BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

Estudio de dimensión fractal, entropía y análisis de datos para mamografías con microcalcificación o tumor

Tesis presentada al

Colegio de física

como requisito parcial para la obtención del grado de

LICENCIADA EN FÍSICA APLICADA

Presenta

Minerva Guadalupe Vázquez Domínguez

Asesorada por

Dr. Cristian Heber Zepeda Fernández

Dr. Benito De Celis Alonso

Puebla Pue.
Diciembre 2020

BENEMÉRITA UNIVERSIDAD
AUTÓNOMA DE PUEBLA

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS
LICENCIATURA EN FÍSICA APLICADA

Estudio de dimensión fractal, entropía y análisis de datos para mamografías con microcalcificación o tumor

Tesis presentada al

Colegio de física

como requisito parcial para la obtención del grado de

LICENCIADA EN FÍSICA APLICADA

Presenta

Minerva Guadalupe Vázquez Domínguez

Asesorada por

Dr. Cristian Heber Zepeda Fernández

Dr. Benito De Celis Alonso

Puebla Pue.
Diciembre 2020

Título: Estudio de dimensión fractal, entropía y análisis de datos para mamografías con microcalcificación o tumor.

Estudiante: MINERVA GUADALUPE VÁZQUEZ DOMÍNGUEZ

COMITÉ

Mario Rodríguez Cahuantzi
Presidente

Javier Miguel Hernández López
Secretario

Jorge Velázquez Castro
Vocal

Dr. Cristian Heber Zepeda Fernández
Asesor

Dr. Benito De Celis Alonso
Asesor

Agradecimientos

A las primeras personas que quisiera agradecerles es a mi familia ya que ellos han estado para mí desde el principio de la licenciatura, me han ayudado siempre de todas las formas posibles y han sido un apoyo incondicional. Sin ellos el flujo de sucesos en mi vida no sería el mismo en esta etapa tan importante. Mis Padres influyen en mi vida personal y contribuyen a mi formación y a ellos agradezco su calidez y apoyo que impulsó mi trabajo día a día, brindando las fuerzas para seguir adelante e ir superando mis propias expectativas.

También quisiera agradecer a los profesores que me han brindado las herramientas necesarias para abrirme paso en los estudios y llegar a esta etapa. Sobre todo a mis asesores el Dr. Cristian Heber Zepeda Fernández, que sabe de verdad lo que significa ser Doctor y lo pone en práctica en todo momento para todo el que recurra a él, y el Dr. Benito De Celis Alonso, ellos me han guiado y compartido su experiencia y conocimiento en el desarrollo de mi proyecto de tesis.

A todas estas personas que han tenido la paciencia y la virtud de enseñarme y ayudarme en mi formación académica. Sé que sin ellos nada de esto sería posible y se los agradezco.

Dedicado a...

mis Papás Minerva y Daniel

y mis hermanos Dany y Edwin

las personas más importantes en mi vida.

Índice general

Resumen XVI

Capítulo 1. Introducción 1

1.1 Rayos X y obtención de una Mamografía digital 1

1.2 Microcalcificación y Tumor 13

1.3 La Dimensión Fractal y la Entropía de Shannon 18

1.3.1 Dimensión Fractal 18

1.3.2 Entropía de Shannon 25

Capítulo 2. Metodología 28

2.1 Redigitalización de mamografías 28

2.2 Selección de elementos a analizar 31

2.2.1 Microcalcificación, Candidato y Tejido 31

2.2.2 Tumor y Tejido 33

2.3 Cálculo de la dimensión fractal y la entropía 37

Capítulo 3. Resultados 40

3.1 Análisis de mamogramas con microcalcificación 41

3.2 Análisis de mamogramas con tumor 50

Capítulo 4. Análisis 58

Capítulo 5. Conclusiones 66

Apéndice A. Muestras de mamografías con microcalcificaciones 69

Apéndice B. Muestras de mamografías con tumor 71

Bibliografía

Índice de figuras

Figura 1.1. La ionización de un átomo de carbono mediante un rayo X, deja el átomo con una carga eléctrica neta +1. El átomo ionizado y el electrón liberado se denominan par iónico. · 3
Figura 1.2 Modelos habituales de sistemas de imagen mamográficos. · · · · 4
Figura 1.3 Partes del tubo de Rayos X. · · · · · · · · · · · · · · · · · · · 6
Figura 1.4 Viaje del haz de Rayos X desde el cátodo hasta conseguir el tamaño aparente del punto focal efectivo. · 7
Figura 1.5 A) Proyección de un mamograma frente a la misma anatomía vista con la tomosíntesis mamográfica digital B). · · · · · · · · · · · · · · · · · 9
Figura 1.6 a) Evidencia de imágenes punteadas radiopacas en la axila derecha por antitranspirante, b) artefacto por polvo. Se observan puntos blancos, y c) líneas blancas causadas por polvo en el lector. · · · · · · · · · · · · · · · · · 10
Figura 1.7 A la izquierda artefacto debido a que el sistema lee mal el tamaño del detector, y a la derecha deformación de una parte de la imagen en la proyección. · 11
Figura 1.8 Esquema de los descriptores BI-RADS de distribución de las microcalcificaciones. · 15
Figura 1.9 Una representación de estructura de alambre de la superficie que se puede asociar con una imagen mamográfica de la mama. · · · · · · · · · · 19
Figura 1.10 Gráfico que muestra las tres dimensiones de un mapa de intensidades. · 20
Figura 1.11 Medición de la longitud de la costa utilizando una regla de 16 Km (izquierda) y 8 Km (derecha). · 22

Figura 1.12 Logaritmo de las longitudes aproximadas de diversas costas y fronteras en función del logaritmo de la escala. · · · · · · · · · · · · · · · · · 24
Figura 2.1 A la izquierda la imagen de mamografía y a la derecha la nueva imagen en escala de grises. · 30
Figura 2.2 Mamograma normalizado y en escala de grises. Se muestran los ejes y la escala que relaciona el color con la intensidad, a la derecha. · · · · 31
Figura 2.3 Imagen de una mama con la MC encerrada en el círculo azul y el C en el blanco. · 32
Figura 2.4 MC-área (izquierda) y C-área (derecha), de la muestra de la Figura 2.3. · 33
Figura 2.5 Área del tejido de la muestra de la Figura 2.3. · · · · · · · · · · 34
Figura 2.6 Mamograma y redigitalización de muestra con T y con artefacto encerrado en un círculo. · 35
Figura 2.7 Acercamiento al T (izquierda) y al tejido (derecha) de la muestra con T. · 36
Figura 3.1 Dimensión fractal vs Entropía para cada cuadrante de cada muestra con MC. · 42
Figura 3.2 Histograma con las intensidades del cuadrante que abarca a la MC. 43
Figura 3.3 Histograma con las intensidades del cuadrante que abarca al C. · · 44
Figura 3.4 Histograma con las intensidades del cuadrante con tejido para la muestra con MC. · 45
Figura 3.5 Histograma con las intensidades de todo el mapa de intensidades, de la muestra de la Figura 3.2. · 48
Figura 3.6 Mapa de intensidades de la muestra de la Figura 3.2, para los valores mayores de 0.7 en la escala de intensidad. · · · · · · · · · · · · · · 48
Figura 3.7 Histograma con las intensidades de tofo el mapa en escala de grises con MC, incluyendo las intensidades cero, de la muestra de la Figura 3.3. · · 49
Figura 3.8. Muestra de mamografía con MC elevada a la potencia b) tercera, c) sexta y d) decimoquinta. La imagen original se muestra en a). · · · · · · · 50
Figura 3.9 Dimensión fractal vs Entropía para cada cuadrante de cada muestra con T. · 51
Figura 3.10 Histograma con las intensidades del cuadrante con T para la muestra con T. · 52
Figura 3.11 Histograma con las intensidades del cuadrante con tejido para la muestra con T. · 53
Figura 3.12 Histograma con las intensidades de todo el mapa en escala de grises con T, incluyendo las intensidades cero, de la Figura 3.10. · · · · · · 55

Figura 3.13 Muestra redigitalizada de la muestra de la Figura 3.10 para los valores mayores de 0.7 en la escala de intensidad. · · · · · · · · · · · · · · 55
Figura 3.14 Resultados al elevar a la a) tercera, b) quinta y c) décima potencia la muestra de la Figura 3.14. · 56
Figura 4.1 A la izquierda un cuadrante con el parámetro color ordenado y a la derecha desordenado. · 59
Figura 4.2 a) Mamograma con nódulo, b) selección del cuadrante que encierra a la lesión, c) segmentación, d) contorno de lesión y enrejado para cálculo de dimensión fractal por conteo de cajas. · 62
Figura 4.3 a) Imagen original. b) Imagen segmentada preprocesada, reajustada, limpia de artefactos y con las microcalcificaciones detectadas en su tamaño real. c) Imagen con microcalcificaciones realzadas. · · · · · · · · · 62
Figura 4.4 Secciones de mamogramas originales (primera fila), procesadas por la transformada Top-hat (segunda fila). · · · · · · · · · · · · · · · · · · · 62

Índice de tablas

Tabla 3.1 En la primera columna aparece el número de muestra con MC y en las siguientes la dimensión fractal y entropía para cada elemento. · · · 41
Tabla 3.2 En la primera columna aparece el número de muestra, en las siguientes tres el valor promedio de cada elemento analizado para las cinco muestras con MC. · 46
Tabla 3.3 En la primera columna aparece el número de muestra, en las siguientes dos los valores de la distribución de intensidad para el tejido, y en los últimos el valor máximo de intensidad del área de la MC y C. · · · 46
Tabla 3.4 En la primera columna aparece el número de muestra con T y en las siguientes la dimensión fractal y entropía para cada elemento. · · · · · 51
Tabla 3.5 En la primera columna aparece el número de muestra, en las siguientes dos los valores de la intensidad promedio que presenta el área de T y tejido. · 53
Tabla 3.6 En la primera columna aparece el número de muestra, en las siguientes dos los valores de la distribución de intensidad para el tejido, y en la última el valor máximo de intensidad del área de T. · · · · · · · · · 54
Tabla 4.1 En la primera columna aparece el número de muestra con T y en las siguientes el error porcentual entre las coordenadas dadas en el diagnóstico y las obtenidas mediante el método presentado. · · · · · · · 64

Resumen

Para este trabajo se realizaron estudios con imágenes de mamografías en donde se presentan microcalcificaciones (MC) o tumores (T). El objetivo principal fue identificar el área de tumor (T-área) y microcalcificación (MC-área) del resto de la imagen de mamografía, la cual llamamos área del tejido (t-área), y conocer su ubicación, de una manera fácil y con una alta precisión.

La manera de obtener los parámetros que distinguen cada elemento, consistió primero en transformar las imágenes de mamografía en mapas de intensidades o mapas en escala de grises, los cuales fueron normalizados. Utilizamos un programa escrito en C++ para dicha transformación. Esto hizo que se pudieran identificar el MC-área y T-área, debido a que éstos presentan el valor máximo en la escala de grises, y por ende corresponden a las zonas más brillantes en la imagen reprocesada. Esto da una distinción visual notable y una distinción numérica para el MC-área y T-área.

Asociamos dos cantidades a cada área para poder diferenciar MC y T con el resto de la mamografía: dimensión fractal (DF) y entropía de Shannon (S). Se calcularon para las regiones que pertenecen a T o MC y tejido del mapa de densidades. Lo que se obtuvo fue que los valores de DF, son mayores para MC-área y T-área, mientras que para t-área son menores. Además los valores de S para MC-área y T-área son menores que los de t-área.

Finalmente, realizamos un análisis de datos utilizando el conjunto de valores de intensidad de las áreas previamente mencionadas. Concluimos que los valores en MC-área y T-área se encuentran a más de 2-σ del valor central de t-área, y por ello podemos

así distinguir las distintas áreas mediante los valores de intensidad que presenten cada una.

Este estudio fue realizado para mamografías con diagnóstico, por lo cual puede ser aplicado a nuevas imágenes para la mejora de un diagnóstico en el cual visualmente no se puede determinar la ubicación de MC o T.

Cabe aclarar que este trabajo no permite diferenciar de un T maligno o benigno.

Palabras clave: *Dimensión fractal, Entropía, Mamografía Digital.*

Capítulo 1

Introducción

En este capítulo se presenta la teoría revisada y utilizada para la realización de este trabajo y para las interpretaciones dadas a los resultados que posteriormente se muestran.

1.1 Rayos X y obtención de una Mamografía digital

La energía emitida y transferida en el espacio se denomina radiación y es básicamente energía electromagnética que incluye a los Rayos X (RX), las ondas de radio, las microondas y la luz ultravioleta, infrarroja y visible. Todas estas conforman al espectro electromagnético. La luz visible ocupa un espacio muy pequeño del espectro. A energías mayores a esta, se encuentran los rayos ultravioleta, los responsables de que nuestra piel se queme en un día soleado. Son tan energéticos que tenemos que protegernos con filtros especiales. Después vienen los RX de los que hablaremos adelante. Luego están los Rayos Gamma que son el tipo de luz más energética que existe, afortunadamente el sol produce una pequeña cantidad de esta radiación. Se usa para radioterapia y esterilización de productos, además se producen en explosiones nucleares. Por el contrario, hacia energías menores a la luz visible, está

la luz infrarroja. Esta ayuda a medir la temperatura de los cuerpos. Por debajo están las microondas que se usan para calentar nuestra comida a través de las moléculas de agua presentes en ella. Por último, las ondas menos energéticas son las ondas de radio que se usan para las telecomunicaciones. Cada clasificación tiene su longitud de onda característica, que entre mayor sea, menor será la energía de este tipo de radiación. Además tiene una frecuencia que crece conforme disminuye la longitud de onda. Especialmente, los RX abarcan aproximadamente desde los 10keV hasta los 50MeV, y su longitud de onda asociada es aproximadamente de 10^{-10}m hasta 10^{-14}m. La frecuencia de estos fotones varía desde unos 10^{16} Hz hasta 10^{25} Hz. En medicina se utilizan para realizar estudios donde se generan imágenes de los tejidos y las estructuras dentro del cuerpo cuando estos son irradiados. Los técnicos encargados de estos estudios son los radiólogos, y tienen una gran responsabilidad en la realización de exámenes radiológicos, utilizando estándares de protección de la radiación establecidos para la seguridad de los pacientes y del personal médico, pues en periodos prolongados de tiempo o a grandes dosis de radiación, el cuerpo humano sufre daños severos por ser radiación altamente energética. Si se controlan estos factores, la radiación es muy útil. En los exámenes radiológicos, el paciente está irradiado con radiación ionizante, que es un tipo especial de radiación que incluye a los RX, la cual es capaz de retirar un electrón orbital del átomo con el que interactúa. A esta interacción entre la radiación y la materia se denomina ionización. Esta ocurre cuando los RX pasan cerca de un electrón orbital de un átomo, proporcionándole suficiente energía para salir de éste. De hecho la radiación ionizante puede interaccionar e ionizar átomos adicionales, separando electrones orbitales y átomos, los cuales se denominan par iónico. El electrón es un ion negativo y el átomo restante un ion positivo. Por tanto, cualquier tipo de energía capaz de ionizar la materia se conoce como radiación ionizante. Los RX, los rayos gamma y la luz ultravioleta son las únicas formas de radiación electromagnética con suficiente energía para ionizar. Algunas partículas de movimiento rápido (partículas con elevada energía cinética) también son capaces de producir ionización.

Ejemplos de estas partículas son la alfa y la beta. La *Figura 1.1* representa la interacción entre un RX y un átomo de carbono, un constituyente básico del tejido humano. El RX transfiere su energía a un electrón orbital y hace saltar ese electrón del átomo. El RX puede desaparecer formando un par ion. El átomo resultante es ahora un ion positivo porque contiene una carga positiva más que negativa.

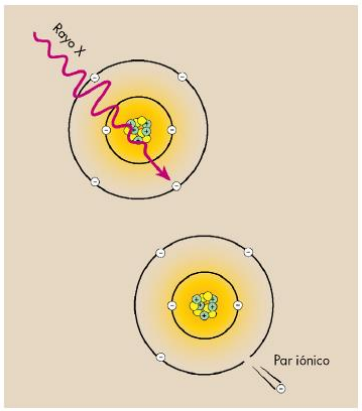

Figura 1.1. La ionización de un átomo de carbono mediante un rayo X, deja el átomo con una carga eléctrica neta +1. El átomo ionizado y el electrón liberado se denominan par iónico.

La fuerza que mantiene unido a cada electrón con el núcleo, antes de que sufra ionización, se llama *energía de enlace electrónico* E_b y aumenta conforme el electrón se acerca al núcleo del átomo. Para el caso de la capa K que está más cerca del núcleo que la capa L, tiene una energía de enlace mayor que la segunda, y a su vez esta tiene una mayor energía que la capa M, y así sucesivamente. Es decir, cuanto más grande y complejo es el átomo, mayor es E_b para los electrones en cualquiera de las capas, y el átomo es grande cuanto mayor número de protones tenga. Por ejemplo, para ionizar el tungsteno (elemento principal para generar RX en un estudio de radiografía), con número atómico de 74, arrancando un electrón de la capa K, la energía mínima necesaria es de $E_b = 69$ keV, por el contrario para el carbono que es un componente importante del tejido humano con número atómico de 6, $E_b = 10$ eV aunque se necesitarán unos 34 eV para ionizar átomos del tejido. El valor de 34 eV se denomina

potencial de ionización. La diferencia (24 eV) ocasiona múltiples excitaciones de los electrones, que en última instancia producen calor [1].

Además del tamaño del átomo y la energía E_b necesaria, los fotones (componentes de la radiación) interaccionan más fácilmente con la materia cuando esta es aproximadamente del mismo tamaño que su longitud de onda. Por ejemplo, la luz visible interacciona con las células vivas, como los bastones o los conos del ojo y tiene longitudes de onda de alrededor de nm, la luz ultravioleta interactúa con las moléculas y tiene también longitudes de onda del orden de nm pero menores a la de la luz visible, mientras que los RX, interactúan con los electrones y los átomos. En general las estructuras que absorben los RX se llaman radiopacas, y las que los atenúan se llaman radiolúcidas. Hablando del cuerpo humano y el estudio de radiografías, el hueso es un elemento radiopaco, mientras que el tejido pulmonar y la mayoría de los tejidos blandos como la mama, son radiolúcidos.

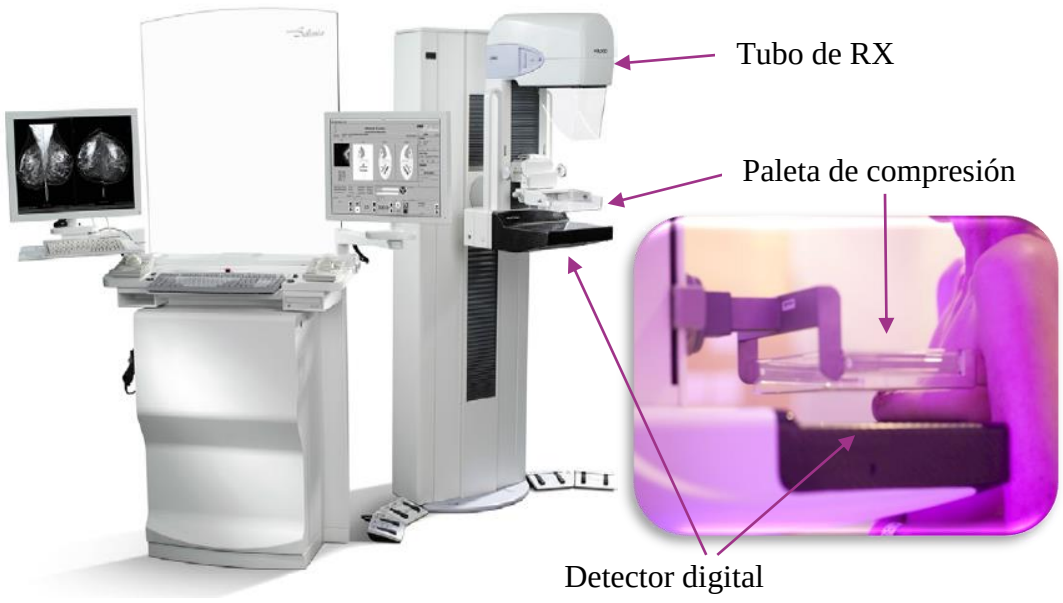

Figura 1.2 Modelos habituales de sistemas de imagen mamográficos.

Ya hemos hablado de cómo los RX ionizan a los átomos con los que interaccionan, pero ahora expliquemos cómo se crean estos RX. En medicina, para la mamografía (examen radiológico donde se estudia la mama), se utilizan sistemas de imágenes mamográficos como el que se muestra en la *Figura 1.2*. Estos cuentan con un tubo de RX, en donde se genera esta radiación, con un detector digital donde se reciben los RX para después generar una imagen, y una paleta de compresión donde se comprime la mama para una mejor visualización como veremos más adelante.

El tubo de RX (*Figura 1.3*), es una ampolla de vidrio con dos extremos de un cable conductor de cargas eléctricas diferentes llamadas cátodo, que es el que tiene carga negativa, y ánodo el de carga positiva. El cátodo se compone de una copa de enfoque (hecha de niquel) y un filamento. Este último es un espiral de aleación de tungsteno y cesio, con punto de fusión a 3,410°C, esto es muy alto para que cuando el filamento se caliente por la energía eléctrica recibida, no se evapore y el vapor llene el tubo contribuyendo al mal funcionamiento del mismo. Mide 2mm de diámetro y de 1cm a 2cm de longitud. Por emisión termoiónica expulsa electrones de las últimas capas de los átomos que lo componen, que viajan hacia el ánodo por la diferencia de potencial. La copa de enfoque se mantiene con carga negativa por lo que repele los electrones que salen del filamento, y los redirige hacia el ánodo. Los valores de voltaje normales para el estudio de mamografía son de 20kVp a 35kVp, y los valores de amperaje son de 5mAs a 400mAs. Entonces los electrones llegan al ánodo donde está el blanco o diana, que es la zona donde se producen los choques entre electrones del filamento, con los de los átomos que componen al material del mismo blanco. Para el caso de mamografía, la diana se compone de molibdeno por su bajo número atómico y por la baja energía característica de su capa K. Mide 15cm de largo y para ánodos rotatorios, gira de 3,600 rpm a 10,000 rpm. Al ser rotatorio, lo que hace es que el haz de electrones interactúe con un área mayor, por lo tanto el calor producido de estas interacciones en el ánodo, que puede llegar hasta los 2,800°C, no se confina en un solo punto, y es que es importante mencionar que de toda la energía cinética que tienen los

electrones, solo el 1% se convierte en RX mientras que el 99% se convierte en energía térmica o calor. A los RX producidos se les llama *bremsstrahlung*, y es la radiación resultante del frenado de electrones proyectil por el núcleo del átomo de molibdeno. Estos rayos tienen energías de 17keV a 24keV que resultan de interacciones K. Así pues, la función principal de un sistema de imágenes mamográfico, es convertir energía eléctrica en energía electromagnética.

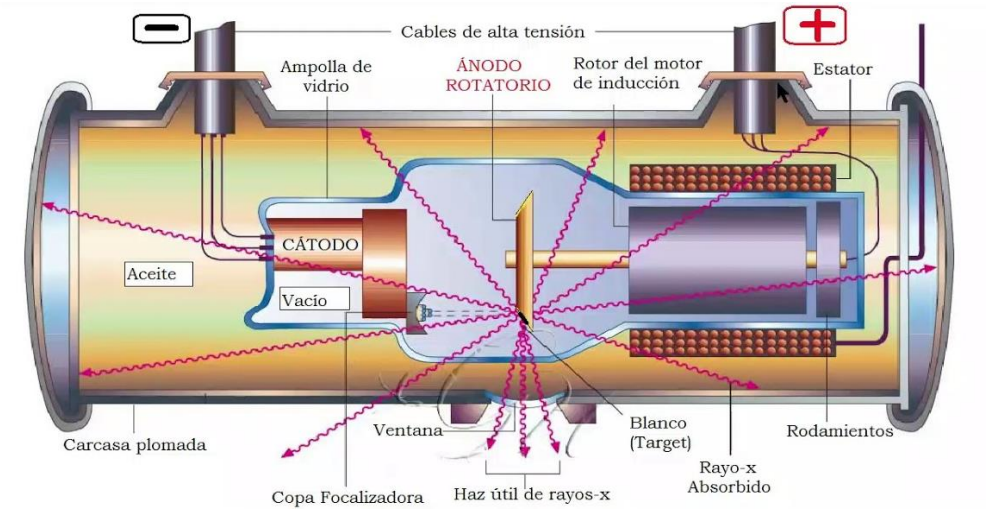

Figura 1.3 Partes del tubo de Rayos X.

Cuando lo que se va a explorar mediante radiación son tejidos blandos como la mama, se requieren técnicas específicas que difieren de las usadas en la radiografía convencional. Estas diferencias técnicas se deben a las distintas peculiaridades de la anatomía que se está explorando. En la radiografía convencional, el contraste material es elevado debido a las grandes diferencias de densidad másica y de número atómico entre los tejidos óseo, muscular, adiposo y pulmonar. En radiografía de tejidos blandos, sólo las estructuras musculares y adiposas se exploran. La mama se forma en gran medida únicamente de tejido fibroglandular y graso. Por tanto, las técnicas radiográficas para estos elementos están diseñadas para optimizar la absorción diferencial en estos tejidos tan parecidos. Se elige molibdeno para la composición de la diana pues este material incrementa la absorción diferencial en el tejido de la mama

y maximiza el contraste radiográfico, ya que los RX producidos por este material son los que están en el intervalo de energía necesario para este tipo de estudios. El molibdeno tiene un número atómico de 42 y el tungsteno, material usado en la radiografía convencional, de 74. Esta diferencia de número atómico es la responsable de las distinciones en el espectro de emisión. Por ello en este tipo de radiografías se utiliza el molibdeno en lugar del tungsteno. Además, una característica importante de la diana en mamografía, o también llamado punto focal, es su tamaño. Debido a la gran demanda de resolución espacial que demandan estos estudios, formar imágenes de microcalcificaciones requiere puntos focales pequeños que se consiguen inclinando el ánodo como se muestra en la *Figura 1.4.* Normalmente los tubos de RX en mamografía utilizan puntos focales de entre 0.3mm y 0.1mm.

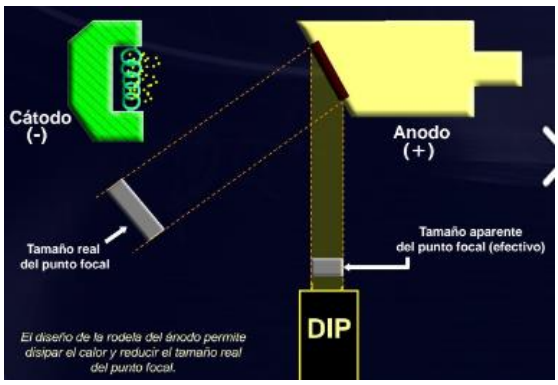

Figura 1.4 Viaje del haz de Rayos X desde el cátodo hasta conseguir el tamaño aparente del punto focal efectivo.

Una vez que los RX se producen, viajan en todas direcciones, por lo que el tubo se blinda con plomo en el exterior con una carcasa, excepto en una ventana por la que sí queremos que salgan los RX para que lleguen a la mama de la paciente. La ventana está hecha de berilio o borosilicato, materiales que no atenúan los RX, los cuales viajan de 50cm a 80cm y llegan al detector eléctrico o "pantalla de fósforo" que convierte a los RX en luz, que pasa por cables de fibra óptica hacia un dispositivo que convierte la luz en señal eléctrica para luego ver la imagen en el monitor. A la imagen se le llama mamograma o imagen de mamografía.

Cuando se realiza el estudio, se comprime la mama del paciente. Esto es importante pues la compresión brinda numerosas ventajas como generar un grueso más uniforme y, por tanto, la densidad óptica de la imagen también es más uniforme, además los tejidos cerca de la pared torácica tienen menos probabilidades de ser subexpuestos y los tejidos cerca del pezón tienen menos probabilidades de ser sobreexpuestos. También la compresión extiende el tejido y por lo tanto reduce la superposición de estructuras en el mismo.

Para observar la imagen de mamografía, comúnmente se han utilizado cuatro tipos de receptores de imagen: película de exposición directa, xerorradiografía, pantalla-película y detectores digitales como ya mencionamos. Los últimos dos se usan actualmente. En general, la radiografía digital recibió un gran impulso a finales de la década de 1990 con la aplicación de la radiografía digital RD a la mamografía, la llamada mamografía digital. Sin embargo, a partir de los resultados del estudio llamado *Ensayo de estudio de imágenes de mamografía digital*, DMIST por sus siglas en inglés, que se publicó a finales de 2005, la mamografía digital recibió un impulso mayor. En este estudio se incluyeron las imágenes de casi 50,000 mujeres con mamografías convencionales. El objetivo fue determinar si la mamografía digital era tan buena como la mamografía convencional, dado que la resolución espacial de la RD está limitada por el tamaño del píxel, dando lugar a que ningún sistema de RD puede igualar la resolución espacial de la radiografía convencional, no obstante la resolución espacial de la mamografía digital fue mucho menor que la de la mamografía de pantalla-película. Según la interpretación de los radiólogos, los resultados demostraron que no sólo la mamografía digital era igual a la mamografía convencional en todas las pacientes, sino que era mejor en caso de un tejido glandular mamario denso. Este resultado indica que la resolución con contraste es más importante que la resolución espacial para la mamografía y posiblemente para todas las imágenes médicas.

Figura 1.5 A) Proyección de un mamograma frente a la misma anatomía vista con la tomosíntesis mamográfica digital B).

Otro avance significativo es la tomosíntesis mamográfica digital, DMT por sus siglas en inglés, la cual es una herramienta complementaria de la mamografía digital. Con la DMT, una superficie del haz de RX interacciona con el receptor de imágenes de la mamografía digital, produciendo un mamograma digital. Se diferencia de la mamografía digital básicamente por un tubo móvil de Rayos X que hace múltiples disparos de bajas dosis de radiación separados por 1 mm y centrados en la mama con diferente angulación, los cuales posteriormente son reconstruidos con algoritmos para obtener una imagen bidimensional llamada mamografía sintetizada, como se muestra en la *Figura 1.5*. Fue desarrollada para reducir la superposición de estructuras y así disminuir la tasa de falsos positivos de la mamografía convencional.

En cuanto a las imágenes médicas, hasta mitad de la década de 1990, la práctica totalidad de las imágenes se presentaban en forma de placas o películas fotográficas y la imagen se interpretaba a partir de estas películas. La tomografía computarizada (1974) y la resonancia magnética (1980) fueron las primeras imágenes médicas digitales que se utilizaron de forma generalizada. Hoy en día, prácticamente todas las imágenes digitales se interpretan a partir de su presentación digital, lo que genera una ventaja fundamental sobre la imagen convencional, y es la capacidad de manipular la imagen antes de representarla, lo que llaman preprocesado, o después de representarla,

el posprocesado. El preprocesado de la imagen digital, implica la calibración electrónica para reducir las diferencias de respuesta píxel a píxel, fila a fila y columna a columna. Se aplican de manera automática en la mayoría de los equipos y consisten en procesos de interpolación de píxeles, de corrección de demora y de corrección de ruido. La interpolación de píxeles se hace puesto que una imagen digital tiene millones de píxeles y alguno puede no responder como debería, en tal caso se promedia la señal recibida en los píxeles vecinos y se le asigna al píxel defectuoso. En cambio, en el posprocesado de la imagen digital, se incluye cualquier modificación que pueda realizarse en una imagen digital después de haber sido obtenida por el sistema de imagen. Requiere la intervención del técnico radiólogo y del radiólogo. Está diseñado para producir imágenes digitales sin los llamados artefactos, los cuales son inconsistencias en la imagen y se clasifican de la siguiente manera:

• Artefactos asociados con el paciente: típicamente incluyen el cabello, el movimiento, las partículas de desodorante o talco, que pueden ser vistas como microcalcificaciones, y el tamaño del pequeño seno que a la hora de comprimirse se adelgaza hasta los 2 cm. Un ejemplo de esto se muestra en la *Figura 1.6 a).*

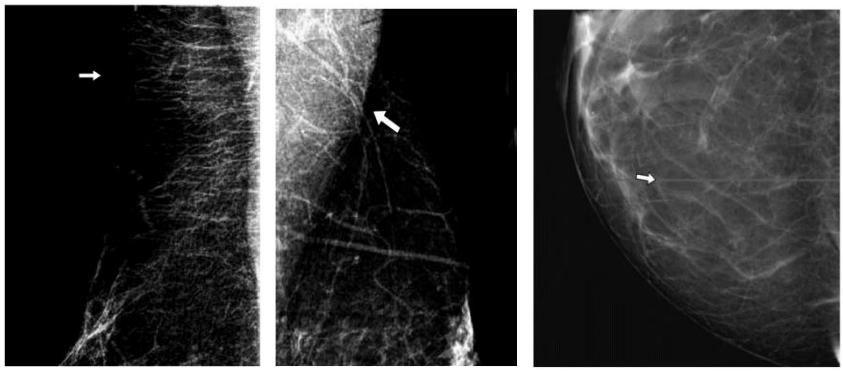

Figura 1.6 a) Evidencia de imágenes punteadas radiopacas en la axila por antitranspirante, b) artefacto por polvo. Se observan puntos blancos, y c) líneas blancas causadas por polvo en el lector.

• Artefactos relacionados con la tecnología: por incorrecta manipulación del sistema, limpieza inadecuada, configuración inadecuada de los parámetros de imagen. Un ejemplo de esto se muestra en la *Figura 1.6 b).*

• Artefactos relacionados con la unidad de mamografía: relacionados con el detector o el lector electrónico, con la imagen en la placa y el lector de las placas. Los artefactos relacionados con el tubo de RX se encuentran en esta categoría. Un ejemplo de esto se muestra en la *Figura 1.6 c).*

• Artefactos relacionados con el software: tienen que ver con el procesamiento. Son generados por un mal funcionamiento o una mala configuración de los algoritmos de procesamiento de la imagen, o son introducidos por el software durante la adquisición. Ejemplos de esto se muestran en la *Figura 1.7.*

• Artefactos relacionados con la visualización: problemas en la calibración del monitor, errores en el software de visualización y problemas con el medio ambiente donde se lee la imagen.

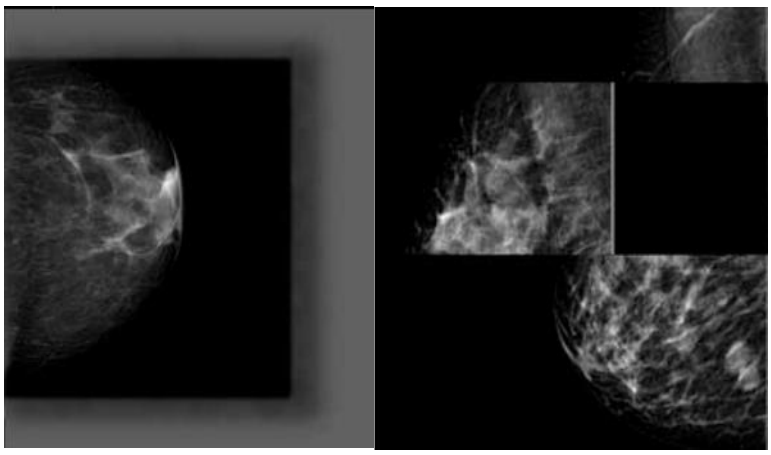

Figura 1.7 A la izquierda artefacto debido a que el sistema lee mal el tamaño del detector, y a la derecha deformación de una parte de la imagen en la proyección.

El posprocesado de la imagen digital optimiza el aspecto de la imagen y así se detecta mejor cualquier anormalidad. Esto se logra además, cuando el tamaño de los píxeles de la pantalla digital es muy pequeño, logrando así un aumento de una región de una imagen para definir los detalles visibles más pequeños. Cuanto mayor es el tamaño de la matriz en las pantallas digitales, mejor resolución espacial presentan, ya que tienen píxeles de menor tamaño.

Por otro lado, como ya se dijo antes la radiación de RX puede lesionar a los humanos, así que algunos de los sistemas de protección principal de la radiación en la medicina son:

- La filtración: en el interior del tubo de RX, se insertan filtros de metal (normalmente de aluminio o de cobre), de manera que los rayos de baja energía con escaso valor diagnóstico sean absorbidos antes de que alcancen al paciente.
- Colimación: restringe el haz de RX útil a la parte del cuerpo de la que debe tomarse una imagen y, por tanto, protege los tejidos adyacentes de una exposición innecesaria.
- Ropa de protección: los guantes y delantales que usan los técnicos radiólogos durante algunos procedimientos radiográficos se fabrican con un material impregnado de plomo.
- Escudo o blindaje gonadal: fabricados con material impregnado de plomo que deben usar todas las personas en edad fértil cuando las gónadas están en o cerca del haz útil de RX, y cuando la utilización de estos escudos no interfiera con el valor diagnóstico de la exploración.
- Barreras protectoras: la consola de control radiográfico siempre está localizada tras una barrera protectora. Con frecuencia la barrera es de barras de plomo y está equipada con una ventana de vidrio plomado. Bajo circunstancias normales, el personal permanece detrás de la barrera durante el examen radiográfico, como se ve en la *Figura 1.2*.

También deben seguirse otros procedimientos como por ejemplo, las exploraciones de RX abdominales de una gestante nunca deben realizarse durante el primer trimestre a menos que sea absolutamente necesario, además deben realizarse todos los esfuerzos posibles para asegurar que una exploración no tendrá que repetirse debido a un error técnico, ya que las exploraciones repetidas someten al paciente al

doble de la radiación necesaria, y como comentamos antes, la sobreexposición a los RX es dañino para una persona debido a las altas energías que estos presentan, por lo que ahora hablaremos de la cantidad de RX que recibe un paciente en el estudio de mamografía.

La intensidad del haz de RX del tubo de RX se mide en roentgens (R o mGya), miliroentgens (mR) o milisievert (mSv), y es lo mismo hablar de cantidad de RX, exposición de radiación, o intensidad del haz de RX. El roentgen (mGya) es una medida del número de pares de iones producidos en el aire por una cantidad de RX, pues la ionización del aire aumenta al elevar el número de RX del haz. En cambio, el sievert (Sv) es la unidad en el SI de equivalencia de dosis de radiación ionizante. Pretende representar la probabilidad de cáncer inducido por la radiación y el daño genético. Un sievert lleva consigo una probabilidad del 5,5% de desarrollar cáncer. En los estudios de mamografía, en promedio la dosis total por mamograma a dos tomas para cada seno es de aproximadamente 0.3 mSv, que es casi la misma que una persona promedio recibe de la radiación de fondo en un mes.

La cantidad de RX producidos en una mamografía aumenta proporcionalmente cuando aumentan los mAs, porque aumentan los electrones proyectil que salen del filamento, y cuando aumentan los kVp en el tubo de RX. Por el contrario, disminuye con la distancia entre la mama y la fuente de RX. Los mamogramas exponen los senos a pequeñas cantidades de radiación, sin embargo los beneficios de la mamografía superan cualquier posible daño que pudiera causar la exposición a la radiación.

1.2 Microcalcificación y Tumor

El Instituto Nacional del Cáncer define a la microcalcificación como:

"depósito pequeño de calcio en la mama que no se puede sentir, pero que se puede detectar en una mamografía. Un conglomerado de estas partículas muy pequeñas del calcio puede indicar la presencia de cáncer".

La aparición de microcalcificaciones es común en las imágenes de mamografía y su frecuencia aumenta con la edad del paciente. Estas pueden verse como diminutos puntos brillantes comparado con el tejido circundante, y pueden estar aislados o agrupados en racimos. Miden entre 0.05 mm y 2 mm de tamaño, y en dependencia de su tamaño, número y disposición puede deducirse su naturaleza benigna o maligna [3]. El léxico BI-RADS (Breast Imaging Reporting and Data System), del Colegio Americano de Radiología, divide las calcificaciones por tipos morfológicos en tres categorías: típicamente benignas, sospecha intermedia y alta probabilidad de malignidad [4]. En general, las calcificaciones de aspecto benigno son grandes y toscas, poseen bordes suaves, en comparación con las calcificaciones malignas que usualmente son pequeñas y difíciles de percibir. En estos casos suele realizarse una biopsia del tejido donde se encuentra la lesión, sin necesidad de ingreso hospitalario y sin dejar cicatrices. Solo en algunos casos podría ser necesario obtener la muestra en quirófano bajo sedación. El reto en el diagnóstico se presenta con las calcificaciones que son ocasionadas por causas tanto benignas como malignas, y que no presentan una morfología característica, como ocurre con las calcificaciones indeterminadas. Es importante diferenciar las microcalcificaciones de origen benigno de las sospechosas puesto que un 55% de los cánceres no palpables se diagnostican mediante la presencia de microcalcificaciones y porque estas son la principal forma de manifestación del carcinoma ductal in situ (CDIS) [4], que es una lesión mamaria precancerosa presente en los conductos lácteos. Su aspecto es el de una agrupación de microcalcificaciones.

La mejor manera de detectar estas lesiones es con una mamografía digital, aunque la tomosíntesis está ganando terreno. En caso contrario, cuando se usa la ecografía la gran mayoría de las microcalcificaciones no son visibles y solo se logra

detectar aquellas de gran tamaño o que están asociadas a nódulos o quistes. Aun cuando las calcificaciones se logren evidenciar como imágenes ecogénicas, no es posible caracterizarlas adecuadamente por este método [4].

En cuanto a la composición, si bien se conoce que está formada por diferentes sales de calcio, con metales como el zinc, magnesio y hierro, estos últimos sobre todo en calcificaciones malignas, la fisiopatología de las calcificaciones es aún no bien comprendida. Prácticamente cualquier tejido que constituye a la mama será capaz de presentar calcificación, y suelen deberse a los cambios evolutivos que afectan a las glándulas mamarias a lo largo de la vida de la mujer, pero el consumo de calcio que ingerimos a través de la alimentación no es la causa. La diferencia de densidad entre las calcificaciones benignas y malignas está dada principalmente por los distintos compuestos químicos que prevalecen en cada una. Mientras las calcificaciones benignas están conformadas principalmente por oxalato de calcio, las calcificaciones malignas están conformadas predominantemente por fosfato de calcio. Ambos tipos pueden coexistir y no es posible determinar sus componentes por mamografía, requiriéndose estudios químicos para estos fines [4].

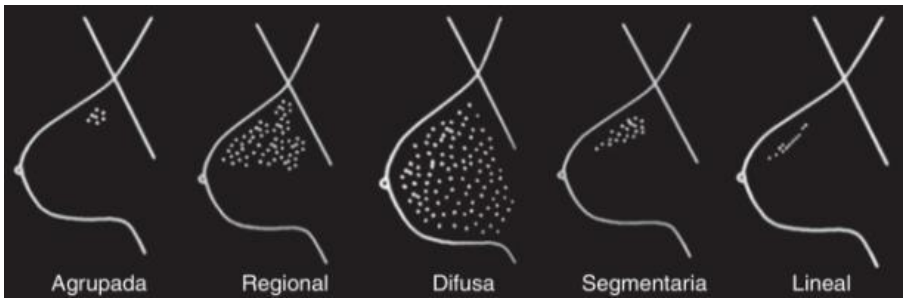

Figura 1.8. Esquema de los descriptores BI-RADS de distribución de las microcalcificaciones.

Existen descriptores BI-RADS de distribución de las microcalcificaciones que se refieren a la disposición de las calcificaciones en el interior de la mama. Los tipos se puede ver en la *Figura 1.8* y son: la distribución difusa, antes llamada dispersa, que se refiere a calcificaciones distribuidas aleatoriamente dentro de la mama. Las

calcificaciones puntiformes y amorfas en esta distribución habitualmente son benignas; la distribución regional describe calcificaciones en un área extensa, mayor a 2 cm en su dimensión mayor. Ya que pueden abarcar más de un cuadrante su riesgo de malignidad es bajo, sin embargo, debe considerarse la morfología para establecer el grado de sospecha. Se describe una probabilidad de malignidad de alrededor de un 26% [4]; la distribución agrupada se refiere a pocas calcificaciones en un área pequeña de tejido. El límite inferior de este descriptor son 5 calcificaciones en 1 cm o cuando exista un patrón definible. El superior se refiere a cuando se encuentra un número mayor de microcalcificaciones dentro de 2 cm [4]. Esta distribución requiere evaluación complementaria con proyecciones magnificadas y debe ser una agrupación en ambas proyecciones para considerarla como tal, puesto que si solo se ve una agrupación en una proyección, pueden corresponder a sobreposición de calcificaciones en distintas posiciones. Éstas se consideran benignas o sospechosas según la morfología de cada grupo. La distribución lineal ocurre cuando las microcalcificaciones se disponen en un trayecto lineal que puede ramificarse, lo que sugiere depósito de calcio dentro de un conducto. Se describe una probabilidad de malignidad de alrededor de un 60% [4]. Finalmente la distribución segmentaria sugiere el depósito de calcio en los ductos y sus ramas, siguiendo la forma anatómica de un lóbulo mamario, es decir, en forma de triángulo cuyo vértice se dirige hacia el pezón. Si bien puede presentarse en patología benigna, su presentación puede deberse a un cáncer extenso o multifocal. Se describe una probabilidad de malignidad de alrededor de un 62% [4].

Retomando el caso del CDIS o carcinoma, éste no es un caso de emergencia pero se convertirá tarde o temprano en cáncer. Si se llega a detectar uno, se debe realizar una biopsia donde se extraiga tejido dañado para luego ser estudiado por un patólogo y determinar su peligrosidad. Cabe aclarar que estos tumores pueden estar en la escala de milímetros en etapas preliminares. En un CDIS mayor a 5 cm, el riesgo de que el cáncer haya ido más allá de los conductos lácteos asciende al 50% [5]. Siempre hay riesgo de que las células tumorales salgan de los conductos lácteos y se extiendan por

todo el cuerpo. Cuanto mayor sea el tamaño del carcinoma, mayor es el riesgo de que sea un carcinoma invasivo. Tan pronto estas células llegan al tejido conjuntivo circundante, también pueden llegar a alcanzar los vasos linfáticos sanguíneos y con ello existe el peligro de metástasis. Esta se presenta cuando las células cancerosas se desprenden del tumor original (primario), viajan por el cuerpo a través de la sangre o el sistema linfático y forman un tumor nuevo en otros órganos o tejidos. El líquido transparente dentro de los vasos linfáticos, llamado linfa, contiene productos derivados de los tejidos y materia de desecho, así como células del sistema inmunitario. Los vasos linfáticos transportan líquido linfático fuera de los senos. En el caso de cáncer de seno, las células cancerosas pueden ingresar en los vasos linfáticos y comenzar a crecer en los ganglios linfáticos. La mayoría de los vasos linfáticos del seno drenan hacia los ganglios linfáticos localizados debajo del brazo (ganglios axilares), los ganglios linfáticos que rodean la clavícula, y los ganglios linfáticos que se encuentran en el interior del tórax y cerca del esternón (ganglios linfáticos mamarios internos). Mientras más ganglios linfáticos haya con células cancerosas del seno, mayor es la probabilidad de encontrar cáncer en otros órganos. Debido a esto, encontrar cáncer en uno o más ganglios linfáticos a menudo afecta su plan de tratamiento. Generalmente, se necesita cirugía para extirpar uno o más ganglios linfáticos para saber si el cáncer se ha propagado. Finalmente el cáncer de seno se puede propagar cuando las células cancerosas alcanzan la sangre o el sistema linfático y llegan a otras partes del cuerpo. Cabe aclarar que el nuevo tumor metastásico es el mismo tipo de cáncer que el tumor primario. Sobre esto, el Instituto Nacional del Cáncer precisa que los tumores son crecimientos anormales en el cuerpo y pueden ser benignos o malignos. Los tumores benignos no son cáncer, pero los malignos sí. Los tumores benignos solo crecen en una parte del cuerpo. No pueden aparecer ni invadir otras partes del cuerpo. Para este trabajo se analizaron imágenes de mamografía con presencia de tumores malignos, por lo que nos centraremos en estos. De este modo, el instituto Nacional del Cáncer define al cáncer mamario como el que se forma en los tejidos del seno (mama). Se presenta tanto en hombres como en mujeres, aunque el masculino es poco frecuente. En México

este tumor maligno es la primera causa de muerte. Se estiman alrededor de 1.7 millones de casos nuevos cada año y fallecen 522,000 mujeres por esta enfermedad [5]. Se diagnostica a una edad media de 52 años, aunque la incidencia aumenta proporcionalmente al incremento de la edad. Los factores que propician la enfermedad son la ingesta de grasas animales, el consumo de tabaco y alcohol, el sedentarismo y sobrepeso, retraso en el inicio de la reproducción (>30 años) y ausencia de lactancias, y uso de agentes hormonales en la menopausia. El riesgo aumenta si el paciente sufrió una lesión en alguna biopsia o si recibió radioterapia en el tórax en la infancia o adolescencia.

El uso de los estudios de imagen como la mastografía o mamografía, el US (ultrasonido mamario) y la resonancia magnética (RM), permiten detectar, caracterizar y evaluar la extensión de la enfermedad y dar seguimiento a lesiones mamarias. Las biopsias percutáneas con aguja de corte y sistemas de corte por aspiración con guía por RX o por US son los métodos de elección en lesiones no palpables con sospecha de malignidad. En especial, la mastografía es el único método de imagen que disminuye la mortalidad por cáncer de mama en un 21%. Además presenta una sensibilidad diagnóstica de 77% a 95% [6]. En específico, la mamografía digital mejora la detección del cáncer de mama en un 27% y reduce la dosis de radiación hasta en un 39% [6].

1.3 La Dimensión Fractal y la Entropía de Shannon

1.3.1 Dimensión Fractal

El concepto fractal se ha aplicado en los campos de la medicina y en particular, en la descripción de imágenes radiográficas [7] [19] [20] [21]. En estas aplicaciones, la dimensión fractal D se ha calculado para caracterizar la textura inherente en algunas

regiones de interés de una imagen. Por ejemplo, una dimensión fractal más baja corresponde a una textura más suave. En el caso de este trabajo, la dimensión fractal calculada de una imagen de mamografía, será un parámetro cuantitativo útil para representar el grado de brillo que presenta la imagen de mamografía. Como se ve en la *Figura 1.9* que es la representación de estructura de alambre de la superficie de una imagen de una mama, ésta tendrá una textura con picos bajos cuando la imagen asociada tenga un grado menor de densidad. De manera similar, cuando hay un alto grado de brillo o intensidad, la imagen aparecerá más irregular, con regiones de picos altos y regiones de picos muy bajos con respecto a los anteriores. Con esto podemos ver la textura de una imagen de mamografía, y por ende la textura del interior de un seno de un paciente.

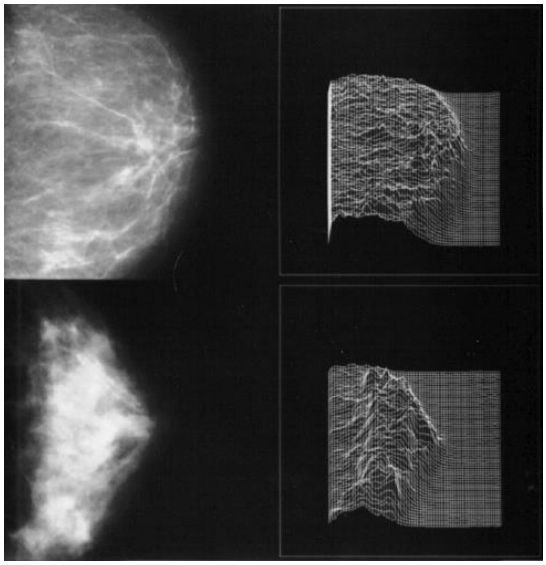

Figura 1.9. Una representación de estructura de alambre de la superficie que se puede asociar con una imagen mamográfica de la mama.

Para calcular la dimensión fractal de un mamograma con tumor o microcalcificaciones, se puede realizar un cálculo análogo a la técnica de *box-counting* [22], que es una de las formas de determinar la dimensión fractal de un objeto, mediante la medición de las irregularidades del borde del objeto, con rejillas de diferentes

tamaños. En este caso, los cuadrantes de cada rejilla estarán determinados por los pixeles de la imagen de mamografía, cada uno cuenta con un brillo i, y una coordenada (x, y) de la imagen digital bidimensional. Justamente esta intensidad i puede tratarse como una dimensión en un espacio fase, que acompaña a la asignación bidimensional de píxeles de tamaño $\varepsilon \times \varepsilon$ en el proceso de digitalización de la mamografía, como se ve en la *Figura 1.10*.

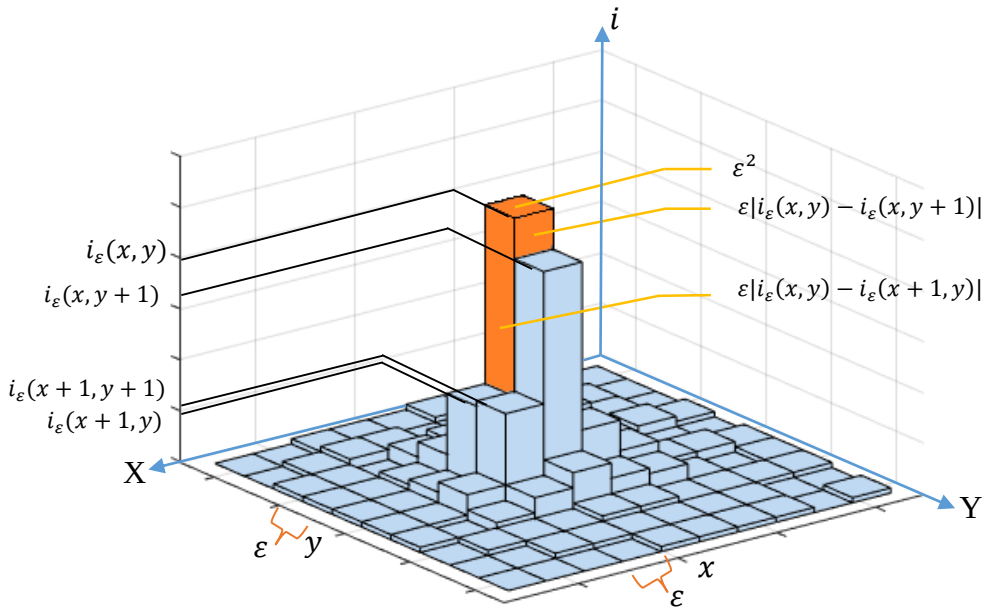

Figura 1.10 Gráfico que muestra las tres dimensiones de un mapa de intensidades.

En esta figura se muestran barras de diferentes alturas i, de ancho ε. La ubicación de cada barra corresponde a un pixel o bin en la imagen de mamografía, con una intensidad i correspondiente y un ancho de ε. Tomando lo dicho en [7], todos tienen el mismo ancho pero diferente intensidad que vamos a representar como

$$i_\varepsilon\,(x,y),\quad i_\varepsilon\,(x,y+1),\quad i_\varepsilon\,(x+1,y),\quad etc.$$

Ahora, para determinar el área $A(\varepsilon)$, que es el área "visible" de cada barra para un tamaño ε de píxel dado, necesitamos primero la suma de las contribuciones de los

lados “expuestos” de las barras, esto es, la diferencia de altura o intensidad entre píxeles vecinos, que se representan como

$$|i_\varepsilon(x,y) - i_\varepsilon(x+1,y)| \quad , \quad |i_\varepsilon(x,y) - i_\varepsilon(x,y+1)|$$

donde $(x,y),(x+1,y),(x,y+1)$ son coordenadas de pixeles vecinos, más el área superior de la barra, que será el área del píxel, esto es ε^2. El ejemplo visual de esto se ve en la *Figura 1.10* donde la parte color naranja es el área $A(\varepsilon)$ para la barra con intensidad $i_\varepsilon\ (x,y)$. Entonces $A(\varepsilon)$ es la suma de la cara superior, esto es ε^2 (pues el ancho del bin mide ε), y de cada lado tiene un área igual a la diferencia de intensidades entre las barras que la rodean consigo misma, multiplicado por el ancho del bin, esto se representa como

$$\varepsilon|i_\varepsilon(x,y) - i_\varepsilon(x+1,y)| \quad , \quad \varepsilon|i_\varepsilon(x,y) - i_\varepsilon(x,y+1)|$$

así el área visible se puede escribir como

$$\varepsilon^2 + \varepsilon\,|i_\varepsilon(x,y) - i_\varepsilon(x,y+1)| + |i_\varepsilon(x,y) - i_\varepsilon(x+1,y)|$$

para cada barra en cada punto (x,y) de toda la imagen. Finalmente la suma de las contribuciones de cada barra queda como

$$A(\varepsilon) = \sum_{x,y} \varepsilon^2 + \sum_{x,y} \varepsilon\ [|i_\varepsilon(x,y) - i_\varepsilon(x,y+1)| + |i_\varepsilon(x,y) - i_\varepsilon(x+1,y)|].$$

El matemático Benoit Mandelbrot presentó en 1982 su idea de “fractal” (que explicaremos más adelante), y mostró que para estas estructuras existe una relación de ley de potencia entre $A(\varepsilon)$, estando el exponente relacionado con la dimensión fractal. Específicamente, para una imagen bidimensional, la dimensión fractal viene dada por

$$D = 2 - \frac{\log[A(\varepsilon)]}{\log(\varepsilon)}. \tag{1.1}$$

La medición de fractales involucra conceptos diferentes a los conceptos geométricos clásicos. Como un fractal se construye de elementos cada vez más pequeños, la longitud es un concepto que no cabe en estas figuras. Si quisiéramos medir un segmento de una de éstas figuras, siempre habrá patrones u objetos menores que escaparán de la sensibilidad del instrumento con el que se mida. Un ejemplo de esto es la longitud de una costa como se menciona en [8]. Si tuviéramos un mapa de la costa a cierta escala y después, con una regla de cierta medida, contamos cuantas veces cabe nuestra regla en el perímetro de la costa, para saber su longitud L, esta será el número de segmentos contados (n) multiplicada por la escala de la regla en kilómetros $(escala)$, esto es

$$L = (escala)n.$$

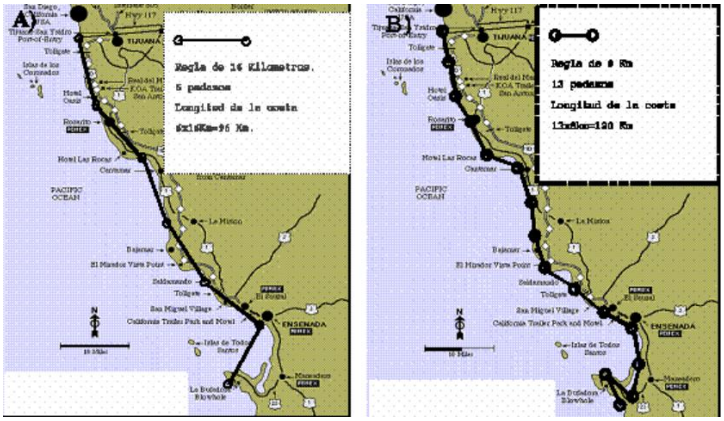

Figura 1.11 Medición de la longitud de la costa utilizando una regla de 16 Km (izquierda) y 8 Km (derecha).

Por ejemplo, si la escala de la regla es 16 km, y se obtienen seis pedazos como se puede ver a la izquierda de la *Figura 1.11*, entonces L será igual a 96 Km. Sin embargo, al agrandar el mapa, muchos detalles no pueden medirse en esta escala tan grande, ya que su tamaño es menor a la regla utilizada, pero podemos utilizar una regla

más pequeña para medir los detalles, por ejemplo, una regla de 8 Km, y entonces el conteo de n es de 15 pedazos como se ve a la derecha de la *Figura 1.11*, siendo L =120km. A medida que medimos con mayor precisión, pueden verse más detalles de la costa, por lo cual la longitud crece. Así, la longitud de una costa depende de la escala en la cual se mida. La forma de esta dependencia puede obtenerse inmediatamente si graficamos, tal y como se muestra en la *Figura 1.12* para diferentes costas y fronteras, el $Log_{10}(L)$ contra el $Log_{10}(escala)$. Puede observarse que todas son líneas rectas que se caracterizan por diferentes pendientes, cuyo valor denotaremos por -DF y ordenadas al origen que llamaremos $Log_{10}(L_0)$. Las ecuaciones de estas rectas son del tipo,

$$Log_{10}(L) = -DF\ Log_{10}(escala) + Log_{10}(L_0)$$

$$\Rightarrow L = L_0(escala)^{-DF} \tag{1.2}$$

así, obtenemos la dependencia de la longitud de la costa en términos de la escala.

El exponente DF que caracteriza la rugosidad de la costa resultó ser mayor para costas muy rugosas, y menor para costas poco accidentadas [8].

Si ahora llevamos estas ideas a la geometría euclidiana, podemos repetir el mismo procedimiento para medir una costa totalmente recta de tamaño L. Si escogemos reglas de tamaños L/n donde n es un número entero, entonces el número de veces que cabe la regla de escala (L/n) en L es simplemente n. La longitud de la recta está dada por

$$(escala)(n) = (L/n)n = L$$

de donde obtenemos que,

$$n = L(escala)^{-1}. \tag{1.3}$$

Si ahora medimos el área de un cuadrado con lados de tamaño L, contando cuántos cuadrados de lado (L/n) caben en el cuadrado grande, es claro que ahora caben n^2, entonces se tiene que

$$n^2 = L^2 \left(\frac{L}{n}\right)^{-2}$$

de donde

$$n^2 = L^2 \, (escala)^{-2}. \tag{1.4}$$

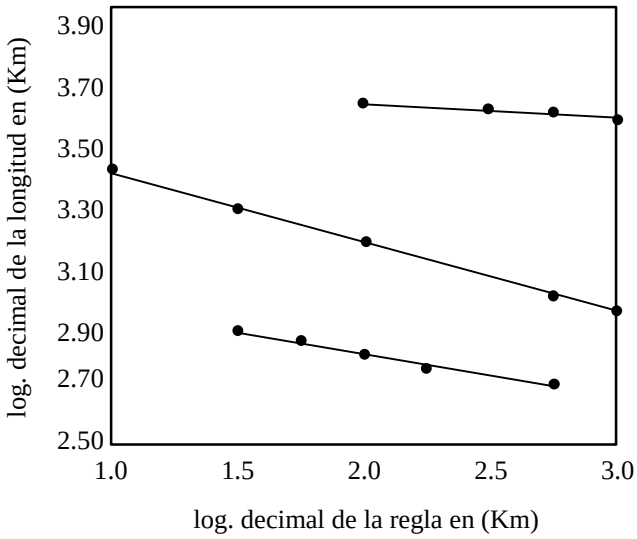

Figura 1.12 Logaritmo de las longitudes aproximadas de diversas costas y fronteras en función del logaritmo de la escala.

De acuerdo a los exponentes de las ecuaciones (1.3) y (1.4) y el exponente de la ecuación (1.2), es claro que la longitud de una recta tiene una DF $= 1$, mientras que un cuadrado tiene DF $= 2$. Ahora recordamos que la recta es una figura geométrica unidimensional, siendo el cuadrado una figura en dos dimensiones. Siguiendo esta idea, Mandelbrot presentó que el fractal es un objeto irregular que tiene la propiedad de autosimilitud, que consiste en estar formado por partes también irregulares que, si son aumentadas de tamaño, se muestran prácticamente iguales a su todo y a su vez están formadas de partes más pequeñas sucesivamente con la misma propiedad, a lo que se

conoce como reiteración. Además propuso identificar a la DF con la dimensión de los objetos. Esta dimensión no es igual a la dimensión del espacio donde se sitúa la figura. El número DF puede ahora ser un número fraccionario por lo cual se le llamó dimensión fractal. Una figura con dimensión fractal $\mathrm{DF} = 1.3$ es una curva irregular, ya que tiene una dimensión entre una recta y un plano. Si la curva tiene dimensión $\mathrm{DF} = 2$, sería tan irregular que alcanzaría a cubrir de manera uniforme el plano, mientras que una curva con dimensión fractal 1 es una recta [7]. Lo importante de este número es que pone de manifiesto una clase de simetría, que consiste en que muchos objetos son similares a sí mismos aunque cambiemos la escala con la cual los miremos. Un árbol es el ejemplo cotidiano más sencillo que podamos imaginar. Su tronco se subdivide en ramas más pequeñas, y cada rama en otra rama y así sucesivamente. Sin embargo, si amplificamos suficientemente una sola rama junto con sus ramas más pequeñas, su aspecto general será la del mismo árbol. Así, podríamos decir que las partes del todo están formadas por versiones más pequeñas del todo [8].

1.3.2 Entropía de Shannon

Desde que las telecomunicaciones se hicieron presentes en la civilización, el concepto técnico de lo que es la información se tornó importante. Se hizo a su vez necesaria una medida que cuantificara este concepto. En 1928 Hartley define una primera medida de la cantidad de información de una secuencia de datos. Consideró un mensaje constituido por símbolos, cada uno con una probabilidad S de ocurrencia. De esta forma, en una palabra de n símbolos, y no tomando en cuenta las reglas sintácticas, hay s^n posibles mensajes. Aunque lo que Hartley buscaba era una medida que se incrementara con la longitud de mensaje, notó que tomar $H = s^n$ era poco realista. Su búsqueda estaba más enfocada hacia medidas del tipo $H = kn$, con k constante. Escogió finalmente

$$H = n\,log(s)$$

la cual resultaba en principio adecuada ya que cuando se tiene sólo un símbolo no se tiene información porque $H = 0$ (se conoce el símbolo que se va a recibir), y lo mismo ocurre cuando $n = 0$, es decir, cuando no hay mensaje. Shannon, en su famoso artículo en el que introdujo las bases de la teoría de la información, adapta esta medida para símbolos de diferente probabilidad de ocurrencia $p(i)$ de la siguiente forma

$$S = -\sum_{i=1}^{n} p(i)\, Log_2\big(p(i)\big). \tag{1.5}$$

Esta cantidad se llama entropía y puede verse entonces, como una medida de incertidumbre, ya que su máximo se da cuando todos los símbolos tienen igual probabilidad de ocurrencia, mientras que cuando un símbolo tiene mayor probabilidad de aparecer que el resto, sabremos instantáneamente que obtendremos este símbolo la mayor parte de las veces.

Las imágenes pueden también ser consideradas como portadoras de información. Si convertimos una imagen cualquiera en un mapa en escala de grises, en lugar de probabilidades de ocurrencia de letras en un mensaje, se tendría distribución de valores de gris. A modo de ejemplo, es sencillo inferir que una imagen casi uniforme en tonos de gris será poseedora de muy poca información. La distribución de probabilidad de los valores de gris se puede estimar contando el número de veces que aparece cada valor de gris en la imagen y dividiendo el conteo por la cantidad total de pixeles. De esta manera una imagen con una gran cantidad de sus pixeles con la misma intensidad va a tener una baja entropía, mientras que con una distribución menos uniforme de los valores de gris se obtiene una alta entropía. De esta manera puede verse la entropía de Shannon como una medida de la distribución de probabilidad de los valores de gris. Lo mismo sucede con la imagen de mamografía, pues en cada punto

tiene una tonalidad de gris, o un cierto nivel de brillo como se habló antes. De este modo, para el caso tratado en el presente trabajo, seguimos considerando la intensidad i de cada píxel en cada punto (x, y) de la imagen bidimensional de mamografía. Entonces usamos la entropía de Shannon donde $p(i)$ es el valor de la intensidad que tiene el i-ésimo pixel y n es el número total de pixeles de la imagen. Así, lo que se hace para calcular S es sumar las intensidades de cada coordenada (x, y) y multiplicarlas por el logaritmo base dos de esa i-ésima intensidad, como se menciona en la ecuación 1.5.

Estas dos cantidades, DF y S, serán calculadas en el MC-área, el T-área y el t-área de cada muestra, y dependiendo de la correlación existente entre cada una, podremos distinguir los elementos que presenta cada muestra con MC o T.

Capítulo 2

Metodología

En este capítulo se presentan la serie de pasos que se siguieron para obtener el cálculo final de la entropía y dimensión fractal de cada muestra, además del tratamiento que se le dio a cada imagen de mamografía.

2.1 Redigitalización de mamografías

Lo primero que se realizó para lograr el objetivo de este trabajo fue digitalizar nuevamente la imagen de mamografía. Ésta ya se presenta en formato digital una vez realizado el estudio al paciente. Para el presente trabajo se redigitalizaron imágenes en formato .PNG o .JPG mediante un código de programación escrito en C++, mediante el cual dichas imágenes se convirtieron en imágenes normalizadas en escala de grises. A estas se les llama mapa de intensidades o mapa en escala de grises.

El hecho de crear una nueva imagen de la mama, a partir del mamograma, con la que podamos estudiar las MC y T, consiste en asignarle a cada bin de la imagen de

mamografía, una tonalidad de la escala de grises (del color oscuro al claro), que dependerá de que tan intenso sea el brillo de dicho bin, en la imagen original. Si un bin es muy brillante se le asignarán colores claros en la escala, y además se le asociarán valores altos de intensidad. Como la escala está normalizada, los valores van de cero a uno. Lo contrario sucede con bines oscuros, a los cuales se les asignarán valores menores de intensidad. Es decir, las zonas más brillantes en la imagen de mamografía (las zonas blancas), tendrán valores más altos en la escala de intensidad y por consiguiente tonos claros en la escala de grises, mientras que las zonas opacas, se asociarán a valores bajos de intensidad y por ende, a tonalidades oscuras en la escala de grises. Esto pasará para todos y cada uno de los bins del mamograma, entonces a los que presentan un brillo intermedio, se les asociará su correspondiente valor intermedio de intensidad y su correspondiente tonalidad de la escala de grises. La escala de intensidad que se relaciona con la escala de grises, va desde el valor cero hasta el valor uno, siendo las zonas brillantes las más intensas y las cercanas al valor uno, y las zonas oscuras las cercanas al valor de cero. Como ejemplo, el resultado de este proceso se ve en la *Figura 2.1*, donde se hace la comparación de la mama en la imagen de mamografía (izquierda), contra el mapa en escala de grises (derecha). En esta figura se aprecia la correspondencia del brillo y el color entre una imagen y otra, además se ve que se mantiene la estructura de la mama tras este tratamiento.

Ahora expliquemos cómo se normalizó el mapa de intensidades. Con normalizar nos referimos a que se ajustó un sistema de coordenadas (sin unidades), con origen en el centro de la imagen, y la longitud de cada semieje se hizo de una unidad, así, las coordenadas x y y de cualquier punto en la imagen, van desde el valor menos uno hasta el valor uno. Esto porque no tiene relevancia la medida de la imagen redigitalizada, ya que no es una imagen a escala ni a tamaño real del seno del paciente. Más bien lo que importa es la relación que se hace entre un punto del mapa en escala de grises y el mamograma original, la cual permite ubicar un elemento en la imagen

redigitalizada en un punto (x, y) y poder transformarlo a un punto (x, y) ahora en el mamograma original, dando así la ubicación de cualquier elemento deseado, en nuestro caso una MC o un T. Entonces, si sabemos las coordenadas de la MC o del T en el mapa en escala de grises, sabemos la ubicación de estos elementos en el mamograma, y por consecuente el médico sabrá la ubicación exacta de los mismos en el seno del paciente. Para que quede claro este punto, pongamos el ejemplo de las muestras con T. Se detallará el mamograma original más adelante, pero para efectos de este ejemplo podemos decir que cada imagen de mamografía mide 1024 píxeles de cada lado, y nuestra imagen en escala de grises mide dos unidades de cada lado (de $x = -1$ hasta $x = 1$, y de $y = -1$ hasta $y = 1$), entonces si un elemento se encontrase en las coordenadas (x, y) del mapa de intensidades, en la mamografía se encontraría en el punto $(\frac{1024x}{2}, \frac{1024y}{2})$. Así, todas las imágenes se hacen del mismo tamaño y resulta práctica la ubicación de un determinado punto en ellas.

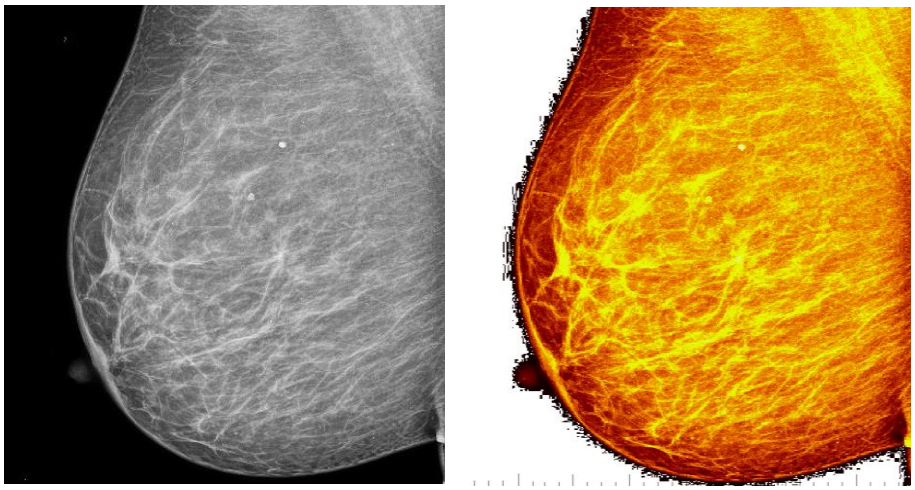

Figura 2.1 A la izquierda la imagen de mamografía y a la derecha la nueva imagen en escala de grises.

Todo lo antes mencionado forma parte del primer paso del método que es la redigitalización de la imagen de mamografía. La muestra de esto se ve en la *Figura 2.2*, y es sumamente importante porque de esta manera podemos tener caracterizados todos los puntos de una muestra, debido a que la redigitalización, en resumen, le dio

tres propiedades a cada punto del mamograma, la coordenada en y, la coordenada en x, y el valor de intensidad i, que es la tercera coordenada. Así, para identificar un punto basta con conocer las tres coordenadas (x, y, i) de este. Los datos resultantes del proceso explicado hasta ahora, nos servirán para los siguientes pasos.

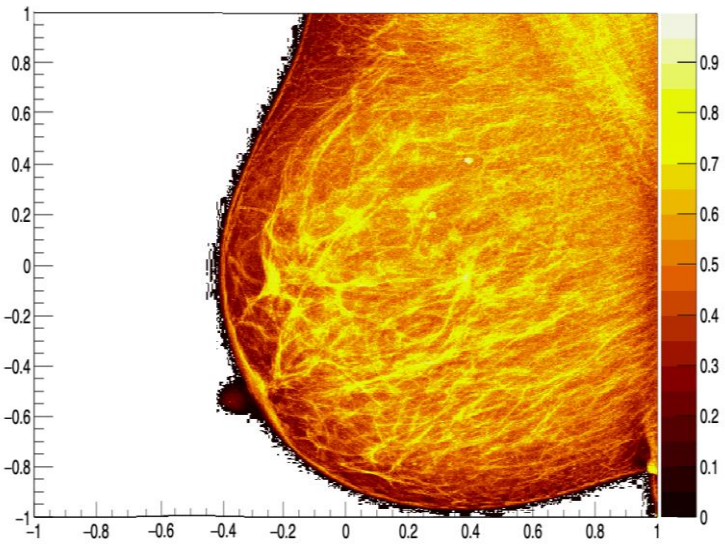

Figura 2.2 Mamograma normalizado y en escala de grises. Se muestran los ejes y la escala que relaciona el color con la intensidad, a la derecha.

2.2 Selección de elementos a analizar

2.2.1 Microcalcificación, Candidato y Tejido

Para el estudio de MC, se tuvo acceso a cinco mamogramas hechos en el Hospital de la Mujer Puebla, todos con MC ya identificadas y ubicadas gracias al diagnóstico de la Doctora Karla Herrera quien proporcionó estas muestras.

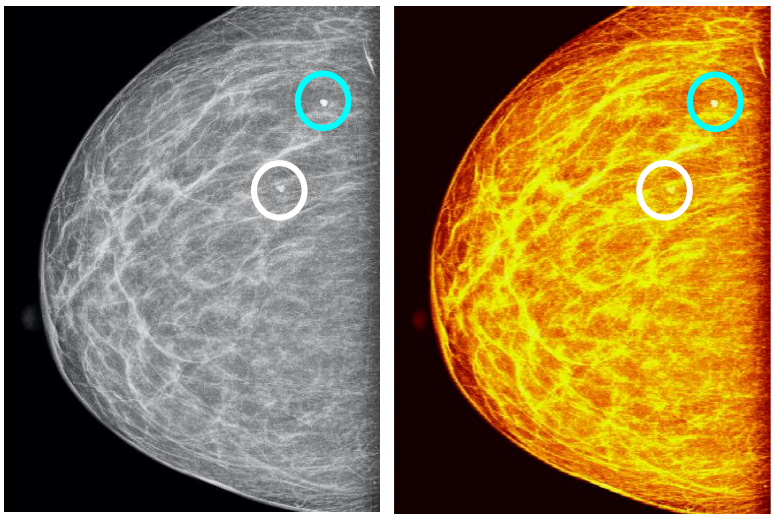

Figura 2.3 Imagen de una mama con la MC encerrada en el círculo azul y el C en el blanco.

Cada una se sometió a la redigitalización y se visualizaron áreas pequeñas muy brillantes que rondan el color blanco en la escala de grises y los valores máximos de intensidad de la imagen total. Además, gracias a que la imagen se normalizó, se ubicó el punto más intenso de toda la imagen, un punto con coordenadas $(x, y, i_{máx})$, al cual se le llamó MC-área. Sin embargo la imagen lógicamente puede contener algunas otras áreas que igualmente tengan valores de intensidad muy por encima del valor promedio de la imagen total, es decir, que visualmente haya más de un área o punto blanco que resalte del resto. Para estos casos de múltiples máximos tomamos los dos puntos más intensos. Al más intenso se le identificó como MC-área y al siguiente lo nombramos Candidato (C-área). Entonces de las cinco imágenes de mamografía, los elementos que se tomaron en cuenta para ser estudiados fueron la MC-área, el C-área y todos los puntos restantes de la mama que no pertenecen a estos elementos, a los que se les llamo tejido-área. En la *Figura 2.3* el diagnóstico del médico determinó que el seno tiene MC y por ello se pueden apreciar puntos blancos que se señalan con un círculo, no obstante, a estas alturas del proceso no podemos asegurar que todos los puntos se deban a MC, por lo que el encerrado con color azul es MC-área y al otro se identificó como C-área.

Para tener mayor control de los elementos mencionados, éstos se aislaron del resto de la imagen. Se tomó el área que encierra a la MC (a la cual nos referiremos como MC-área), mediante cortes seleccionados por nosotros mismos al considerar que se encierran los bines más brillantes (los valores más altos de intensidad). Estos puntos brillantes representan MC y lo del rededor es tejido circundante. Esto se puede afirmar porque si el centro de la mancha presenta una intensidad $i_{máx}$ y los bines claros tienen intensidades similares, además el rededor tiene intensidades bastante bajas en comparación con la $i_{máx}$, se puede asumir que toda la zona brillante es la MC. Un ejemplo de esto se muestra a la izquierda de la *Figura 2.4*, que es el MC-área de la muestra de la *Figura 2.3*, que se encerró en un círculo azul.

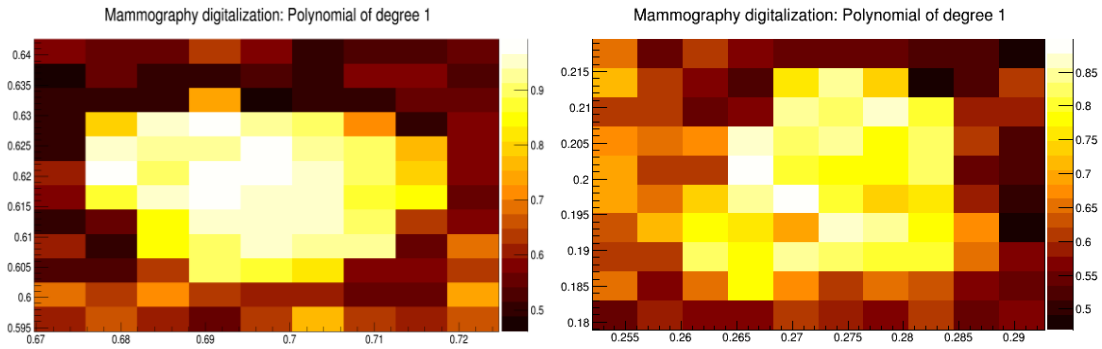

Figura 2.4 MC-área (izquierda) y C-área (derecha), de la muestra de la Figura 2.3.

Una vez seleccionado el cuadrante de la MC, se tomó el área que encierra al C (C-área) y se visualizó algo similar a la MC-área, una zona pequeña y brillante. Un ejemplo de esto se muestra a la derecha de la *Figura 2.4*, en donde se observa el C-área de la muestra de la *Figura 2.3*, que se encerró en un círculo blanco. Es importante mencionar que el cuadrante del C tiene el mismo número de bins que el cuadrante que encierra a la MC, para que la estadística que se le hizo a los datos de ambos elementos fuera igual para todos y los resultados no dependieran del tamaño del cuadrante seleccionado.

Después se seleccionó el cuadrante de tejido (t-área), el cual se tomó de una parte de la imagen que no abarcara ningún bin perteneciente al C o a la MC. El cuadrante se tomó del mismo tamaño que el de los otros dos elementos y el resultado fue una sección con diferentes intensidades distribuidas de manera no uniforme, al contrario del caso de MC-área o C-área, en donde las intensidades altas se concentran en el centro y las bajas en el rededor. Un ejemplo de cuadrante de tejido se ve en la *Figura 2.5*, en donde se presenta tejido de la muestra de la *Figura 2.3*.

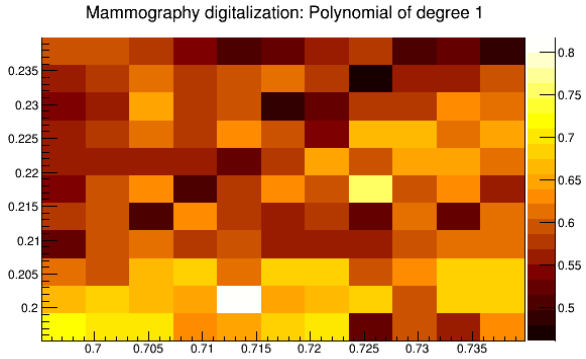

Figura 2.5 Área del tejido de la muestra de la Figura 2.3.

En resumen, en este paso lo que se logró fue identificar las coordenadas (x, y, i) de cada bin de la imagen, para después en base a las intensidades máximas, ubicar las coordenadas correspondientes a la MC y al C, y entonces poder seleccionar un cuadrante de tejido libre de algún otro elemento. Esta selección es un recorte de la imagen total, que abarca la anormalidad aislada en un cuadrante de aproximadamente su tamaño y con un número menor de bines que el total de toda la imagen, gracias al cual fue posible saber el rango que abarca la MC y el C en los ejes x y y.

2.2.2 Tumor y Tejido

En este caso se trabajó con muestras tomadas de la página Pilot European Image Processing Archive (PEIPA): la base de datos de mamografías miniMIAS [9]. Esta

página es un archivo relacionado con el procesamiento y análisis de imágenes con énfasis en la visión por computadora. Su objetivo principal es proporcionar información, conjuntos de datos y software que permitan medir y comparar la efectividad de algoritmos asociados a diferentes áreas. En cuanto al diagnóstico de radiografías por computadora, PEIPA pone a disposición una lista de imágenes de mamografía con T y su diagnóstico correspondiente que incluye: carácter del tejido de fondo (graso, graso-glandular o denso-glandular), gravedad de la anormalidad (benigno o maligno), coordenadas (x, y) del centro de la anormalidad y el radio aproximado (en píxeles) de un círculo que encierra completamente al tumor. De dicho archivo se seleccionaron diez mamogramas con T malignos y se redigitalizaron, con lo cual se obtuvo el punto más brillante con coordenadas $(x, y, i_{máx})$. La imagen resultante se ve como a la derecha de la *Figura 2.6*, en donde se observa que el T es un área brillante de gran tamaño en comparación con la MC, de ahí la importancia de darle un radio que encierre la anormalidad.

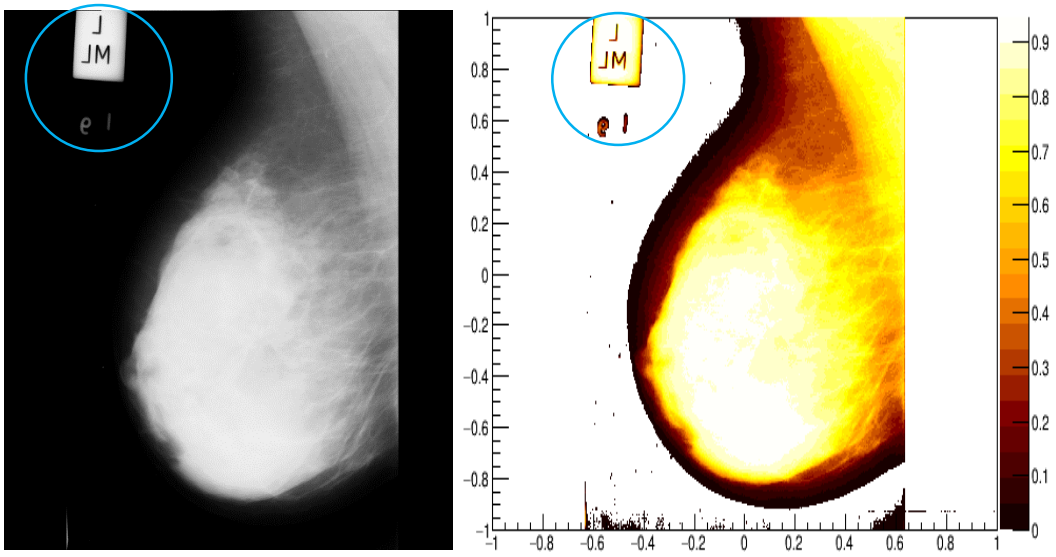

Figura 2.6 Mamograma y redigitalización de muestra con T y con artefacto encerrado en un círculo.

Un problema que se presentó en estas muestras fue la aparición de los artefactos. Como se observa en la *Figura 2.6* (una de las muestras tomadas para mamografía con T), en el borde superior izquierdo se encuentra encerrada una imagen ajena al seno del paciente. Lo que se hizo fue aislar el artefacto de la imagen completa para poder trabajar con la parte restante, así nuevamente se realizó la redigitalización de la imagen. En general esto se realizó con todas las muestras que presentaron este problema. Igualmente si el artefacto hubiese sido un pixel dañado en la zona de la mama, puede aislarse y decirle al programa que analice el resto de la imagen.

Una vez resueltos estos problemas en cada muestra, se aisló al T y al tejido. El T se encerró en un cuadrante (T-área) que midiera de cada lado una longitud igual ala del radio establecido en el diagnóstico. La imagen resultante es como el ejemplo de lado izquierdo de la *Figura 2.7*, en donde se ve que los bins con intensidades altas se presentan en el centro de la imagen y las zonas más oscuras alrededor de la zona amorfa brillante. Todos estos bins pertenecen al T-área y se sabe porque presenta intensidades similares a las del punto central del T y además tiene tejido alrededor cuyas intensidades son muy por debajo de las intensidades del T.

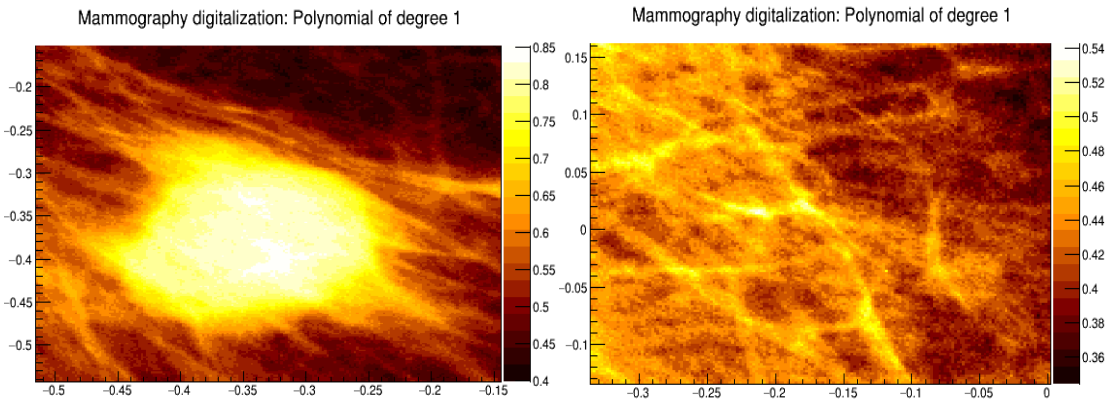

Figura 2.7 Acercamiento al T (izquierda) y al tejido (derecha) de la muestra con T.

El caso contrario se muestra en el ejemplo de lado derecho de la *Figura 2.7*, en donde el color está igualmente distribuido en toda la imagen, tanto las zonas claras

como las oscuras están en cualquier lado de la selección. Para este caso se seleccionó una zona del mismo tamaño que el T-área y que no abarcara ningún bin de este elemento.

Por otro lado, mencionábamos anteriormente que en los mamogramas con MC identificamos a un C, porque resultó que habían varias zonas muy brillantes, y es que generalmente en una mama pueden haber varias MC, al contrario del caso de los T que no se encuentran varios en una misma, entonces en estas muestras no identificamos un C, aunque sí se presentaron igualmente varios máximos que se debieron a artefactos como ya se mencionó antes, pero fueron descartados.

Finalmente en esta sección lo que se obtuvo fueron los valores de intensidad para cada bin de la imagen completa, incluyendo el T-área, con lo cual se pudo saber la ubicación exacta del centro de este elemento. Con el radio dado por el diagnóstico se pudo encerrar el T en un cuadrante y con ello también se pudo seleccionar el cuadrante de tejido.

2.3 Cálculo de la dimensión fractal y la entropía

En la sección anterior se explicó cómo se obtuvieron los elementos a analizar separados en distintos cuadrantes. Se trabajó con tres para cada una de las cinco muestras con MC, y dos para cada una de las diez muestras con T. Lo que se pretende explicar ahora es el proceso del cálculo de DF y S para cada uno de los cuadrantes mencionados.

Para el caso del cálculo de DF, debemos ver el mapa de intensidades como se muestra en la *Figura 1.10*. Los ejes x, y abarcan un rango de $(-1,1)$, y el eje que sale

del plano es el eje de las intensidades, con un rango de (0,1). En esta figura se muestran barras de diferentes alturas i, de ancho ε. Para nuestro caso la ubicación de cada barra corresponde a un bin en el mapa en escala de grises, con una intensidad i correspondiente y un ancho de bin de ε. Así pues, el cálculo de DF para cada cuadrante, por separado, se calculó con la ecuación 1.1. El bin para todas las imágenes mide $\varepsilon = 0.007$ y se usó un ciclo iterativo para sumar el área "visible" de cada barra, con la diferencia de intensidades entre la $i_{máx}$ que presenta el cuadrante y las de su alrededor. Para el caso de MC-área o C-área, la $i_{máx}$ será la del punto central del mismo elemento, y esta se restará con las intensidades del resto de la MC o C y del tejido que los rodea. Para el caso del T sucede lo mismo, la $i_{máx}$ es la del centro del T, y ésta se usó para obtener la diferencia de intensidades con el resto del T-área y con el tejido que lo rodea. Para el caso del tejido se tomó la intensidad más alta que presentó el cuadrante y de ahí se sumaron las contribuciones de cada bin de toda la selección.

Es importante aclarar que no se debe confundir el área geométrica de la selección o cuadrante a analizar para cada elemento (MC-área, C-área, T-área y t-área), con el valor de $A(\varepsilon)$ de la ecuación 1.1. El primero representa el tamaño de la sección elegida usando la métrica euclidiana en el espacio de fase, mientras que $A(\varepsilon)$ se determina por el área "visible" del esquema en que se manejan las coordenadas x, y, i en este trabajo, y por ende su valor depende de la intensidad i y el tamaño ε del píxel de la imagen.

Para el caso del cálculo de S usamos la fórmula de Shannon ecuación 1.5, donde $p(i)$ es el valor de la intensidad que tiene el i-ésimo bin del cuadrante de cada elemento a analizar y n es el número total de bins contenidos en el cuadrante. He aquí la importancia de que los cuadrantes de cada elemento de la misma muestra midan lo mismo o tengan el mismo número de bins, si el MC-área es más pequeño que el

cuadrante del tejido (de la misma muestra), este último tendrá un mayor número de bins que el primero, lo que puede provocar un cambio en el cálculo de S.

Finalmente, con lo explicado en esta sección se obtuvieron valores de DF y de S para cada uno de los elementos (MC-área, C-área y t-área) de las cinco muestras con MC y para cada uno de los elementos (T-área y t-área) de las diez muestras con T.

Capítulo 3

Resultados

En este capítulo se presentan los datos resultantes gracias al proceso que se le hizo a las imágenes de mamografía, explicado en el capítulo anterior. Además se hará un análisis de los datos mediante histogramas, para obtener información nueva, aclarar algunos resultados y crear conclusiones sobre lo obtenido.

El primer resultado importante que obtuvimos es que a partir del mapa de intensidades, es posible distinguir entre MC-área y T-área del resto de la mama, porque el respectivo valor de mayor intensidad de toda la imagen se ubica en estas dos áreas. Cuando un radiólogo examina la imagen de mamografía, lo que busca son áreas o tipos de tejido que se vean diferentes al tejido normal y en algunos casos se debe precisamente a las partículas de calcio (MC), que aparecen como pequeños puntos muy brillantes, o pueden ser debidas a tumores que aparecen como zonas amorfas igualmente brillantes. El hecho de este fenómeno se debe a la densidad del elemento analizado, como se explica en el capítulo 1. Por ende en el MC-área y en el T-área estarán los valores más altos de intensidad de la imagen redigitalizada, por ser las zonas más brillantes de todo el mamograma.

Para todas las muestras con MC, la intensidad mayor de toda la imagen, con coordenadas $(x, y, i_{\text{máx}})$ correspondió precisamente a la ubicación dada por la Doctora Herrera para las distintas MC. En los cinco mamogramas ocurrió lo mismo, los resultados del método que empleamos en esta tesis concuerdan con los del diagnóstico. Para todas las muestras con T, igualmente la mayor intensidad de toda la imagen correspondió al centro del T dado en el diagnóstico. Esto nos ayuda a ubicar estos elementos en la imagen de mamografía.

3.1 Análisis de mamogramas con microcalcificación

Con cada cuadrante seleccionado como en la sección 2.2.1, se realizó el cálculo de D y S. Los valores obtenidos para cada área de las 5 muestras se indican en la *Tabla 3.1*. En la *Figura 3.1* se muestra la gráfica de los datos de esta tabla. Cada línea de color en la gráfica pertenece a los datos de una muestra diferente. Cada línea tiene señalados los valores del tejido, MC y C, obtenidos en dicha muestra. Se puede observar que, para todas las muestras, la entropía del MC-área es menor en comparación con la obtenida para el t-área. La dimensión fractal por el contrario, es menor para el tejido que para el MC- área, en todos los casos de nuestras muestras.

Tabla 3.1 En la primera columna aparece el número de muestra con MC y en las siguientes la dimensión fractal y entropía para cada elemento.

Muestra	Dimensión fractal			Entropía		
	MC	Candidato	Tejido	MC	Candidato	Tejido
1	1.5738	1.4889	1.52464	18.4916	19.7352	26.781
2	1.55654	1.53991	1.34919	16.3955	17.2283	22.0053

3	1.84757	1.56296	1.74482	17.9145	17.9677	21.0235
4	2.19221	2.14136	1.8985	296.044	329.451	340.664
5	1.61021	1.56922	1.44533	22.4045	24.1162	37.9038

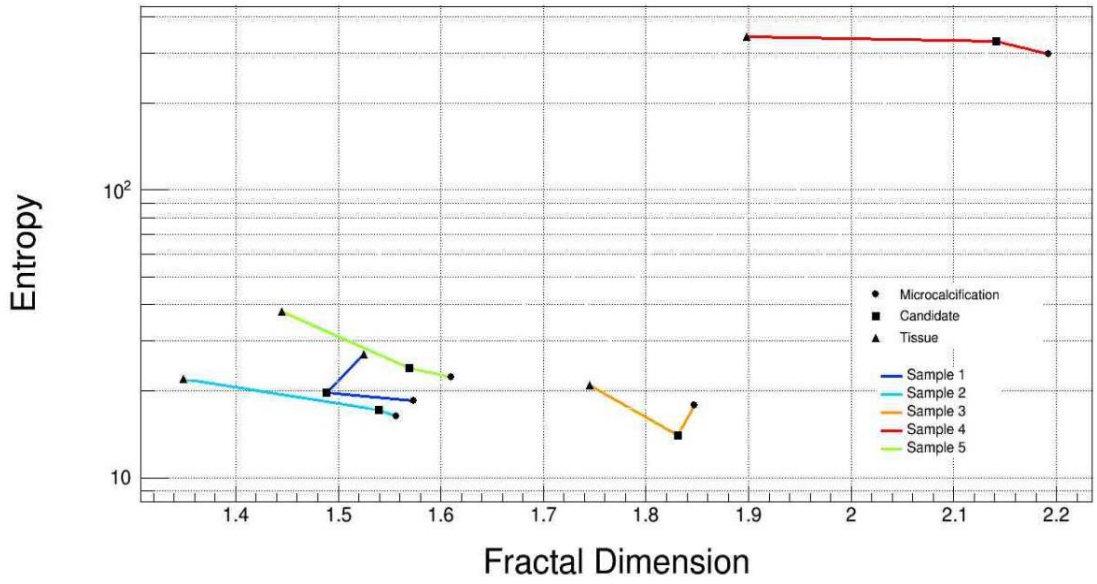

Figura 3.1 Dimensión fractal vs Entropía para cada cuadrante de cada muestra con MC.

Se tomaron los valores de intensidad de cada punto en el MC-área, C-área y t-área, y se construyó un histograma para cada elemento con estos datos, en donde el eje x corresponde a la escala de intensidades (de cero a uno), y el eje y corresponde al número de bins con determinada intensidad. Los histogramas hechos con los datos del cuadrante que presenta MC se ven como en la *Figura 3.2* el cual viene de la muestra de la *Figura 2.4*. El título del gráfico muestra el rango de valores que toma la coordenada x y la coordenada y en dicho cuadrante es decir, que muestra dónde se hizo el corte en la imagen completa. Como se puede observar, el histograma muestra dos gaussianas, lo que se refiere a que en esa área se encuentran dos conjuntos de intensidades, un cierto número de bins tiene intensidades altas y otros intensidades bajas. Debido a esto, se ajustaron todos los datos a dos funciones gaussianas. El ajuste

se muestra en color rojo en la *Figura 3.2*. Esto nos ayuda a saber el valor promedio de cada distribución e identificar a cada una con un valor de sigma, este valor indica qué tan concentrados están los datos alrededor de una barra central con un valor de intensidad promedio. Entonces en la etiqueta del gráfico podemos ver estos datos.

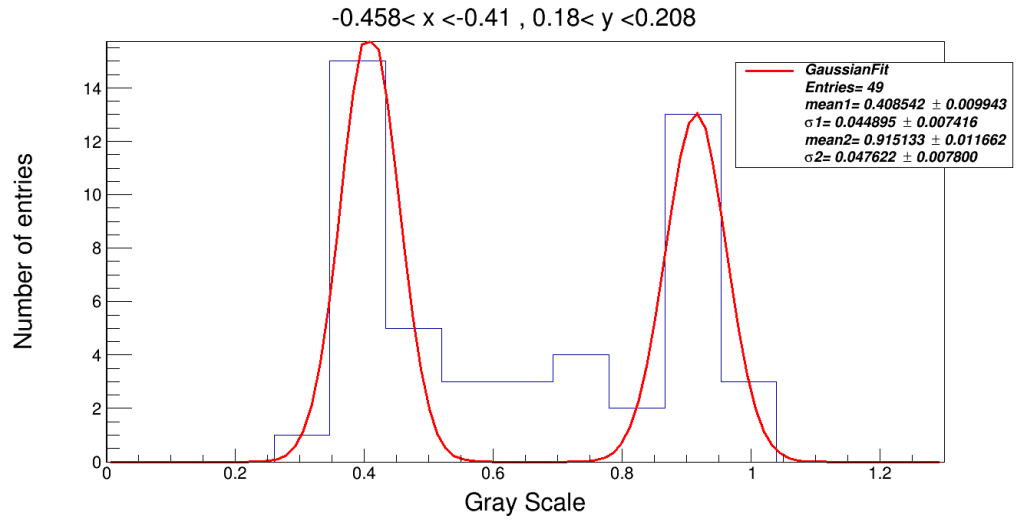

Figura 3.2 Histograma con las intensidades del cuadrante que abarca a la MC.

- *Entries* indica el número de bins que tiene el cuadrante, con los cuales se hizo el histograma. Debido a que los tres cuadrantes de cada muestra se hicieron del mismo tamaño, el número *Entries* de cada uno de ellos debe ser muy similar uno del otro.
- *mean1* indica el valor promedio de intensidad que presentan los datos de la derecha, los de intensidades altas.
- $\sigma 1$ señala el valor de sigma correspondiente a la función con *mean1*.
- *mean2* indica el valor de intensidad promedio que presentan los datos de la izquierda, los de intensidades bajas.
- $\sigma 2$ señala el valor de sigma correspondiente a la función con *mean2*.

Como sabemos, la MC tiene la intensidad más alta de toda la imagen, y ya que la *Figura 3.2* tiene los datos del cuadrante con MC, entonces el bin con $i_{máx}$ estará contabilizado en la distribución de datos con intensidades altas, el conjunto de datos de la derecha. Dado que el grupo al que pertenece este bin tiene una distribución separada del resto de datos, podemos afirmar que los bins que se ajustaron a la campana con *mean1* pertenecen todos a la MC. De ahí que caractericemos a este elemento con el valor de la sigma que tiene su ajuste de datos.

Para el caso del C, de la misma muestra, sucedió algo muy similar. Se realizó su histograma y se obtuvo la *Figura 3.3*. Igualmente los datos se pudieron ajustar a dos funciones gaussianas, el número de bins con los que se hizo el gráfico es el mismo que en el cuadrante de la MC-área porque los dos miden lo mismo, y además podemos ver que el rango que abarcan las coordenadas son diferentes del caso anterior, es decir, la MC y el C están ubicados en zonas diferentes. Al igual que en el análisis anterior, el valor de $i_{máx}$ corresponde al centro del C, por lo que el conjunto de datos de la derecha corresponde a todos los bins que forman parte del C y tienen un valor de sigma en su ajuste de datos.

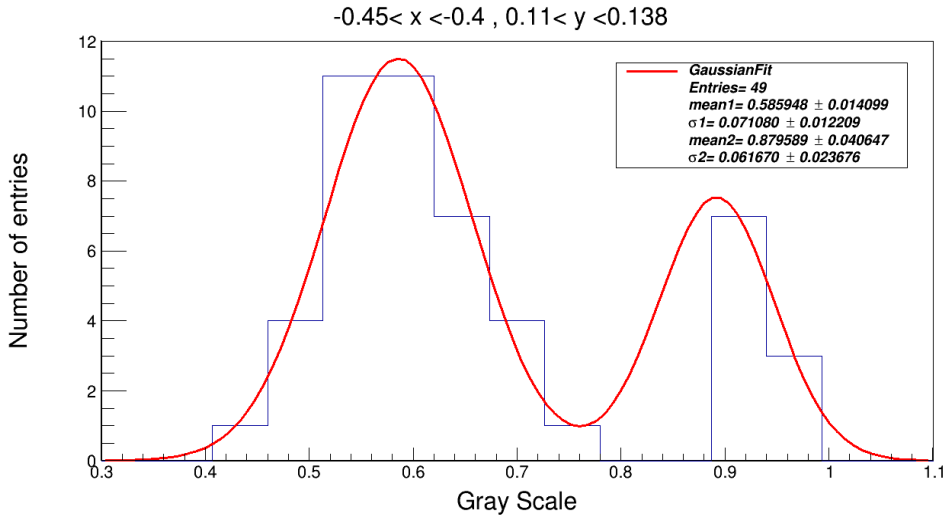

Figura 3.3 Histograma con las intensidades del cuadrante que abarca al C.

Ahora bien, la *Figura 3.4* muestra el histograma con los valores para el tejido de la mama, los del t-área de la misma muestra de la de *Figuras 3.2* y *3.3*. En este podemos ver que los datos se ajustan a una sola función gaussiana pues todos los bins de esta sección corresponden a un único elemento. También se observa que el número de entradas es el mismo que en los cuadrantes anteriores (por ser todos del mismo tamaño y por tratarse de la misma muestra). Todos los cuadrantes, en todas las muestras, tienen menos de mil entradas, por lo que la estadística es poca pero suficiente para realizar un análisis. El valor *mean* de esta zona es menor que el de los elementos anteriores, como era de esperarse ya que la MC y C son muy brillantes. Cabe aclarar que cuando decimos *mean* nos referimos al valor medio dado por el ajuste gaussiano de los datos. El valor *mean* de la MC o C es el valor asociado a la campana de intensidades altas en los histogramas de las *Figuras 3.3 y 3.2*. En cambio, en la *Tabla 3.2* se muestran los valores promedio de cada cuadrante de las cinco muestras analizadas. El valor de este promedio abarca también los bins te tejido circundante a la MC y C.

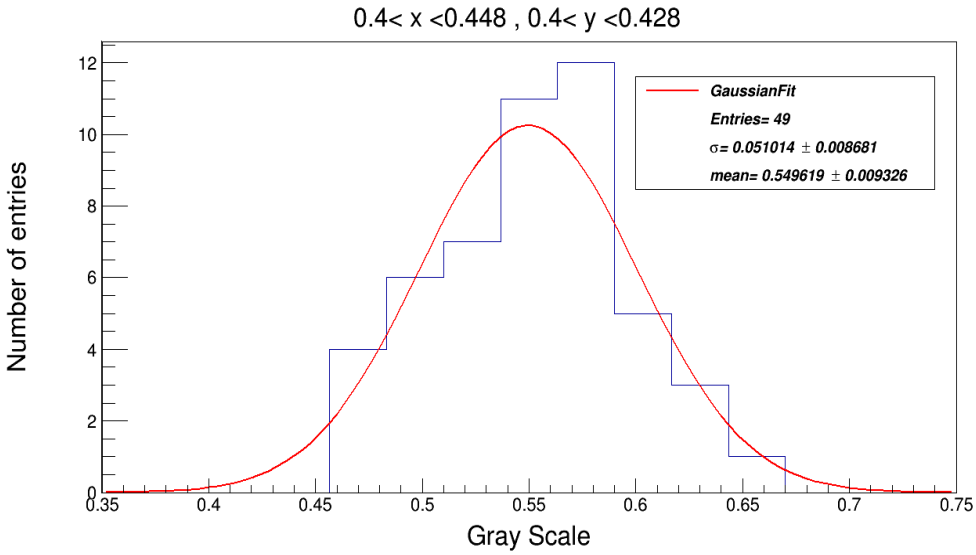

Figura 3.4 Histograma con las intensidades del cuadrante con tejido para la muestra con MC.

Tabla 3.2 En la primera columna aparece el número de muestra, en las siguientes tres el valor promedio de cada elemento analizado para las cinco muestras con MC.

Muestra	Promedio del área de MC	Promedio del área de C	Promedio del área de tejido
1	0.870495±0.0217284	0.804044±0.0086690	0.613301±0.0139076
2	0.915133±0.0116628	0.879589±0.0233282	0.549619±0.0093257
3	0.734961±0.1308161	0.725522±0.2005841	0.47934±0.00835189
4	0.797968±0.0081869	0.819147±0.0208582	0.1889±0.001314934
5	0.898861±0.0409994	0.803359±0.1028973	0.557635±0.0051492

En la *Tabla 3.3* están los valores de *mean* encontrados para el ajuste de la distribución de intensidad para el tejido (*mean* relacionado con la campana de intensidades bajas en los histogramas como los de las *Figuras 3.2* y *3.3*), y el valor de intensidad máxima de la zona de MC y del C. En la *Figura 3.4* (muestra 5 en la *Tabla 3.3*) el tejido tiene un valor para *mean* de 0.5496 y $\sigma = 0.0510$, entonces el valor máximo de toda la imagen, que corresponde al centro de la MC, está a más de 7-σ del valor promedio del tejido. Este proceso se repitió para todas las muestras y en general, se obtuvieron los resultados de la *Tabla 3.3* donde se ve que los valores máximos para MC se ubican a más de 3-σ del valor medio en la zona de tejido de su misma muestra.

Tabla 3.3 En la primera columna aparece el número de muestra, en las siguientes dos los valores de la distribución de intensidad para el tejido, y en los últimos el valor máximo de intensidad del área de la MC y C.

Muestra	*Mean* del área de tejido	σ del área del tejido	Valor máximo de intensidad en el área del C	Valor máximo de intensidad en el área de MC
1	0.816406±0.01390	0.0636±0.01211	0.898438	0.996094
2	0.667969±0.00933	0.0510±0.00868	0.957031	0.964844
3	0.617188±0.00835	0.0511±0.00706	0.929688	0.941406
4	0.304688±0.00131	0.0329±0.00109	0.902344	0.988281
5	0.549601±0.00933	0.0510±0.00807	0.941406	0.992188

Si quisiéramos realizar el análisis anterior pero para los datos de la imagen completa, es decir, tomando en consideración para los histogramas, todos los valores de intensidad de todos los bin del mapa de intensidades completo, se necesitarían considerar ciertos aspectos y dificultades a la hora de obtener resultados de esta manera.

Como ya mencionamos antes, la mayoría de los cuadrantes seleccionados en estas muestras tienen menos de 100 entradas. Estas son muy pocas en consideración del total de bins de toda la mama completa, por lo que al hacer un histograma de todos los datos de la mama, se obtiene una imagen como la de la *Figura 3.5*. En base a la *Figura 3.2*, el rango de intensidades que presenta la MC es de 0.7-1, por lo que en la *Figura 3.5* los bins pertenecientes a la MC se encuentran en este rango de intensidades, pero no es posible distinguir su distribución de datos como en la *Figura 3.2* en donde la selección o cuadrante es de menor tamaño que la imagen completa. Esto dificulta el poder ajustar los datos de la zona de MC y C a una función para poder determinar su intensidad promedio y poder aislar cada elemento. En el caso de la *Figura 3.5* el total de entradas es de más de 12 mil, por lo que las 49 entradas de la *Figura 3.2* son imperceptibles. Incluso si se marcara bien la distribución de los datos para los valores de intensidad altos (de la MC y C), como ambos elementos se encuentran en el mismo rango de intensidades, los bins de ambos se combinarían haciendo una única distribución, entonces no sería posible aislar de esta manera a cada elemento y tener tres distribuciones de datos. Para ejemplificar y explicar mejor esta idea, hicimos que el programa nos muestre las zonas con intensidad i en el rango de la intensidad de la MC. Como ejemplo, en la *Figura 3.6* se le pidió al programa que mostrara únicamente las zonas con intensidad mayor a 0.7, pues es la muestra de la *Figura 3.2*, y como vemos se encuentran distribuidas en toda la mama, la MC y el C están encerrados en el círculo 1 y 2 respectivamente. Esto muestra que lógicamente estos bins estarán combinados con los de la MC y C en el histograma de la imagen completa, haciendo imposible que se presenten por separado las distribuciones de cada elemento en el

histograma. Sin embargo esto se menciona para mostrar el contraste de entradas entre los del tejido y los de la MC y C. Por ello se decidió aislar los datos con los que se trabajaría en secciones aparte, dependiendo del elemento a estudiar y hacer con ellos el histograma y ajuste correspondiente.

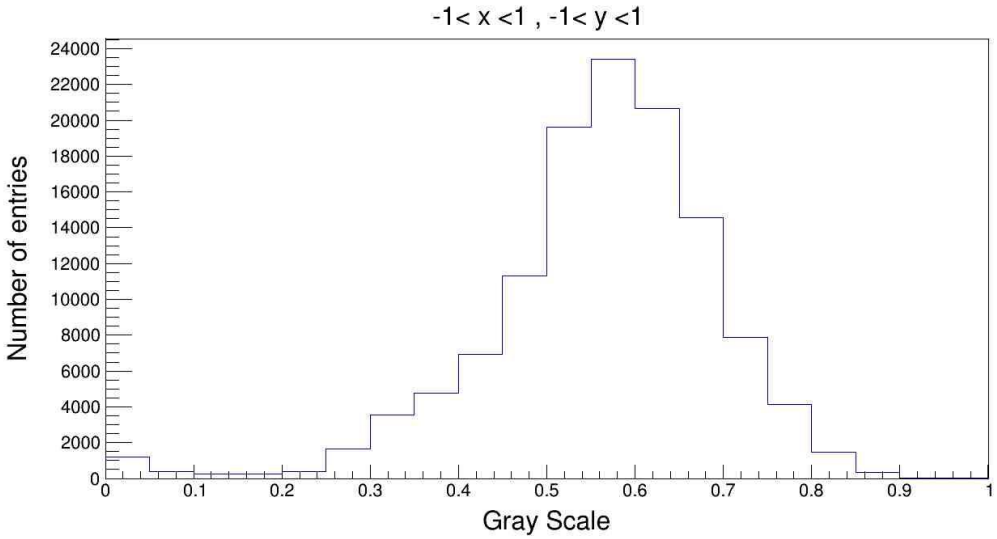

Figura 3.5 Histograma con las intensidades de todo el mapa de intensidades, de la muestra de la Figura 3.2.

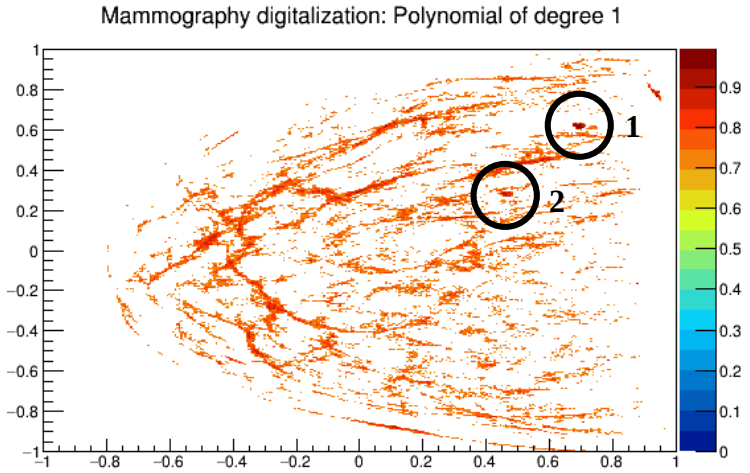

Figura 3.6 Mapa de intensidades de la muestra de la Figura 3.2, para los valores mayores de 0.7 en la escala de intensidad.

En el histograma de la *Figura 3.5* se omitieron los bins con intensidad cero, que son todos aquellos que rodean el seno, los correspondientes al aire y que por procesos físicos aparecen completamente de color negro en el mamograma. La *Figura 3.7* muestra los datos de la *Figura 3.5* incluyendo los de intensidad cero. Como es de esperarse la distinción de diferentes elementos en un histograma así es difícil, debido a la gran cantidad de entradas que presenta.

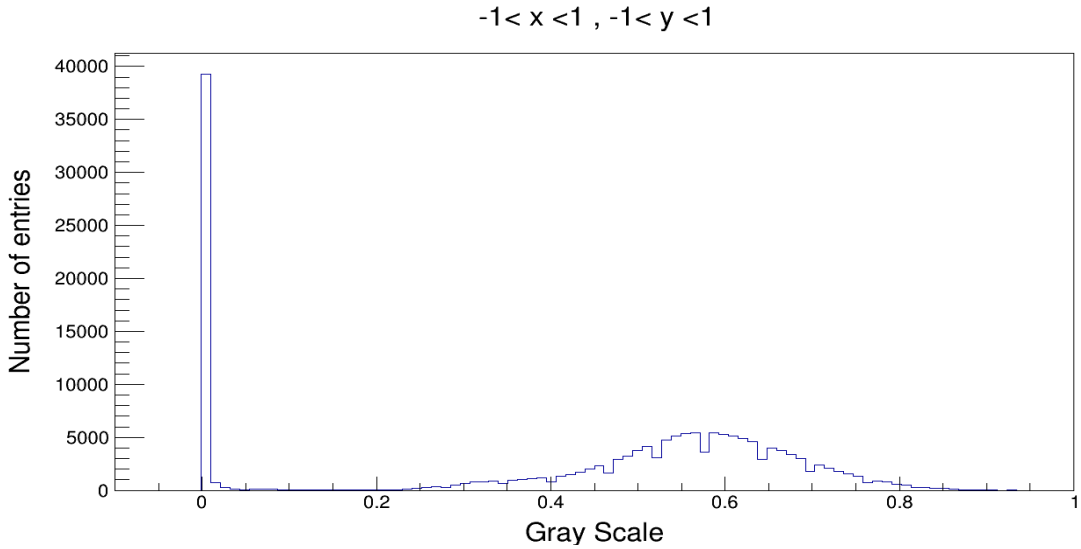

Figura 3.7 Histograma con las intensidades de tofo el mapa en escala de grises con MC, incluyendo las intensidades cero, de la muestra de la Figura 3.3.

Por otro lado, como se mencionó antes, el punto con el valor máximo de intensidad corresponde a la MC. En algunos casos es posible identificar estos elementos de manera visual, pero en otros no tanto, por ejemplo cuando la mama es muy densa, como se mencionó en el capítulo 1. Sin embargo podemos elevar a alguna potencia la intensidad de cada bin, lo que hará que las zonas con intensidades bajas tengan un valor más bajo aún y se oscurezcan, y los bins intensos sobresalgan del resto. Además se reafirma la localización de MC-área y de C-área en la imagen. Como ejemplo, en la *Figura 3.8* se observa la muestra de la *Figura 3.6*, elevada a la tercera, sexta y decimoquinta potencia en b), c) y d) respectivamente. Con ello visualmente la

ubicación de MC-área y de C-área se van acentuando. El MC-área se encerró en el cuadro 1, el C-área en el cuadro 2 y el t-área en el cuadro 3. En cada imagen podemos ir viendo que es más fácil la ubicación de cada elemento.

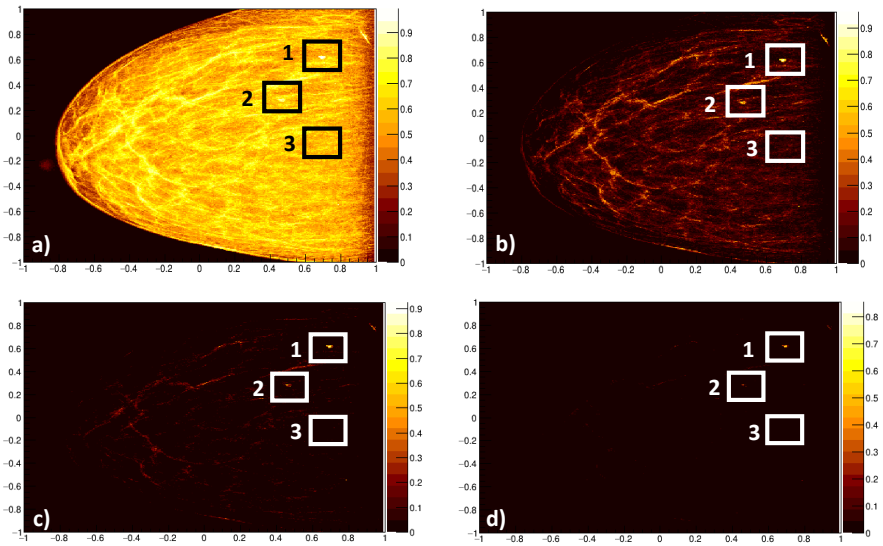

Figura 3.8. Muestra de mamografía con MC elevada a la potencia b) tercera, c) sexta y d) decimoquinta. La imagen original se muestra en a).

3.2 Análisis de mamogramas con tumor

Las imágenes de la mamografía original, de la base de datos, se dan en una matriz de 1024 × 1024 píxeles en formato .PGM, ubicando el origen en la parte inferior izquierda. Exportamos las imágenes a formato .PNG para su tratamiento con el método presentado. Los resultados de DF y S para t-área y T-área de las 10 muestras con T se presentan en la *Tabla 3.4*. La gráfica de estos resultados se encuentra en la *Figura 3.9.* Cada color de línea aborda los datos de una muestra diferente y los datos para tejido y T vienen señalados también.

Tabla 3.4 En la primera columna aparece el número de muestra con T y en las siguientes la dimensión fractal y entropía para cada elemento.

Muestra	Dimensión fractal		Entropía	
	Tumor	Tejido	Tumor	Tejido
1	2.64528	2.56735	8593.83	13343.5
2	2.76996	2.61889	4112.7	9371.07
3	2.46696	2.36638	4365.68	14947.3
4	2.49713	2.34575	3863.93	4850.95
5	2.78622	2.61102	14410.3	22058.8
6	2.69651	2.4714	8342.44	10039.8
7	2.49814	2.2836	4127.82	7082.47
8	2.76827	2.67171	14200.7	14351.7
9	1.9429	1.82621	426.211	555.947
10	2.84105	2.74235	16397	21513.8

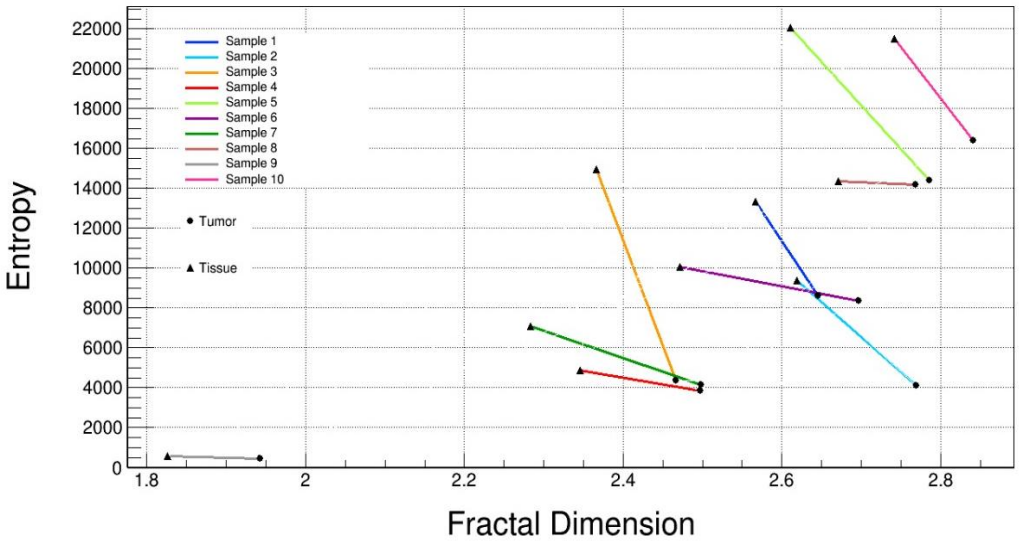

Figura 3.9 Dimensión fractal vs Entropía para cada cuadrante de cada muestra con T.

Los histogramas hechos para las muestras con Tumor se realizaron de la misma manera que para los elementos de MC, C y tejido de la sección anterior, el eje x muestra la escala de valores de intensidad posibles y el eje y el número de bins o entradas con determinada intensidad. Las etiquetas en cada histograma tienen el número *Entries* correspondiente y los parámetros de cada ajuste. El histograma del T-área se muestra en la *Figura 3.10*, y podemos ver en primer lugar, que el número de entradas es mucho mayor en comparación con el cuadrante de MC, esto se ve en el rango tomado para x y y que se muestra en el título del gráfico.

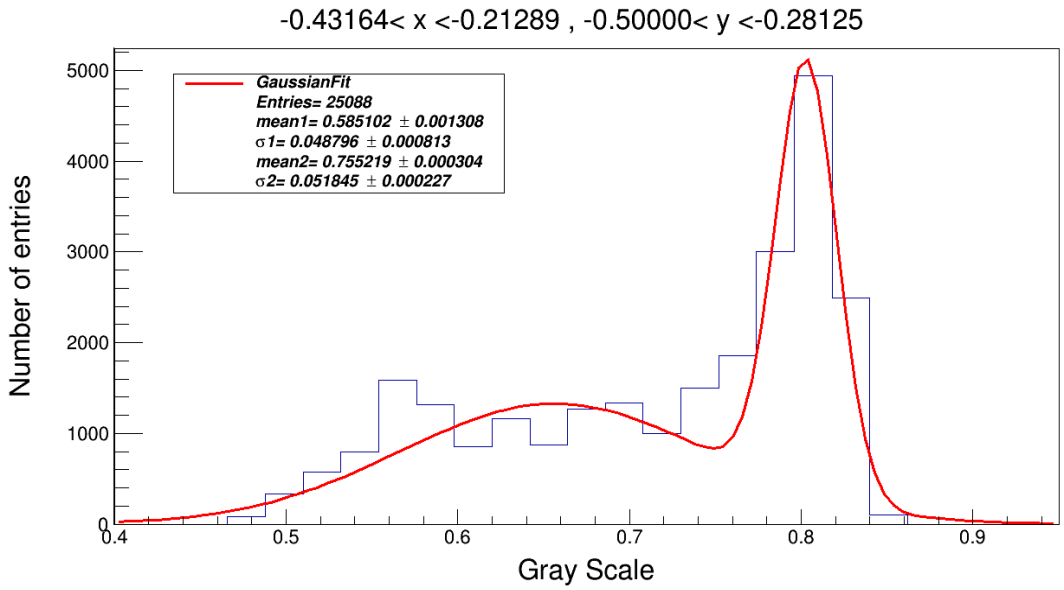

Figura 3.10 Histograma con las intensidades del cuadrante con T para la muestra con T.

El histograma para la zona de tejido se muestra en la *Figura 3.11*. En esta se observa que el número de entradas es similar que para la zona de T (por ser del mismo tamaño ambos cuadrantes), y la distribución de los datos se puede ajustar a una función gaussiana. La etiqueta muestra el valor promedio de esta zona y el valor de σ del ajuste.

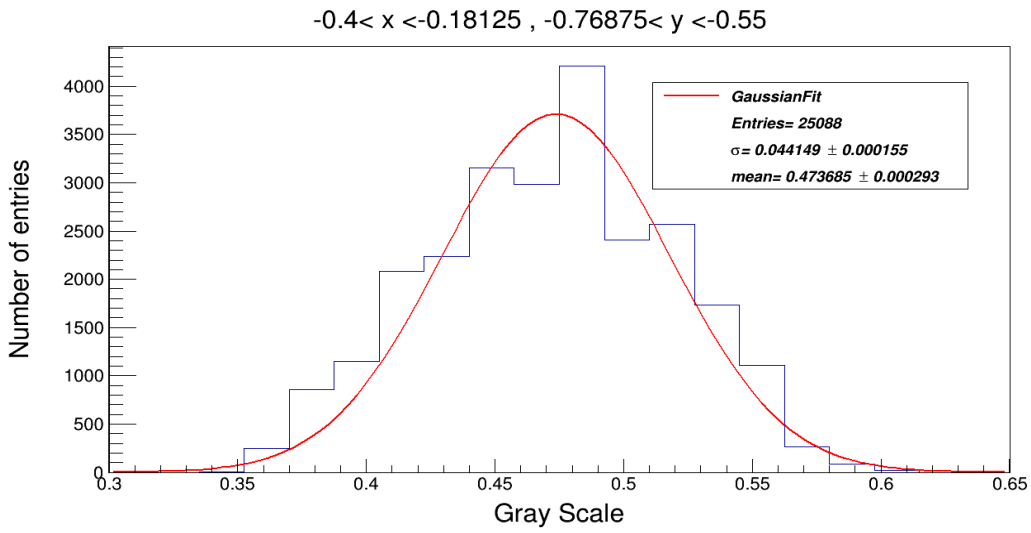

Figura 3.11 Histograma con las intensidades del cuadrante con tejido para la muestra con T.

Los valores promedio de intensidad para T y tejido se escribieron en la *Tabla 3.5*, en las columnas dos y tres, en la primera columna se anotó el número de muestra (recordemos que se analizaron 10). El valor promedio de la zona de T es mayor que el valor promedio del tejido, y a su vez en general, comparando la *Tabla 3.2* con la *Tabla 3.5*, el valor promedio del cuadrante de T de las diez muestras, es menor que el valor promedio de MC de las cinco muestras analizadas.

Tabla 3.5 En la primera columna aparece el número de muestra, en las siguientes dos los valores de la intensidad promedio que presenta el área de T y tejido.

Muestra	Promedio del área de T	Promedio del área de tejido
1	0.755219±0.00127021	0.430663±0.00016187
2	0.754016±0.00130635	0.49157±0.000193344
3	0.911547±0.00019624	0.840697±0.00045030
4	0.753818±0.00058737	0.511244±0.00032434
5	0.760541±0.00062243	0.322015±0.00277352
6	0.730885±0.00038544	0.62915±0.000390161
7	0.764449±0.00061100	0.481209±0.00022869
8	0.65051±0.000198105	0.655631±0.00016777

9	0.723961±0.00107212	0.588583±0.00059157
10	0.784384±0.00198347	0.633504±0.63350400

Ahora bien, en la *Tabla 3.6* se encuentran los valores de *mean* y σ para la distribución de datos del área de tejido y el valor máximo de intensidad de toda la imagen, que corresponde al área del T. A modo de ejemplo podemos tomar la muestra dos, con una intensidad máxima de 0.890625 justo en el centro del T, y un valor medio de intensidad de 0.499031 para el tejido. Así, el valor máximo de intensidad de T está a 31-σ del valor central del tejido. Se realizó este cálculo con los datos de la *Tabla 3.6*, para todas las muestras y en general se obtuvo que el valor de intensidad máxima (ubicado en el área del T) es mayor que 3-σ de la media de intensidad en el área de tejido.

Tabla 3.6 En la primera columna aparece el número de muestra, en las siguientes dos los valores de la distribución de intensidad para el tejido, y en la última el valor máximo de intensidad del área de T.

Muestra	*Mean* del área de tejido	σ del área del tejido	Valor máximo de intensidad en el área del T
1	0.523438±0.0002	0.0224±0.0001	0.851565
2	0.499031±0.0007	0.0126±0.0004	0.890625
3	0.796202±0.0003	0.0241±0.0002	0.945313
4	0.599844±0.0003	0.0222±0.0002	0.820313
5	0.660156±0.0005	0.0169±0.0004	0.886719
6	0.703125±0.0004	0.0245±0.0002	0.902344
7	0.554688±0.0003	0.0269±0.0002	0.886719
8	0.765625±0.0002	0.0259±0.0001	0.835938
9	0.644531±0.0006	0.0189±0.0004	0.808594
10	0.746094±0.0002	0.0240±0.0001	0.914063

Por otro lado, al realizar un histograma con los datos de la imagen completa se obtuvieron resultados como los del ejemplo de la *Figura 3.12*. Se podría pensar que en base al histograma de la *Figura 3.10*, los bins con valores i entre 0.75 y 1, de la *Figura*

3.12 corresponden todas a los datos del T, pero no es así pues sucede algo similar al caso de MC. En el mapa de intensidades hay zonas que tienen un valor de intensidad en el rango del T, por lo que todos estos bins estarían combinados en la distribución de datos con intensidades mayores a 0.75, haciendo que no se pueda aislar el T del tejido desde un histograma como el de la *Figura 3.12*.

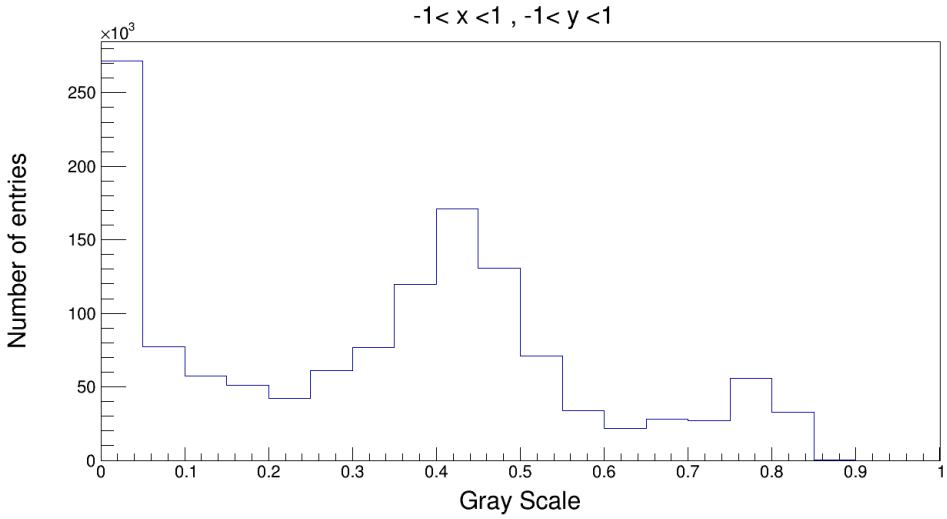

Figura 3.12 Histograma con las intensidades de todo el mapa en escala de grises con T, incluyendo las intensidades cero, de la muestra de la Figura 3.10.

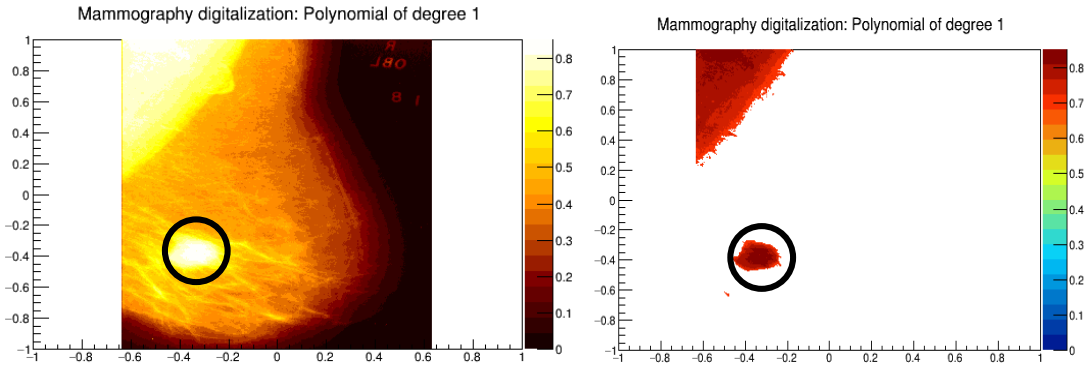

Figura 3.13 Muestra redigitalizada de la muestra de la Figura 3.10 para los valores mayores de 0.75 en la escala de intensidad.

En la *Figura 3.13* se observa la muestra de la *Figura 3.10*, y se encerró al T. En la figura de la izquierda se mantuvieron todos los datos, y en la de la derecha se le pidió al programa que mostrara las zonas que tuvieran intensidad mayor a 0.75, esto en base a la *Figura 3.10*. Aquí se puede ver que hay diferentes zonas con la misma intensidad que el T. Los bins de estas se combinan con los del T en el histograma de la *Figura 3.12* y por eso no se puede determinar de manera exacta la parte de T.

Finalmente, con estas muestras se pudo elevar a alguna potencia las intensidades del mamograma para distinguir mejor la zona de T, como se muestra en el ejemplo de la *Figura 3.14.*

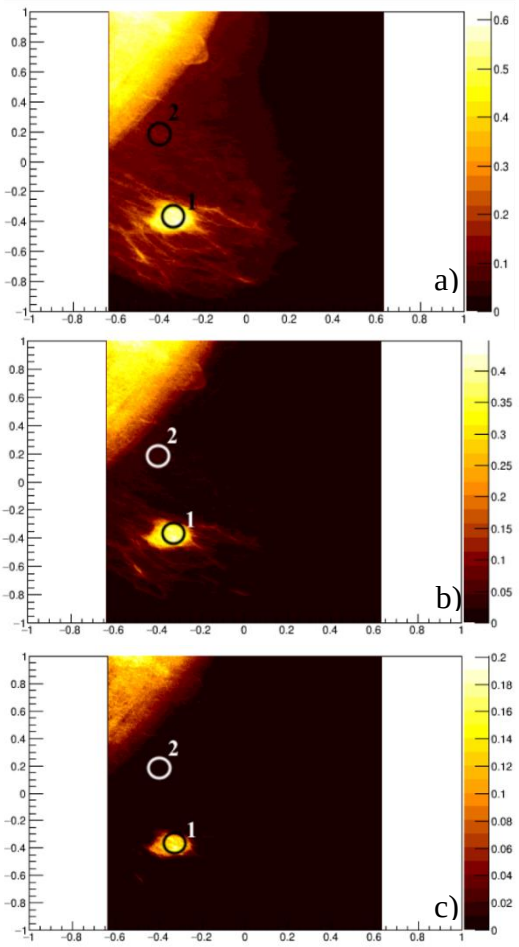

Figura 3.14 Resultados al elevar a la a) tercera, b) quinta y c) décima potencia la muestra de la Figura 3.13.

En resumen, en esta parte del proceso que se explica en las secciones 3.1 y 3.2, se utilizaron los datos de cada cuadrante seleccionado en el capítulo anterior, para realizar un histograma. El programa aisló de la base de datos de toda la imagen, únicamente los puntos que se encontrasen en el cuadrante especificado. La coordenada x y y de estos datos estaría dentro del rango que abarca el recorte seleccionado. Este fue el parámetro que se utilizó para filtrar los datos a utilizar. Lógicamente en el recorte no se puede seleccionar únicamente a la MC sin tomar un poco de tejido, debido a que este elemento no es un cuadrado perfecto que se pueda seleccionar con precisión y por ende no se puede determinar un rango específico para las coordenadas x y y de tal manera que sean únicamente las que pertenezcan a la MC sin que falte ninguna. En general, por más que se recorte la imagen, siempre habrá bins que pertenezcan al tejido pero serán menos que los correspondientes al propio elemento (MC, C o T), así pues, el histograma mostrará valores de intensidad que pertenecen a ambos elementos. Esto realmente no importó porque los datos se pudieron ajustar a funciones, una para los datos de MC, C o T y otra para los del tejido. Con esto se obtuvo un valor promedio de intensidad para cada elemento y se logró aislar uno de otro.

Capítulo 4

Análisis

El hecho de que para todas las muestras analizadas en este trabajo, la entropía de la MC y del T es menor en comparación con la obtenida para el tejido, indica que la MC y T son zonas más ordenadas en cuanto a intensidad, la distribución de tonalidades que presentan, como se explicó en el capítulo 1, se encuentran de manera más uniforme que en la zona de tejido, que presenta intensidades muy diferentes para cada bin. En la *Figura 2.4* y a la izquierda de la *Figura 2.7* se muestra el MC-área, C-área y T-área, respectivamente. En estos cuadrantes, las intensidades altas se encontraron en el centro del área seleccionada, y se presenta una diferencia marcada con los bordes del rededor de baja intensidad. Esta es la manera en que se ordenan las tonalidades en la escala de gris conforme a las intensidades de la MC, C y T. En cambio en la *Figura 2.5* y a la derecha de la *Figura 2.7*, se muestran las áreas de tejido para la muestra con MC y T respectivamente. En ambos casos el cuadrante presenta distintas intensidades distribuidas de manera no uniforme por toda la selección, visualmente estos cuadrantes muestran distintos bins con distintas tonalidades de gris repartidos por todo el área. La diferencia entre las tonalidades e intensidades en cada grupo de elementos indica que

la entropía es mayor para el área de tejido y menor para el MC-área y T-área. No obstante, para el caso del C-área algunas veces la entropía que presentó era mayor que la de la MC-área y otras veces era menor, sin embargo siempre fue menor que la entropía del área del tejido, esta relación se mantuvo en todas las muestras analizadas.

Supongamos que le calculamos la entropía al cuadrante de la *Figura 4.1* con el color como parámetro de medición, entonces el cuadrante de la derecha tendrá una entropía mayor que el de la izquierda pues los colores se encuentran de manera desordenada. Justamente eso sucede con los cuadrantes de las *Figuras 2.4, 2.5* y *2.7*, pero aquí el parámetro a medir es la intensidad del bin, en cambio del color.

Figura 4.1 A la izquierda un cuadrante con el parámetro color ordenado y a la derecha desordenado.

En cuanto a la dimensión fractal, esta es menor para el tejido que para la MC o T. Lo que determina este resultado es el término $A(\varepsilon)$ de la ecuación 1.1. Como habíamos mencionado en el capítulo 1, este término es el área "visible" en un gráfico como el de la *Figura 1.10*. Si vemos de esta manera a cada cuadrante seleccionado, es decir como un gráfico tridimensional, podemos analizar la diferencia de $A(\varepsilon)$ entre el T-área, MC-área y tejido-área. Así pues, entre mayor sea la cantidad de área "visible" que se presenta en el gráfico, mayor será el valor de $A(\varepsilon)$. Ahora, esta área se determina gracias a la altura de cada barra de la *Figura 1.10*, que para nuestro caso tienen una altura correspondiente con el valor i. Si las coordenadas i de cada bin no varían mucho entre sí, $A(\varepsilon)$ será pequeña ya que no hay una gran área "visible" entre cada barra porque todas tienen una intensidad similar. Por ello, dependiendo del valor de $A(\varepsilon)$ la

dimensión fractal será mayor o menor. Para el caso del MC-área, las barras centrales de intensidades altas que pertenecen a la MC, difieren mucho con el tejido que la rodea, tanto como para hacer que $A(\varepsilon)$ sea mayor que la del tejido-área, en donde sí hay variación de intensidades pero no difieren tanto como las intensidades de la MC y su tejido circundante. Es por ello que la dimensión fractal de la MC es mayor que la del tejido. De hecho, en todas las muestras, tanto las de T como las de MC, resultó que esta área $A(\varepsilon)$ es mayor para el T o para la MC según sea el caso, que para el respectivo tejido de la muestra. Es por ello que la dimensión fractal de estas zonas es mayor que la de la zona de tejido. En cuanto a la dimensión fractal del C, esta era algunas veces mayor y otras menor que la del área de tejido, pero siempre resultó ser menor que la de su respectiva MC-área.

Está claro que las relaciones de S y D se mantienen para todos los elementos analizados de las muestras estudiadas, sin embargo algo que se observa en la *Figura 3.1* es que los valores de S y D varían para cada muestra y no se mantienen en algún rango específico. Lo mismo pasa con las muestras con T y se observa en la *Figura 3.9.* Por ejemplo, el MC-área de la muestra 1 tiene una S de 18.49 y una DF de 1.57, mientras que el MC-área de la muestra 4 tiene una S de 296.04 y una DF de 2.19, según los datos de la *Tabla 3.1.* Igualmente el T-área de la muestra 4 tiene una S de 3863.93 y una DF de 2.49, mientras que el T-área de la muestra 10 tiene una S de 16397 y una DF de 2.84, según los datos de la *Tabla 3.4.* Como podemos ver, los valores de cada muestra son muy distantes entre sí, por lo menos para S. Estos valores dependen de la distribución de las intensidades de los bins en la imagen redigitalizada. A su vez, el valor de intensidad se debe al brillo en la mamografía digital, entonces el valor de S y DF se debe a la distribución de brillo en la imagen de mamografía, es decir, a cómo se distribuyen las tonalidades de la escala de gris en la imagen de la mama. Esta distribución es diferente para cualquier mamografía, pues depende de la constitución de la mama, que es diferente en todas las pacientes. Principalmente lo que lo determina

es la cantidad y distribución de tejido fibroglandular y graso de la mama. En [7] se menciona que a mayor área con brillo, en nuestro caso, mayor área con intensidades altas, mayor será la probabilidad de adquirir cáncer de mama. En este artículo se refieren a dicha área como densidad mamográfica y presentan un estudio donde se menciona que el riesgo relativo de cáncer de mama para categorías de densidad alta, mayores al 75%, fue 5.3 veces mayor que el de las mujeres cuyos senos estaban en la categoría de densidad más baja. La manera de caracterizar las variaciones de brillo, o la densidad mamográfica en las imágenes digitales y así poder clasificarlas de acuerdo al riesgo de presentar cáncer de mama, fue mediante el cálculo de la DF, como en el presente trabajo, de toda la imagen de mamografía. Esto nos dice que el cálculo de DF funciona para crear una relación entre distintas partes de interés en la mama y así identificar distintos elementos en ella, pero además sirve para determinar la densidad de la mama y determinar una probabilidad de que el paciente obtenga cáncer, si se le aplica el cálculo a toda la imagen. Así mismo, en [10] se digitalizaron imágenes de mamografía para analizar imágenes nodulares y por medio de la DF determinar si son benignas o malignas. Primero se determinó la región de interés donde visualmente se detectó la lesión y se separó de la imagen completa (*Figura 4.2 b)*), después ésta se segmentó mediante el establecimiento de un umbral inferior y un umbral superior que permite eliminar aquellos niveles de grises de la imagen que no son de interés. La distribución de estos niveles fue valorada por medio del histograma, donde los modos determinan las tonalidades predominantes en la imagen (*Figura 4.2 c)*). Luego se detectó el borde de la lesión por medio del operador Laplaciano, y se extrajo la lesión nodular empleando el algoritmo de clúster, donde se agrupan los pixeles y sus pixeles circunvecinos en base a su tonalidad de gris (*Figura 4.2 d)*). Finalmente se calculó la dimensión fractal par conteo de cajas. La mediana de la dimensión fractal para las lesiones benignas fue menor que para las lesiones malignas.

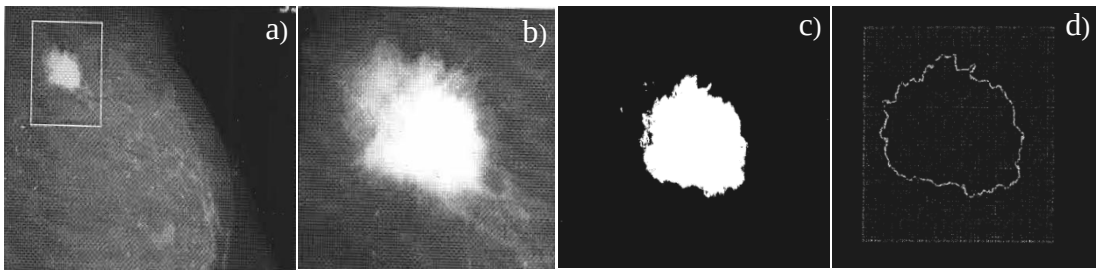

Figura 4.2 a) Mamograma con nódulo, b) selección del cuadrante que encierra a la lesión, c) segmentación, d) contorno de lesión y enrejado para cálculo de dimensión fractal por conteo de cajas.

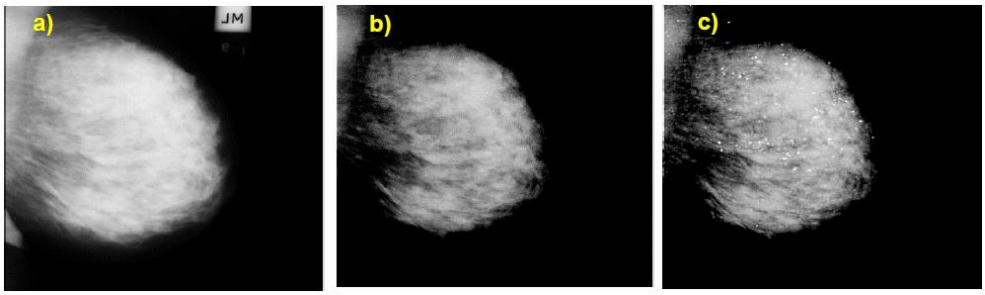

Figura 4.3 a) Imagen original. b) Imagen segmentada preprocesada, reajustada, limpia de artefactos y con las microcalcificaciones detectadas en su tamaño real. c) Imagen con microcalcificaciones realzadas.

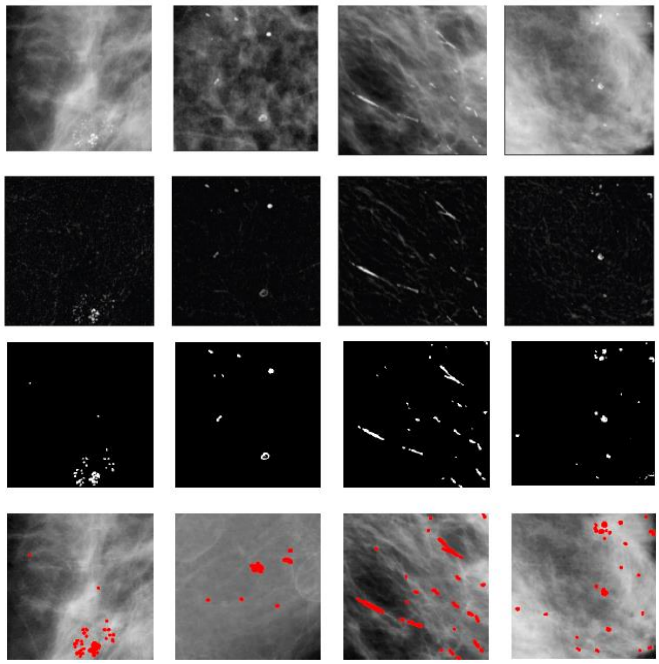

Figura 4.4 Secciones de mamogramas originales (primera fila), procesadas por la transformada Top-hat (segunda fila).

Existen diferentes métodos para poder ubicar mejor alguna anormalidad como las MC. En [3] se propone un sistema de detección asistido por computadora CAD, que mejora la calidad de la imagen mamográfica digital, basado en un procesamiento digital, para después sobre estas imágenes mejoradas, segmentar la glándula y resaltar la presencia de MC. El CAD mejora el contraste de la imagen mediante filtros de convolución, elimina artefactos del rededor del seno mediante apertura y cierre morfológico y filtrado Laplaciano, utiliza métodos de umbralado basados en entropía para segmentar la glándula mamaria y filtrado morfológico y reajuste de histograma para realzar las MC. El proceso que sufre la imagen se muestra en la *Figura 4.3.* Sin embargo este método no asegura que todos los puntos brillantes resaltados sean efectivamente MC. En el presente trabajo proponemos, en cambio, detectar visualmente las MC elevando a alguna potencia el mamograma redigitalizado, entonces se obtendrá una imagen como la de la *Figura 3.14*. Además se propone detectar las MC mediante las intensidades de cada bin, lo cual requiere pocos segundos de trabajo computacional igualmente, y gracias a las relaciones de S y DF se confirma la identidad de cada punto brillante. En [11] se presenta también una manera de mejorar el contraste de los mamogramas mediante la transformada Top-hat, la cual, se define como un tipo de filtro residual, debido a que, se compone de una diferencia de entre dos transformaciones básicas que permiten resaltar aquellas regiones de la imagen que han sido eliminadas por la operación de apertura. Por medio de la elección apropiada del tamaño y forma de un elemento y aplicando una operación de apertura se obtiene una imagen modificada, en la cual, se eliminan algunos elementos determinados de la imagen original. Después, por medio de la diferencia entre la imagen original y la imagen modificada, se aumenta de forma considerable el contraste de las regiones eliminadas. De acuerdo con lo anterior, la transformada Top-hat resalta las MC con respecto al tejido circundante como se observa en la segunda fila de la *Figura 4.4*. Después se segmenta una imagen en forma binaria, donde los patrones clasificados como MC tienen el valor de 1 mientras que los patrones clasificados como tejido

circundante tienen el valor de 0, para poder apreciar el resultado de la detección de forma visual como se observa en las últimas dos filas de la *Figura 4.4*. En [12] igualmente se detectan MC usando técnicas de inteligencia artificial, en ese caso se procesan imágenes mediante un filtro gaussiano, el cual permite realzar el contraste entre las MC y el tejido sano de la mama, para luego aplicar un algoritmo de identificación de las MC sospechosas basado en su morfología. En este estudio se identifican las MC en base a 7 características extraídas de cada una de las MC sospechosas, que finalmente conformaron el vector de prueba o test para el clasificador.

Por su parte, el método mostrado en el presente trabajo, arroja la ubicación (x, y) de una MC o T simplemente con la redigitalización de la imagen de mamografía. Este es el principal y más importante resultado obtenido con esta investigación. Para las muestras con T, es posible comparar este resultado obtenido con el que nos proporciona el diagnóstico de la base de datos. Como ejemplo, para la muestra 1 de mamogramas con T, el diagnóstico da las coordenadas de $(338, 314)$ para el centro de T y las coordenadas en nuestro sistema, para el valor de intensidad más alto son $(-0.320313, -0.400391)$, entonces son consistentes. Obtenemos la misma consistencia para las otras muestras. En la *Tabla 4.1* se muestra el error porcentual para las coordenadas x, y usando

$$P_{error} = \frac{|m_i - d_i|}{d_i}$$

donde m es el valor obtenido por nuestra transformación, d es el valor dado por el diagnóstico e $i = x, y$.

Tabla 4.1 En la primera columna aparece el número de muestra con T y en las siguientes el error porcentual entre las coordenadas dadas en el diagnóstico y las obtenidas mediante el método presentado.

Muestra	x (%)	y (%)

1	2.36	5.75
2	2.93	6.42
3	1.77	0.35
4	1.06	4.12
5	0.18	8.51
6	2.36	0.02
7	0.02	3.45
8	4.02	0.01
9	1.85	2.65
10	1.42	4.16

Para el caso de mamogramas con MC, no podemos comparar las coordenadas que nos da el método para el centro de la MC, con las coordenadas de la imagen original, como en el caso de los mamogramas con T, debido a que el diagnóstico fue hecho por la Doc. Herrera. Sin embargo la localización de las MC concuerda con el diagnóstico.

Finalmente, como trabajo futuro se pretende estudiar la correlación que se presenta entre la DF y S en mamogramas con T benignos, realizar los histogramas para el análisis de los datos como para el caso de T malignos. Además, estudiar la relación que puedan tener las imágenes y los parámetros propuestos, entre mamogramas con T malignos y mamogramas con T benignos. Igualmente como trabajo futuro se puede hacer una reconstrucción completa de T en base a las intensidades presentadas.

Capítulo 5

Conclusiones

Lo que se logró con este trabajo fue poder identificar y ubicar una MC o T en una imagen de mamografía digital, de manera visual como se muestra en las *Figuras 2.4, 2.5* y *2.7,* o a partir de la intensidad más alta de toda la imagen. Este método sirve incluso en mamografías de mamas muy densas donde no es posible visualizar a simple vista algunos elementos, ya que el contraste entre éstos y el tejido mamario suele ser muy bajo. Siempre es posible elevar a alguna potencia la imagen para una mejor visualización.

Una manera de complementar la ubicación de un elemento en la imagen es haciendo histogramas. En definitiva es bastante importante y útil realizar el histograma ya que, de esta manera podemos darnos cuenta de si hay alguna anormalidad o simplemente toda la mama tiene solo tejido. Así, si quisiéramos saber si hay alguna MC o la ubicación del T, fácilmente podríamos hacer todos los pasos explicados en este trabajo, esto es, primero el programa nos da el valor de la intensidad mayor y sus coordenadas (x, y) y los valores de la zona intensa circundante al valor máximo de intensidad. Después, haciendo el histograma del recorte notaríamos los valores de

intensidad cercanos a las intensidades características de la MC o T, para determinar el rango de intensidades donde se encuentra cada elemento. Una vez hechos los ajustes y dependiendo del valor medio de intensidad que presente el tejido, y su sigma, podemos decir a que elemento pertenece cada distribución de datos, si pertenecen al tejido, o a una MC o T. Siempre que el valor máximo de intensidad a analizar, esté a 3-σ o más del valor medio para el tejido podremos decir que ese elemento ya no es tejido. Esto podría hacerse en mamogramas con MC o T incluso sin ver la imagen.

Una vez que conocemos el rango de valores para la intensidad del T o MC podemos hacer un mapa de densidades pero ahora solamente para valores i entre el rango de intensidades de la MC o T, y veremos mejor estos elementos en la imagen, como en la *Figura 3.8*.

Finalmente se puede determinar la identidad de cada elemento mediante la relación de entropía y dimensión fractal encontrada en este trabajo. Si la dimensión fractal del elemento es mayor que la del tejido, y si la entropía del elemento a analizar es menor que la del tejido, estaremos hablando de una MC o T. La diferencia entre estos dos últimos radica en la cantidad de entradas que construyan el histograma. Con poca estadística seguramente estaremos hablando de una MC.

Este trabajo es muy importante además por la facilidad para obtener una conclusión, pues no se necesita nada más que el código y la imagen de mamografía en un formato como .PNG o .JPG, que después será redigitalizada y será con esta última con la que se trabaje para todos los pasos siguientes. Y al estar normalizada es fácil pasar de un punto de esta imagen al mamograma original para dar una ubicación real de la MC o T, en la mama del paciente.

Como se puede ver en el capítulo anterior, se reportan muchos métodos para visualizar MC y T en la imagen de mamografía, por asistencia computarizada. El presente trabajo propone uno donde es posible visualizar las MC y T, mediante un

código que requiere unos pocos segundos de trabajo computacional, sencillo y fácil de usar que no requiere más que elevar a alguna potencia la intensidad de cada bin. Además se confirma la identidad de cada zona brillante mediante las relaciones de S y DF, y se diferencia del tejido mediante el valor de σ, producto de los histogramas. Son tres métodos que nos ayudan a caracterizar y aislar T malignos y MC en la imagen de mamografía. La consistencia de nuestros resultados se muestra en la *Tabla 4.1* donde se presenta el error porcentual entre las coordenadas que ubican el centro del T, dadas en el diagnóstico de la base de datos, y las coordenadas obtenidas por el método mostrado en este trabajo. En cuanto a las muestras con MC, en todas las imágenes, el elemento que se identificó como MC fue el elemento que la Dra. Herrera diagnosticó como MC, acertando así, en la ubicación de este elemento en todas las muestras estudiadas.

En conclusión es un método confiable para encontrar una MC o T en una mamografía, que no depende más que de parámetros propios de la imagen y no de la interpretación del médico radiólogo, sin embargo no se pretende sustituir el papel del médico, sino brindar apoyo en el diagnóstico. El desarrollo del método contribuye en la detección temprana del cáncer de mama, pues entre un 30% y 50% de los casos diagnosticados como cáncer de mama a nivel mundial, han sido a través de la detección de MC [13]. Además es una herramienta que proporciona información objetiva de las imágenes que se analizan y de las características que de ellas se puedan extraer mediante el uso del código programado, es decir, que mediante el uso de las matemáticas, podemos identificar MC y T en un mamograma.

Apéndice A

Muestras de mamografías con microcalcificación

En este apéndice se incluye la imagen de mamografía original junto con su mapa de intensidades, para las cinco muestras con MC, y se encerró en el círculo 1 a la MC, al C en el círculo 2 y al área que se tomó de tejido en el círculo 3. El tamaño del círculo no es del tamaño del cuadrante seleccionado, solo sirve para señalar el elemento.

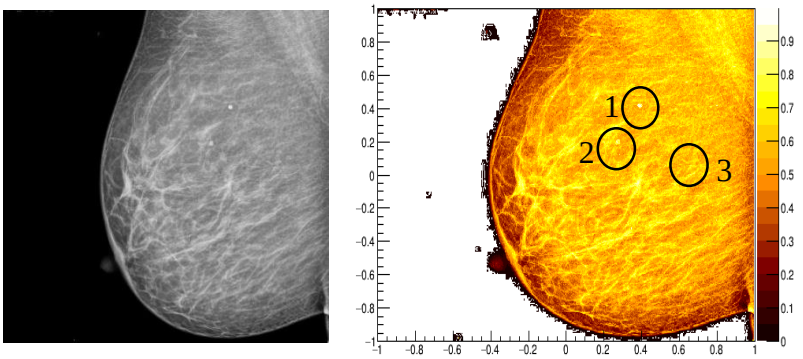

Figura A1 Muestra 1 de mamografía con MC.

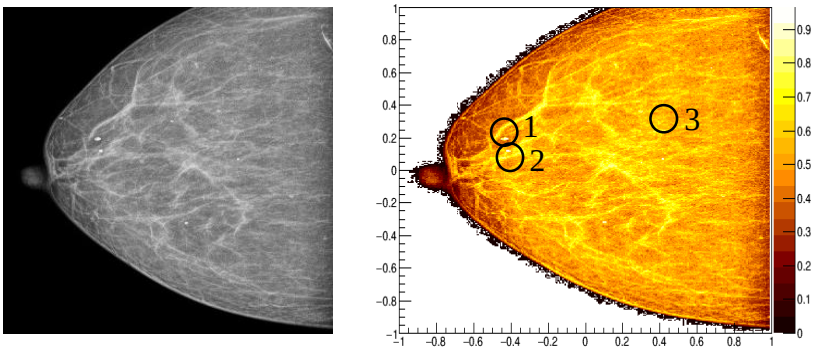

Figura A2 Muestra 2 de mamografía con MC.

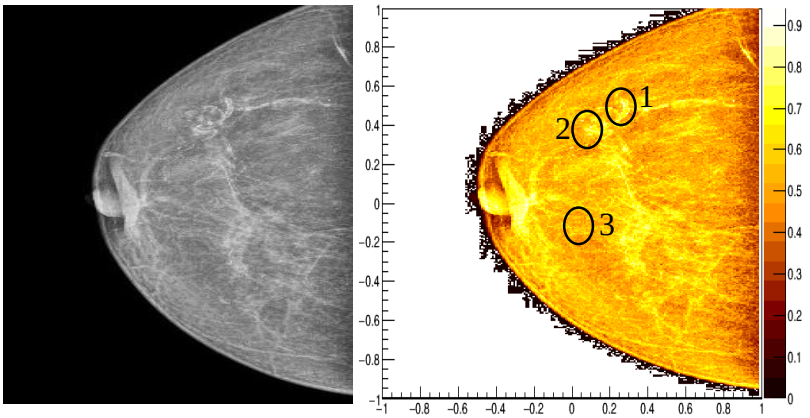

Figura A3 Muestra 3 de mamografía con MC.

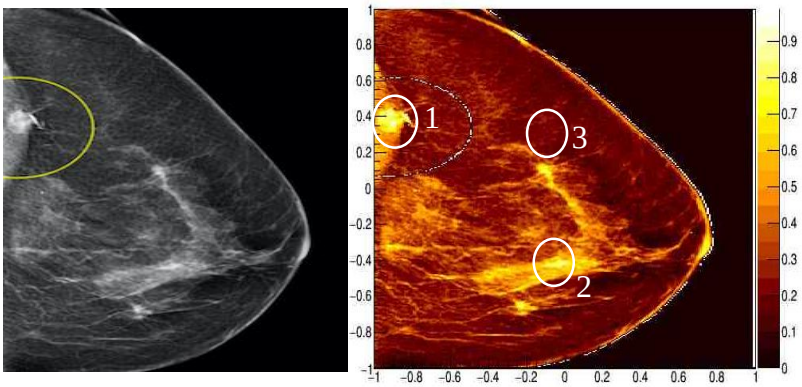

Figura A4 Muestra 4 de mamografía con MC.

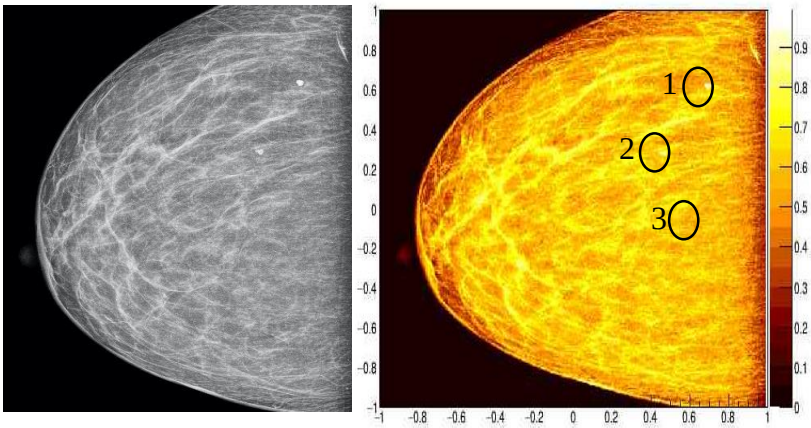

Figura A5 Muestra 5 de mamografía con MC.

Apéndice B

Muestras de mamografías con tumor

En este apéndice se incluye la imagen de mamografía original junto con su mapa de intensidades, para las diez muestras con T.

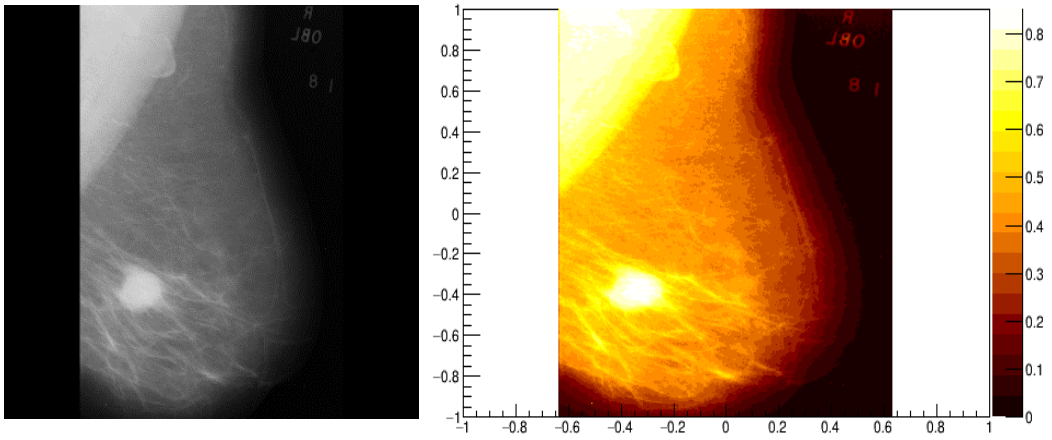

Figura B1 Muestra 1 de mamografía con T.

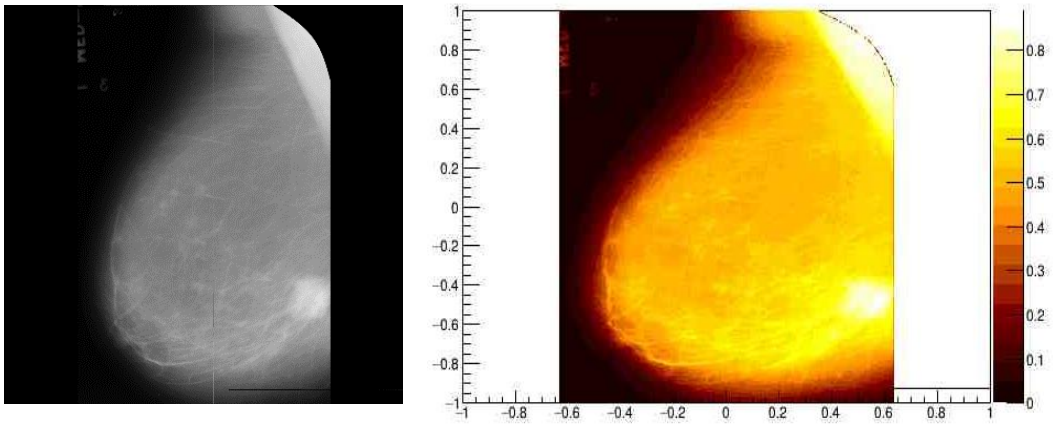

Figura B2 Muestra 2 de mamografía con T.

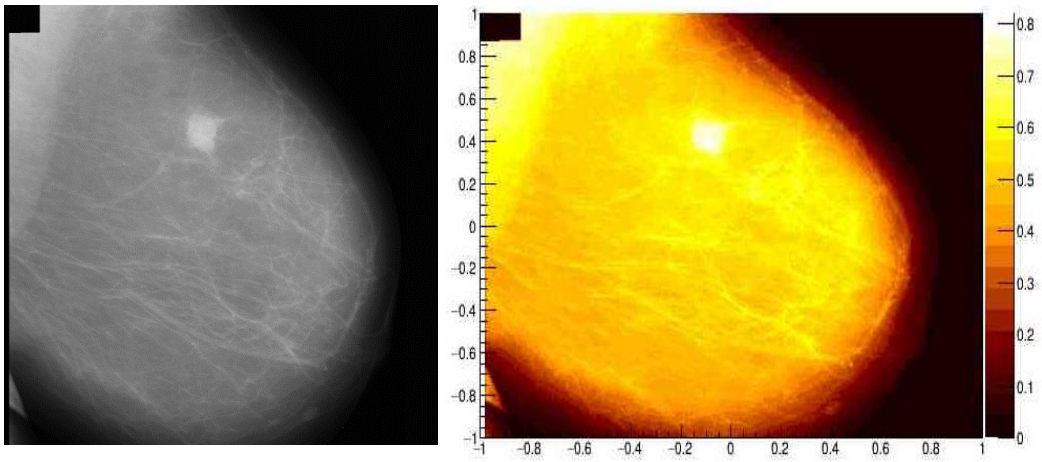

Figura B4 Muestra 4 de mamografía con T.

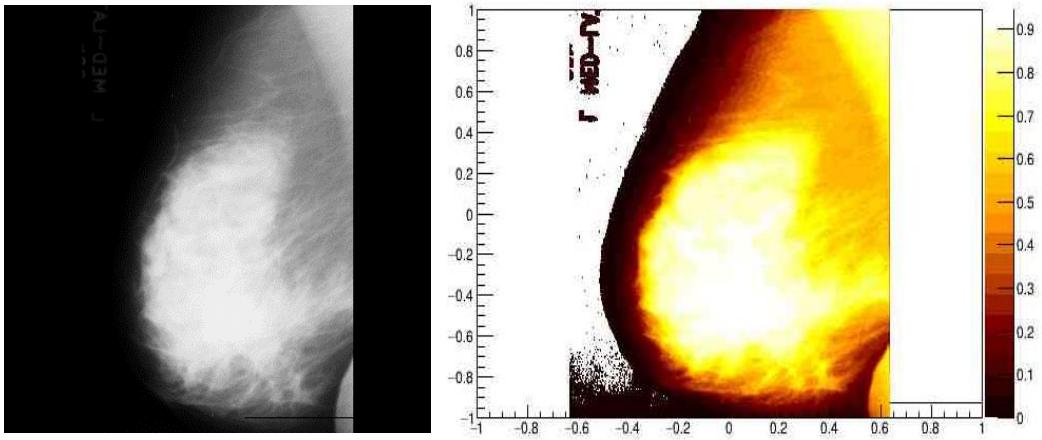

Figura B3 Muestra 3 de mamografía con T.

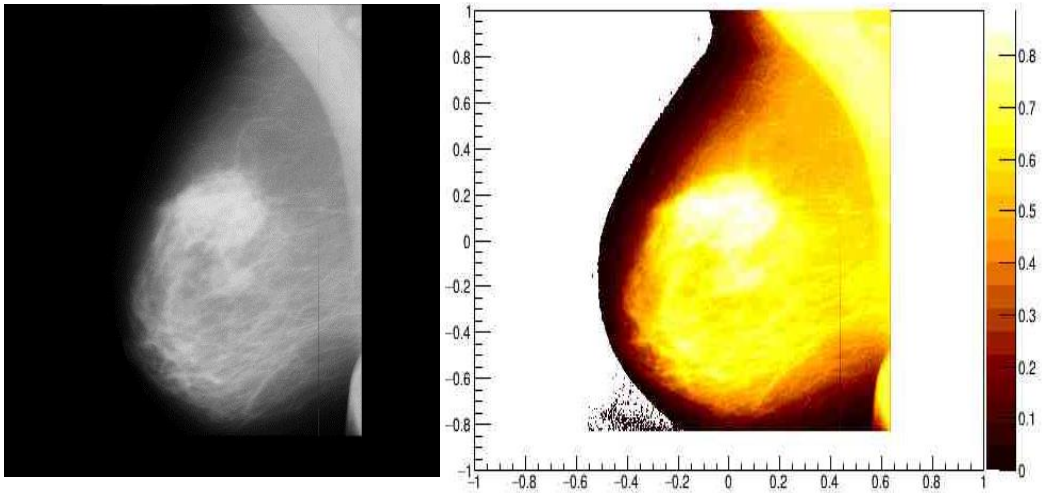

Figura B5 Muestra 5 de mamografía con T.

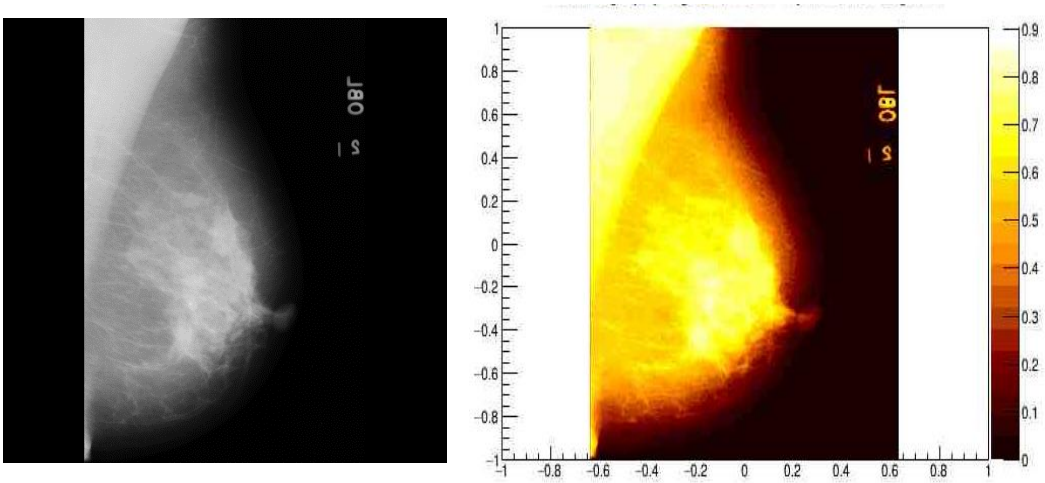

Figura B6 Muestra 6 de mamografía con T.

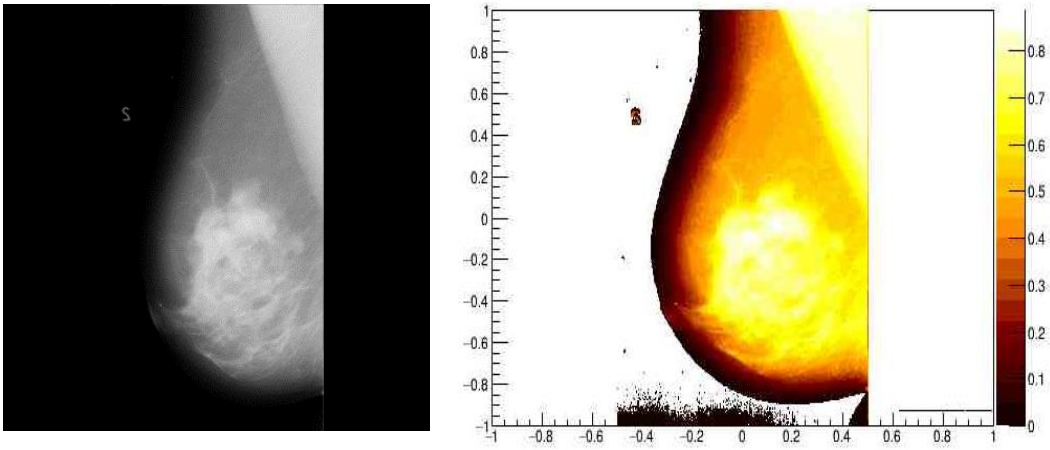

Figura B7 Muestra 7 de mamografía con T.

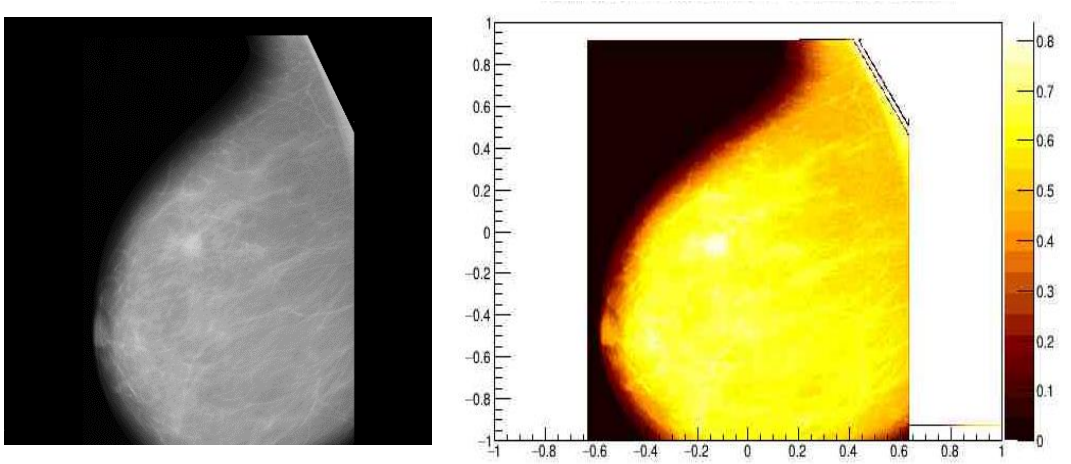

Figura B8 Muestra 8 de mamografía con T.

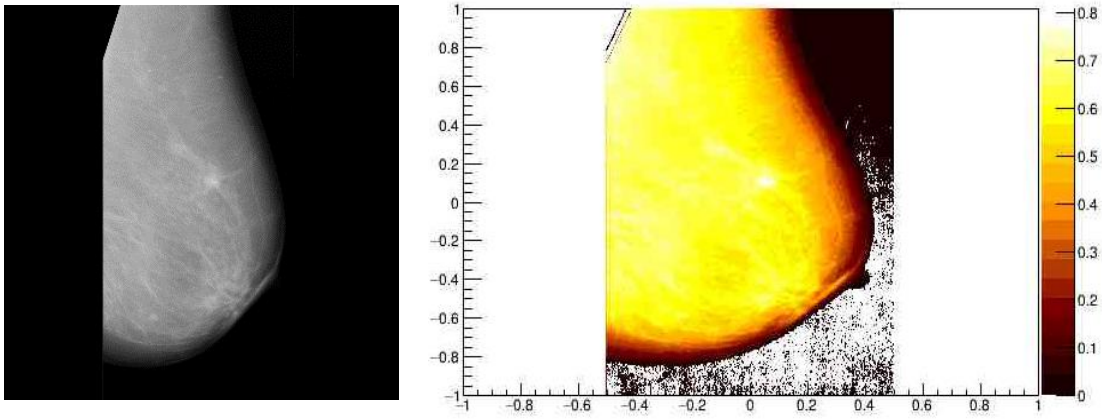

Figura B9 Muestra 9 de mamografía con T.

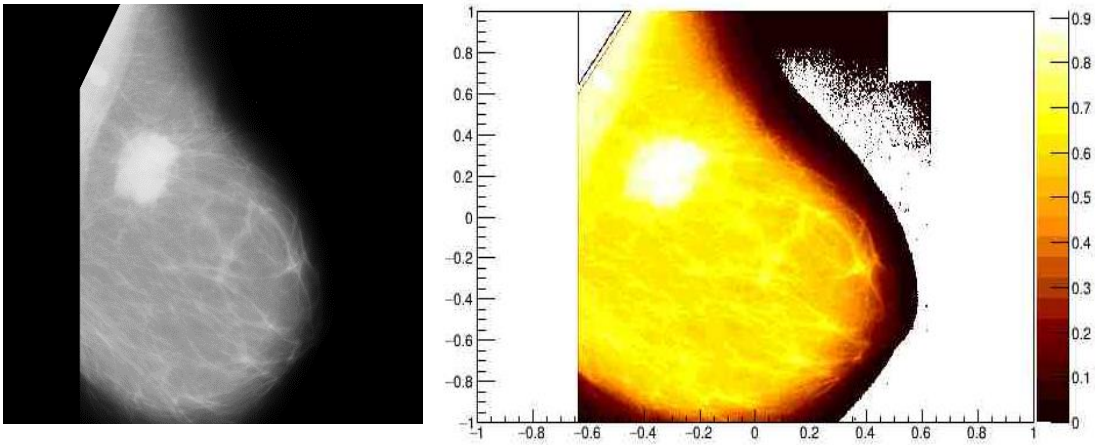

Figura 10 Muestra 10 de mamografía con T.

Bibliografía

[1] Bushong, S. C. (2010). Radiologic Science for Technologists: Physics, Biology, and Protection (9.a ed.). Barcelona, España: Mosby-Year Book.

[2] R. Padilla, A., Jaynes, R. M., & Ruprecht, A. (2010). Aparato de rayos x Generación de rayos x. Recuperado de https://issuu.com

[3] Pérez Díaz, M., Orozco Morales, R., & Suárez Aday, E. D. (2019). Detección automática de microcalcificaciones en mamografía digital. Universidad Central “Marta Abreu” de las Villas, Cuba. Recuperado de https://scholar.google.es

[4] Arancibia Hernández, P., Taub Estrada, T., López Pizarro, A., Díaz Cisternas, M., & Sáez Tapia, C. (2016, abril). Breast calcifications: description and classification according to BI-RADS 5th Edition. Revista Chilena de Radiología, 22(2). Recuperado de https://www.sciencedirect.com

[5] DW. (17-08-17). Microcalcificación sospechosa [Archivo de vídeo]. Recuperado de https://p.dw.com/p/2iO21

[6] Cárdenas-Sánchez, J., Erazo-Valle-Solís, A. A., Arce-Salinas, C., Bargalló-Rocha, E., Bautista Piña, V., Cervantes-Sánchez, G., Valero Castillo, V. (2019). Consenso Mexicano sobre diagnóstico y tratamiento del cáncer mamario. Octava revisión. Colima 2019. Gaceta Mexicana de Oncología, 18(3), 1. https://doi.org/10.24875/j.gamo.m19000180

[7] Byngy, J. W., Boydz, N. F., Fishellx, E., Jongk, R. A., & Yaffe, M. J. (1995, 28 abril). Automated analysis of mammographic densities. IOPscience. Recuperado de http://iopscience.iop.org

[8] García Naumis, G. (2002). Los fractales: una nueva geometría para describir el espacio geográfico. Recuperado de https://www.fisica.unam.mx/personales/naumis/fractales.php in Phys. Res. A 389 (1997) 81-86. See also https://root.cern/

[9] The mini-MIAS database of mammograms. http://peipa.essex.ac.uk/info/mias.html

[10] Carrero, G. (2007). Caracterización de imágenes nodulares en Mamografía por medio de dimensión fractal (Tesis de Maestría). Universidad Central de Venezuela Facultades de Ciencias y Medicina, Caracas Venezuela.
[11] Cerrato Vargas, J. L., & Quintanilla Domínguez, J. (2014). Sistema de detección de microcalcificaciones en mamografía digitalizada aplicando procesamiento de imagen e inteligencia artificial. Revista de Divulgación Científica Jóvenes en la Ciencia, 1(1). Recuperado de http://www.jovenesenlaciencia.ugto.mx
[12] Madrigal-González, Carlos A., & Prada-Vásquez, Ronny, & Fernández-McCann, David S. (2013). Detección Automática de Microcalcificaciones en una Mamografía Digital, Usando Técnicas de Inteligencia Artificial. TecnoLógicas,743-756. ISSN: 0123-7799. Disponible en: https://www.redalyc.org/articulo.oa?id=3442/344234341057.
[13] Rene Brun and Fons Rademakers, ROOT - An Object Oriented Data Analysis Framework, Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996, Nucl. Inst. & Meth.
[14] Bonsignore, F., (2004). Aplicación de las técnicas de la Teoría de la Información en el registro de Imágenes Médicas. XIII Seminario de Ingeniería Biomédica. Facultades de Medicina e Ingeniería, Univ. de la República Oriental del Uruguay.
[15] Cruz Morales, R. A., Villaseñor Navarro, Y., Pavón Hernández, C., Pérez Badillo, M., Aguilar Cortázar, L., & Pérez Zúñiga, I. (2010). Microcalcificaciones de la mama, un reto para el diagnóstico. Gaceta Mexicana de Oncología, 11(4), 251-259. Recuperado de https://www.elsevier.es/es-revista-gaceta-mexicana-oncologia-305
[16] Sociedad Americana Contra El Cáncer. (2019, 18 septiembre). ¿Qué es el cáncer de seno? Recuperado de https://www.cancer.org/es/
[17] Palazuelos, G., Trujillo, S., & Romero, J. (2014, 29 abril). Breast Tomosynthesis: The New Age of the Mammography. Revista colombiana de radioterapia. Recuperado de https://www.webcir.org
[18] Patiño, P., Restrepo, M., & Santamaría, J. A. O. (2013, 10 mayo). Devices in Digital Mammography. Revista colombiana de radiología. Recuperado de http://contenido.acronline.org
[19] Lundahl T, Ohley W J and Kuhlinski W S et al 1985 Analysis and interpolation of angiographic images by use of fractals Trans. IEEE 355–8
[20] Majumdar S and Prasad R R 1988 The fractal dimension of cerebral surfaces using magnetic resonance images Comput. Phys. 2 69–73

[21] Chen C C, Daponte J S and Fox M D 1989 Fractal feature analysis and classification in medical imaging IEEE Trans. Med. Imaging 8 133–42
[22] Li, J., Du, Q., & Sun, C. (2009). An improved box-counting method for image fractal dimension estimation. Elsevier, 2460-2469. https://doi.org/10.1016/j.patcog.2009.03.001