$

Entonces el hyperplano no lineal de separación puede ser encontrado como la solución de:

$$\begin{aligned} &Mx\ w(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ &\sum_{j=1}^{l} \alpha_i \alpha_j = 0\ 0 \leqslant \alpha_i \leqslant C, \mathrm{i} = 1, ...\mathrm{l} \end{aligned} \tag{6.24}$$

Así la función de decisión será:

$$\mathrm{f(x)} = \mathrm{sign}((w \bullet z + b) = sign\left(\sum_{i=1}^{l} \alpha_i y_i K(x_i, x_j) + b\right) \tag{6.25}$$

7. Materiales

7.1. Base de datos Mini-Mias

Para el desarrollo de este trabajo se utiliza una versión reducida de la base de datos MIAS *(The Mammographic Image Analysis Society)*, la cual es ampliamente usada en el procesamiento de imágenes mamográficas, ésta contiene 322 imágenes digitalizadas; 204 normales y 118 con algún hallazgo, de las cuales 66 son benignas y 52 malignas; la resolución de las imágenes es de 200 micrómetros por pixel y 8 bits por pixel, de modo que el tamaño es de 1024 x 1024 [35]. A continuación se muestra una imagen de esta base de datos.

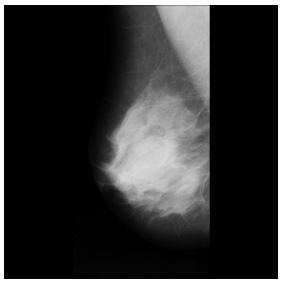

Figura 7.1: Mamografía Normal

La base de datos proporciona un archivo txt con información detallada de cada una de las imágenes, organizada en 7 columnas, a continuación se muestra la información relacionada en cada una de ellas.

1. Columna 1: Número de referencia de la imagen en la base de datos MIAS.
2. Columna 2: Caracter que corresponde al tipo de tejido de la mama.
   - F *(Fatty)*, tejido graso.
   - G *(Fatty-glandular)*, tejido mixto, graso-glandular.
   - D *(Dense-glandular)*, tejido glandular denso.
3. Columna 3: Clase de anormalidad presentada en la imagen.
   - CALC (*Calcification*), calcificación.
   - CIRC (*Well-defined/circumscribed masses)*, masa bien definida o circunscrita.
   - SPIC *(Spiculated masses)*, masas espiculadas.
   - MISC (*Other, ill-defined masses)*, otros o masas mal definidas
   - ARCH *(Architectural distortion)*, distorsión de arquitectura.

- ASYM *(Asymmetry)*, asimetría.
- NORM Normal.

4. Columna 4: Severidad de la anormalidad.
   - B (Benign), benigna.
   - M (Malignant), maligna.
5. Columnas 5 y 6: coordenadas x,y del centro de la anormalidad.
6. Columna 7: Radio aproximado en píxeles que encierra la anormalidad.

Se debe tener en cuenta para trabajar con esta base de datos: la lista está ordenada en pares de imágenes que corresponden al estudio de un mismo paciente, el número par corresponde a la proyección de la mama izquierda y el número impar a la mama derecha.

Cuando en la imagen se presenten calcificaciones, las coordenadas del centro (columnas 5 y 6) y el radio que encierra a la anormalidad (columna 7), se aplican a las calcificaciones agrupadas.

El origen del sistema de coordenadas está ubicado a partir de la esquina inferior izquierda.

7.2. Base de datos de registros mamográficos del eje cafetero

La construcción de la base de datos se realiza con imágenes proporcionadas por la Fundación Alejandro Londoño FAL de la ciudad de Armenia, Quindío, que corresponden a estudios de mamografías de pacientes del eje cafetero, sexo femenino, con una edad comprendida entre los 18 y los 80 aproximadamente.

La mayoría de las pacientes se realizaron el estudio de mamografía como cita de control, cumpliendo con la sugerencia de la liga Nacional contra el Cáncer [36] que indica que las mujeres mayores de 40 años deben realizarse mamografías cada uno o dos años con el fin de encontrar signos precoces de cáncer de mama. Solo en algunos casos, las pacientes se realizaron el estudio por molestias en sus mamas.

7.2.1. Autorización

Para el préstamo por parte de la FAL y para la utilización de las imágenes en este trabajo, fue necesario contar con la autorización escrita (consentimiento informado) de cada paciente. Para obtener los consentimientos se asistió varios días a la clínica,sin embargo, la mayoría de ellos fueron obtenidos mediante visitas autorizadas por las pacientes al lugar de residencia.

En la figura 7.2 se muestra una imagen del consentimiento informado.

En el Anexo 1 se incluye una carta del Director de la Fundación Alejandro Londoño indicando su colaboración en el proceso de análisis y préstamo de imágenes de las pacientes que cumplieran con el consentimiento informado. En el Anexo 2 se incluyen los formatos de autorización diligenciados por las pacientes.

Armenia, _____

Yo _____ identificada con C.C. _____, autorizo a la Fundación Alejandro Londoño, a las Universidades Tecnológica de Pereira y del Quindío, para utilizar las imágenes mamográficas que serán obtenidas de mi estudio en proyectos de investigación con fines académicos, en ellos no será utilizada mi información personal.

_____

Firma

Figura 7.2: Consentimiento informado

7.2.2. Adquisición de las imágenes y almacenamiento

La mama es una región anatómica que requiere de técnicas radiográfica altamente especializada para su análisis, pues los tejidos que la componen (glandular, conjuntivo, epitelial, graso, etc.) presentan pocas diferencias de absorción fotoeléctrica al haz de radiación; y el resto de las estructuras mamarias, como vasos sanguíneos o conductos galactóforos, son de muy pequeño tamaño [37].

La adquisición de las imágenes se realiza en la FAL con el Mamógrafo Mammo Diagnos UC de marca Philips. En la figura 7.3 se muestra el mamógrafo.

Los registros de mamografía fueron almacenados en un computador con la ayuda del digitalizador Agfa CR 35-X, el cual cuenta con placas y casetes específicos para cada aplicación [38]. Para el caso de las mamografías se utiliza el chasis Agfa CRMM 3.0 mammo. En la figura 7.4 se muestra el digitalizador.

Como resultado de cada estudio se obtienen 4 imágenes en formato DICOM (*Digital Imaging and Communication in Medicine*) correspondientes a las vistas CC y MLO de cada mama con las siguientes características:

- Tipo de color: Escala de grises
- Bits de profundidad: 12

Figura 7.3: Mamográfo Mammo Diagnos UC

Figura 7.4: Digitalizador CR 35-X

- Tamaño de la imagen: 4640 x 3560 píxeles

La base de datos de registros mamográficos del eje cafetero cuenta con 139 estudios de mamografía. Dos estudios están compuestos por 2 imágenes, el resto de ellos (137 casos) contienen 4 imágenes, para un total de 552 registros mamográficos.

7.3. Caja de herramientas

En esta sección se discute alrededor de las librerías y toolbox utilizados para el desarrollo del proyecto.

7.3.1. Toolbox de Procesamiento de imágenes

El toolbox de procesamiento de imágenes y videos de *MathWorks MATLAB* agrupa gran cantidad de algoritmos para el procesamiento de archivos digitales de imágenes, es compatible con varios tipos de archivo conocidos, incluido el formato DICOM *(Digital Imaging and Communication in Medicine)* internacionalmente conocido para el intercambio de archivos médicos.
Dentro de las características más importantes de la herramienta se encuentan:

- Realce de imágenes y filtrado.
- Análsis de imágenes y segmentación.
- Morfología matemática y extracción de características.
- Transformaciones geométricas y registro de imágenes.
- Visualizador de imágenes y compatibilidad con archivos DICOM enditemize

7.3.2. Librerías para la caracterización de análisis de textura

Libreria LBP

La libreria LBP (Local Binay Pattern) es una herramienta para la extracción de características cuyo enfoque se basa en el análisis de textura describiéndola en términos microprimitivas (textones) y sus reglas estadísticas. Los archivos .m [39] se instalan en la carpeta de toolbox de Matlab.

Libreria SFTA

La libreria SFTA *(Segmentation-based Fractal Texture Analysis)* es una herramienta para la extracción de características cuyo enfoque se basa en el análisis de textura describiéndola en términos de características fractales de un conjunto de imágenes binarias resultantes luego de procesar una imagen original en escala de grises. [40] que se instalan en la carpeta de toolbox de Matlab.

7.3.3. PRTools para MATLAB

PRTools es una caja de herramientas de reconocimiento de patrones para matlab, con más de 300 rutinas para tareas tradicionales de reconocimiento de patrones. PRTools incluye procedimientos para generación de datos, entrenamiento de clasificadores, combinación de clasificadores, selección de características, extracción lineal y no lineal de características, estimación de densidad al igual que analisis evaluación y visualización de clusters [41].

8. Metodología

8.1. Base de datos de registros mamográficos del eje cafetero

8.1.1. Epidemiología de cáncer de mama en el Quindío

Según datos reportados hasta el año 2013 por el departamento de salud de la cuidad de Armenia, Quindío-Colombia, existen $162{,}935$ personas adscritas al régimen de Salud. De este total de personas, existen aproximadamente $23{,}933$ personas de género femenino con edad mayor o igual a 40 años, para las cuales como método preventivo se inicia el estudio de mamografía anual. Siendo la población de estudio relacionada en este trabajo de 139 pacientes, el índice muestral corresponde al $5{,}8\,\%$ de la población estudio. Esto evidencia un buen elemento muestral con el cual realizar un estudio epidemiológico (hallazgos de patologías en imágenes de mamografía).

8.1.2. Análisis y validación médica de las imágenes

En esta etapa, se analiza con el acompañamiento de un especialista en radiología de la FAL, la información contenida en el reporte médico escrito que se entrega a las pacientes luego de un primer análisis de las imágenes. En el Anexo 3 se encuentran los reportes de cada estudio.

Se realiza el análisis de las 552 imágenes de la base de datos del eje cafetero que corresponden a 139 estudios de mamografía, en el cual se valida información relacionada con los antecedentes de cáncer de la pacientes, el tipo de tejido, los hallazgos detectados, la clasificación Bi-Rads y la sugerencia de realizar otros estudios o exámenes complementarios. Los conceptos médicos aquí tratados se encuentran explicados en 6.1.

Antecedentes de cáncer:

En esta parte, se revisan los reportes de las mamografías de años anteriores para determinar si existen antecedentes familiares de cáncer o si la paciente ha sido intervenida quirúgicamente por la misma causa.

Tipo de tejido:

Según la composición del tejido predominante en la mama observada en las mamografías de las pacientes, el especialista indica el tipo de tejido al que corresponde: graso, mixto (graso y glandular) o glandular.

En la figura 8.1 se muestran los tres tipos de tejido en imágenes de pacientes de la FAL.

Figura 8.1: Tipos de tejido: graso, mixto y glandular

Hallazgos detectados:

Los hallazgos que pueden presentarse en una mamografía se explican en 6.1, para el caso de las imágenes de la base de datos del eje cafetero se manejarán los siguientes:

- Normal
- Masa circunscrita
- Masa mal definida
- Masa espiculada
- Calcificación
- Asimetría
- Distorsión de arquitectura.

Los hallazgos fueron encerrados mediante cuadrantes en las imágenes y almacenadas nuevamente en formato jpg, esto con el objetivo de hacer el etiquetado de las imágenes según el criterio médico.

Las figuras 8.2 y 8.3 muestran los hallazgos que pueden presentarse en pacientes del eje cafetero.

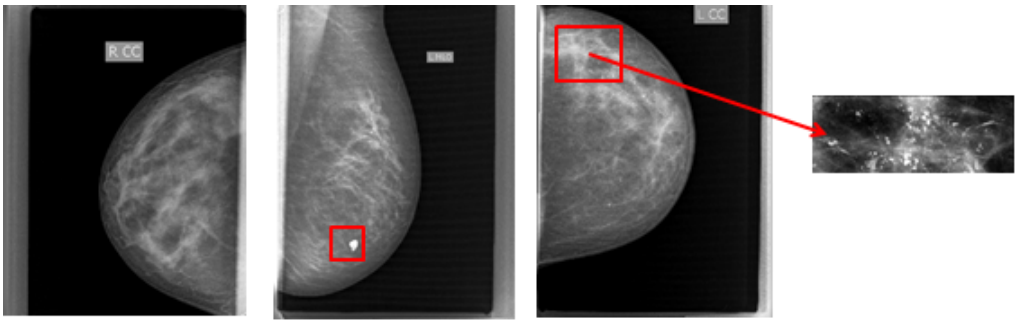

Figura 8.2: Mamografía normal, con calcificación benigna y con mcals calcificaciones

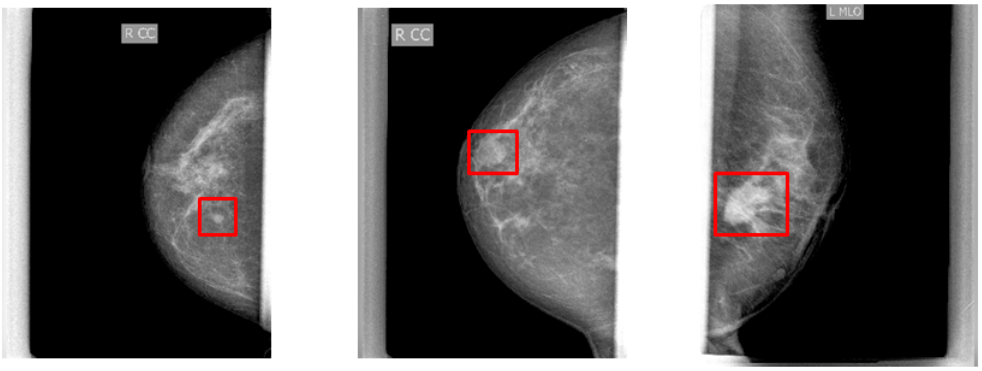

Figura 8.3: Mamografías con masa circunscrita, mal definida y espiculada

Clasificación Bi-Rads:

Luego de analizar con detenimiento las características de los hallazgos detectados y su posible severidad se asigna al estudio la categoría según la clasificación Bi-rads, la cual ha sido explicada en 6.1.

Exámenes complementarios:

En algunos casos donde exista una posible anomalía que no sea visible o no esté definida con claridad, el especialista asigna el Bi-Rads 0, y suguiere comparar con mamografías anteriores o realizar exámenes adicionales como: ecografía o vistas especiales de mamografía, para reclasificar según la categoría Bi-Rads.

Cuando el estudio de la paciente sea clasificado como Bi-Rads 3, donde los hallazgos son probablemente benignos se sugiere hacer seguimiento a los 6 meses mediante exámenes de mamografía y/o ecografía.

Si el estudio es clasificado como Bi-rads 4 o 5, se sugiere realizar biopsia, con esta prueba determinante se conocerá de manera definitiva la severidad del hallazgo.

8.1.3. Etiquetado

Siguiendo como referente la estructura de la base de datos internacional MIAS explicada en 7.1, la información suministrada por el especialista y la confirmación del resultado de las biopsias en los casos que lo necesitaron, se realiza el etiquetado de las imágenes para la creación de la base de datos de registros mamógráficos del eje cafetero. En el Anexo 4, se encuentra la información de cada estudio, así como se describe a continuación:

1. **Código asignado a cada imagen:** Inicialmente, se modifica el nombre de los archivos del estudio de cada paciente que en origen corresponde al ID asignado en la FAL y se asigna un nombre de archivo que inicia con la letra C (caso), seguido de número consecutivo de tres cifras asignado a cada estudio en la base de datos del eje cafetero, seguido del tipo de vista (CC o MLO) y el seno analizado (D=derecho, I=izquierdo). A continuación se muestran las etiquetas asociadas al caso 001:

   - $C\_001\_CC\_I$ corresponde a la proyección cráneo-caudal de la mama izquierda.
   - $C\_001\_MLO\_I$ corresponde a la proyección medio lateral oblicua de la mama izquierda.
   - $C\_001\_CC\_D$ corresponde a la proyección cráneo-caudal de la mama derecha.
   - $C\_001\_MLO\_D$ corresponde a la proyección medio lateral oblicua de la mama derecha.

2. **Código de cada estudio:** código de identificación asignado a cada estudio, inicia con la palabra Caso, seguido de tres números, los cuales inician para el primer estudio con el 001 y continúan así sucesivamente hasta el 139 que corresponde al total de estudios de mamografías. Esta etiqueta es utilizada en la carpeta que almacena las imágenes en formato digital.

3. **Clasificación Bi-Rads:** número de la categoría según el Bi-Rads, explicado en 6.1.5.

4. **Tipo de tejido:** Las etiquetas asociadas al tipo de tejido corresponden con la composición del mismo. Se diferenciará entre tres tipos de tejido: graso, mixto y glandular.

5. **Clase de anormalidad o hallazgo**: las etiquetas asociadas al tipo de hallazgo corresponden a las lesiones analizadas con el especialista: Normal, Masa circunscrita, Masa mal definida, Masa espiculada, Calcificación, Asimetría, Distorsión de arquitectura.

6. **Tipo de severidad:** Según la severidad del hallazgo se han asignado las letras B y M según corresponda a la lesión Benigna o maligna. Para el caso de los estudios clasificados como Bi-Rads 0, se asocia una etiqueta con las letras SC, que corresponde a sin clasificar, dado que faltan estudios complementarios para reasignar el Bi-rads.

7. **Ubicación de los hallazgos:** coordenadas en x y en y (en píxeles) del centro de la anormalidad.

8. **Radio: corresponde al radio en píxeles que encierra la anormalidad**

Para determinar el centro de la anormalidad en píxeles y el radio que la encierra, se realizó un algoritmo de procesamiento que utiliza las imágenes resultantes del proceso de análisis por parte del especialista. En la figura 8.4 se muestra dicho procedimiento.

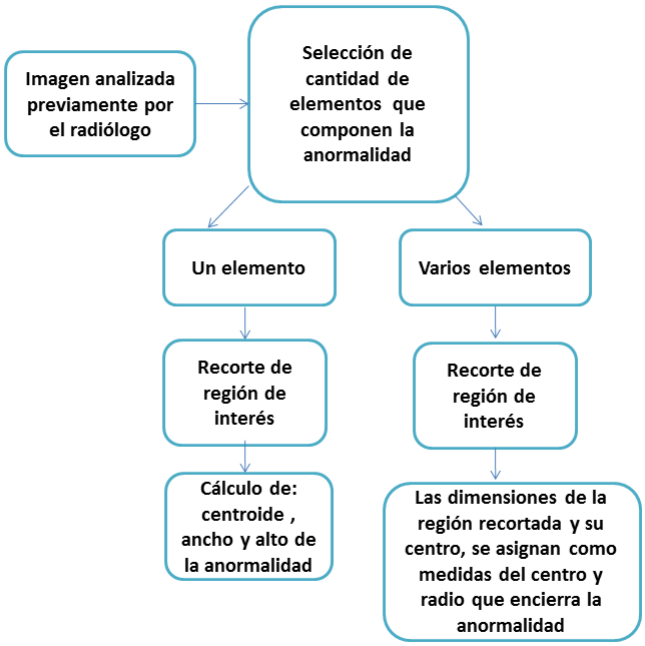

Figura 8.4: Cálculo del centro (x,y) y radio que encierra una anormalidad

El procedimiento de asignación de etiquetas a las imágenes de mamografía de la base de datos de la FAL se muestra en la figura 8.5.

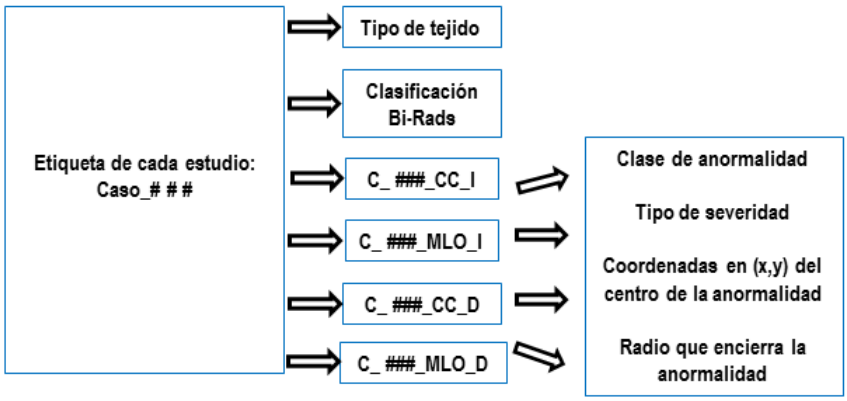

Figura 8.5: Etiquetado de imágenes de la base de datos del eje cafetero

8.1.4. Observaciones de la base de datos de registros mamográficos del eje cafetero

1. El sistema de coordenadas para la ubicación del centro de un hallazgo, se inicia en la parte superior izquierda.

2. Las imágenes categorizadas en Bi-Rads 2 que no contengan masas, calcificaciones, distorsión de arquitectura o asimetría, pueden considerarse como normales. Estas han sido incluidas en Bi-Rads 2, debido a que contienen generalmente ateromatosis, que es un rasgo normal para la edad de las pacientes.

3. Existen estudios de pacientes clasificados mediante Bi-Rads 2, 3, 4 o 5, en los cuales solo una mama contiene uno de los hallazgos mencionados anteriormente, la otra mama no contiene ninguno de ellos, en este caso se ha colocado la frase "No se observa anormalidad en esta imagen". Dichas imágenes pueden considerarse normales por no contener hallazgos como masas, calcificaciones, distorsión de arquitectura o asimetría.

4. La etiqueta "No posee imagen en esta proyección", corresponde a pacientes que no poseen una de sus mamas, por lo tanto el estudio contiene dos imágenes de mamografía.

5. Los estudios clasificados en Bi-Rads 4 y 5, que necesitaron de estudios complementarios para conocer la severidad de los hallazgos y obtuvieron resultados negativos, en la base de datos no fueron recategorizados en otro Bi-rads, pero la severidad fue etiquetada como Benigna.

8.1.5. Resultados de exámenes complementarios

Los estudios categorizados en Bi-Rads 4 y 5, en la mayoría de los casos requieren de una prueba invasiva (biopsia) o de un estudio complementario como ecografía para comprobar la severidad de los hallazgos detectados. A los estudios de mamografía de FAL Bi-Rads 4 y 5 se les realizó seguimiento

para conocer el resultado de los exámenes complementarios. En la tabla 8.1 se describen estos resultados.

Tabla 8.1: Resultados exámenes complementarios

Exámenes complementarios				
Bi-Rads	No.ecografía	No.Biopsia	Resultado negativo	Resultado positivo
4	2	8	9	1
5	0	11	1	10

8.2. Segmentación de zona mamaria

Inicialmente se realiza la binarización de la imagen original para utilizar operaciones posteriores que requieren dicho formato, además buscando eliminar elementos que no proporcionan información relevante en la caracterización de mcals como la etiqueta y bordes. Se binariza la imagen utilizando como valor de umbral 0.04. En la siguiente figura se muestra la imagen original y luego del proceso de binarización.

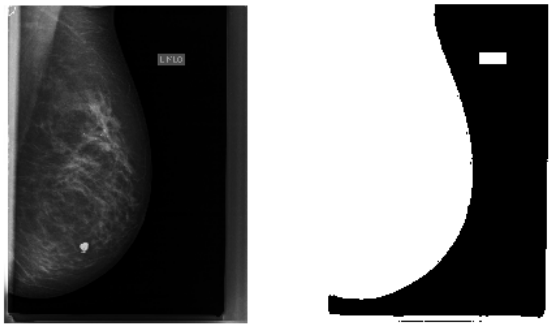

Figura 8.6: Imagen original y binarizada

Luego se aplica la operación morfológica erosión con el elemento estructurante line de una longitud de 35 píxeles y una inclinación de 0 grados, mejorando el contorno de la mama y eliminando pequeños objetos del fondo innecesarios en la detección de calcificaciones.

Por último, mediante el etiquetado de regiones conexas y de contorno, se determinó y seleccionó el objeto de mayor área que correspondía a la mama.

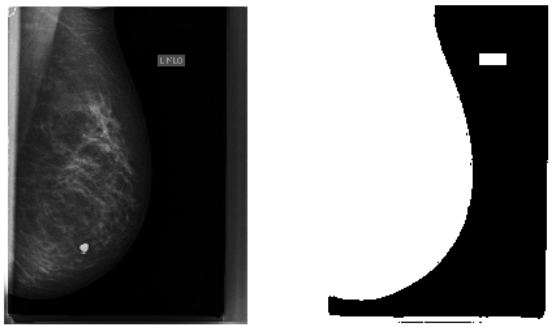

Figura 8.7: Imagen original y binarizada

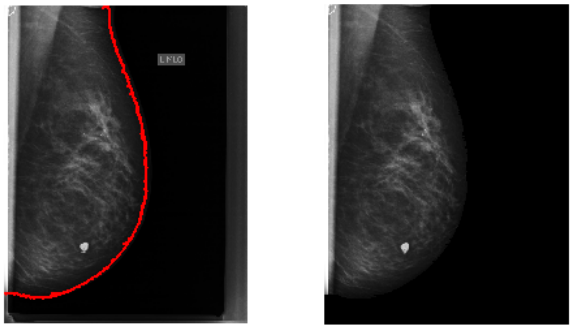

Figura 8.8: Segmentación de zona mamaria

En la figura 8.9 se muestra el proceso completo de la segmentación de la zona mamaria en una imagen de mamografía.

Figura 8.9: Proceso de segmentación de zona mamaria en una mamografía

Para la segmentación de la mama en las imágenes de la base de datos MINI-MIAS se utilizan los mismos algoritmos de procesamiento, los cuales funcionan adecuadamente logrando segmentar la zona mamaria de estas imágenes. En la figura 8.10 se muestra una imagen de la base de datos MINI-MIAS original y segmentada.

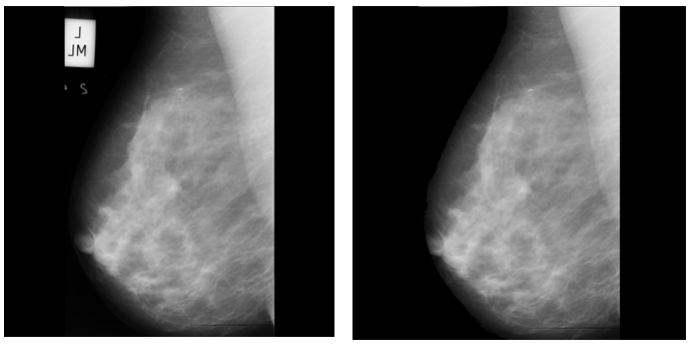

Figura 8.10: Proceso de segmentación en una mamografía de la base de datos MINI-MIAS

8.3. Módulo de Filtrado

Este módulo se utiliza para disminuir el ruido que es introducido durante la adquisición y digitalización de las imágenes el cual puede interferir con las características de interés e información propia de una mamografía [5]. Como el ruido se presenta con pequeños puntos en la imagen, puede entorpecer etapas posteriores del procesamiento y el diagnóstico dado por el especialista. Por lo tanto, se utilizan técnicas de filtrado espacial que ayuden a reducir el ruido presente en las mamografías.

Luego de la segmentación de la zona mamaria se realiza la reducción de ruido mediante los filtros de media, gaussiano, mediana y unsharp, modificando sus parámetros para obtener los mejores resultados en cuanto a disminución de ruido se refiere.

Estas técnicas se aplican a las imágenes de las bases de datos MINI-MIAS y la de registros mamográficos del eje cafetero teniendo en cuenta la proyección (MLO y CC) y la mama a la que corresponde. La base de datos MINI-MIAS no contiene imágenes de las proyecciones Cráneo - Caudal por lo tanto,

los métodos de filtrado se aplican a las proyecciones medio lateral oblicua.

8.3.1. Filtros implementados para la reducción de ruido

Los parámetros de los filtros espaciales implementados se modifican hasta obtener los mejores resultados en la etapa de reducción de ruido, a continuación se describen brevemente los parámetros que se ajustan en cada máscara de filtrado.

- Filtro de media: Inicialmente se utiliza una ventana de filtrado tamaño de 3 x 3 píxeles y se van incrementando sus dimensiones hasta obtener una ventana de análisis de 30 x 30 píxeles.

- Filtro gaussiano: Se filtran las imágenes con ventanas de 3 x 3 píxeles y desviación estándar de 0.2, incrementando sus dimensiones hasta un kernel de 30 x 30 píxeles y una desviación de 4.0.

- Filtro de mediana: Las imágenes se filtran con una ventana de tamaño de 3 x 3 píxeles, incrementando sus dimensiones hasta una de 30 x 30 píxeles.

- Filtro unsharp: Se filtran las proyecciones de la mama con una ventana de tamaño de 3 x 3 píxeles y un alpha de 0.1, incrementando las dimensiones de la ventana hasta de 30 x 30 píxeles y alpha hasta 1.0. Este parámetro alpha determina la forma del Laplaciano aplicado en el filtrado.

8.3.2. Métrica Q y SSIM

El cálculo de las métricas Q y SSIM permite determinar cuaáles de los parámetros ajustables en cada tipo de filtro son los adecuados para obtener la reducción de ruido en las mamografías.

Métrica Q *(Quality)*:

La particularidad de esta métrica es que no necesita imagen de referencia. La métrica Q se basa en la descomposición de los valores singulares de la matriz gradiente de la imagen, proporcionando una medida cuantitativa del verdadero contenido de la imagen (nitidez y contraste) ante la presencia de ruido y otras perturbaciones [6].

La métrica Q se define en el contenido de una imagen como:

$$Q = s_1 \frac{s_1 - s_2}{s_1 + s_2} \tag{8.1}$$

Donde s1 y s2, representan la energía en las direcciones de los vectores V1 y V2 que representan la orientación del campo gradiente local.

La coherencia que permite analizar las características locales de una imagen, se define como:

$$R = \frac{s_1 - s_2}{s_1 + s_2} \tag{8.2}$$

Métrica ssim (*(structural-similarity Index)*):

La métrica ssim mide la Similitud Estructural entre las señales correspondientes a dos imágenes. Si una de ellas se considera perfecta, la métrica se puede ver como una medida de la calidad de la otra [8], [7].

La métrica ssim es la combinación f(.) de la variación de luminancia l(x,y), contraste c(x,y) y estructura s(x,y). Las funciones f(.), es decir combinación de las variaciones, deben cumplir ciertas condiciones como: ser simétrica (da igual que imagen se compare con cual), estar limitada superiormente (<= 1 donde 1 indica exactamente iguales) y tener un máximo único (solo hay una forma de ser exactamente iguales).

$$SSIM(x,y) = [l(x,y)]^{\alpha}.[c(x,y)]^{\beta}.[s(x,y)]^{\gamma} \tag{8.3}$$

Donde los exponentes, alfa, beta y gamma son parámetros para ajustar la importancia relativa de los tres componentes de la variación sufrida.

En la figura 8.11 se muestran las etapas del módulo de filtrado.

Figura 8.11: Proceso de filtrado

Selección de parámetros de cada filtro:

El proceso para la selección de parámetros adecuados de cada filtro, se realiza con el cálculo de la métrica Q y se realiza de la siguiente forma:

1. Se ajustan los parámetros del filtro y se aplica a una imagen.
2. Se calcula el valor de la métrica después de filtrar cada imagen.
3. Luego de filtrar todas las imágenes, se obtiene el promedio de la métrica.

4. Se continúa modificando nuevamente los valores ajustables del filtro y se repiten los 2 y 3.
5. Por último, se analiza el mayor valor de la métrica y se identifican los parámetros del filtro utilizados cuando se obtuvo ese valor, estos corresponden a los mejores parámetros de filtrado según la métrica.

Después de la selección de los parámetros, se realiza la implementación de los filtros con éstos y se aplican a las imágenes. Se continúa con el cálculo de las métricas Q y ssim para determinar cuál es el mejor filtro en cuanto a reducción de ruido se refiere, para ello:

1. Se ajustan los parámetros adecuados del filtro.
2. Se realiza el filtrado de cada imagen y se calcula el valor de la métrica.
3. Luego de filtrar todas las imágenes, se obtiene el promedio de la métrica para cada tipo de filtro.
4. Por último, se comparan los resultados de la métrica y se identifica para cual filtro se obtuvo el mayor resultado, el cual corresponde al mejor filtro para la disminución de ruido en las imágenes.

8.4. Módulo de extracción de características

Para la correcta detección de hallazgos, especialmente de calcificaciones que puedan presentar malignidad en las mamografías es fundamental la extracción de características que permitan la identificación de éstas con la mayor precisión.

El análisis de textura es una técnica importante utilizada en segmentación, identificación de objetos o regiones de interés en una imagen y obtención de forma. Los modelos activos son un método estadístico de ajuste entre un modelo (capaz de representar forma y apariencia) y una imagen dada. [25]

La base de datos de registros del eje cafetero está conformada por diferentes anormalidades, sin embargo, el sistema de extracción de características y clasificación se utiliza para los siguientes tres conjuntos de datos de imágenes:

- Calcificaciones benignas: contiene 95 imágenes con calcificaciones macro y micro benignas. Los casos de microcalcificaciones sugestivos de malignidad fueron comprobados con biopsia y su resultado fue negativo.
- Calcificaciones malignas: contiene 22 imágenes con microcalcificaciones malignas confirmadas con biopsia.
- Sin calcificaciones: contiene 435 imágenes con hallazgos diferentes a calcificaciones e imágenes normales.

  La base de datos Mini-Mias cuenta con las siguientes imágenes:

  - Calcificaciones benignas: contiene 12 imágenes.

- ◇ Calcificaciones malignas: contiene 13 imágenes.

- ◇ Sin calcificaciones: contiene 297 imágenes con hallazgos diferentes a calcificaciones e normales.

Para la extracción de características se utilizan regiones de interés (ROI) de la zona donde se concentra el hallazgo, los métodos descritos en 6.2.3: el operador LBP *(Local Binary Pattern)* y el SFTA *(Segmentation-based Fractal Texture Analysis)* se aplican a las ROI.

8.4.1. Operador LBP

La extracción de características mediante el operador LBP se realiza mediante la selección de la cantidad de muestras P a tomar en una circunferencia de Radio R.

Para la extracción de las características de cada ROI se utilizan 8 muestras en una circunferencia de radio R = 1, el proceso inicia con la resta entre los 8 píxeles de la muestra y el píxel central.

Se continúa con la umbralización de éstos, a los valores ≥ 0 se les asigna 1, en caso contrario se les asigna 0.

Por último, se multiplican los valores del resultado de la umbralización por 2^n y se suman. Este valor proporciona mediante un patrón binario local información sobre la textura de una región.

El vector de características está formado por el histograma de los valores obtenidos anteriormente.

En las figuras 8.12, 8.13 y 8.14 se muestran los gráficos con las características correspondientes a 3 grupos de ROIS de la base de datos de registros del eje cafetero, cada uno de ellos contiene 20 cúmulos de calcificaciones malignas, 20 muestras de calcificaciones benignas y 20 regiones sin presencia de calcificaciones.

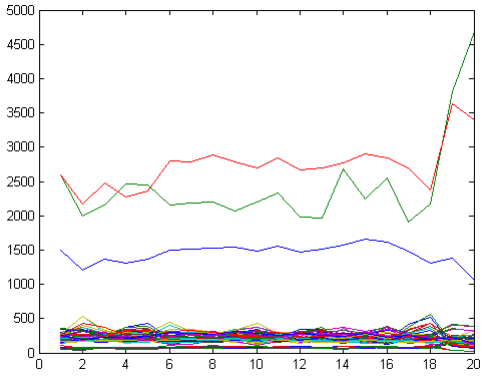

Figura 8.12: Extracción de características mediante LBP a ROIs con calcificaciones malignas

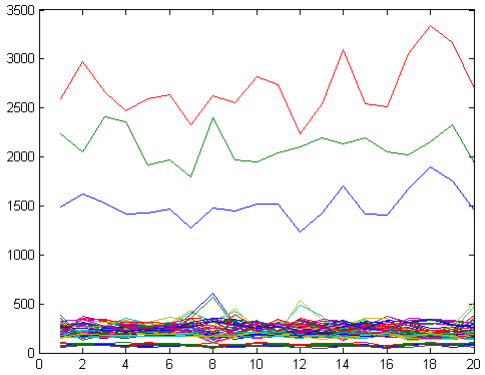

Figura 8.13: Extracción de características mediante LBP a ROIs con calcificaciones benignas

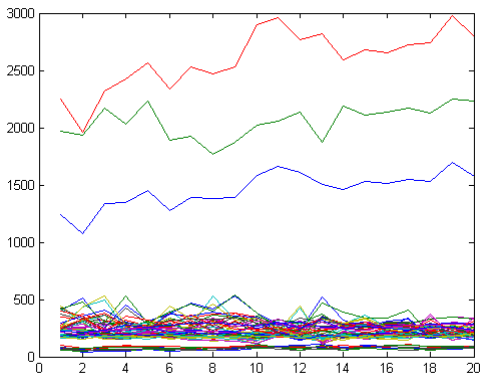

Figura 8.14: Extracción de características mediante LBP a ROIs sin calcificaciones

Para el caso de la base de datos de la MINI-MIAS, las gráficas de los vectores de características se muestran en 8.15, 8.16 y 8.17, donde cada figura corresponde a las calcificaciones

benignas, malignas e imágenes normales y con hallazgos diferentes.

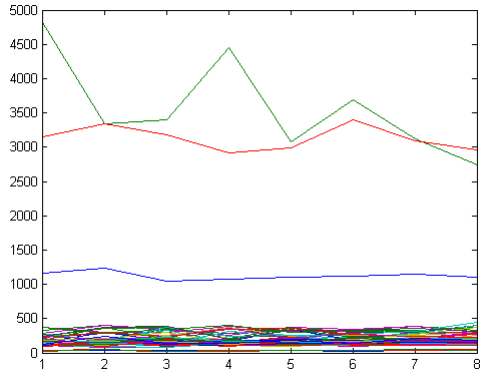

Figura 8.15: Extracción de características mediante SFTA a ROIs con calcificaciones benignas

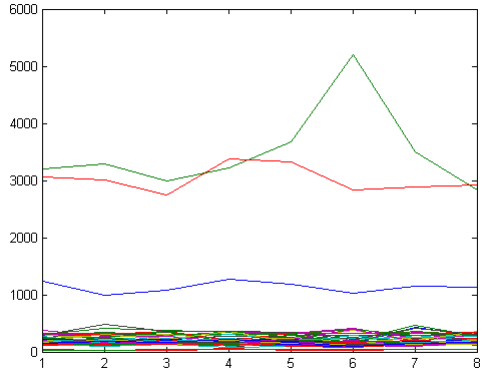

Figura 8.16: Extracción de características mediante SFTA a ROIs con calcificaciones malignas

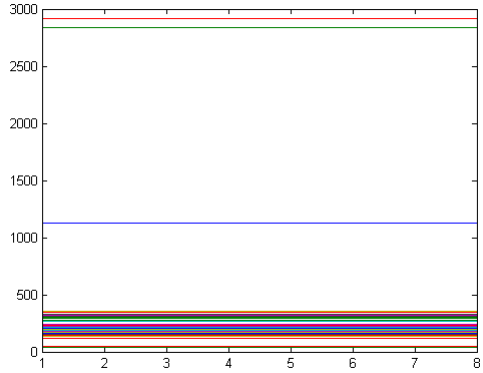

Figura 8.17: Extracción de características mediante SFTA a ROIs sin calcificaciones

8.4.2. Operador SFTA

En la extracción de características mediante SFTA debe seleccionarse el valor deseado de nt que determinará la cantidad de imágenes en las que se descompone la imagen de entrada, para este caso se utiliza nt=4, por lo tanto, se obtienen 2*nt imágenes de salida binaria.

Cada imagen de entrada se descompone en 2 x nt, en este caso 8 imágenes binarias.

Para obtener cada imagen binaria, se calcula un conjunto T de valores umbral que son obtenidos seleccionando los niveles de gris igualmente espaciados, luego se aplica la segmentación con pares de umbrales de T.

A cada imagen binaria se le extraen tres características: Área, media de niveles de gris y dimensión fractal del contorno.

Número de características imagen original = 2 x nt x (3 características) = (8 imágenes binarias) x (3 características)

Número de características imagen original = 24 características

En las figuras 8.18, 8.19 y 8.20 se muestran los gráficos con las características correspondientes a 3 grupos de ROIS, cada uno de ellos contiene 20 cúmulos de calcificaciones malignas, 20 muestras de calcificaciones benignas y 20 regiones sin presencia de calcificaciones.

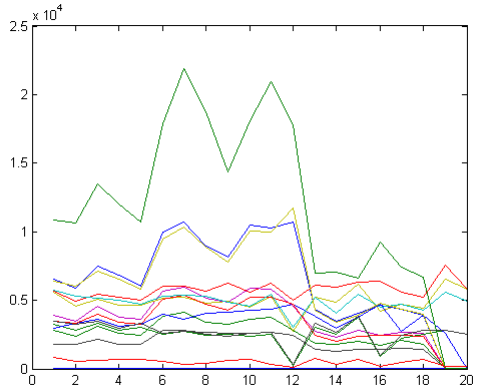

Figura 8.18: Extracción de características mediante SFTA a ROIs con calcificaciones malignas

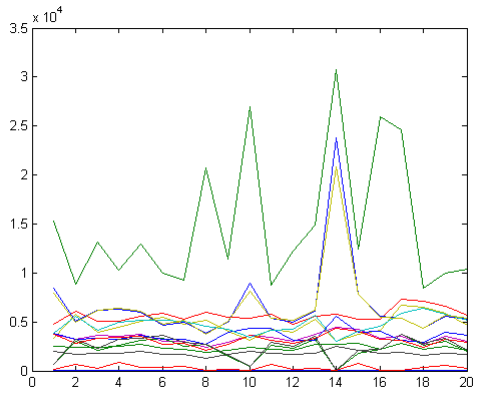

Figura 8.19: Extracción de características mediante SFTA a ROIs con calcificaciones benignas

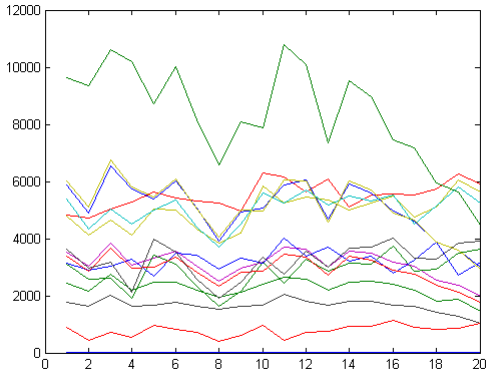

Figura 8.20: Extracción de características mediante SFTA a ROIs sin calcificaciones

Los resultados de extracción de características que corresponden a los tres grupos de datos (calcificaciones malignas, benignas e imágenes sin calcificaciones) de la base de datos MINI-MIAS, mediante la técnica SFTA se muestran respectivamente en las figuras 8.21, 8.22 y 8.23.

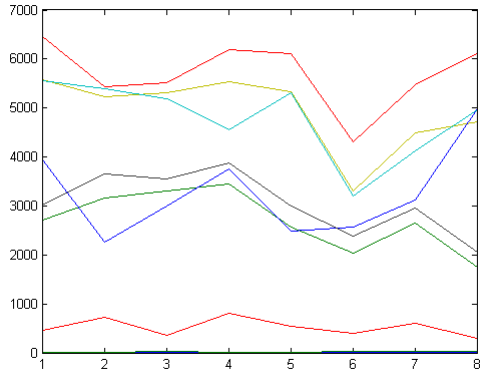

Figura 8.21: Extracción de características mediante SFTA a ROIs con calcificaciones malignas

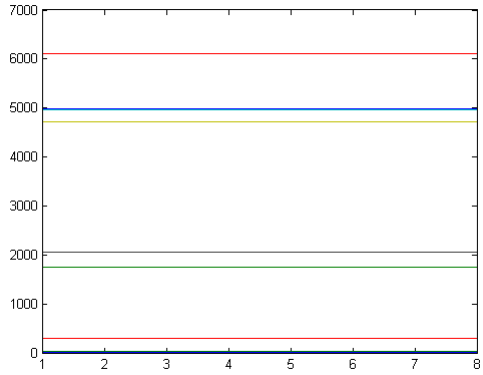

Figura 8.22: Extracción de características mediante SFTA a ROIs con calcificaciones benignas

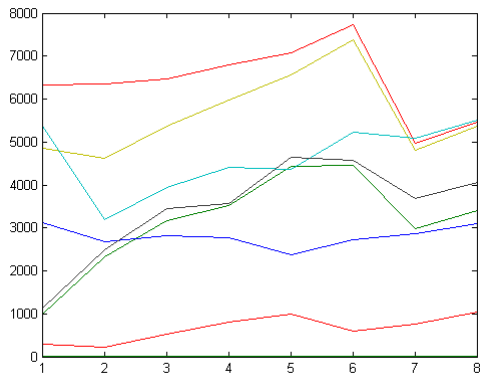

Figura 8.23: Extracción de características mediante SFTA a ROIs sin calcificaciones

8.5. Módulo de clasificación

La toma de decisiones es un tema crítico en el ámbito médico ya que un diagnóstico erróneo puede provocar complicaciones para el paciente o episodios de estrés innecesarios. En el caso de las patologías relacionadas con cáncer de mama, el aprendizaje de máquina y el reconocimiento de patrones son herramientas fundamentales para el análisis de las características de los hallazgos presentes en las mamografías.

En este trabajo se realizan pruebas en imágenes de dos bases de datos: Mini-MIAS y la que contiene registros mamográficos del eje cafetero para determinar la severidad (benigna o maligna) de las mcals detectadas.

El sistema debe clasificar entre tres conjuntos de datos:

- ◇ Calcificaciones benignas: contiene 128 imágenes con calcificaciones macro y micro benignas. Los casos de microcalcificaciones sugestivos de malignidad fueron comprobados con biopsia y su resultado fue negativo.
- ◇ Calcificaciones malignas: contiene 20 imágenes con microcalcificaciones malignas confirmadas con biopsia.
- ◇ Sin calcificaciones: contiene 404 imágenes con hallazgos diferentes a calcificaciones.

La clasificación se realiza utilizando una máquina de soporte vectorial, para la validación de la información se utiliza el esquema Hold Out validation, usando el 80 % de los datos para el entrenamiento y el 20 % para validación.

Las pruebas se realizan con los siguientes cuatro conjuntos de imágenes:

- ◇ Primer grupo de datos (calcificaciones benignas y malignas) se trabajó con 40 imágenes en total.
- ◇ Para el segundo grupo de datos (calcificaciones benignas y sin calcificaciones) se trabajó con 140 imágenes en total.
- ◇ Para el tercer grupo de datos (calcificaciones malignas y sin calcificaciones) se trabajó con 40 imágenes en total.
- ◇ Para el cuarto grupo de datos (calcificaciones benignas, malignas y sin calcificaciones) se trabajó con 60 imágenes en total.

En el siguiente diagrama se muestran las etapas para la detección de calcificaciones benignas (macro y micro), mcals malignas e identificación de imágenes sin ningún tipo de calcificación:

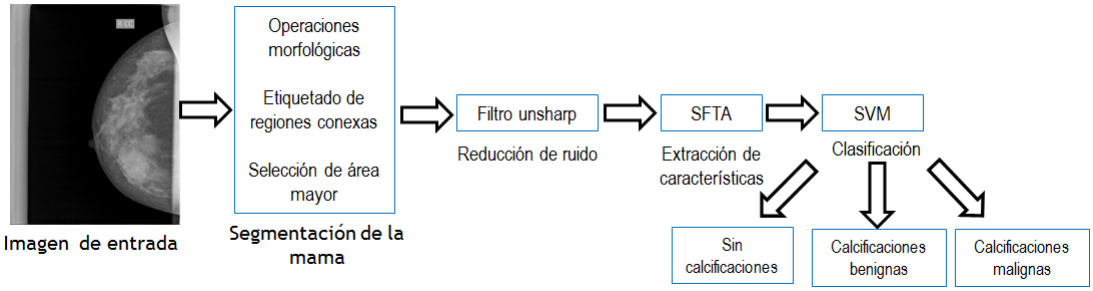

Figura 8.24: Etapas para la detección de mcals

9. Resultados

En el presente capítulo se exponen los resultados obtenidos a partir de la metodología desarrollada para la construcción de una base de datos de imágenes de mamografía que permite la identificación de microcalcificaciones. Los resultados son obtenidos al probar el sistema con dos base de datos: La versión reducida de la Mias y la base de datos de registros mamográficos del eje cafetero.

9.1. Base de datos de registros de mamografía del eje cafetero

La base de datos local cuenta con 552 imágenes que corresponden a estudios de 139 pacientes del eje cafetero, de sexo femenino y edad comprendida entre los 18 y 80 años aproximadamente.

Se obtuvieron 137 consentimientos informados. En el caso de 2 pacientes, se realizaron dos estudios en fechas diferentes, por lo que no fue necesaria una nueva autorización.

Mediante el acompañamiento de un especialista de la FAL se realizó el análisis de las 552 imágenes, validando la información suministrada en el reporte radiológico escrito asignado a cada estudio. La información analizada corresponde a los antecedentes de cáncer de las pacientes, el tipo de tejido, los hallazgos detectados, la clasificación Bi-Rads y la sugerencia de realizar otros estudios o exámenes complementarios. Además, se generó un archivo con las etiquetas asociadas a cada uno de los estudios e imágenes según lo explicado en 8.1.3

En la tabla 9.1 se relaciona la cantidad de hallazgos detectados en las imágenes de la base de datos del eje cafetero, cabe anotar que hay imágenes que presentan más de un hallazgo.

Tabla 9.1: Cantidad de hallazgos detectados en la base de datos

Hallazgo	No. de hallazgos
Normal	387
Calcificación	152
Masa circunscrita	40
Masa mal definida	30
Masa espiculada	11
Distorsión de arquitectura	2
Asimetría	14

El número de hallazgos normales detectados incluye todas las imágenes que no presentan ningún hallazgo relacionado con masas, calcificaciones, distorsión de arquitectura y asimetría.

Las imágenes que corresponden a cada categoría de Bi-Rads se muestra en la tabla 9.2.

Tabla 9.2: Cantidad de estudios según la categoría Bi-Rads.

Bi-Rads	No. estudios
0: Sin clasificar	3
1: Normal	4
2: Hallazgos benignos	87
3: Hallazgos probablemente benignos	21
4: Hallazgos con sospecha de malignidad	12
5: Hallazgos muy sospechosos de malignidad	12
Total de estudios	139

En la tabla 9.3, se muestra el número de imágenes que contienen calcificaciones benignas y malignas, mientras que en la tabla 9.4, se incluyen las calcificaciones detectadas en todas las imágenes que contienen este tipo de hallazgo.

Tabla 9.3: Número de imágenes con calcificaciones benignas y malignas

Calcificación	No. de imágenes
Benigna	95
Malignas	22

Tabla 9.4: Cantidad de calcificaciones detectadas en imágenes de la base de datos

Calcificación	No. detecciones
Benigna	130
Malignas	22

9.2. Módulo de Filtrado

La segmentación de la mama se realiza con técnicas sencillas en términos de implementación y tiempo de procesamiento pero que resultan eficientes para obtener de la mamografía la zona mamaria de interés.
Luego de la segmentación se realiza la reducción de ruido mediante los filtros de media, gaussiano, mediana y unsharp. Estas técnicas se aplicaron a las imágenes de las bases de datos MINI-MIAS y de registros mamográficos del eje cafetero.

La figura 9.1 muestra la aplicación del filtro de media a una imagen de la base de datos regional.

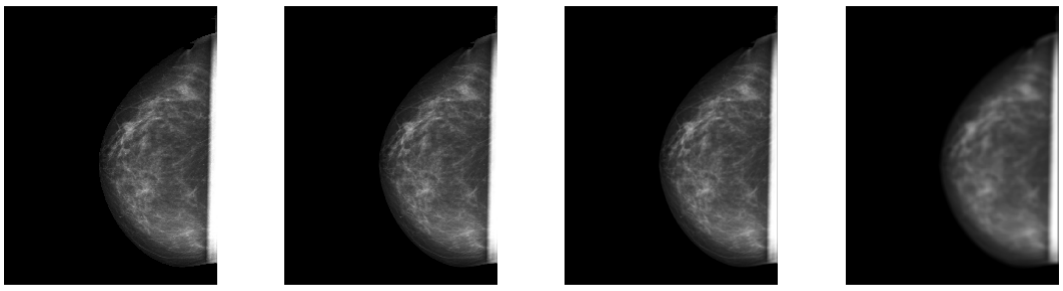

Figura 9.1: Imagen original y filtrada con máscara de media de 3 x 3, 6 x 6 y 20 x 20

El filtro gaussiano también se utiliza en la etapa del filtrado, en la figura9.2 se muestra su aplicación en una imagen.

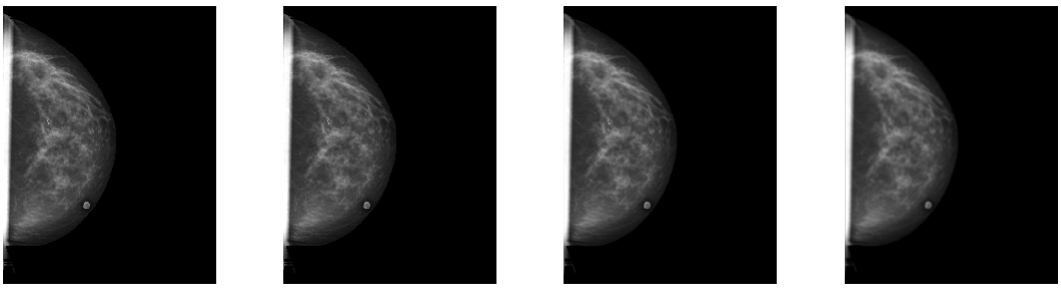

Figura 9.2: Imagen original y filtrada con máscara gaussiana de 3 x 3 $\sigma = 0{,}2$, 7 x 7 $\sigma = 1{,}5$ y 20 x 20 $\sigma = 3$

En la figura9.3, se muestra un registro de mamografía luego de aplicar el filtro de mediana.

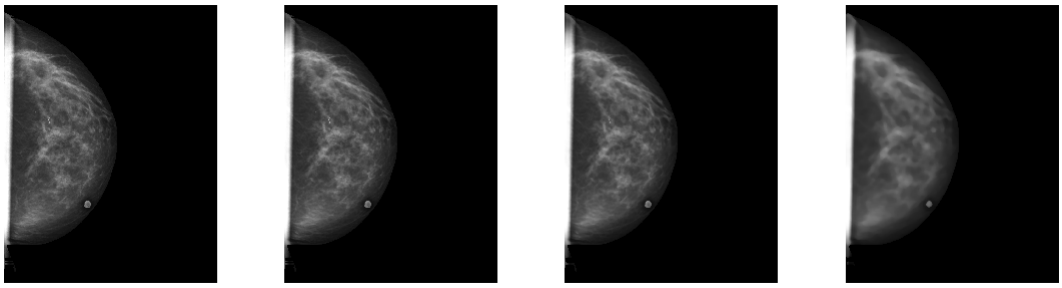

Figura 9.3: Imagen original y filtrada con máscara de mediana de 3 x 3, 7 x 7 y 20 x 20

Por último, se aplica el filtro unsharp a una imagen de prueba, en la figura9.4 se muestran los resultados.

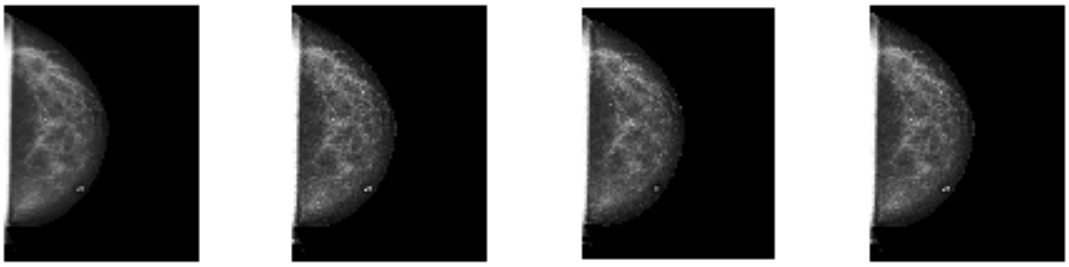

Figura 9.4: Imagen original y filtrada con máscara unsharp con alpha de 0.2, 0.7 y 1.0

A continuación se muestra la selección de parámetros para las máscaras de filtrado que se aplican a las imágenes de la base de datos del eje cafetero, para ello, se utiliza el cálculo de la métrica Q.

En la figura 9.5 se muestra el resultado de la métrica Q, aplicado a imágenes con proyecciones MLO y CC de cada mama luego de filtrarlas con la máscara de media.

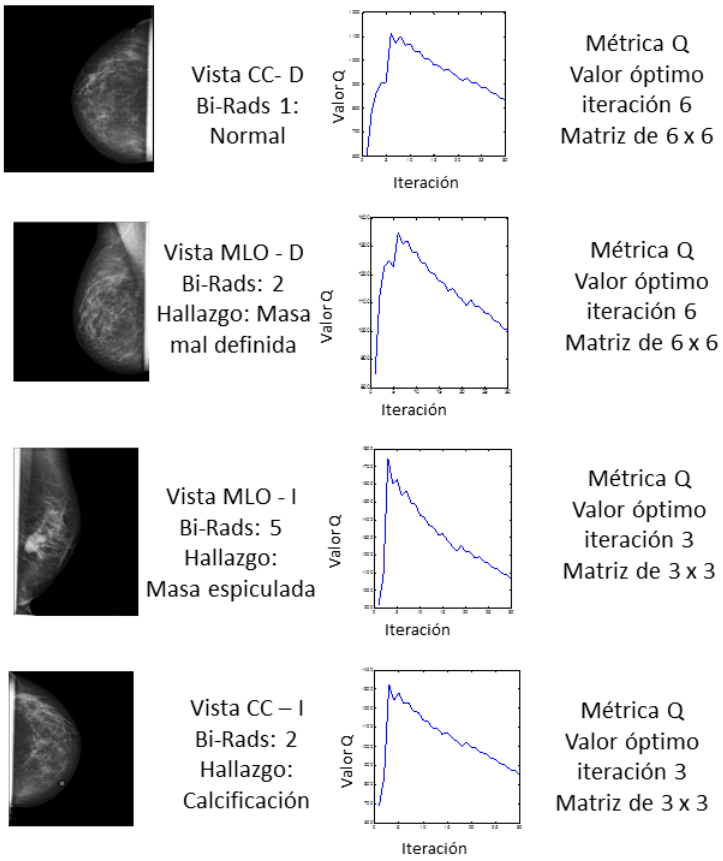

Figura 9.5: Selección de parámetros filtro de media mediante la métrica Q

Se aplica el filtro gaussiano a las imágenes, modificando los valores del tamaño de la máscara de filtrado y la desviación estándar, los resultados de la métrica Q aplicados a un par de imágenes de diferente proyección y mama se muestra en la figura 9.6.

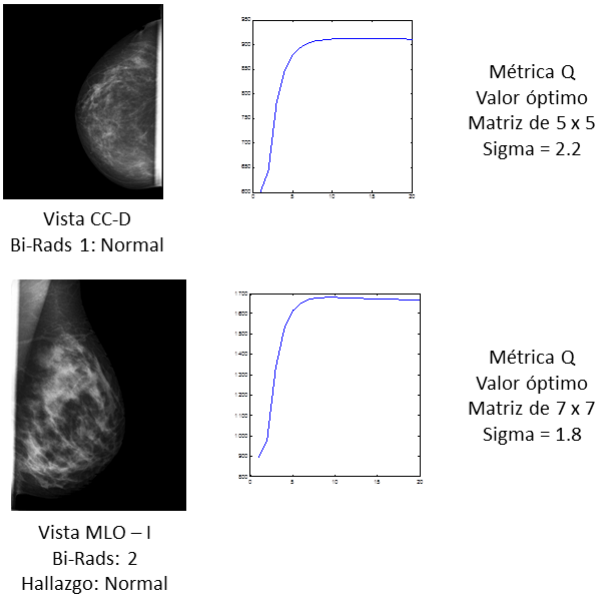

Figura 9.6: Selección de parámetros filtro gaussiano mediante la métrica Q

Los mejores parámetros al aplicar el filtro de mediana y luego de calcular la métrica Q se muestran en la figura 9.7.

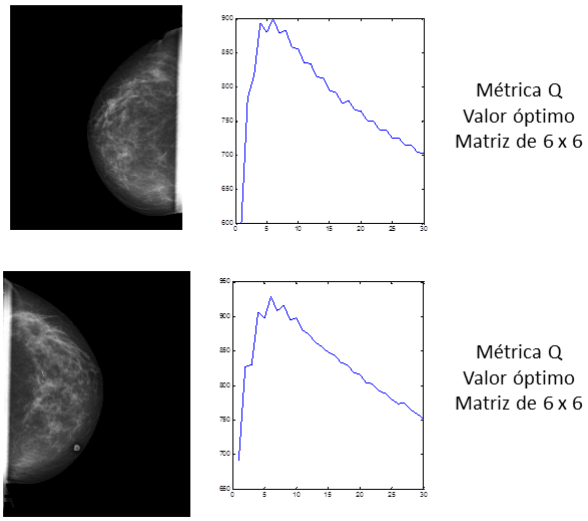

Figura 9.7: Selección de parámetros filtro de mediana mediante la métrica Q

El cálculo de la métrica Q al aplicar el filtro unsharp no determina una estabilidad como en los filtros anteriores, por lo tanto, se utiliza el valor máximo que puede tomar alpha. En la figura 9.8.

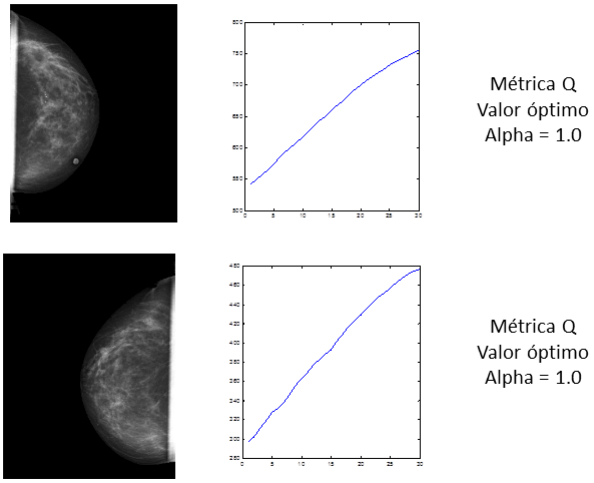

Figura 9.8: Selección de parámetros filtro unsharp mediante la métrica Q

En la siguiente tabla, se muestran los mejores parámetros de cada filtro aplicado a las proyecciones: CC derecha, MLO derecha, CC izquierda y MLO izquierda de la base de datos de registros del eje cafetero, luego de aplicar la métrica Q.

Tabla 9.5: Mejores parámetros de filtro utilizando la métrica Q

Métrica Q				
Tipo de vista	Filtro de media	Gaussiano	Mediana	Unsharp
CC_D	6 x 6	5 x 5, $\delta = 2{,}2$	6 x 6	1.0
MLO_D	6 x 6	5 x 5, $\delta = 2{,}2$	6 x 6	1.0
CC_I	3 x 3	7 x 7, $\delta = 1{,}5$	6 x 6	1.0
MLO_I	3 x 3	7 x 7, $\delta = 1{,}5$	6 x 6	1.0

Luego de ajustar los parámetros adecuados de los filtros, se aplican a los registros mamográficos. Las imágenes filtradas se utilizan para calcular la métrica Q. Además, utilizando cada imagen original y filtrada se encuentra el valor de la métrica ssim, los resultados de las métricas son determinantes para la selección del filtro. La métrica Q proporciona una medida cuantitativa del verdadero contenido de la imagen (nitidez y contraste) ante la presencia de ruido, por lo tanto, el mayor valor de ésta, corresponde al mejor filtro. Por su parte, la métrica SSIM mide la similitud estructural entre las dos imágenes, por ello, el mayor valor corresponde a la mejor máscara de filtrado. Los resultados de las métricas Q y SSIM se muestran en las tablas 9.6 y 9.7 respectivamente.

Tabla 9.6: Resultados de filtrado - métrica Q

Métrica Q				
Tipo de vista	Filtro de media	Gaussiano	Mediana	Unsharp
CC_D	1075	892	805	618
MLO_D	1389	1155	1021	828
CC_I	1505	1415	1123	1065
MLO_I	1710	1582	1224	1257

La tabla 9.6 muestra que se obtiene un mayor valor de métrica Q para la máscara de filtrado de media, luego de procesar las cuatro proyecciones de la mama de la base de datos.

Tabla 9.7: Resultados de filtrado - métrica SSIM

Métrica SSIM				
Tipo de vista	Filtro de media	Gaussiano	Mediana	Unsharp
CC_D	0.9875	0.9950	0.9987	0.9904
MLO_D	0.9827	0.9931	0.9981	0.9961
CC_I	0.9982	0.9930	0.9983	0.9873
MLO_I	0.9979	0.9917	0.9978	0.9847

Revisando los valores de la tabla 9.7 se puede observar que el mayor valor de métrica ssim se obtiene al filtrar los tres primeros tipos de vista con la máscara de mediana y la última vista con el filtro de media.

Se realiza una prueba adicional, filtrando una región de interés que contiene mcals con los 4 tipos de filtros y utilizando los parámetros adecuados acorde a los resultados de la métrica Q. Sin embargo, en las imágenes resultantes se notó que a pesar de la reducción de ruido, hay pérdida de información importante para la detección de microcalcificaciones. En la siguiente imagen se muestra la región de interés luego de aplicar las diferentes técnicas de filtrado.

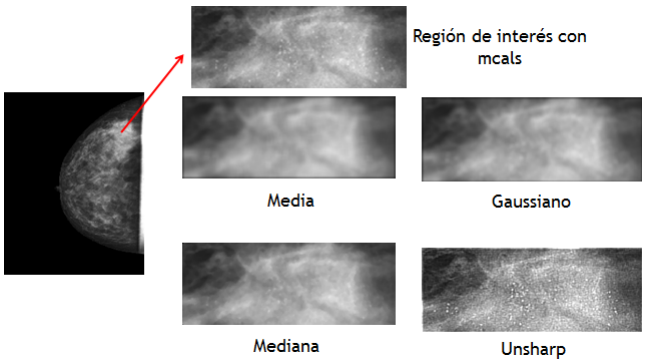

Figura 9.9: Imágenes luego de aplicar técnicas de filtrado a una ROI

Por lo anterior, se observa que la mejor técnica de filtrado espacial corresponde a la máscara unsharp pues destaca los elementos importantes de la imagen y no proporciona pérdida de información.

Se continúa con el proceso de reducción de ruido, pero ahora se realiza la selección de parámetros de los filtros aplicados a las imágenes de la base de datos MINI-MIAS, el procedimiento es análogo al aplicado a las imágenes del eje cafetero, con la diferencia de que esta base de datos no contiene registros de la proyección cráneo-caudal, en la tabla 9.8 se muestran los mejores parámetros según la métrica Q.

Tabla 9.8: Mejores parámetros de filtro utilizando la métrica Q

Métrica Q				
Tipo de vista	Filtro de media	Gaussiano	Mediana	Unsharp
MLO_I	3 x 3	5 x 5, $\delta = 1{,}1$	8 x 8	1.0
MLO_D	3 x 3	7 x 7, $\delta = 0{,}5$	7 x 7	1.0

Los parámetros de los filtros son ajustados de manera consecuente con el resultado de la métrica Q, buscando mejorar el nivel de ruido de las imágenes. Las imágenes filtradas que corresponden a las proyecciones MLO_D y MLO_I se utilizan para calcular la métrica Q y establecer la mejor máscara de filtrado para este tipo de imágenes, con el mismo objetivo, se calcula además, el valor de la métrica ssim, utilizando dos imágenes (original y con reducción de ruido). Los mayores valores de las métricas Q y SSIM tienen concordancia con los mejores filtros implementados. El resultado de la métrica se muestra en la tabla 9.9.

Tabla 9.9: Resultados de filtrado - métrica Q

Métrica Q				
Tipo de vista	Filtro de media	Gaussiano	Mediana	Unsharp
MLO_I	3.60	3.64	3.73	4.91
MLO_D	4.07	3.94	3.87	5.28

En la tabla 9.9 se observa que el valor más alto de métrica Q se obtiene al filtrar las imágenes con la máscara unsharp.

Al calcular los valores de la métrica SSIM se obtuvo siempre un valor de 1.0, dicho parámetro no fue determinante para la selección del filtro. Por lo tanto, se utilizan los resultados de la métrica Q obtenidos en la tabla 9.9 y se realiza el filtrado a una región de interés con los filtros de media, gaussiano, mediana y unsharp.

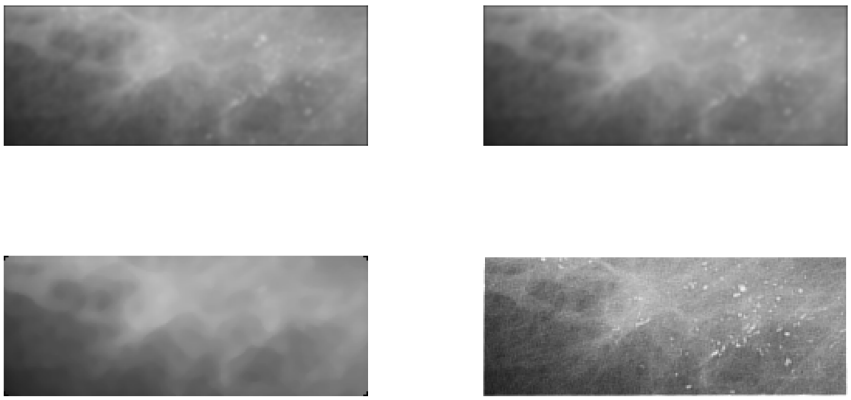

Figura 9.10: Imágenes luego de aplicar técnicas de filtrado (media, gaussiano, mediana y unsharp) a una ROI

En la figura 9.10 se comprueba que el filtro unsharp realza las características de la imagen mejorando el contraste y realizando la reducción de ruido, lo que proporciona unos adecuados resultados para las etapas posteriores de extracción de características y clasificación de las mcals.

9.3. Identificación de microcalcificaciones utilizando descriptores de textura

En esta parte, se realiza la extracción de características aplicando las técnicas SFTA y el operador LBP. Estos métodos utilizan un análisis de textura del tejido que compone la mama, aprovechando que sus características suelen cambiar con la presencia de hallazgos y su grado de severidad.

La extracción de características se realiza a cuatro conjuntos de imágenes de la base de datos del eje cafetero. Los cuales están conformados como se describe a continuación:

- El primer grupo de datos contiene imágenes de macro y microcalcificaciones benignas y mcals malignas, el número total de imágenes es 40.
- El segundo grupo de datos está compuesto por 140 imágenes, donde 70 imágenes corresponden a calcificaciones benignas y el resto a imágenes sin presencia de ellas.
- El tercer grupo contiene 20 imágenes de mcals malignas y 20 sin calcificaciones, para un total de 40 imágenes.
- El último grupo contiene 20 muestras de cada uno de los 3 tipos de imágenes, con calcificaciones benignas, malignas y sin calcificaciones, para un total de 60 imágenes.

Los primeros resultados de clasificación se obtienen utilizando la técnica SVM para la clasificación, el vector de características es obtenido con el operador LBP, en la tabla 9.10 se muestran los resultados.

Tabla 9.10: Porcentajes de acierto luego de aplicar LBP y SVM

Tipo de imágenes analizadas	Porcentaje de acierto
Calcificaciones benignas y malignas	72.50
Calcificaciones benignas e imágenes sin calc	73.21
Calcificaciones malignas e imágenes sin calc	70.00
Calcificaciones benignas, malignas y sin calc	65.00

Se extraen características aplicando la técnica de análisis de textura fractal SFTA, las cuales sirven para la clasificación de hallazgos mediante la máquina de soporte vectorial. En la tabla 9.11 se muestran los porcentajes de acierto de la clasificación de las imágenes aplicando estas técnicas.

Tabla 9.11: Porcentajes de acierto luego de aplicar SFTA y SVM

Tipo de imágenes analizadas	Porcentaje de acierto
Calcificaciones benignas y malignas	83.75
Calcificaciones benignas e imágenes sin calc	92.50
Calcificaciones malignas e imágenes sin calc	80.63
Calcificaciones benignas, malignas y sin calc	83.13

Los registros mamográficos de la base de datos MINI-MIAS se agrupan en cuatro conjuntos de datos, de la misma manera que las mamografías del eje cafetero. La extracción de características se realiza inicialmente mediante el operador LBP y luego mediante la técnica SFTA para luego ser clasificados usando la SVM.

- ◇ El primer grupo de datos contiene imágenes de macro y microcalcificaciones benignas y mcals malignas, el número total de imágenes es 16.
- ◇ El segundo grupo de datos está compuesto por 16 imágenes, donde 8 imágenes corresponden a calcificaciones benignas y el resto a imágenes sin presencia de ellas.
- ◇ El tercer grupo contiene 8 imágenes de mcals malignas y 8 sin calcificaciones, para un total de 16 imágenes.
- ◇ El último grupo contiene 8 muestras de cada uno de los 3 tipos de imágenes, con calcificaciones benignas, malignas y sin calcificaciones, para un total de 24 imágenes.

Mediante la aplicación del operador LBP se extraen características a las imágenes de la MINI-MIAS, para luego ser clasificados mediante una máquina de soporte vectorial. En la tabla 9.12 se muestran los resultados.

Tabla 9.12: Porcentajes de acierto luego de aplicar LBP y SVM

Tipo de imágenes analizadas	Porcentaje de acierto
Calcificaciones benignas y malignas	58.75
Calcificaciones benignas e imágenes sin calc	68.75
Calcificaciones malignas e imágenes sin calc	66.25
Calcificaciones benignas, malignas y sin calc	68.33

Se extraen características fractales aplicando la técnica, las cuales sirven para la clasificación de hallazgos mediante la máquina de soporte vectorial. En la tabla 9.13 se muestran los porcentajes de acierto de la clasificación de las imágenes aplicando estas técnicas.

Tabla 9.13: Porcentajes de acierto luego de aplicar SFTA y SVM

Tipo de imágenes analizadas	Porcentaje de acierto
Calcificaciones benignas y malignas	91.25
Calcificaciones benignas e imágenes sin calc	91.25
Calcificaciones malignas e imágenes sin calc	91.25
Calcificaciones benignas, malignas y sin calc	85.00

La técnica de extracción de características de medidas fractales SFTA proporciona una mejor caracterización de la textura que compone la mama y los hallazgos presentes en ella, esto puede observarse en las tablas de resultados 9.11 y 9.13 donde los mejores resultados de clasificación de mcals están asociados a ella.

10. Conclusiones y trabajos futuros

En este trabajo se desarrolló una metodología para la construcción de una base de datos con registros mamográficos de pacientes de la región del Eje Cafetero. Por lo tanto se generó una base de datos con 552 imágenes de mamografía de 139 estudios de pacientes locales, cumpliendo con el mínimo del índice muestral del $5\,\%$. Los registros adquiridos, evidencian una población amplia de micro-calcificaciones con lo cual la se pueden desarrollar diferentes metodologías para la identificación de este tipo de hallazgos.

Además, en este trabajo se desarrolló una metodología para la identificación de mcals basada en descriptores de textura. Los resultados evidenciaron que la metodología de analisis de textura fractal (SFTA) mostró ser más relevante que la metodología de patrones binarios locales (LBP) en la extracción de rasgos característicos en una determinada región de interés con presencia de mcals. La metodología desarrollada presenta un nivel de robustez adecuado para validar los hallazgos reportados por el personal médico en cada una de las imágenes recopiladas en la base de datos.

Como trabajos futuros se plantea el continuo mejoramiento de la base de datos, además del registro de la información y el desarrollo de una plataforma para el libre acceso a la comunidad científica.

Además, se plantea el desarrollo de otras metodologías de extracción de información de textura, como modelos estadísticos que sean capaces de identificar un determinado rasgo característico en una región de interés, a partir del conocimiento a-priori de la información de textura en una determinada imagen de mamografía.

Bibliografía

[1] Fernando Andrés Angarita y Sergio Andrés Acuña. Presentación inicial de las pacientes con diagnóstico de cáncer de seno en el centro javeriano de oncología, hospital universitario san ignacio, 2010. 4, 5, 7, 8, 10, 12

[2] Fernando Andrés Angarita y Sergio Andrés Acuña. Cáncer de seno: de la epidemiología al tratamiento, 2008. 4, 5, 7, 8, 10, 12

[3] National Cancer Institute. Lo que usted necesita saber sobre el cáncer de seno, 2013. 4

[4] Sonia Elias, Alvaro Contreras, and C Elias S Llanque. Cáncer o carcinoma de mama. *Rev Paceña Med Fam*, 5(7):14–23, 2008. 4

[5] Marta Lucía Guevara G Damián Alberto Álvarez G. Detección de microcalcificaciones en mamografías digitales, 2006. 4, 6, 7, 8, 9, 10, 12, 19, 20, 21, 22, 23, 24, 25, 28, 52

[6] Sukhamwang N Muttarak M, Kongmebhol P. Breast calcifications: which are malignant?, 2009. 5, 7, 10

[7] A. Papadopoulos, D.I. Fotiadis, and L. Costaridou. Improvement of microcalcification cluster detection in mammography utilizing image enhancement techniques. *Computers in Biology and Medicine*, 38(10):1045 – 1055, 2008. 5, 11, 12, 15, 18

[8] Arnau Oliver, Albert Torrent, Xavier LladÃ^3, $MeritxellTortajada, LidiaTortajada, Melcio$
$n s, JordiFreixenet, andReyerZwiggelaar. Automaticmicrocalcificationandclusterdetec$
$68 - -75, 2012. 5$

[9] X.-S. Zhang and Hua Xie. Discriminant subspace learning for microcalcification clusters detection. *Physics Procedia*, 24, Part C(0):2237 – 2244, 2012. International Conference on Applied Physics and Industrial Engineering 2012. 5, 11, 12, 15, 17, 18

[10] Xinsheng Zhang and Xinbo Gao. Twin support vector machines and subspace learning methods for microcalcification clusters detection. *Engineering Applications of Artificial Intelligence*, 25(5):1062 – 1072, 2012. 5, 11, 12, 15, 17, 18

[11] Mammographic image analysis homepage. 5, 7, 8, 12

[12] Xiang Zhu and P. Milanfar. Automatic parameter selection for denoising algorithms using a no-reference measure of image content. *Image Processing, IEEE Transactions on*, 19(12):3116–3132, Dec 2010. 6, 53

[13] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. 6, 54

[14] A.F. Costa, G. Humpire-Mamani, and A. J M Traina. An efficient algorithm for fractal analysis of textures. In *Graphics, Patterns and Images (SIBGRAPI), 2012 25th SIBGRAPI Conference on*, pages 39–46, Aug 2012. 6, 32, 33

[15] Jinshan Tang, Rangaraj M Rangayyan, Jun Xu, Issam El Naqa, and Yongyi Yang. Computer-aided detection and diagnosis of breast cancer with mammography: recent advances. *Information Technology in Biomedicine, IEEE Transactions on*, 13(2):236–251, 2009. 8, 10

[16] Juan José Bedolla Solano. Sistema automatizado diagnostico lesiones imágenes de mamografias. tercera parte. metodologia. 8, 9, 11

[17] José Avelino Manzano Lizcano, Laura Moyano Pérez, and Carmen Sánchez Avila. Sistema para la deteccion automatica de microcalcificaciones en mamografía digitalizada. 8, 9, 10, 11, 12

[18] Brijesh Verma, Peter McLeod, and Alan Klevansky. Classification of benign and malignant patterns in digital mammograms for the diagnosis of breast cancer. *Expert Systems with Applications*, 37(4):3344 – 3351, 2010. 8, 9, 12

[19] Shin-Yuan Hung and Chin-Yu Chen. Mammographic case base applied for supporting image diagnosis of breast lesion. *Expert Systems with Applications*, 30(1):93 – 108, 2006. Intelligent Bioinformatics Systems. 9, 10, 14

[20] M. Sameti, R.K. Ward, J. Morgan-Parkes, and B. Palcic. Image feature extraction in the last screening mammograms prior to detection of breast cancer. *Selected Topics in Signal Processing, IEEE Journal of*, 3(1):46–52, 2009. 9, 11, 17, 18

[21] A Karahaliou, S Skiadopoulos, I Boniatis, P Sakellaropoulos, E Likaki, G Panayiotakis, and L Costaridou. Texture analysis of tissue surrounding microcalcifications on mammograms for breast cancer diagnosis. *British journal of radiology*, 80(956):648–656, 2007. 11

[22] Alain Tiedeu, Christian Daul, Aude Kentsop, Pierre Graebling, and Didier Wolf. Texture-based analysis of clustered microcalcifications detected on mammograms. *Digital Signal Processing*, 22(1):124 – 132, 2012. 11, 12, 17, 18

[23] Hui Zhu, Francis H.Y. Chan, and F.K. Lam. Image contrast enhancement by constrained local histogram equalization. *Computer Vision and Image Understanding*, 73(2):281 – 290, 1999. 11

[24] Julián Dondero. Modelos activos de apariencia y máquinas de soporte vectorial para reconocimiento de expresiones faciales en tiempo real, 2013. 11, 30, 55

[25] Kehong Yuan, Zhen Tian, Jiying Zou, Yanling Bai, and Qingshan You. Brain {CT} image database building for computer-aided diagnosis using content-based image retrieval. *Information Processing Management*, 47(2):176 – 185, 2011. 14

[26] ¿qué indican las estadísticas clave sobre el cáncer de seno? 19

[27] Protocolo diagnóstico - mamografía screenig. 19

[28] Informes de mamogramas – bi-rads. 26

[29] Segmentación de imágenes. 27

[30] Universidad Nacional de Quilmes – Ing. en Automatización y Control Industrial. Filtrado espacial. 28

[31] Jesús M. De La Cruz Garcia y Antonio García 2ª Edició Gonzálo Pajares Martin sanz. In *Visión por computador. Imágenes Digitales y Aplicaciones*, pages 588–593, 2008. 28, 29, 30

[32] Sánchez Serralde. El algoritmo lbp, 2012. 30

[33] García Arias Hernán Felipe. Identificacion de artefactos en bases de datos de señales eeg, mediante tecnicas espacio-temporales. 35

[34] Adrian F. Clark. Information of the mini-mias database of mammograms, 2012. 39

[35] Liga contra el cáncer. Mamografía. 40

[36] Miguel Alcaraz Baños. El mamógrafo. bases de la mamografía. principios diagnósticos diferenciales. 41

[37] Digitalizador cr-35x. 41

[38] Marko Heikkilä and Timo Ahonen. Libreria lbp. 43

[39] Costa Alceu Ferraz. Libreria sfta. 43

[40] R. P. W. Duin. 43

[41] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, April 2004. 54