



Benemérita Universidad Autónoma de Puebla

---

Facultad de Ciencias Físico Matemáticas

---

SEPARACIÓN DE REGIONES DE INTERÉS PARA EL  
DIAGNÓSTICO Y TEJIDO DE FONDO EN  
MAMOGRAFÍAS MEDIANTE ANÁLISIS DE DATOS.

Tesis presentada al

**Colegio de Física**

como requisito parcial para la obtención del grado de

**LICENCIADO EN FÍSICA APLICADA**

por

Alejandra Daniela López Solís

Asesorado por

Dr. Cristian Heber Zepeda Fernández / Dr. Benito de Celis Alonso

Puebla Pue.  
8 de abril de 2022





Benemérita Universidad Autónoma de Puebla

---

Facultad de Ciencias Físico Matemáticas

---

SEPARACIÓN DE REGIONES DE INTERÉS PARA EL  
DIAGNÓSTICO Y TEJIDO DE FONDO EN  
MAMOGRAFÍAS MEDIANTE ANÁLISIS DE DATOS.

Tesis presentada al

**Colegio de Física**

como requisito parcial para la obtención del grado de

**LICENCIADO EN FÍSICA APLICADA**

por

Alejandra Daniela López Solís

Asesorado por

Dr. Cristian Heber Zepeda Fernández / Dr. Benito de Celis Alonso

Puebla Pue.  
8 de abril de 2022



**Título:** SEPARACIÓN DE REGIONES DE INTERÉS PARA EL DIAGNÓSTICO Y TEJIDO DE FONDO EN MAMOGRAFÍAS MEDIANTE ANÁLISIS DE DATOS.

**Estudiante:** ALEJANDRA DANIELA LÓPEZ SOLÍS

COMITÉ

---

Dr. Javier Miguel Hernández López  
Presidente

---

Dr. Jorge Velázquez Castro  
Secretario

---

Dr. Arturo Fernández Téllez  
Vocal

---

Dr. Guillermo Tejeda Muñoz  
Vocal

---

Dr. Cristian Heber Zepeda Fernández / Dr. Benito de Celis Alonso  
Asesor



# Dedicatoria

A mamá y papá que han sido la mayor motivación para alcanzar mis logros. A mi hermana que me brinda su apoyo incondicional y siempre cree en mí.



# Agradecimientos

A la Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias Físico Matemáticas, institución en la que finalmente he logrado realizarme como profesional. A todos quienes estuvieron durante mi carrera universitaria, Asesor de Tesis, Dr. Cristian Heber Zepeda Fernández por motivarme a desarrollar este trabajo y por sus orientaciones y paciencia; profesores, les debo mis conocimientos.

Dar las gracias a toda mi familia, mis padres Francisco López Velasco y Julia Solís Cid y hermana Sandra Gabriela López Solís, soy quién soy y he conquistado mis metas con total seguridad gracias a ellos. A mi madrina Victorina Solís Cid, por creer en mí y apoyarme siempre y a John porque su ayuda fue muy importante para la realización de este trabajo.

Finalmente, también a manera de agradecimiento, menciono a la base de datos de libre acceso para la investigación científica MINI-mias pues este análisis no habría sido posible sin las mamografías que proporciona.



# Índice general

<b>Introducción</b>	<b>XXIII</b>
<b>1. Planteamiento del problema</b>	<b>1</b>
1.1. Tema de investigación . . . . .	1
1.2. Descripción del problema . . . . .	1
1.3. Formulación del problema . . . . .	3
1.4. Delimitación del problema . . . . .	3
1.5. Justificación del problema . . . . .	3
1.6. Hipótesis . . . . .	4
1.7. Objetivos . . . . .	4
1.7.1. Generales . . . . .	4
1.7.2. Particulares . . . . .	5
1.8. Contenido . . . . .	5
<b>2. Marco Teórico</b>	<b>7</b>
2.1. Antecedentes . . . . .	7
2.2. La mama . . . . .	8
2.2.1. Anatomía de la mama . . . . .	8
2.2.2. Tipos de tejido mamario . . . . .	9
2.2.3. Tipos de lesiones mamarias . . . . .	12
2.2.4. El cáncer de mama . . . . .	17
2.3. Técnicas de exploración . . . . .	17
2.3.1. Radiografía de tejidos blandos . . . . .	17
2.3.2. La mamografía . . . . .	20
2.4. La imagen digital . . . . .	22
2.5. Fundamento matemático . . . . .	22
2.5.1. Conceptos . . . . .	22
2.5.2. Errores . . . . .	23
2.5.3. Distribución de probabilidad normal . . . . .	23
2.6. Conclusiones del capítulo . . . . .	24
<b>3. Metodología</b>	<b>25</b>
3.1. Recopilación de datos . . . . .	25
3.2. Procesamiento de imágenes . . . . .	29
3.2.1. Redigitalización de imágenes . . . . .	29
3.2.2. Conversión de imágenes a Histogramas 2D . . . . .	35
3.3. Caracterización de tejido . . . . .	38
3.3.1. Delimitación de tejido y creación de histogramas . . . . .	38
3.3.2. Ajuste . . . . .	42
3.3.3. Análisis de datos . . . . .	43

<b>4. Resultados y Discusión</b>	<b>55</b>
4.1. Estadística descriptiva de las medias de las distribuciones de intensidad . . . . .	55
4.2. Calculo del "valor umbral" para cada tipo de tejido . . . . .	56
4.2.1. Tejido F . . . . .	59
4.2.2. Tejido G . . . . .	59
4.2.3. Tejido D . . . . .	60
4.3. ROI aislada . . . . .	60
4.4. Discusión . . . . .	61
<b>5. Conclusiones</b>	<b>81</b>
<b>A. Imágenes obtenidas de la base de datos mini-MIAS empleadas para el estudio</b>	<b>83</b>
A.1. Muestra de tejido F . . . . .	83
A.2. Muestra de tejido G . . . . .	83
A.3. Muestra de tejido D . . . . .	83
<b>B. Códigos en ROOT</b>	<b>87</b>
B.1. Código para la conversión de imágenes a mapas de colores . . . . .	87
B.2. Código para la creación de histogramas de frecuencia de intensidades . . . . .	89
<b>C. Determinación del área debajo de la curva de Gauss</b>	<b>93</b>
<b>D. Resultados</b>	<b>95</b>
D.1. Tejido F . . . . .	95
D.2. Tejido G . . . . .	95
D.3. Tejido D . . . . .	107

# Índice de tablas

2.1. Densidad másica y número atómico de los tejidos humanos. Datos tomados de Manual de radiología para técnicos, por Carlyle Bushong S., 2010, ELSEVIER. . . .	20
3.1. Lista de mamografías utilizadas para el estudio que presentan tejido graso . . . . .	27
3.2. Lista de mamografías utilizadas para el estudio que presentan tejido glandular . .	28
3.3. Lista de mamografías utilizadas para el estudio que presentan tejido denso . . . . .	29
3.4. Valor del bin de intensidad máxima para cada una de las mamografías utilizadas en el estudio que presentan tejido graso . . . . .	33
3.5. Valor del bin de intensidad máxima para cada una de las mamografías utilizadas en el estudio que presentan tejido glandular . . . . .	34
3.6. Valor del bin de intensidad máxima para cada una de las mamografías utilizadas en el estudio que presentan tejido denso . . . . .	35
3.7. Datos obtenidos para la muestra de tejido graso (F). Primera columna: mamografía. Siguiendo columna: Número de identificación en la base de datos. Dos siguientes columnas: media y desviación estándar del ajuste de la distribución de intensidad en el área de tejido sano seleccionada. . . . .	46
3.8. Datos obtenidos para la muestra de tejido glandular (G). Primera columna: mamografía. Siguiendo columna: Número de identificación en la base de datos. Dos siguientes columnas: media y desviación estándar del ajuste de la distribución de intensidad en el área de tejido sano seleccionada. . . . .	47
3.9. Datos obtenidos para la muestra de tejido denso (D). Primera columna: mamografía. Siguiendo columna: Número de identificación en la base de datos. Dos siguientes columnas: media y desviación estándar del ajuste de la distribución de intensidad en el área de tejido sano seleccionada. . . . .	48
3.10. Valores de intensidad máxima ubicados en el ROI son mayores que el valor medio de la distribución de fondo más $3\sigma$ para la muestra de tejido graso . . . . .	50
3.11. Valores de intensidad máxima ubicados en el ROI son mayores que el valor medio de la distribución de fondo más $3\sigma$ para la muestra de tejido glandular . . . . .	50
3.12. Valores de intensidad máxima ubicados en el ROI son mayores que el valor medio de la distribución de fondo más $3\sigma$ para la muestra de tejido denso . . . . .	51
3.13. Valor medio de la distribución de fondo más $3\sigma$ para las mamografías que no presentan anomalía de la muestra original . . . . .	52
3.14. Valores $\mu + 4\sigma$ calculados la muestra de tejido glandular . . . . .	53
3.15. Valores $\mu + 4\sigma$ calculados la muestra de tejido denso . . . . .	54
4.1. Datos estadísticos de los valores promedio obtenidos de las distribuciones de intensidades para la muestra de tejido graso (F) . . . . .	55
4.2. Datos estadísticos de los valores promedio obtenidos de las distribuciones de intensidades para la muestra de tejido glandular (G) . . . . .	55

4.3. Datos estadísticos de los valores promedio obtenidos de las distribuciones de intensidades para la muestra de tejido denso (D) . . . . .	56
---	----

# Índice de figuras

2.1. Anatomía de la mama. De cancer.org , por American Cancer Society ( <a href="https://www.cancer.org/es/cancer/cancer-de-seno/pruebas-de-deteccion-y-deteccion-temprana-del-cancer-de-seno/mamogramas/la-densidad-de-los-senos-y-el-informe-de-su-mamograma.html">https://www.cancer.org/es/cancer/cancer-de-seno/pruebas-de-deteccion-y-deteccion-temprana-del-cancer-de-seno/mamogramas/la-densidad-de-los-senos-y-el-informe-de-su-mamograma.html</a> ) . . . . .	9
2.2. Categorías de densidad mamaria de acuerdo a BIRADS. Imágenes obtenidas de The mini-MIAS database of mammograms: MAMMOGRAPHIC IMAGE ANALYSIS SOCIETY MiniMammographic Database . . . . .	11
2.3. Morfología de una masa . . . . .	13
2.4. Contorno de una masa. De ondasyparticulas.com, por Gabriela López ( <a href="https://ondasyparticulas.com/2020/12/10/sistema-bi-rads-como-se-interpretar-una-mamografia/">https://ondasyparticulas.com/2020/12/10/sistema-bi-rads-como-se-interpretar-una-mamografia/</a> ) . . . . .	13
2.5. Densidad de una masa. De ondasyparticulas.com, por Gabriela López ( <a href="https://ondasyparticulas.com/2020/12/10/sistema-bi-rads-como-se-interpretar-una-mamografia/">https://ondasyparticulas.com/2020/12/10/sistema-bi-rads-como-se-interpretar-una-mamografia/</a> ) . . . . .	14
2.6. Mamografía mdb_248 de la base de datos MINI-mias, presenta calcificaciones de tipo benigno . . . . .	15
2.7. Mamografía mdb_219 de la base de datos MINI-mias, presenta calcificaciones de tipo maligno . . . . .	15
2.8. Mamografía mdb_170 de la base de datos MINI-mias, presenta una distorsión arquitectónica maligna . . . . .	16
2.9. Mamografía mdb_111 de la base de datos MINI-mias, presenta una asimetría maligna . . . . .	17
3.1. Patologías más comunes en mamas . . . . .	26
3.2. Redigitalización de la imagen mdb 057 . . . . .	30
3.3. Imagen mdb 057 redigitalizada con sistema de coordenadas . . . . .	31
3.4. Área del seno de la imagen mdb 057 aislada . . . . .	32
3.5. Imagen redigitalizada de la MI mdb 057 cuyos valores de intensidad han sido elevados a la quinta potencia . . . . .	32
3.6. Distribución de valores de intensidad de la Figura 3.4 . . . . .	36
3.7. MRI de la mamografía mdb 028, clasificada como tejido F con una masa bien definida . . . . .	37
3.8. Distribución de valores de intensidad de la Figura 3.7 . . . . .	38
3.9. Muestra de tejido sano seleccionada para una mamografía . . . . .	39
3.10. Muestra de tejido sano seleccionada para una mamografía . . . . .	39
3.11. Transformación a histograma de la información correspondiente a un área de tejido sano en una mamografía de tejido graso normal (mdb 009) . . . . .	40
3.12. Transformación a histograma de la información correspondiente a un área de tejido sano en una mamografía de tejido graso que presenta una masa mal definida maligna (mdb 134) . . . . .	40
3.13. Transformación a histograma de la información correspondiente a un área de tejido sano en una mamografía de tejido glandular normal (mdb 008) . . . . .	41

3.14. Transformación a histograma de la información correspondiente a un área de tejido sano en una mamografía de tejido glandular que presenta una masa circunscrita benigna (mdb 015) . . . . .	41
3.15. Transformación a histograma de la información correspondiente a un área de tejido sano en una mamografía normal de tejido denso (mdb 039) . . . . .	41
3.16. Transformación a histograma de la información correspondiente a un área de tejido sano en una mamografía de tejido denso que presenta calcificaciones malignas (mdb 249) . . . . .	42
3.17. Ajuste gaussiano de la distribución de intensidades obtenida para una región de tejido sano de la mamografía mdb 009 . . . . .	43
3.18. Ajuste gaussiano de la distribución de intensidades obtenida para una región de tejido sano de la mamografía mdb 134 . . . . .	44
3.19. Ajuste gaussiano de la distribución de intensidades obtenida para una región de tejido sano de la mamografía mdb 008 . . . . .	44
3.20. Ajuste gaussiano de la distribución de intensidades obtenida para una región de tejido sano de la mamografía mdb 015 . . . . .	45
3.21. Ajuste gaussiano de la distribución de intensidades obtenida para una región de tejido sano de la mamografía mdb 039 . . . . .	45
3.22. Ajuste gaussiano de la distribución de intensidades obtenida para una región de tejido sano de la mamografía mdb 249 . . . . .	49
4.1. Se presentan los valores de la media obtenidos para cada una de las mamografías de la muestra F con sus respectivas barras de error. La línea horizontal negra corresponde a la media de estos datos y las líneas horizontales rojas presentan el primer intervalo de confianza de los mismos. . . . .	56
4.2. Se presentan los valores de la media obtenidos para cada una de las mamografías de la muestra G con sus respectivas barras de error. La línea horizontal negra corresponde a la media de estos datos y las líneas horizontales rojas presentan el primer intervalo de confianza de los mismos. . . . .	57
4.3. Se presentan los valores de la media obtenidos para cada una de las mamografías de la muestra D con sus respectivas barras de error. La línea horizontal negra corresponde a la media de estos datos y las líneas horizontales rojas presentan el primer intervalo de confianza de los mismos. . . . .	58
4.4. Distribución de caja para la densidad media de las tres muestras . . . . .	59
4.5. Resultados para la mamografía con ID mdb 005 . . . . .	62
4.6. Resultados para la mamografía con ID mdb 012 . . . . .	63
4.7. Resultados para la mamografía con ID mdb 025 . . . . .	64
4.8. Resultados para la mamografía con ID mdb 028 . . . . .	65
4.9. Resultados para la mamografía con ID mdb 069 . . . . .	66
4.10. Resultados para la mamografía con ID mdb 134 . . . . .	67
4.11. Resultados para la mamografía con ID mdb 231 . . . . .	68
4.12. Resultados para la mamografía con ID mdb 015 . . . . .	69
4.13. Resultados para la mamografía con ID mdb 211 . . . . .	70
4.14. Resultados para la mamografía con ID mdb 213 . . . . .	71
4.15. Resultados para la mamografía con ID mdb 218 . . . . .	72
4.16. Resultados para la mamografía con ID mdb 219 . . . . .	73
4.17. Resultados para la mamografía con ID mdb 063 . . . . .	74
4.18. Resultados para la mamografía con ID mdb 102 . . . . .	75
4.19. Resultados para la mamografía con ID mdb 111 . . . . .	76
4.20. Resultados para la mamografía con ID mdb 170 . . . . .	77
4.21. Resultados para la mamografía con ID mdb 249 . . . . .	78

A.1. Imágenes recopiladas para la muestra de tejido F . . . . .	84
A.2. Imágenes recopiladas para la muestra de tejido G . . . . .	85
A.3. Imágenes recopiladas para la muestra de tejido D . . . . .	86
C.1. Área bajo la curva de Gauss. De ResearchGate , por Alfredo Baro- nio ( <a href="https://www.researchgate.net/figure/Figura-213-Areas-bajo-la-curva-normal_fig8_317800983">https://www.researchgate.net/figure/Figura-213-Areas-bajo-la-curva-normal_fig8_317800983</a> ) . . . . .	94
D.1. Resultados para la mamografía con ID mdb 010 . . . . .	96
D.2. Resultados para la mamografía con ID mdb 075 . . . . .	97
D.3. Resultados para la mamografía con ID mdb 080 . . . . .	98
D.4. Resultados para la mamografía con ID mdb 092 . . . . .	99
D.5. Resultados para la mamografía con ID mdb 095 . . . . .	100
D.6. Resultados para la mamografía con ID mdb 141 . . . . .	101
D.7. Resultados para la mamografía con ID mdb 238 . . . . .	102
D.8. Resultados para la mamografía con ID mdb 248 . . . . .	103
D.9. Resultados para la mamografía con ID mdb 252 . . . . .	104
D.10.Resultados para la mamografía con ID mdb 256 . . . . .	105
D.11.Resultados para la mamografía con ID mdb 267 . . . . .	106
D.12.Resultados para la mamografía con ID mdb 013 . . . . .	108
D.13.Resultados para la mamografía con ID mdb 017 . . . . .	109
D.14.Resultados para la mamografía con ID mdb 019 . . . . .	110
D.15.Resultados para la mamografía con ID mdb 021 . . . . .	111
D.16.Resultados para la mamografía con ID mdb 023 . . . . .	112
D.17.Resultados para la mamografía con ID mdb 030 . . . . .	113
D.18.Resultados para la mamografía con ID mdb 032 . . . . .	114
D.19.Resultados para la mamografía con ID mdb 090 . . . . .	115
D.20.Resultados para la mamografía con ID mdb 209 . . . . .	116
D.21.Resultados para la mamografía con ID mdb 264 . . . . .	117
D.22.Resultados para la mamografía con ID mdb 265 . . . . .	118
D.23.Resultados para la mamografía con ID mdb 270 . . . . .	119
D.24.Resultados para la mamografía con ID mdb 058 . . . . .	120
D.25.Resultados para la mamografía con ID mdb 105 . . . . .	121
D.26.Resultados para la mamografía con ID mdb 110 . . . . .	122
D.27.Resultados para la mamografía con ID mdb 236 . . . . .	123
D.28.Resultados para la mamografía con ID mdb 239 . . . . .	124
D.29.Resultados para la mamografía con ID mdb 241 . . . . .	125
D.30.Resultados para la mamografía con ID mdb 253 . . . . .	126



# Resumen

La mamografía es el método de diagnóstico por imagen que se obtiene por la interacción de los rayos X con el seno y actualmente es una de las herramientas imprescindibles para la detección temprana de patología mamaria. No obstante, este método de diagnóstico no es infalible, pues en ocasiones su sensibilidad se ve limitada por el hecho de que los tejidos que estudia presentan características radiológicas similares. Siendo que, en pacientes con mamas densas, se estima que el 10% de las lesiones no son visibles en este estudio. En el presente trabajo para la obtención de grado se propone que el análisis de datos de intensidad de los píxeles que conforman la imagen mamográfica es una herramienta que podría utilizarse para aislar aquellas regiones de alta densidad que pudieran ser de interés para el diagnóstico (ROI) del resto de la mama (fondo).

Se describe un método simple de análisis que busca caracterizar cuantitativamente los tipos de tejido sano a partir del análisis de su distribución de grises en la mamografía, haciendo uso de los parámetros media y desviación estándar. Se recopilaron un total de 90 mamografías obtenidas de la base de datos mini-MIAS, que presentan diferentes clasificaciones de acuerdo a su diagnóstico. Dichas imágenes se convirtieron en histogramas 2D, utilizando el sistema ROOT V. 6.23/01 para análisis de datos, a fin de ser examinadas de acuerdo a su distribución de valores de intensidades, los cuales fueron normalizados. Este procedimiento, llevado a cabo sobre áreas representativas de tejido sano llevaría a la posterior identificación de las zonas que no lo son.

Esta técnica permitió observar que los valores de intensidad ubicados en las áreas que corresponden a lesiones son mayores que el valor medio de la distribución del tejido de fondo  $+3\sigma$ . De este modo, utilizando las herramientas de la estadística se obtuvo un valor de intensidad umbral para cada tipo de tejido normal que separaría las ROI al mostrar únicamente aquellos bins que superaran el valor umbral. Finalmente, los valores de intensidad de los bins persistentes se elevan a alguna potencia, lo que permite resaltar visualmente el área aislada, para facilitar su comparación con el diagnóstico presente. En general, la región establecida como anomalía por la evaluación médica se encuentra presente en los píxeles persistentes, por consiguiente este análisis cumple sus objetivos.

Palabras clave: análisis de datos, media, desviación estándar.



# Abstract

Mammography is the diagnostic imaging method obtained by the interaction of X-rays with the breast and is currently one of the essential tools for the early detection of breast pathology. However, this diagnostic method is not infallible, as its sensitivity is sometimes limited by the similar radiological characteristics of the studied tissue. Being that, in patients with dense breasts, it is estimated that 10% of the lesions are not visible in this study. In the present work, it is proposed that the analysis of intensity data of the pixels of the mammographic image is a tool that could be used to isolate those regions of high density that could be of interest for the diagnosis (ROI) of the rest of the breast (bottom).

A simple method of analysis is described that seeks to quantitatively characterize the types of healthy tissue from the analysis of their gray distribution in mammography, using the mean and standard deviation parameters. A total of 90 mammograms obtained from the mini-MIAS database were collected, which present different classifications according to their diagnosis. These images were converted into 2D histograms, using the ROOT V. 6.23/01 system for data analysis, in order to be examined according to their distribution of intensity values, which were normalized. This procedure, carried out on representative areas of healthy tissue, would lead to the subsequent identification of areas that are not.

This technique allowed us to observe that the intensity values located in the areas corresponding to lesions are greater than the mean value of the background tissue distribution  $+3\sigma$ . In this way, using the statistical tools, a threshold intensity value was obtained for each type of normal tissue that would separate the ROIs by showing only those bins that exceeded the threshold value. Finally, the intensity values of the persistent bins are raised to some power, allowing the isolated area to be visually highlighted, to facilitate its comparison with the present diagnosis. In general, the region established as an abnormality by medical evaluation is present in persistent pixels, therefore this analysis fulfills its objectives.

Keywords: data analysis, mean, standard deviation.



# Símbolos y Abreviaturas

2D - Dos dimensiones

3D - Tres dimensiones

A - Posible región de anomalía

ACR - American College of Radiology

ARCH - Distorsión arquitectónica

ASYM - Asimetría

b - Background, región de tejido de fondo

B - Benigno

BIRADS - Breast Imaging Reporting and Database System

CALC - Calcificación

CC - Creneocaudal

CIRC - Masas bien definidas / circunscritas

D - Denso-glandular

F - Graso

FFDM - Full-field digital mammography

G - Graso-glandular

M - Maligno

MD - Doctor en medicina

MI - Mammographic images, imágenes mamográficas

MIH - 2D histogram from the MI, Histograma 2D de la imagen mamográfica

MISC - Otras masas mal definidas

MRI - Imagen por resonancia magnética

MQSA - Mammography Quality Standards Act

NORM - Normal

OML - Oblicua lateral media

OMS - Organización mundial de la salud

RMI - Re-digitized mammography image

SFM - Screen-film mammography

# Introducción

El cáncer de mama representa un importante problema social. Su incidencia ocupa el primer lugar de los cánceres en las mujeres y es su principal causa de muerte entre los 35 y 55 años. Según la Organización Mundial de la Salud (OMS), en 2020 se registraron 2,26 millones de nuevos casos de cáncer de mama y la mortalidad afectó a 685,000 personas. De estos datos se desprende la importancia de la enfermedad y del interés porque su diagnóstico sea lo más temprano posible, pues aún cuando se identifican algunos factores de riesgo, la OMS hace énfasis en que los conocimientos actuales sobre las causas del cáncer de mama son insuficientes [1], por lo que no se han logrado establecer las medidas necesarias para la prevención de esta patología, de ahí que la detección precoz siga siendo la mejor oportunidad de quienes luchan contra esta enfermedad.

El diagnóstico inicial de esta enfermedad neoplásica mejora las cifras de morbimortalidad a medio y a largo plazo, de tal modo que en los últimos años los esfuerzos se encaminan a efectuar el diagnóstico de cáncer mamario tan pronto como sea posible, es por ello que las organizaciones de salud pública en todo el mundo, han promovido las campañas de información respecto al tema recomendando la autoexploración de mama como primer método de detección precoz de anomalías, seguida de la exploración clínica periódica por el médico y el tamizaje por medio de la mastografía en mujeres en edades de riesgo, incluso si no presentan síntomas de la enfermedad. Además, con ayuda de la nueva tecnología, se ha mejorado la capacidad del diagnóstico por imagen, con técnicas tan sofisticadas como la mamografía digital, la ecografía tridimensional y la resonancia magnética.

Actualmente, la mamografía digital es por excelencia la prueba de diagnóstico mamario en la mayoría de los casos, ya que permite detectar lesiones en la mama hasta dos años antes de que sean palpables y cuando aun no han invadido en profundidad ni se han diseminado a los ganglios ni a otros órganos. Existen dos tipos de exámenes que se realizan mediante la mamografía: la detección y el diagnóstico. La mamografía de detección se realiza en pacientes que no presentan ningún síntoma de patologías mamarias mediante un protocolo de dos proyecciones, normalmente la oblicua lateral medial (OML) y la craneocaudal (CC), para detectar un cáncer no sospechado. Mientras que el diagnóstico tiene como objetivo examinar a pacientes que han presentado anomalías con anterioridad. Dada la importancia del diagnóstico precoz de esta enfermedad, se recomienda altamente realizar una doble lectura de los resultados, llevados a cabo por dos diferentes especialistas del campo correspondiente para disminuir tanto como sea posible casos de cáncer no detectados, pues en ocasiones la correcta detección de lesiones mamarias se ve limitada por la densidad mamográfica representada como una gran cantidad de tonos de gris en este tipo de imágenes radiográficas.

La digitalización de las mamografías ha sido de gran utilidad para el perfeccionamiento de la detección de lesiones mamarias que pudieran corresponder a un posible cáncer, pues favorece la facilidad de manipulación de las imágenes obtenidas utilizando softwares de procesamiento de imágenes para ayudar al radiólogo a evaluar los mamogramas en busca de las más mínimas sospechas, que serían difícilmente detectadas a simple vista si el tejido circundante presenta una alta densidad. De modo que, a fin de mejorar la capacidad de diagnóstico en este medio, se han propuesto diferentes métodos y algoritmos computarizados que permiten llevar a cabo el tratamiento de las imágenes mamografías (MI) para su estudio, tal es el caso del procesamiento mediante análisis de datos desarrollado en el presente trabajo para la obtención de grado. Está ampliamente documen-

tado que la densidad de la mama y su expresión en la escala de grises en las MI es un factor de riesgo de malignidad y un parámetro importante en la evaluación diagnóstica de la imagen, por lo que se propone llevar a cabo la caracterización de los diferentes tipos de tejido mamario sano mediante sus respectivas distribuciones de intensidades, obtenidas del adecuado procesamiento de MI disponibles en una base de datos pública para la investigación científica, utilizando el lenguaje de programación C++. La correcta caracterización cuantitativa del tejido sano proporcionará los parámetros necesarios para medir un índice de la densidad característica de cada tipo de tejido de la mama en mamografía y dará pie a la posterior identificación de áreas de interés para el diagnóstico de algunas de las patologías conocidas.

# Capítulo 1

## Planteamiento del problema

### 1.1. Tema de investigación

Separación de regiones de interés para el diagnóstico (ROI), que por la intensidad de sus píxeles, podrían representar lesiones de mama, y ruido o tejido de fondo (b) en MI mediante el análisis de valores de intensidad para los diferentes tipos de tejido sano.

### 1.2. Descripción del problema

La patología mamaria representa uno de los principales motivos de consulta en la práctica diaria de los médicos especialistas en esta área y es que si bien existen enfermedades benignas de la mama, el principal objetivo de una consulta es descartar cuanto antes la presencia de alteraciones malignas como el cáncer mama. De acuerdo con la OMS [1], el cáncer de mama es la enfermedad neoplásica más frecuente en la población mundial, con más de 2,2 millones de casos en 2020. Esta patología puede estar presente tanto en hombres como en mujeres, sin embargo, se trata de la principal causa de muerte por cáncer entre la población mundial femenina pues se estima que una de cada 12 mujeres enfermarán de cáncer de mama a lo largo de su vida [1]. Lamentablemente, a pesar de que se conoce la magnitud del problema, se trata de una enfermedad que no ha sido posible prevenir hasta ahora. Dada esta limitación, se ha hecho énfasis en la importancia de mejorar la capacidad de diagnóstico para favorecer las oportunidades de supervivencia de la población afectada; una detección rápida de las diversas anomalías es de suma importancia, pues el tratamiento del cáncer de mama puede ser muy eficaz, en particular cuando la enfermedad se detecta temprano.

Al tratarse de una de las pocas enfermedades cancerosas que tarda en presentar síntomas y que puede ser detectada antes de que lo haga, las estrategias de detección precoz recomendadas son: la autoexploración, que en México, el instituto mexicano del seguro social [2] recomienda realizar continuamente a partir de los 19 años y el tamizaje mediante mamografía, este tipo de examen se encuentra dentro de las recomendaciones de la sociedad americana contra el cáncer [3], que sugiere realizar un mamograma de referencia como primera exploración radiográfica de las mamas antes de cumplir los 40 años. Los radiólogos lo utilizarán para compararlo con futuros mamogramas que, para mujeres con riesgo promedio de padecer cáncer de mama deben llevarse a cabo de acuerdo a la siguiente información:

- Las mujeres de 40-44 años tienen la opción de iniciar la detección con un mamograma cada año.
- Las mujeres de 45-54 años deben someterse a un mamograma cada año.

## CAPÍTULO 1. PLANTEAMIENTO DEL PROBLEMA

### 1.2. DESCRIPCIÓN DEL PROBLEMA

---

- A partir de los 55 años, las mujeres puede cambiar a un mamograma cada 2 años, o pueden optar por continuar sus mamogramas anualmente. Los exámenes de detección deben continuar siempre y cuando la mujer se encuentre en buen estado de salud.

En el margen del propósito de detección, se considera que una mujer tiene riesgo promedio de cáncer de seno, si no presenta antecedentes personales de este tipo de cáncer, un antecedente familiar significativo del mismo o una mutación genética conocida que aumente el riesgo del desarrollo de dicha enfermedad y además que no recibió radioterapia dirigida al tórax antes de los 30 años de edad, en otras palabras, se trata de mujeres que no tienen síntomas ni signos de cáncer de mama.

A pesar de que los avances técnicos en los últimos años y la experiencia adquirida en la interpretación mamográfica han mejorado de forma considerable la detección y caracterización de lesiones mamarias, la mamografía no es infalible, ya que puede pasar por alto algunos cánceres. Además de que, con frecuencia, una mujer necesitará más exámenes para averiguar si una anomalía encontrada en un mamograma representa un problema maligno, es por ello que, a las mujeres que presentan alto riesgo de cáncer de seno, basándose en marcadores establecidos, se les recomienda someterse a una imagen por resonancia magnética (MRI) de los senos cada año a partir de los 30 años además de la toma de su respectiva mastografía, pues debe recalarse que una MRI debe hacerse en conjunto con, y no en sustitución de un mamograma de detección. Lo anterior debido a que, no hay suficiente evidencia que justifique una preferencia por la MRI para la detección anual en las mujeres que presentan alto riesgo, incluso considerando el factor de densidad mamográfica que, se sabe, se asocia a un aumento del riesgo de cáncer de mama y dificulta la detección de anomalías mediante mamografía, ya que aún así, un estudio de resonancia podría no detectar anomalías que una mamografía detectaría. Por ejemplo, para el caso de las microcalcificaciones, que constituyen la mayor parte de los hallazgos radiológicos en el caso de lesiones mamarias asintomáticas o sin señales visibles, se tiene que sólo son susceptibles de ser diagnosticadas radiológicamente y es que si bien la mayoría de las microcalcificaciones son originadas por patología benigna, algunos patrones agrupados específicos pueden ser indicios de patologías malignas. En un tercio de los casos, la aparición de depósitos cálcicos de características sospechosas conlleva a la existencia de un proceso maligno y además, se ha reportado que aproximadamente el 55 % de los cánceres de mama no palpables presentan microcalcificaciones notorias [4].

A pesar de que la mamografía se ha convertido en un estudio de detección fundamental de la práctica clínica diaria en gran parte del mundo, este tipo de examen de diagnóstico no es 100 % exacto, pues pueden mostrar resultados falsos de diagnóstico tanto positivos, como negativos [5] :

- Un mamograma con resultado negativo falso se ve normal aun cuando existe la presencia de anomalías de seno. En general, la sociedad americana contra el cáncer [3] reporta que los mamogramas de detección no encuentran la anomalía en alrededor de 1 de 5 casos de cáncer de seno. Además, las mujeres con senos densos son más propensas a recibir este tipo de resultados.
- Un mamograma con resultado positivo falso se aprecia como de tipo anormal, sin la existencia de patologías en el seno. Para verificar el diagnóstico, los mamogramas con resultados anormales se someten a pruebas adicionales (ecografía, MRI o incluso una biopsia de seno) para comprobar si se trata de algún padecimiento maligno. Los resultados positivos falsos son más comunes en mujeres más jóvenes, debido a que la composición de sus senos es más densa.

De acuerdo a lo anterior, una alta densidad mamográfica representa una limitante para el diagnóstico, pues la gran cantidad de tonos de gris representados en las imágenes de mamografía son difícilmente percibidos por el ojo humano incluso si la tecnología permite obtener imágenes digitales con amplio rango dinámico; es por ello que se busca desarrollar métodos de análisis que faciliten la detecciones de lesiones mamarias.

### **1.3. Formulación del problema**

La alta incidencia del cáncer de mama en la actualidad ha motivado el continuo desarrollo y mejora de la mamografía como una técnica simple y precisa para la detección de esta enfermedad. Se trata de un procedimiento de exploración radiográfica de tejidos blandos que se intentó realizar por primera vez en la década de 1920 y desde entonces ha sufrido numerosos cambios, pues se sabe que requiere técnicas específicas que difieren de las usadas en la radiografía convencional. Estas diferencias técnicas se deben a que los tejidos explorados (estructuras musculares y adiposas) tienen similares números atómicos efectivos y presentan densidades másicas con las mismas características lo que dificulta la correcta distinción entre estructuras y por ende, la eficaz detección de anomalías. Es por ello que, las técnicas radiográficas enfocadas al estudio de tejidos blandos están diseñadas para optimizar la absorción diferencial en estos tejidos tan parecidos, con el objetivo de facilitar la detección de regiones de interés, mejorando el contraste y el detalle gracias al resaltado de la interfase entre diferentes tejidos.

No obstante, en una imagen mamográfica, las variaciones de densidad de tejido están representadas por los niveles de brillo de los píxeles que la conforman, es decir por el valor numérico correspondiente, que en imágenes monocromáticas va desde 0 = negro hasta hasta 1 = blanco. De modo que, una mayor densidad mamográfica, asociada a un aumento del riesgo de padecer cáncer de mama, también enfatiza una importante limitación para el diagnóstico mediante este método de imagen, pues aún cuando la tecnología actual permite obtener imágenes digitales con un amplio rango dinámico, se tienen miles de tonos intermedios entre los cuales resulta difícil detectar anomalías a simple vista. Es por ello que en este estudio se presenta un método de tratamiento de datos dirigido a la identificación de áreas de mayor densidad en mamografías digitales, con el fin de ayudar a detectar eficazmente las posibles zonas de anormalidades en las mamas partiendo de la caracterización cuantitativa de tejido sano.

### **1.4. Delimitación del problema**

- Área académica: Física Médica
- Línea de investigación: Análisis de datos
- Sublínea de investigación: Estadística
- Delimitación espacial: La investigación se realizará sobre una muestra de 90 mamografías obtenidas de la base de datos pública mini-MIAS para la investigación científica.
- Delimitación temporal: La investigación se desarrolló en el periodo comprendido entre enero y julio del año 2021, la posterior redacción de esta tesis se realizó durante el segundo semestre de dicho año.

### **1.5. Justificación del problema**

En México, al igual que en una gran cantidad de países, el cáncer de mama es una de las enfermedades más relevantes en términos de salud pública. Los estudios sobre cáncer mamario en nuestro país han demostrado que el número de casos nuevos reportados va en aumento, no obstante, de acuerdo con la OMS, en las últimas décadas se ha registrado una importante mejora en la supervivencia de la población afectada en países que cuentan con programas de detección precoz combinados con diferentes tipos de terapias para contener la enfermedad invasiva y es que, la probabilidad de curación del cáncer de mama desciende a medida que el tumor crece con el paso del tiempo. Por tal motivo, en las últimas décadas se ha optado por la mejorarla eficacia en la detección de lesiones mamarias mediante, el método más efectivo para la detección de las mismas,

las mamografías, implementando la frecuente utilización del procesamiento digital de imágenes. Y es que si bien, debido a los avances tecnológicos actualmente se utilizan mamografías digitales, existen factores que dificultan significativamente la detección de anomalías, tal es el caso de las formas irregulares y las sutiles variaciones de densidad entre las estructuras de los tejidos estudiados que al estar representadas por variaciones de brillo de los píxeles que forman la imagen se observan como una imagen en escala de grises en la que es difícil diferenciar entre las diferentes estructuras, incluso tratándose de imágenes de alta resolución, por lo que la labor de un radiólogo al interpretar mamografías es desafiante. Diversos estudios reportan que el desempeño de los radiólogos al realizar un análisis mamográfico es de aproximadamente el 75 % debido a que los tipos de tejidos de las mamas varían considerablemente, incluso en la misma persona a lo largo de su vida.

Para ayudar en esta problemática, el presente trabajo para la obtención de grado plantea el desarrollo de un método de análisis de datos para la caracterización cuantitativa de los diferentes tipos de tejido sano presentes en las mamas, este consiste en el estudio de las diferentes distribuciones de intensidad obtenidas para zonas sin anomalías o bien, zonas de tejido de fondo *background* o *b* como se denominarán de ahora en adelante, para después excluirlas utilizando los parámetros que caractericen a la distribución obtenida y así mantener únicamente las posibles regiones de interés para el diagnóstico (ROI), es decir, aquellas que pudieran presentar una anomalía (A). Con lo anterior se entregaría una herramienta para justificar cuantitativamente la clasificación del parénquima mamario según su densidad y así, posteriormente, poder utilizarlo en pro del diagnóstico asilando zonas de alta densidad .

## 1.6. Hipótesis

- La imagen radiográfica de la mama está conformada por diferentes valores en un mapa de intensidad que se pueden analizar estadísticamente al representarlos gráficamente como un histograma.
- El análisis de la distribución de intensidades para histogramas de regiones de tejido sano permitirá arribar conclusiones importantes para la caracterización cuantitativa de los diferentes tipos de este tejido mamario, esto de acuerdo a los parámetros de las distribuciones obtenidas. Se esperan distribuciones de tipo gaussiano.
- La media  $\mu$  y la desviación estándar  $\sigma$  son excelentes índice cuantitativos del tipo de tejido de la mama según su densidad.
- Conocido el tipo de distribución (gaussiana) y los valores de los parámetros ( $\mu$  y  $\sigma$ ) que caracterizan las intensidades de los bins pertenecientes a tejido sano (t) en mamografías digitales, es posible separar este de las ROI estableciendo un valor de intensidad umbral.
- Dada la intensidad umbral para bins de tejido sano, es posible reconstruir únicamente la imagen del área que podría presentar anomalías, los resultados obtenidos deben coincidir con los datos de diagnósticos presentados en la base de datos de dónde se obtuvieron las mamografías.

## 1.7. Objetivos

### 1.7.1. Generales

Se busca mostrar que es posible caracterizar el tejido sano presente en la mama a fin de, posteriormente distinguirlo de las ROI para el diagnóstico, es decir, aquellas regiones que presentan una mayor densidad y que por ende podrían presentar una anomalía (A). Lo anterior, utilizando un método computacional de análisis de datos haciendo uso de un método cuantitativo aplicado a

la mamografía, buscando de este modo, disminuir la subjetividad en la clasificación de la densidad del parénquima mamario saludable debida a la variación de tipo interobservador.

### 1.7.2. Particulares

- Realizar un análisis de las imágenes diagnósticas de distintas características obtenidas de la base de datos utilizada.
- Crear los códigos computacionales para llevar a cabo el análisis.
- Caracterizar los diferentes tipos de tejido mamario.
- Evaluar la calidad de los resultados obtenidos.

## 1.8. Contenido

Este trabajo se encuentra dividido en 5 capítulos y 4 anexos. El capítulo 1, esta dedicado a extender un panorama general del trabajo, sirviendo como una revisión sobre el problema que representa la rápida y correcta identificación de lesiones mamarias, siendo que podría tratarse de un fuerte indicador de una enfermedad tan grave como lo es el cáncer de mama. Se presentan también la hipótesis y los objetivos del trabajo.

Debido a que esta tesis presenta un análisis basado en investigaciones realizadas anteriormente, el capítulo 2 presenta los respectivos antecedentes, así como una revisión de los conceptos más importantes respecto a la mama, lesiones mamarias y las técnicas de exploración mamaria, específicamente la mamografía, finalmente, se da una breve recopilación de la información necesaria para entender cada uno de los pasos llevados a cabo en la metodología.

De acuerdo a la disposición de la teoría expuesta en el capítulo 2, el capítulo 3 explica en detalle el proceso que se llevó a cabo para este estudio, desde la obtención de las mamografías a estudiar (muestra), su posterior procesamiento para la conversión en una nueva forma de representación de la información (histogramas) y el análisis de las distribuciones obtenidas, lo que dio paso a la caracterización de los tipos de tejido según los cuales las mamografías se clasifican en la base de datos utilizada y culmina con la separación de las ROI del resto de tejido de fondo al definir un valor umbral de intensidad de acuerdo a la caracterización.

Para mostrar la viabilidad de la técnica propuesta, el capítulo 4, muestra los resultados experimentales que consisten en los datos estadísticos que caracterizan a las distribuciones de cada tipo de tejido sano y las reconstrucciones de imágenes que sólo muestran los píxeles que superan el valor umbral definido para cada tejido, estos representan las ROI y se comparan con las regiones de anomalías establecidas en el diagnósticos, se presenta una discusión al respecto.

Finalmente. En el capítulo 6 se dan las conclusiones y las sugerencias del trabajo de tesis, así como un posible trabajo posterior.

Debido a la cantidad de mamografías estudiadas, en el primer apéndice se muestran todas y cada una de las mismas, clasificadas según la densidad de su tejido, el segundo apéndice muestra los códigos C++ creados para el análisis, el tercero se refiere a la teoría de la curva de gauss utilizada para este análisis y el cuarto contiene todos los resultados obtenidos, incluyendo aquellos que no fueron satisfactorios; cada uno de ellos es referido durante el texto cuando se requiere.



## Capítulo 2

# Marco Teórico

En este capítulo se presentan los antecedentes, fundamentos y bases teóricas del presente trabajo de Tesis. Se detallan las características de la mama y los distintos tipos de lesiones que pueden afectar a la misma, así como el examen de exploración diagnóstica utilizado para detectarlas, la mamografía o mastografía cuyo desarrollo se fundamentó en la comprensión de las bases físicas de la interacción de la radiación con la materia, específicamente con los tejidos blandos. De igual manera, se profundiza en el fundamento matemático utilizado para este análisis de datos, específicamente conceptos importantes de estadística y las propiedades de las curvas de Gauss, finalmente se presenta las conclusiones correspondientes.

### 2.1. Antecedentes

En 1976 el M.D. John N. Wolfe publicó un artículo en el que clasificó 6 años de exámenes de mamografías (7.214 casos, de enero de 1967 a enero de 1973) en cuatro categorías de patrones parenquimatosos que ideó empíricamente [7]. En este aseguraba que "la apariencia radiográfica del tejido mamario provee un método para predecir quién desarrollará cáncer de mama", de este modo, fue pionero en el estudio de la relación entre los patrones parenquimatosos de la mama, como se muestran radiográficamente y la probabilidad del desarrollo de cáncer de mama. Los métodos de medición de la densidad mamaria y la asociación de medidas cuantitativas de densidad con riesgo de cáncer de mama cobraron importancia desde entonces. Actualmente, la densidad mamaria es considerada como un factor de riesgo moderado independiente para el desarrollo de cáncer de mama. En su mayoría, estudios realizados posteriores a la contribución de Wolfe, han mostrado un incremento de 4 a 6 veces en el riesgo de cáncer de las mujeres con mamas densas comparativamente con las mujeres con mamas con densidad menor al 50%. De modo que, el porcentaje de densidad mamográfica se asocia fuertemente con un incremento del riesgo de cáncer mamario [6] y este riesgo es incluso mayor que el que se atribuye a factores como que una mujer no haya tenido hijos o bien al inicio temprano de la menstruación. Conociendo esta relación vino la necesidad de disminuir la subjetividad en la cuantificación y clasificación de la densidad del tejido mamario, a fin de homogeneizar la variabilidad del diagnóstico de un observador a otro, de modo que, se planteó la importancia del uso de métodos cuantitativos en las MI, algunos de los cuales se basan en el análisis de la distribución de intensidades presentes en las mismas; pues dado que la imagen radiográfica digital está compuesta por un conjunto de píxeles en escala de grises, la información que contiene se puede representar gráficamente como el número de píxeles asociado a cada tono de gris que corresponde con las diferentes densidades de los tejidos en la mama en una imagen, en particular: en histogramas de frecuencia. Y en vista de que, la distribución de grises en el histograma mamográfico refleja la distribución de tejidos mamarios según sus densidades, se han analizado los parámetros que caracterizan a estas representaciones bajo la hipótesis de que

estos son un excelente índice cuantitativo de la densidad global de la mama; tal es el caso de la desviación estándar [8] y de la densidad media calculada, que se ha correlacionado con los valores correspondientes de la escala BIRADS con el objetivo de investigar la presencia de un valor límite de la densidad media que separe significativamente los casos de BIRADS bajo de los altos y su relación con la edad [9]. Para un análisis del histograma de intensidades correspondiente a una mamografía, debe tenerse en cuenta que dado que las anomalías mamarias (A) son masas de alta densidad, estas se presentan como píxeles de tonos claros en la imagen de modo que, es posible distinguir entre el área A y el fondo (b) desde el histograma de la imagen mamográfica (MIH), porque el valor de intensidad más alto se localizó en el área A [10]. Al respecto conviene decir que se ha propuesto al análisis estadístico de intensidad como método para distinguir entre las áreas A y b. Ajustando las distribuciones de la muestra total a curvas de análisis conocidas encontrando así los valores de los parámetros que los describen y notando que, en general, el valor de intensidad máxima se ubica a  $+3\sigma$  del valor medio del área b, es decir, el valor de intensidad más alto está situado a más del 99,6 % de los valores medios de intensidad en el área b. Por lo tanto, para un valor de intensidad dado mayor que la media ( $\mu$ )  $+3\sigma$ , este muy probablemente se encuentra en el área A, esto independientemente de la selección del área b para estudio pues los resultados son siempre consistentes. Estos antecedentes justifican el análisis de la distribución de grises en el histograma específico de cada imagen mamográfica digital para distinguir entre áreas A y b llevado a cabo en el presente trabajo de tesis. El tratamiento en cuestión pretende inferir un corte en un umbral para excluirla sección del área b y distinguir el área A, que se precisa como un conjunto de píxeles de alta densidad.

## 2.2. La mama

Se define la mama como la estructura anatómica donde se encuentra la glándula mamaria. Las glándulas mamarias se encuentran presentes en ambos sexos, sin embargo, en las mujeres presentan un desarrollo notorio durante la pubertad, es durante esa etapa que comienza su proceso de maduración y posteriormente presenta cambios de acuerdo a la etapa de la vida que la mujer se encuentre atravesando. Su máximo desarrollo se produce durante el embarazo como parte de la preparación para la etapa posterior al parto, la lactancia.

### 2.2.1. Anatomía de la mama

En la mujer, las mamas están situadas en la parte superior delantera del pecho, a derecha e izquierda del esternón, delante de los músculos pectorales mayor y menor, en el área comprendida entre la tercera y séptima costillas. Situadas simétricamente a la altura de los brazos. Estos órganos tienen el aspecto de una media esfera, cuya cara plana se posiciona sobre el tórax y presentan, en la parte más o menos central de su cara convexa, una protuberancia pequeña, denominada pezón. Esta estructura, se considera un órgano dinámico, pues ofrece numerosas variedades influenciadas por diferentes estados fisiológicos presentes a lo largo de la vida de la mujer: embarazos, lactancias y estado de salud por ejemplo.

Dada esta última característica, hay una sorprendente ausencia de una definición clara de la "mama normal", lo que dificulta significativamente la correcta interpretación de imágenes para el diagnóstico de lesiones mamarias, se conocen una gran cantidad de cambios que ocurren en la mama sin que exista una clara delimitación o diferenciación entre lo normal y lo anormal. Además, la anatomía de este órgano glandular está compuesto por varios elementos de pequeñas dimensiones, lo que hace que la obtención de imágenes para su estudio sea difícil. Anatómicamente, la mama se puede subdividir en las siguientes estructuras: lobulillos, conductos, tejidos adiposos y tejido conectivo fibroso [3].

- Los lobulillos son los encargados de la producción de leche y a menudo se les llama tejido glandular.

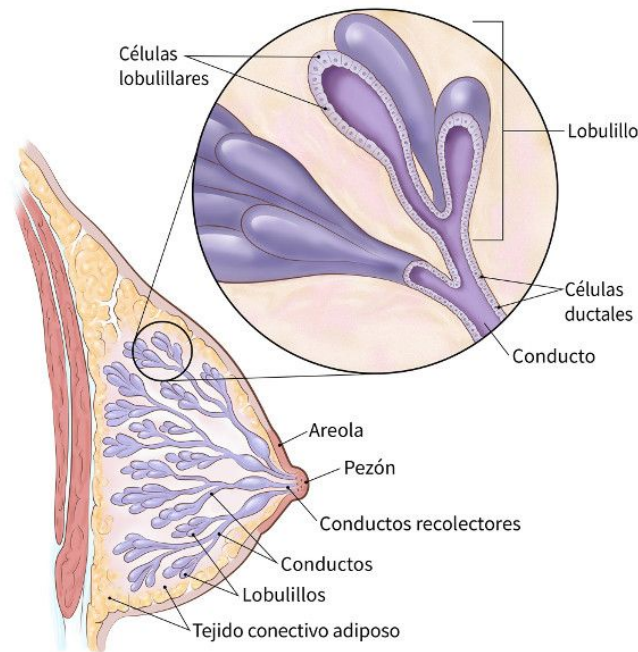


Figura 2.1: Anatomía de la mama.

De cancer.org , por American Cancer Society (<https://www.cancer.org/es/cancer/cancer-de-seno/pruebas-de-deteccion-y-deteccion-temprana-del-cancer-de-seno/mamogramas/la-densidad-de-los-senos-y-el-informe-de-su-mamograma.html>)

- Los conductos son tubos muy pequeños que conducen la leche desde los lobulillos hasta el pezón.
- El tejido fibroso y la grasa son los responsables de mantener las otras estructuras en su lugar y además, dan el tamaño y la forma a los senos.

Las glándulas dentro de las mamas se clasifican como glándulas tubuloalveolares compuestas, estas se separan entre sí por tejidos adiposos y conectivo colagenoso. Esta porción de tejidos de tipo glandular está situada en el centro y se encuentra rodeada y atravesada por tejido de clasificación adiposa; la grasa cuya cantidad varía de una mujer a otra, proporciona a la mama su característica consistencia blanda (Figura 2.1).

La glándula mamaria presenta particularidades estructurales en los diferentes periodos de su evolución: en su periodo de desarrollo, en la pubertad, durante el embarazo, durante la lactancia y durante la menopausia. La mama sufre cambios significativos con la edad de la mujer. El más notorio probablemente es aquel asociado con alteraciones hormonales características del período de la menopausia que comienza aproximadamente en la tercera y cuarta décadas de la vida de la mujer. Es al atravesar esta etapa que, de forma progresiva, la glándula mamaria se reduce a simples conductos excretorios muy atrofiados que se reducen de tamaño y se aprecia entonces un importante reemplazamiento de tipo graso.

### 2.2.2. Tipos de tejido mamario

La composición del tejido mamario refleja la densidad mamaria. La mama normal está constituida principalmente por tres tipos de tejido:

- Fibroso
- Glandular
- Adiposo (Graso)

El tejido de los senos se considera denso si está conformado en su mayoría por tejido fibroso o glandular sin tanta grasa presente. El tejido mamario denso es común. Algunas mujeres tienen más tejido mamario denso que otras. En general, los senos de las mujeres se vuelven menos densos con la edad. En una mujer premenopáusicas, los tejidos fibroso y glandular estructuran los conductos, glándulas y tejidos conjuntivos, estos a su vez, se encuentran rodeados por una fina capa de grasa, mientras que, en la mama de una mujer que ya ha atravesado la menopausia, se registra una degeneración del tejido fibroglandular y un aumento importante del tejido graso [11] .

La densidad de los senos es una característica que únicamente puede ser observada a través de los mamogramas [3]. La apariencia radiográfica de la mama varía de una mujer a otra dependiendo de las cantidades relativas de grasa y tejidos glandulares presentes. Las variaciones en la densidad mamográfica del tejido de la mama se conocen como el patrón de parénquima del seno. De acuerdo con el sistema metodológico uniforme desarrollado por el Colegio Estadounidense de Radiología (American College of Radiology, (ACR)) para que los radiólogos describan los resultados de las mamografías, denominado BIRADS (*the Breast Imaging and Reporting Data System*), la densidad mamaria se clasifica en cuatro categorías (Figura 2.2) , que van desde tejido adiposo casi en su totalidad hasta tejido extremadamente denso con muy poca grasa (A-D) [12] .

- Categoría A: casi completamente adiposos, es decir, de tipo graso
- Categoría B: se presentan áreas dispersas de densidad fibroglandular
- Categoría C: heterogéneamente densa, y esta característica puede oscurecer masas pequeñas
- Categoría D: Extremadamente denso, lo que reduce notablemente la sensibilidad de mamografía. (Casi toda la mama muestra tejido denso excepto por grasa subcutánea).

De modo que, se consideran senos densos si se clasifican como heterogénea o extremadamente densos (las últimas 2 clasificaciones C y D) y senos no densos si son grasos o dispersos (las dos primeras clasificaciones A y B). El radiólogo decide cuál de las 4 categorías describe mejor la densidad de los mismos.

El tipo de tejido se considera un factor determinante para la detección de lesiones. El nivel de dificultad de detección es menor en los tejidos predominantemente adiposos y mayor en los extremadamente densos, además, independientemente del ocultamiento de lesiones producido por el aumento de la densidad mamaria en la mamografía, se sabe que existe una relación entre la densidad del tejido mamográfico y el riesgo de padecer cáncer de mama, que fue propuesta por primera vez por Wolfe [14], quien describió cuatro clasificaciones del tipo de parénquima mamario designados por N1, P1, P2 y DY:

- Categoría N1: indica un seno en el que el parénquima o tejido mamario es radiológicamente lúcido, lo que significa que la apariencia radiográfica del seno es menos brillante y el cáncer de mama el riesgo es el más bajo.
- Categorías P1 y P2: Se refieren a densidades medias asociadas con grados intermedios de riesgo.
- Categoría DY: Describe una mama en la que el parénquima está ocupado por densidades difusas o nodulares que muestra un severo compromiso con células consideradas anormales que oscurecen y ocultan el patrón de conductos prominente. El riesgo de cáncer es para este caso el más alto.

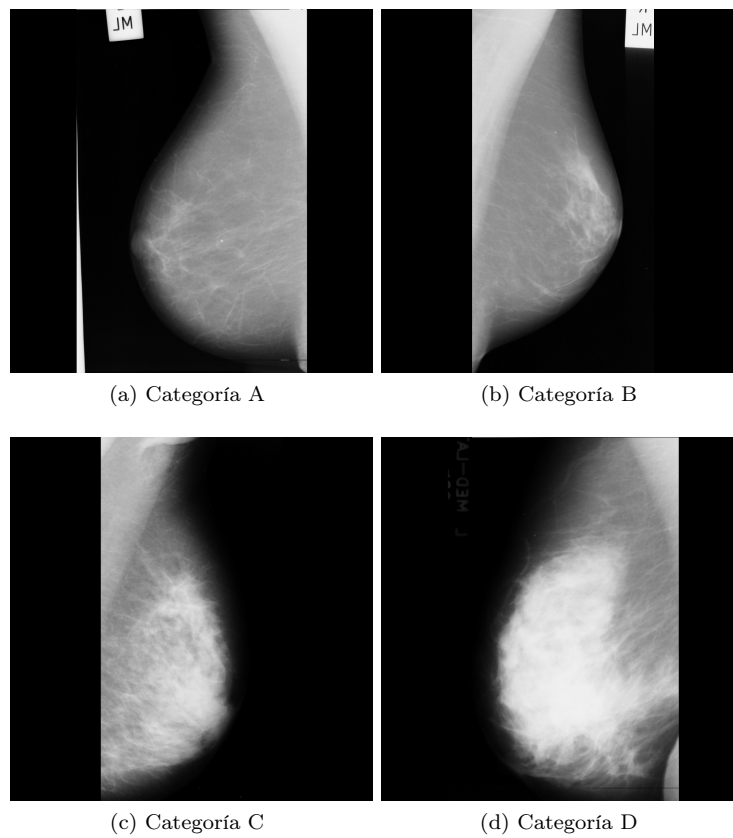


Figura 2.2: Categorías de densidad mamaria de acuerdo a BIRADS. Imágenes obtenidas de The mini-MIAS database of mammograms: MAMMOGRAPHIC IMAGE ANALYSIS SOCIETY MiniMammographic Database

Hoy en día, el porcentaje de densidad mamográfica se asocia fuertemente con un incremento del riesgo de cáncer mamario y existe una gran heterogeneidad en las estimaciones de riesgo informadas [8].

Se sabe que son diversos los factores influyen en la densidad mamaria entre los cuáles se encuentran la edad, el peso, algunos factores hormonales y la herencia principalmente. Hay un potencial uso de la densidad mamaria, en la prevención y diagnóstico del cáncer de mama, pero es necesario mejorar su medición para que esta característica pueda ser clasificada adecuadamente y se logre utilizar como un predictor de riesgo de manera individual.

### **2.2.3. Tipos de lesiones mamarias**

Es relativamente sencillo realizar la correcta identificación de lesiones mamarias en tejido adiposo, pero la tarea se dificulta considerablemente si se trata de tejido denso, esto debido a las pequeñas diferencias de densidad entre el tejido sano y aquel que no lo es. Los errores que se producen en el diagnóstico de patologías mamarias en mamografías se deben principalmente a tres factores: la alta densidad existente en la mamografía analizada, el uso de una técnica poco apropiada y a un error u omisión en la lectura de la misma.

Algunas anomalías relacionadas con el cancer de mama que es posible detectar mediante mamografía son: masas o nódulos, microcalcificaciones, distorsión de la arquitectura y asimetría. Con la finalidad de consolidar la terminología y facilitar descripción de las lesiones mamarias, estas se han clasificado dentro de los siguientes cinco grupos [15]:

- I. Lesiones circulares u ovaladas que pueden ser solitarias o múltiples.
- II. Lesiones estrelladas o irregulares y distorsión arquitectónica
- III. Calcificaciones que bien podrían o no ser asociadas con el desarrollo de un tumor. Un conjunto de calcificaciones pueden constituir un hallazgo relevante.
- IV. Síndrome de piel engrosada: se tiene piel engrosada en la mayor parte de la mama, lo que se asocia con una densidad mayor y un patrón en forma de red es visible en la mamografía.
- V. Cualquier combinación de dos o más de los resultados anteriores.

En segundo lugar, después de la clasificación anterior, las lesiones deben someterse a un análisis más detallado, en donde se evalúan características como: los contornos que presentan respecto a la densidad del tejido mamario del paciente, su morfología y también su densidad. Además, debe ponerse especial atención en las calcificaciones, que se sabe , se pueden presentar por un alteración tanto benigna como maligna de las mamas.

#### **Masas y nódulos**

De acuerdo con [12], un nódulo es una lesión pequeña, concreta y que se puede individualizar de cuanto lo rodea; en otras palabras, es una lesión bien definida; mientras que las masas se definen como lesiones que ocupan un gran volumen (espacio en tres dimensiones) y son visibles al menos en dos proyecciones mamográficas (CC y MLO). Las masas exhiben diferentes variaciones en forma, margen y densidad; según su forma, estas lesiones se pueden dividir en cuatro categorías:

- Circular
- Ovalada
- Lobulada
- Irregular

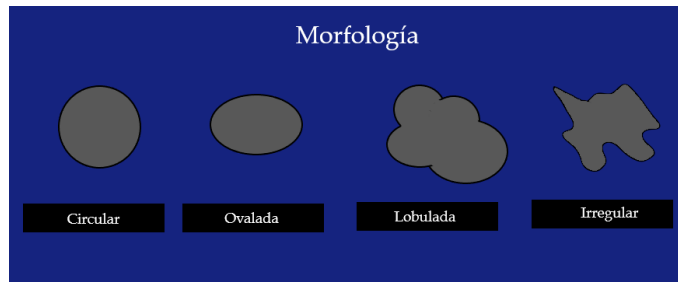


Figura 2.3: Morfología de una masa

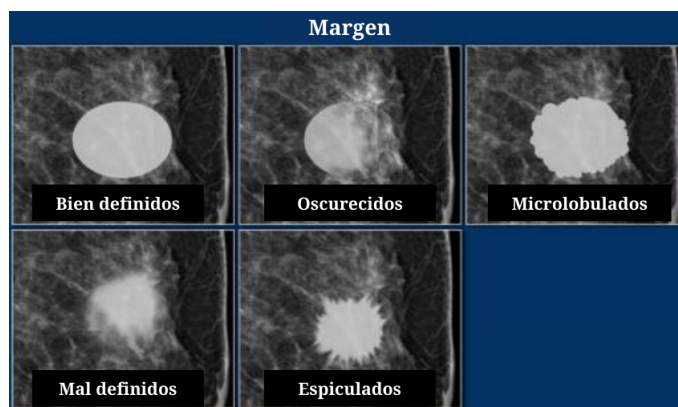


Figura 2.4: Contorno de una masa.

De [ondasyparticulas.com](https://ondasyparticulas.com/2020/12/10/sistema-bi-rads-como-se-interpreta-una-mamografia/), por Gabriela López (<https://ondasyparticulas.com/2020/12/10/sistema-bi-rads-como-se-interpreta-una-mamografia/>)

Se tiene especial interés en la última clasificación, debido a que las lesiones malignas presentan esta característica de manera frecuente.

Respecto a los contornos de la lesión, estos se evalúan con respecto a la densidad del tejido mamario del paciente. Según las fuentes de diagnóstico por imagen citadas en este trabajo, se dividen en cinco categorías:

- **Circunscritas:** Al menos el 75 % de su perímetro debe estar bien delimitado y el resto debe visualizarse sobre el tejido mamario.
- **Microlobuladas:** Entre el 50 % y 75 % de la lesión se observa bien delimitado.
- **Mal definido:** Al menos el 50 % de los márgenes de la lesión no logran delimitarse correctamente.
- **Oscurecidas:** El radiólogo considera que la lesión está bien delimitada pero se aprecia oscurecida por el tejido mamario.
- **Espiculadas:** Los márgenes que este tipo de masas presentan son muy irregulares, este hallazgo está estrechamente relacionado con la predicción de desarrollo de malignidad.

Una masa puede presentar tres diferentes niveles de densidad en comparación con el tejido sano del que se encuentra rodeado: inferior, similar o superior. Las figuras 2.3, 2.4 y 2.5 muestran las posibles clasificaciones de formas, márgenes y densidades de una masa, respectivamente.

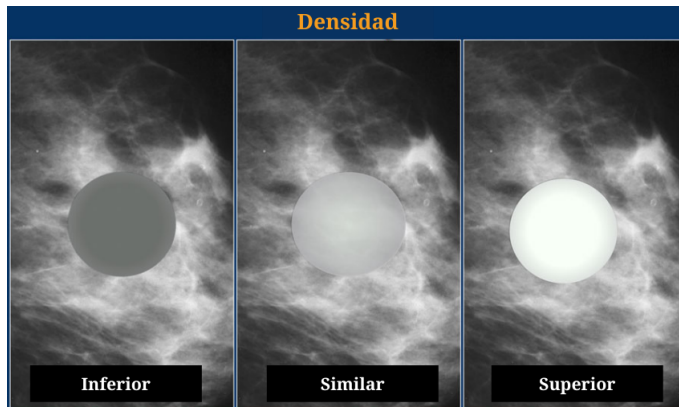


Figura 2.5: Densidad de una masa.

De ondasyparticulas.com, por Gabriela López (<https://ondasyparticulas.com/2020/12/10/sistema-bi-rads-como-se-interpreta-una-mamografia/>)

### Microcalcificaciones

Este tipo de lesiones son las más comúnmente encontradas en exámenes mamográficos. Se trata de minúsculos depósitos de calcio que aparecen en los mamogramas como puntos de brillo intenso. Pueden aparecer tanto en solitario como en grupo. De acuerdo a BI-RADS, las calcificaciones se agrupan en dos tipos:

- Típicamente benignas
- Morfología sospechosa

Las calcificaciones malignas se distinguen del tipo benigno debido al proceso que dio origen a su formación. Si las calcificaciones surgen dentro de estructuras que no contienen epitelio mamario (estroma, piel, vasos sanguíneos, tejido cicatricial), luego no son de tipo maligno[15]. El hallazgo de este tipo de anomalía es de suma importancia es de gran relevancia, pues las microcalcificaciones se forman a menudo como un subproducto de la rápida reproducción de células epiteliales como un tumor maligno. La apariencia mamográfica de las calcificaciones asociadas a patologías malignas depende en gran medida del nivel de malignidad del crecimiento celular circundante y la ubicación de estos depósitos cálcicos. A pesar de su amplia variación en apariencia, las calcificaciones de tipo maligno pueden ser clasificado en cuatro formas básicas de acuerdo a la terminología establecida por el American College of Radiology Breast Imaging-Reporting y el sistema de datos (BI-RADS). Las micalcificaciones benignas suelen ser redondeadas, de densidad muy homogénea, de tamaño uniforme, poco agrupadas y con escasos elementos (Figura 2.6).

Las microcalcificaciones malignas pueden tener variadas densidades, son heterogéneas en tamaño y de forma irregular (Figura 2.7). Aquellas que son alargadas en su forma se asocian a un alto índice de riesgo.

Con respecto al tamaño, se tiene que estas anomalías surgen dentro de los conductos de la mama, y por lo tanto están limitadas en tamaño a estas dimensiones; es por ello que se emplea el término microcalcificaciones. Aquellas que se desarrollan en los lobulillos mamarios tienden a agruparse en colonias de forma desordenada mientras que las localizadas en los conductos lo hacen de forma ramificada, siguiendo el trayecto de los mismos. Dada su composición, se trata de estructuras granulares que presentan una alta densidad comparada con el tejido circundante. Sobre el número de depósitos individuales, este tiene importancia diagnóstica debido a que pequeños grupos de microcalcificaciones irregulares dentro de una mamografía son un signo temprano de un posible cáncer de mama maligno, cuando son más de 15 por  $cm^2$ , suelen tener, para algunos

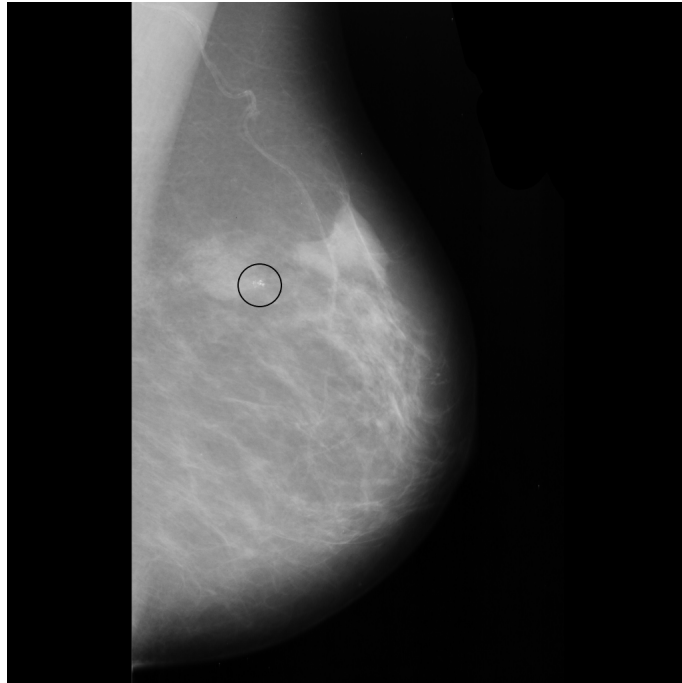


Figura 2.6: Mamografía mdb\_248 de la base de datos MINI-mias, presenta calcificaciones de tipo benigno

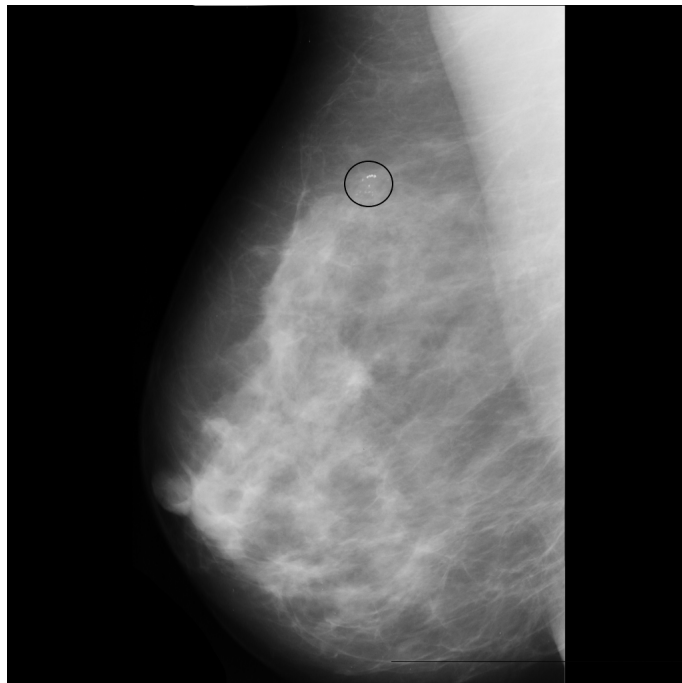


Figura 2.7: Mamografía mdb\_219 de la base de datos MINI-mias, presenta calcificaciones de tipo maligno

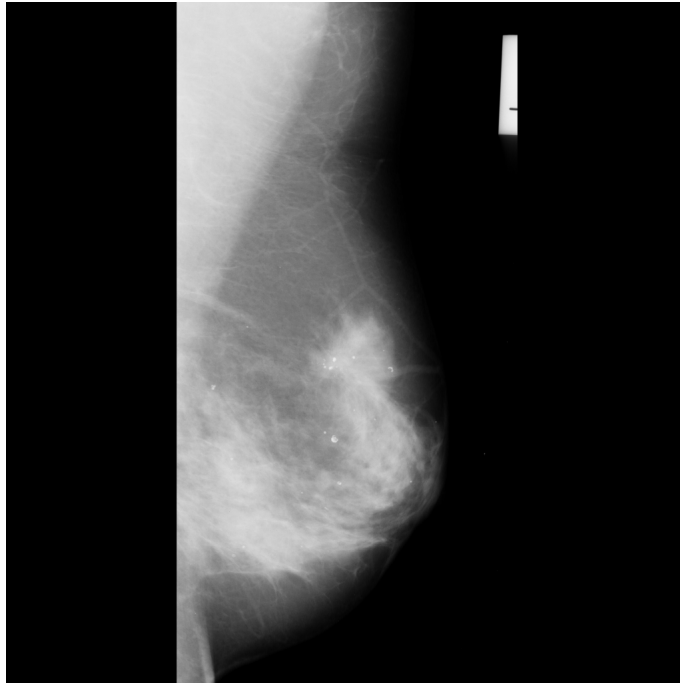


Figura 2.8: Mamografía mdb\_170 de la base de datos MINI-mias, presenta una distorsión arquitectónica maligna

autores, un porcentaje de positividad de hasta el 80 %, mientras que si este número desciende las posibilidades también lo hacen.

### **Distorsión de la arquitectura**

Se trata de una alteración que distorsiona al parénquima mamario sin la presencia de una masa perceptible. Puede estar asociado con otro tipo de anomalías como una masa, una asimetría o una microcalcificación. El aspecto mamográfico de una distorsión de este tipo es una estructura de alto brillo en forma de lesión estrellada, como delgadas líneas irradiadas desde un punto común (Figura 2.8). Una lesión de estas características puede ser causada por una serie de enfermedades de modelo maligno o también de modelo benigno. El 4 % de las neoplasias malignas se manifiestan como una distorsión en la arquitectura del tejido de la mama. La percepción de estas lesiones puede ser difícil, especialmente cuando son pequeñas. Se trata de la anomalía mamográfica que se pasa por alto con mayor frecuencia en exámenes falsos negativos [12].

### **Asimetrías**

Son lesiones que se presentan como áreas de tejido fibroglandular de diferente densidad (asimétrica) a comparación del resto del seno y son visibles en una sola vista mamográfica [12] (Figura 2.9). Una comparación detallada de los senos izquierdo y derecho mejora la detección de estructuras asimétricas, de ahí el nombre que reciben.

Es importante un conocimiento básico de la anatomía y fisiología de la mama y de las complejas estructuras subyacentes en las que tiene lugar los cambios o el desarrollo de anomalías, lo anterior para una clara apreciación del proceso patológico y para una interpretación correcta de la imagen mamográfica.

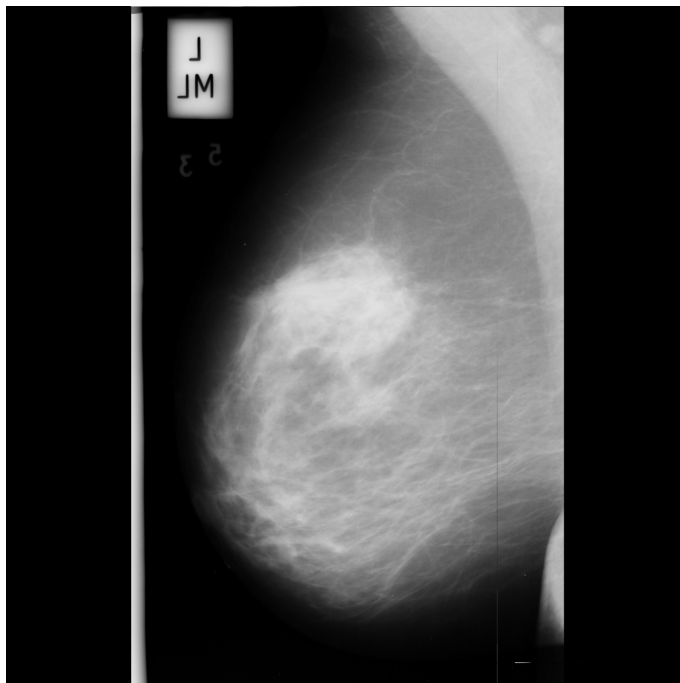


Figura 2.9: Mamografía mdb\_111 de la base de datos MINI-mias, presenta una asimetría maligna

#### 2.2.4. El cáncer de mama

El cáncer de mama constituye uno de los tumores más frecuentes entre la población mundial en la actualidad. Los métodos diagnósticos incluyen preferentemente la mamografía, ecografía y resonancia magnética. El cáncer de mama es una enfermedad que no ha sido posible prevenir hasta ahora. En palabras de la Organización mundial de la salud: "los conocimientos actuales sobre las causas del cáncer de mama son insuficientes, por lo que la detección precoz sigue siendo la piedra angular de la lucha contra esta enfermedad" (OMS, 2020a, párr. 3 [13])

### 2.3. Técnicas de exploración

#### 2.3.1. Radiografía de tejidos blandos

Para entender los principios de la diferencia entre una radiografía de tejidos blandos y una radiografía convencional es necesario entender los mecanismos de interacción de la radiación con la materia.

Los rayos X pueden interactuar con la materia de cinco maneras distintas, dependiendo de su energía. Un rayo X de mayor energía tiene menor longitud de onda, lo que le permite interactuar con estructuras atómicas más pequeñas. Los cinco mecanismos de interacción de la radiación con la materia se enlistan a continuación:

- Dispersión coherente
- Efecto Compton
- Efecto fotoeléctrico
- Producción de pares

- Desintegración fotónica

Únicamente dos de estos procesos son de importancia para la producción de imágenes radiográficas, esto se debe al rango de energía de radiación utilizado en la realización de estos exámenes: el efecto Compton y el efecto fotoeléctrico. De forma similar, únicamente dos métodos de producción de este tipo de radiación: los rayos X bremsstrahlung y los rayos X característicos, son importantes. No obstante, más importantes que los rayos X que interactúan por dispersión Compton o efecto fotoeléctrico, son los rayos X que no presentan interacción alguna con el tejido estudiado, es decir, aquellos que son transmitidos a través del cuerpo sin presentar perturbación. Esto debido a la contribución que cada uno de estos tipos de rayos X tiene en la formación de una imagen radiográfica.

El efecto Compton es un proceso de dispersión, los rayos X dispersados no contribuyen a la obtención de información útil para la imagen. Cuando un rayo X dispersado mediante esta interacción se encuentra con el receptor de imagen, éste no tiene forma de saber de dónde procede ese rayo, así que asume que este proviene de forma directa de la diana del tubo de rayos X, por lo que estos rayos X dispersos únicamente causan ruido en la imagen, el cual puede presentarse como un deslucimiento u oscurecimiento de la misma, ya que estos rayos X que no representan información diagnóstica. Para reducir este tipo de limitante, se han desarrollado técnicas para disminuir el número de rayos X dispersos que alcanzan a interactuar con el receptor de imagen.

Por otro lado, los rayos X que realizan una interacción fotoeléctrica sí proporcionan información de importancia diagnóstica al receptor de imagen. Pues debido a su absorción, estos rayos no alcanzan al receptor de imagen y son entonces representativos de aquellas estructuras anatómicas con características de una elevada absorción de este tipo de radiación, es decir, estructuras anatómicas radioopacas, las cuales se aprecian como áreas claras en la radiografía, por ejemplo las que corresponden a los huesos.

Otros rayos X penetran en el cuerpo y son transmitidos al receptor de imagen sin ningún tipo de interacción. Estos producen las áreas oscuras en una radiografía, que corresponden a estructuras anatómicas a través de las que los rayos X pasan sin encontrarse con obstáculos, estas estructuras se denominan radiolucientes. Es importante mencionar que la probabilidad de interacción disminuye al aumentar la energía de los rayos X, por lo que más rayos X se transmiten sin impedimento alguno.

Una imagen radiológica procede básicamente de la diferencia entre los rayos X absorbidos fotoeléctricamente en el paciente y los rayos X transmitidos sin interacción al receptor de imagen. A esta diferencia se le denomina absorción diferencial. Producir radiografías de elevada calidad es una labor complicada, se sabe que aproximadamente el 1 % de los rayos X incidentes en el paciente llega al receptor de imagen y menos de la mitad de los que lo alcanzan interactúan para formar la radiografía final, de modo que la imagen radiográfica procede aproximadamente de un 0.5 % de los rayos X emitidos por el tubo de rayos X [11]. Por consiguiente, es necesario el correcto control y la selección de este haz; para ello, debe tomarse en cuenta que la absorción diferencial aumenta al reducir el kVp, se requiere una selección adecuada del mismo para obtener la absorción diferencial máxima, esta es la base de las denominadas radiografías de tejidos blandos, como lo es la mamografía. Lo anterior siempre debe realizarse tratando de encontrar un equilibrio para cada exploración, pues al aumentar el contraste de la imagen, la reducción del kVp expone al paciente a una dosis mayor y de acuerdo con el principio ALARA de radiología, la dosis al paciente siempre debe ser tan baja como sea razonablemente posible.

Para la obtención de radiografías convencionales, como es el caso de imágenes de extremidades, se tiene que la estructura ósea logra apreciarse con gran contraste en comparación del tejido circundante, esto debido a que muchos más rayos X son absorbidos fotoeléctricamente en el hueso que en los tejidos blandos, pues se sabe que la probabilidad de que un rayo X se someta a un efecto fotoeléctrico es proporcional a la tercera potencia del número atómico del tejido [11]. De acuerdo con la tabla 2.1, el hueso tiene un número atómico del doble del de los tejidos blandos, por consiguiente, la probabilidad de que un rayo X se someta a una interacción fotoeléctrica es mucho mayor en la primera estructura. Por otro lado, el efecto Compton es independiente del número

atómico del tejido, de modo que, la probabilidad de la dispersión Compton para los átomos de hueso y para los átomos de los tejidos blandos es prácticamente igual y esta disminuye conforme aumenta la energía de los rayos X. No obstante, la disminución en la dispersión Compton en función del aumento de la energía, no es tan rápida como la disminución del efecto fotoeléctrico. La probabilidad del efecto Compton es inversamente proporcional a la energía del rayo X ( $\frac{1}{E}$ ); mientras que, la probabilidad del efecto fotoeléctrico es inversamente proporcional al cubo de la energía del rayo X ( $\frac{1}{E^3}$ ) [11]. Cuando se trabaja con energías reducidas la mayoría de interacciones de los rayos X con los tejidos son de tipo fotoeléctricas, mientras que a energías elevadas predomina la dispersión Compton. Naturalmente, cuando aumentan las energías de los rayos X, disminuye la probabilidad de cualquier interacción de este tipo (una mayor cantidad de radiación se propaga sin interacción). Al aumentar el kVp, más rayos X alcanzan el receptor de imagen y, por este motivo, se precisa una menor cantidad de rayos X (menor mAs de trabajo). Otros dos factores que deben considerarse para realizar una imagen radiológica de calidad son: el espectro de emisión de los rayos X y la densidad de masa del tejido a examinar.

La densidad de masa es la cantidad de materia por unidad de volumen, cuyas unidades son, generalmente, gramos por centímetro cúbico ( $\frac{kg}{m^3}$ ). Esta propiedad está relacionada con la densidad de cada átomo y depende también del nivel de cohesión que existe entre los átomos de una determinada sustancia o material. Intuitivamente se puede asegurar que, la interacción entre los rayos X y los tejidos es proporcional a la densidad de la masa de los últimos, esto sin tener en cuenta el tipo de interacción que se presente, pues un aumento en la densidad de masa nos dice que existe una mayor cantidad de electrones disponibles para interactuar, lo que implica un aumento en la probabilidad de la interacción de los rayos X. La tabla 2.1 presenta también los datos correspondientes a las densidades de masa de algunos tipos de tejido; es sabido que el hueso presenta una densidad de masa superior con respecto a los tejidos blandos, es por ello que se puede obtener la imagen ósea incluso si la absorción diferencial no estuviese relacionada con el Z.

El efecto fotoeléctrico es una interacción de absorción debido a que el rayo X incidente desaparece totalmente en el proceso, mientras que las interacciones en las que los rayos X se absorben únicamente de forma parcial, como el efecto Compton, se denominan procesos de dispersión. La reducción total en el número de rayos X que permanece en un haz de rayos X después de la penetración a través de un espesor de tejido orgánico se denomina atenuación y es producto de los procesos de absorción y dispersión, que experimenta un amplio haz de rayos X cuando incide en cualquier tejido, algunos de los rayos X originales se absorben y otros se dispersan. El resultado es una cantidad reducida de rayos X que alcanza el receptor de imagen y es así como se obtienen las radiografías. Las imágenes obtenidas empleando rayos X se componen de distintos niveles de grises, que se corresponden con las atenuaciones de la radiación al atravesar los distintos tipos de tejido presentes en el cuerpo humano. Debido a las propiedades mencionadas anteriormente, se tiene que, entre el tejido pulmonar, tejidos blandos y estructura ósea, se presenta un gran contraste, sin embargo, distinguir entre tejidos blandos representa una gran limitación para este examen diagnóstico por imagen. Esto se debe a que estos tejidos (estructuras musculares y adiposas) tienen similares números atómicos efectivos y también similares densidades másicas, por lo que la exploración radiográfica de tejidos blandos requiere técnicas específicas de mayor sofisticación que las usadas en la radiografía convencional, las técnicas radiográficas de tejidos blandos están diseñadas para optimizar la absorción diferencial. Tal es el caso de la mamografía, la exploración radiográfica de la mama, cuya técnica se mejoró con el transcurso de los años, encontrando que presenta una alta eficacia si se utilizan valores de kVp bajos y de mAs altos, pues, a pesar de la similitud de densidad entre los tejidos que forman a esta estructura, se tiene que el aspecto radiográfico de los tejidos glandular y conjuntivo es de alta densidad óptica (de apariencia más clara) y que este último es menos denso radiográficamente que el primero, mientras que la grasa es radiológicamente translúcida o clara (apariencia más oscura), así que estas características se aprovechan para obtener el mejor diagnóstico posible.

La mamografía es, por excelencia, el examen diagnóstico para la detección de lesiones mamarias pues ha demostrado ser un método simple y preciso para esta tarea, aunque no es fácil llevarlo

Números atómicos efectivos y densidad de la masa de los tejidos humanos		
Tipo de tejido	Número atómico efectivo	Densidad de la masa [ $Kg/m^3$ ]
Grasa	6.3	910
Tejidos blandos	7.4	1000
Pulmón	7.4	320
Hueso	13.8	1850

Tabla 2.1: Densidad másica y número atómico de los tejidos humanos.  
 Datos tomados de Manual de radiología para técnicos, por Carlyle Bushong S., 2010, ELSEVIER.

a cabo, debido a las limitaciones antes mencionadas para la realización de radiografías de tejidos blandos, por lo que se requiere de los conocimientos y cuidados adecuados. La mamografía ha sufrido numerosos cambios y mejoras gracias al desarrollo de ciencia y tecnología. Actualmente es una aplicación ampliamente utilizada [11].

### 2.3.2. La mamografía

#### Introducción

Se denomina mamografía o mastografía a la imagen del tejido mamario obtenida mediante un procedimiento de tipo radiológico (conlleva radiación ionizante), que se intentó realizar por primera vez en la década de 1920. Este entonces, nuevo tipo de exploración radiográfica tomo interés cuando, treinta años más tarde, a finales de la década de 1950 Robert Egan demostró de que era una técnica eficaz si se utilizaban valores de kVp bajos, de mAs altos, y una exposición directa a la película [11]. No obstante, fue hasta a fines de la década de 1960 cuando la técnica se perfeccionó con la reducción de dosis de radiación al paciente, además, el contraste y el detalle se mejoraron también, generando el resaltado de la interfase entre los diferentes tejidos y permitiendo una mayor eficacia en el diagnóstico de patología mamaria. Esta propiedad se utiliza con frecuencia en el procesamiento de imágenes digitales. Además, se encontró que este método de imagen requería de compresión del órgano estudiado para conseguir imágenes de calidad, esto debido a su forma; una mama comprimida tiene un grueso más uniforme y, por tanto, la densidad óptica de la imagen también es más uniforme. El estudio convencional consta de dos proyecciones de cada mama, una cráneo-caudal (CC) y otra oblicua mediolateral (OML), captándose una estructura en tres dimensiones proyectada sobre un plano (dos dimensiones). La información visual que se presenta en las mamografías permite conocer la estructura interna de los senos para estudiar los tejidos y buscar lesiones en base a la apariencia que presentan debido a su capacidad para atenuar la radiación. La realización de dos proyecciones en cada mama es fundamental para no pasar por alto lesiones que pueden ser visualizadas en una sola proyección, así como para realizar una correcta localización de las mismas.

Durante la última década, la mamografía digital de campo completo, FFDM, por sus siglas en inglés (full-field digital mammography) ha revolucionado las mamografías, provocando el aumento de su utilización en clínicas y hospitales reemplazando a la mamografía de película de pantalla, SFM (screen-film mammography), esto debido la serie de ventajas que una FFDM presenta, incluida una dosis de rayos X más baja, una mayor calidad de imagen, la facilidad de procesamiento posterior del archivo digital, la eficaz transmisión de imágenes, y la ausencia de contaminación química. De modo que, la mamografía digital se ha convertido en parte de la práctica clínica diaria en gran parte del mundo desarrollado. La razón principal para el continuo desarrollo y mejora de este método de estudio por imagen, es la alta incidencia del cáncer de mama y, actualmente, es el más utilizado tanto en el estudio de pacientes con síntomas de anomalías mamarias (mamografía de diagnóstico) como en pacientes asintomáticas (mamografía de detección), con el fin de diagnosticar el cáncer de mama en etapas iniciales, pues permite detectar anomalías de hasta un milímetro de diámetro,

como es el caso de las microcalcificaciones que desde luego no son palpables en un examen de los senos.

Las mamografías son efectivas, aunque pueden tener ciertas limitaciones. Es posible obtener resultados imprecisos. Esto debido a las pequeñas diferencias de densidad en los tejidos de los órganos estudiados, que en las imágenes de detección se presentan como mínimas variaciones de intensidad. Una mastografía tiene una alta sensibilidad (100 %) en la detección de lesiones sospechosas en las mamas compuestas principalmente de tejido graso [16]. Sin embargo, ese valor disminuye significativamente en los senos compuestos por tejido glandular y denso debido a que las lesiones mamarias son frecuentemente similares en densidad radiográfica a estos tipos de tejido. Este hecho dificulta el diagnóstico para los profesionales de la salud encargados de efectuarlo; en algunos casos, una mancha sospechosa en una mamografía puede terminar siendo algo que no representará mayor preocupación, mientras que en otros, se puede pasar por alto un diagnóstico de cáncer de seno. La precisión diagnóstica de la mastografía depende fundamentalmente de la calidad de la toma de los estudios radiológicos, así como de la experiencia del radiólogo para realizar su interpretación.

### **El sistema de imágenes mamográfico**

Los sistemas convencionales de imagen de rayos X no son útiles para adquirir tomas de las mamas, pues estas requieren de sistemas especialmente diseñados para su obtención, denominados mastógrafos. El equipo para la obtención de imágenes mamográficas está constituido por un generador de corriente, una columna con un brazo giratorio en el que se encuentra situado el tubo de rayos X con dianas de tungsteno (W), molibdeno (Mo) o rodio (Rh), un compresor y un receptor de imagen. Actualmente, sólo se utilizan dos tipos de estos receptores: los detectores de pantalla-película y los digitales, en años recientes se ha registrado un importante reemplazo de los primeros por los últimos, debido al constante desarrollo de la tecnología utilizada.

La era moderna de la mamografía comenzó a fines de la década de 1960 cuando la técnica se perfeccionó con equipos dedicados, como el desarrollado por el físico C. Gros. No obstante, llegó a ser clínicamente aceptada con la introducción del molibdeno como diana y filtro hasta 1966 y del receptor de imagen de pantalla-película de emulsión simple en 1972, posteriormente se hizo énfasis en la necesidad de un dispositivo integral de compresión, los generadores de alta frecuencia y la importancia de ofrecer una gran flexibilidad en la ubicación del paciente (el objetivo de la mamografía es demostrar la imagen completa de la mama), además de un tubo de rayos X con puntos focales cada vez más pequeños, debido a la exigencia de una mayor resolución espacial que requiere el tomar imágenes de microcalcificaciones, por ejemplo.

Fue 1992, que el gobierno de Estados Unidos acordó una serie de regulaciones en la Mammography Quality Standards Act (MQSA), que establecen los estándares para la calidad en la imagen, la dosis de radiación, las cualificaciones del personal y los procedimientos de evaluación para esta técnica. La elección de los materiales de las dianas de los tubos de rayos X mamográficos se debe a que los rayos X característicos del molibdeno y del rodio se encuentran dentro del intervalo de energías que es más efectivo para la formación de imágenes de las mamas, debido a que se trata de una radiografía de tejidos blandos, pues esta técnica se realiza normalmente con valores de comprendidos entre 23 y 28 kVp [11]. El bajo kVp se utiliza para conseguir maximizar el contraste radiográfico de dicho tipo de tejidos. En cuanto al receptor de imagen, los equipos de mamografía digital (FFDM) están provistos con un detector que permite la optimización en la detección de rayos X. La imagen obtenida en el detector digital pasa a un sistema digital para su debido procesamiento. La pantalla, generalmente de yoduro de cesio, absorbe los fotones de rayos X y los convierte en luz, posteriormente, el panel, de silicio, absorbe la luz y la convierte en una señal electrónica. Cada fotodiodo representa un píxel o elemento de la imagen final, de tal forma que la carga de píxel es leída por un contador electrónico de bajo ruido, convirtiéndola en datos digitales que son evidenciados en el procesador.

En una mamografía SMF, la película mamográfica cumple tres funciones: de adquisición de imágenes, exhibición y almacenamiento. No obstante, se tienen varias limitaciones como: la dismi-

nación en la visibilidad de las lesiones debido a la composición de la propia película, además, una vez expuesta y procesada la película, la imagen no puede ser alterada significativamente y partes de la mamografía pueden mostrarse con muy bajo nivel de contraste para lo que sólo pueden realizar pequeñas mejoras con una lupa y de no ser posible la realización de mejoras de este modo, es posible que el paciente deba someterse a otra toma de imagen y, en consecuencia, estar expuesto a más dosis de radiación. Otra limitación de la película es que se representan diferentes regiones de la imagen de la mama según la respuesta característica de la película mamográfica; este tipo de cuestiones han provocado el paulatino reemplazo de este tipo de mamografía por la FFDM, pues el procesamiento y la proyección del mamograma son etapas importantes en la mamografía para llevar a cabo un diagnóstico certero. A pesar de las limitaciones mencionadas, la SFM, continúa siendo la modalidad de imagen estándar para detectar lesiones sospechosas en una etapa temprana en los senos de mujeres asintomáticas en varios países, sin embargo, dado el desarrollo tecnológico de los últimos años, la información que contiene la placa revelada es digitalizada para su posterior almacenamiento y procesamiento, esto se hace utilizando un dispositivo digitalizador, como es el caso de las cámaras CCD, que convierten la luz en una señal electrónica que se digitaliza. Obteniendo así, una imagen en formato digital que incluye la información visual de cada pixel en escala de grises.

## 2.4. La imagen digital

Una imagen puede definirse como una función bidimensional  $f(x, y)$  donde las variables  $x$  y  $y$  representan las coordenadas espaciales, y el valor de la función  $f$  en cualquier par de coordenadas  $(x, y)$  es la intensidad de la imagen en dicho punto. Una imagen analógica se caracteriza por ser continua con respecto a  $x$  y  $y$ , y también en intensidad. No obstante, al convertir esta imagen a un formato digital, el proceso requiere que tanto las coordenadas como la intensidad sean digitalizadas. Digitalizar las coordenadas requiere de llevar a cabo un muestreo de las mismas, mientras que a la digitalización de la intensidad se le denomina comúnmente cuantización. Una vez que las cantidades que conforman la imagen son discretas, se dice que la imagen es una imagen digital. El resultado de muestrear y cuantizar es una matriz de números reales, cuyo tamaño, es decir, el número de filas por el número de columnas,  $M \times N$  está definido por el tamaño de la imagen y cada entrada proporciona el valor que ha sido escalado para que representar las intensidades, en el caso de un mapa de intensidad o de una escala de grises. El rango de valores sobre los que el sistema puede dar respuesta se denomina rango de escala de grises. Se describe como el número de niveles de gris que se puede representar en él. Para el caso de las mamografías digitales, las variaciones de densidad de tejido están representadas por los diferentes niveles de brillo de los píxeles que forman la imagen, es decir, a cada bin le corresponde un valor numérico bien definido, que en imágenes monocromáticas según su codificación en bits va desde negro hasta blanco pasando por varios valores de grises discretos, según la codificación utilizada, actualmente en mamografía digital son : 12~14 bit de captura = 4,096~16,384 niveles de gris y 10~12 bit de visualización en pantalla = 1,024~4,096 niveles de gris.

## 2.5. Fundamento matemático

### 2.5.1. Conceptos

Si se busca establecer conclusiones válidas sobre un grupo grande de individuos u objetos, se aplican las herramientas de la estadística a una muestra representativa de esta población. Al proceso de inferir ciertos hechos sobre la población a partir de los resultados que encontramos en la muestra se le denomina inferencia estadística y se emplea frecuentemente en los estudios científicos [17].

El proceso de obtención de la muestra se conoce como muestreo. La composición de la fracción debe reflejar lo más cercanamente posible la composición del total si se dese obtener un resultado relevante y confiable de la población de interés. Se debe asegurar que los elementos seleccionados sean representativos, se requiere la obtención de una muestra aleatoria, elegida sin favoritismos, bajo los mismos criterios.

A partir del estudio de la muestra, se utilizan estadísticas como lo son la media y la desviación estándar para extraer conclusiones sobre la población.

La medida más utilizada para definir el valor central de un conjunto de datos, que generalmente se considera como el que corresponde a la "mejor" estimación del valor real, es la media aritmética o promedio  $\bar{x}$  que se obtiene dividiendo la suma de los datos estudiados entre el total de los mismos:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (2.1)$$

Si los datos se acomodan en orden ya sea creciente o decreciente, el valor central corresponde a la mediana. En casos ideales, la media y la mediana deben ser idénticas. Sin embargo, para un conjunto pequeño de datos, estos valores tienden a ser diferentes.

La desviación estándar  $s$  es una medida del grado de proximidad de los datos alrededor del valor de la media. Si la desviación estándar es pequeña, los datos se agrupan más estrechamente al rededor de  $\bar{x}$

Para una serie infinita de datos, la media se designa con la letra minúscula griega mu  $\mu$  (la media de la población) y la desviación estándar se designa con la letra griega minúscula sigma  $\sigma$  (la desviación estándar de la población). Nunca medimos  $\mu$  y  $\sigma$  pero los valores  $\bar{x}$  y  $s$  se aproximan a ellos a medida que incrementa el número de datos, una mayor cantidad de datos de estudio implica mayor confiabilidad del estudio estadístico.

### 2.5.2. Errores

Las mediciones siempre van acompañadas de errores e incertidumbres, la mayoría de estos son causados por variaciones aleatorias en los resultados que no es posible eliminar. Para operaciones aritméticas con números que tienen asociado un error aleatorio es de esperar que algunos se compensen entre sí [18]. No obstante, es posible calcular la incertidumbre de la respuesta luego de someter a los datos a distintas operaciones, para el caso de las sumas y restas, el error total se obtiene a partir de las incertidumbres absolutas de los términos individuales:

$$e_4 = \sqrt{e_1^2 + e_2^2 + e_3^2} \quad (2.2)$$

### 2.5.3. Distribución de probabilidad normal

Se sabe que la distribución de los datos obtenidos en el estudio de la mayoría de los procesos analíticos cuantitativos tiende a una curva de Gauss. Una serie finita de datos siempre se desviará algo de esta curva teórica, no obstante, la aproximación de la distribución real y la curva mejora conforme más datos se tengan. Como regla general si se tiene más de 30 datos y estos no están claramente sesgados podemos utilizar con certeza una distribución gaussiana.

Este tipo de distribución puede ser descrita por una ecuación que contiene solo dos parámetros, la media poblacional  $\mu$  y la desviación estándar poblacional  $\sigma$ . La ecuación de la curva de Gauss es:

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2.3)$$

El término parámetro se refiere a las cantidades que definen una población o una distribución como  $\mu$  y  $\sigma$ . Los valores tipo  $x$  son las variables observadas y los los valores como  $\bar{x}$  y  $s$  que representan la estimación de parámetro a partir de una muestra finita de datos se denominan datos estadísticos.

Para una curva de Gauss ideal:  $\sigma = 1$  y  $\mu = 0$ . El valor máximo de  $y$  se encuentra en  $x = \mu$  y la curva es simétrica entorno a ese mismo valor. Generalmente, las desviaciones con respecto a la media se expresan en términos de la desviación estándar, pues esto resulta de gran utilidad. De modo que,  $x$  se transforma en  $z$  utilizando la expresión:

$$z = \frac{x - \mu}{\sigma} \approx \frac{x - \bar{x}}{s} \quad (2.4)$$

La probabilidad de medir  $z$  en un intervalo es igual al área de ese intervalo.

Dado que la suma de las probabilidades de todas las medidas es la unidad, el área debajo de la curva desde  $z = -\infty$  a  $z = \infty$  es igual a 1. Esto se ve garantizado en la ec. ?? por el factor  $\frac{1}{\sigma\sqrt{2\pi}}$  que se llama factor de normalización.

La desviación estándar es una medida de la anchura de la curva de Gauss. Cuanto mayor es el valor de  $\sigma$  más ancha es dicha curva. Toda curva de Gauss cumple con su teoría establecida, y es que el 99.7% de los datos no distan de la media en más de tres desviaciones estándar (Véase apéndice C).

## 2.6. Conclusiones del capítulo

En este capítulo se presentan los antecedentes que justifican el análisis de la distribución de grises del histograma obtenido de la mamografía digital para establecer la presencia de regiones de alta densidad. También se presentan los conceptos que serán relevantes para el entendimiento de este trabajo, comenzando por la anatomía de la mama, los distintos tipos de tejido y las lesiones que se pueden presentar así como la técnica de rayos X mediante la que se obtuvieron las muestras que serán analizadas: la mamografía. Además se abordan brevemente las herramientas estadísticas utilizadas en el desarrollo.

## Capítulo 3

# Metodología

En esta sección se presenta la metodología empleada para realizar la separación de ROI y tejido de fondo (b) mediante análisis de datos; este proceso incorpora la adquisición de las MI utilizadas, la adaptación de las mismas de su formato original al formato de datos de entrada requerido para el análisis; el procesamiento necesario para adaptar las imágenes originales a mapas de escala de grises, con el propósito de transformar la información de interés para este trabajo en histogramas 2D; el posterior análisis de este tipo de representación gráfica de acuerdo a la distribución de valores de intensidades para las áreas seleccionadas y finalmente, la caracterización de las representaciones resultantes para realizar cortes en cierto umbral de un valor promedio de bin en tejido sano que concluirá con la identificación y aislamiento de zonas que presentan una alta densidad y pueden ser de interés en el diagnóstico para cualquier mamografía.

### 3.1. Recopilación de datos

Para llevar a cabo el análisis propuesto en el desarrollo de este problema de tesis, se recopiló una muestra de 90 de 322 MI de proyección OML, disponibles en la base de datos de libre acceso para la investigación científica mini-MIAS (Mammographic Image Analysis Society) [20], perteneciente a PEIPA: el archivo piloto europeo de procesamiento de imágenes. Se asume que, por tratarse de una base de datos para la investigación científica citada en múltiples trabajos formales, las MI contenidas en esta base de datos se registraron bajo las mismas condiciones de exposición, es decir, mismo valor kVp y mAs y que además se encuentran normalizadas a una misma escala de grises (12 u 8 bits) y a una altura máxima del número de píxeles (0 a 1).

Se trata de imágenes digitalizadas en formato .pgm (Portable graymap format), este tipo de formato para imágenes almacena la información en forma de texto proporcionando el número de fila y columna para un bin y el número de nivel de gris que este presenta. Para este caso, cada píxel contiene la información de un valor de la escala de gris en 8 bits, es decir, se permiten  $2^8 = 256$  niveles de gris, que van desde 0 (negro) hasta 255 (blanco).

Los archivos PGM se pueden descargar directamente desde la página oficial [20] de un tamaño de  $1024 \times 1024$  píxeles. Además, la página provee una lista completa de los datos descriptivos de cada imagen utilizando los siguientes marcadores:

- **Carácter del tejido de fondo:**
  - Graso (F)
  - Graso-glandular (G)
  - Denso-glandular (D)
- **Clase de anomalía presente:**
  - Calcificación (CALC)

Masas bien definidas / circunscritas (CIRC)  
Otras masas mal definidas (MISC)  
Distorsión arquitectónica (ARCH)  
Asimetría (ASYM)  
Normal (NORM).

- **Gravedad de la anomalía:**
  - Benigno (B)
  - Maligno (M)

Todas las mamografías de la base de datos poseen un número de identificación y se encuentran organizadas en pares, donde cada par representa las mamografías de la mama izquierda (números de nombre de archivo pares) y de la mama derecha (números de nombre de archivo impares) de un solo paciente; también se proporciona información detallada de cada una de las imágenes, obtenida de sus respectivos diagnósticos clínicos, como son:

- Las coordenadas de imagen  $x, y$  del centro de la anomalía, tomando el origen del sistema de coordenadas en la esquina inferior izquierda de la figura.
- El radio aproximado (en píxeles) de un círculo que encierra la anomalía; para el caso en que las calcificaciones se distribuyen ampliamente por la imagen en lugar de concentrarse en un solo sitio, las ubicaciones de los centros y los radios son inapropiados y por ende no se proporciona esta información.

El conjunto de imágenes seleccionadas para este estudio se conforma de 30 mamografías de cada tipo de tejido de fondo, es decir, tres diferentes muestras, cuyos elementos pertenecen o bien a pacientes sanas (mamografías que no presentan irregularidades, denominadas normales) o a quienes presentan diferentes patologías de distinta gravedad (Tablas 3.1, 3.2 y 3.3). Para visualizarlas véase el apéndice A. La figura 3.1 muestra los distintos tipos de de lesiones sobre los que se centrará este trabajo.

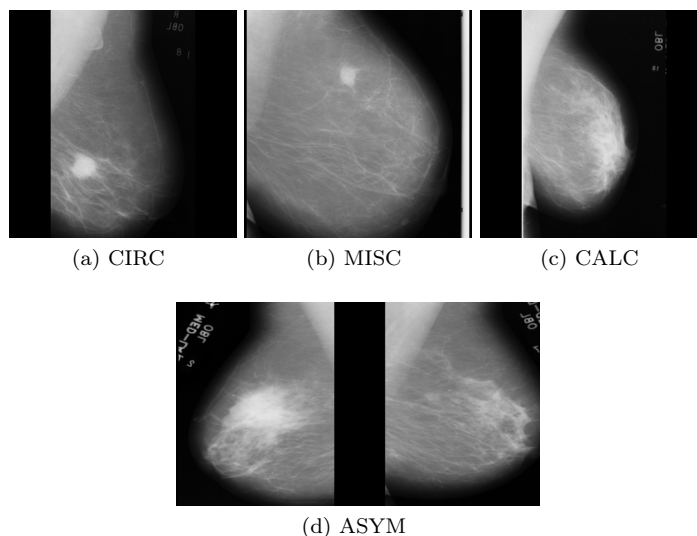


Figura 3.1: Patologías más comunes en mamas

**CAPÍTULO 3. METODOLOGÍA**  
3.1. RECOPIACIÓN DE DATOS

Muestra de tejido graso (F)			
Imagen	Número de identificación	Clase de anomalía	Gravedad de la anomalía
1	mdb 005	CIRC	B
2	mdb 009	NORM	
3	mdb 010	CIRC	B
4	mdb 011	NORM	
5	mdb 012	CIRC	B
6	mdb 025	CIRC	B
7	mdb 026	NORM	
8	mdb 027	NORM	
9	mdb 028	CIRC	M
10	mdb 059	CIRC	B
11	mdb 060	NORM	
12	mdb 069	CIRC	B
13	mdb 070	NORM	
14	mdb 075	ASYM	M
15	mdb 076	NORM	
16	mdb 077	NORM	
17	mdb 078	NORM	
18	mdb 079	NORM	
19	mdb 080	CIRC	B
20	mdb 087	NORM	
21	mdb 092	ASYM	M
22	mdb 095	ASYM	M
23	mdb 134	MISC	M
24	mdb 141	CIRC	M
25	mdb 231	CALC	M
26	mdb 238	CALC	M
27	mdb 248	CALC	B
28	mdb 252	CALC	B
29	mdb 256	CALC	M
30	mdb 267	MISC	M

Tabla 3.1: Lista de mamografías utilizadas para el estudio que presentan tejido graso

**CAPÍTULO 3. METODOLOGÍA**  
**3.1. RECOPIACIÓN DE DATOS**

Muestra de tejido glandular (G)			
Imagen	Número de identificación	Clase de anomalía	Gravedad de la anomalía
1	mdb 007	NORM	
2	mdb 008	NORM	
3	mdb 013	MISC	B
4	mdb 014	NORM	
5	mdb 015	CIRC	B
6	mdb 016	NORM	
7	mdb 017	CIRC	B
8	mdb 018	NORM	
9	mdb 019	CIRC	B
10	mdb 020	NORM	
11	mdb 021	CIRC	B
12	mdb 022	NORM	
13	mdb 023	CIRC	M
14	mdb 024	NORM	
15	mdb 029	NORM	
16	mdb 030	MISC	B
17	mdb 031	NORM	
18	mdb 032	MISC	B
19	mdb 041	NORM	
20	mdb 042	NORM	
21	mdb 072	ASYM	M
22	mdb 090	ASYM	M
23	mdb 209	CALC	M
24	mdb 211	CALC	M
25	mdb 213	CALC	M
26	mdb 218	CALC	M
27	mdb 219	CALC	M
28	mdb 264	MISC	M
29	mdb 265	MISC	M
30	mdb 270	CIRC	M

Tabla 3.2: Lista de mamografías utilizadas para el estudio que presentan tejido glandular

Muestra de tejido denso (D)			
Imagen	Número de identificación	Clase de anomalía	Gravedad de la anomalía
1	mdb 003	NORM	
2	mdb 004	NORM	
3	mdb 033	NORM	
4	mdb 034	NORM	
5	mdb 035	NORM	
6	mdb 036	NORM	
7	mdb 037	NORM	
8	mdb 038	NORM	
9	mdb 039	NORM	
10	mdb 040	NORM	
11	mdb 053	NORM	
12	mdb 054	NORM	
13	mdb 057	NORM	
14	mdb 058	MISC	M
15	mdb 061	NORM	
16	mdb 062	NORM	
17	mdb 063	MISC	M
18	mdb 064	NORM	
19	mdb 065	NORM	
20	mdb 066	NORM	
21	mdb 102	ASYM	M
22	mdb 105	ASYM	M
23	mdb 110	ASYM	M
24	mdb 111	ASYM	M
25	mdb 170	ARCH	M
26	mdb 236	CALC	B
27	mdb 239	CALC	M
28	mdb 241	CALC	M
29	mdb 249	CALC	M
30	mdb 253	CALC	M

Tabla 3.3: Lista de mamografías utilizadas para el estudio que presentan tejido denso

Es importante mencionar que cada uno de los elementos que conforman las muestras se seleccionó bajo los mismos criterios, sin conocer la vista de ninguna MI, esto con el afán de asegurar que el muestreo fuera adecuado para un análisis estadístico, pues se espera que las conclusiones extraídas de la fracción estudiada reflejen lo mejor posible las características de la población total.

Una vez descargadas, cada una de las imágenes elegidas se convirtió a un formato .png para poder ser visualizadas y posteriormente ser sometidas a procesamiento, pues el código C++ utilizado para el análisis (Ver apéndice B.1.) requiere de una imagen en este formato como datos de entrada.

## 3.2. Procesamiento de imágenes

### 3.2.1. Redigitalización de imágenes

Finalizada la adquisición de imágenes, se procedió a realizar el tratamiento necesario para poder operar sobre ellas.

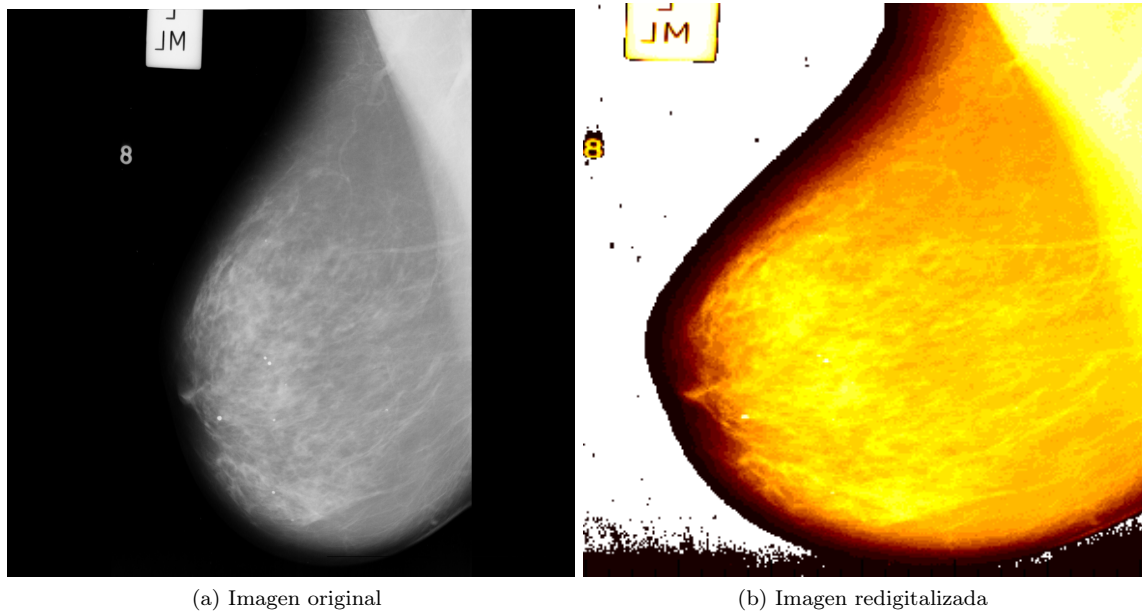


Figura 3.2: Redigitalización de la imagen mdb 057

El análisis comenzó convirtiendo las MI del formato .png en una imagen cuyos valores de píxeles corresponden a una escala de grises, también llamado mapa de intensidades, utilizando la versión del software ROOT 6.23 / 01 [21] para análisis de datos desarrollado por el CERN basado en C++. La clase de referencia utilizada fue TASImage [22], la cual permite transformar la imagen original a un nuevo formato de cuadrícula, en el que a cada píxel de la MI se le asigna un valor de intensidad normalizado, de tal manera que es posible acceder a la información de cada píxel, lo que facilita la manipulación de la imagen. Este formato no posee información de color y es representado por una matriz de datos cuyos valores representan intensidades dentro del rango establecido, cada elemento de la matriz correspondiente a un píxel de imagen que puede ser identificado mediante los dos ejes de un sistema coordenadas geométricas (sin unidades) y cuyo valor de intensidad se normalizó para este caso (0=negro hasta 1=blanco). El mapa de intensidad fue creado por la referencia de clase SetBinContent [23], que corresponde a una matriz 2D que almacena el conjunto de valores de intensidad. A esta nueva representación, la denominaremos imagen mamográfica re-digitalizada (RMI).

Este es el espacio fase en el que se llevó a cabo el análisis posterior, note que el mismo contiene toda la información de la imagen original, pues cada píxel está bien determinado al conocer su ubicación y su valor de intensidad. Cada transformación fue conformada por 1,048,576 píxeles, con un tamaño de píxel de  $0,001953 \times 0,001953$ . En la Figura 3.2 se muestra la MI original y su transformación a mapa de intensidades RMI (el espacio de fase anteriormente mencionado) de una de las imágenes de la base de datos usada (mdb057).

El conjunto de valores de intensidad se encuentra almacenado en una matriz cuadrada donde cada elemento tiene una correspondencia uno a uno con la RMI. Cada entrada en la matriz corresponde a la posición de un píxel de MI.

Esta parte es la parte fundamental de este trabajo, porque de esta será posible realizar el análisis de los datos correspondientes a los valores de intensidad, es decir, se obtuvo una representación visual de la MI en el que los valores de intensidad se aprecian comparando con una paleta de niveles de brillo, para tratamiento cualitativo, además de un arreglo matricial de valores de intensidad de la misma para un análisis cuantitativo.

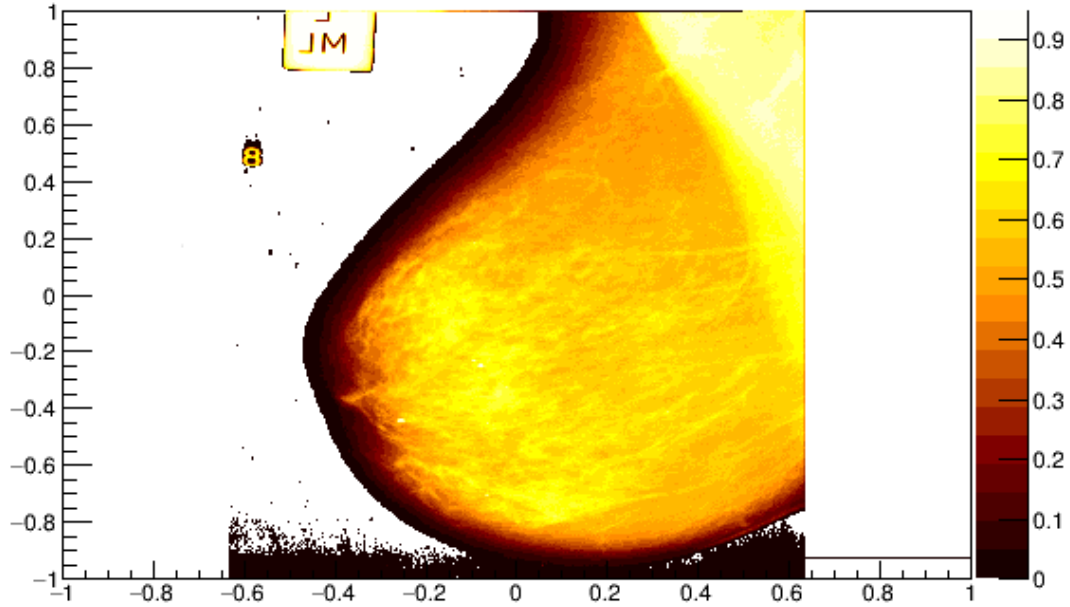


Figura 3.3: Imagen mdb 057 redigitalizada con sistema de coordenadas

Con el fin de comparar las coordenadas de la localización del centro de la anomalía (A) proporcionado por el diagnóstico de cada MI con la ubicación del elemento de valor de intensidad máxima de la matriz, se facilita el elemento de matriz de intensidad que corresponde el mismo, lo anterior, comenzando el conteo de elementos en la esquina superior izquierda del arreglo. Además, se implementó en las RMI un sistema cartesiano con el origen situado en el centro de la imagen cuya escala va de  $[-1,1]$  en ambos ejes y las coordenadas del punto de mayor intensidad también se presentan para este marco de referencia. Es importante mencionar que, el proceso de re-digitalización trae consigo una modificación de la forma de la imagen, esto es debido a que, para la nueva representación, se optó por píxeles cuadrados (Figura 3.3).

Para la mayoría de las imágenes examinadas se tiene que las coordenadas  $(x, y)$  correspondientes al bin con el valor de intensidad más alto coinciden con las coordenadas que el diagnóstico presenta como el centro del área de las diferentes lesiones en las mamografías anormales. No obstante, esta técnica es vulnerable a errores, debido a que existen bins de mayor intensidad en regiones que no son de interés diagnóstica, tal es el caso de los caracteres correspondientes a los datos del paciente en la región superior izquierda de la Figura 3.3 o bien, en la zona correspondiente al músculo pectoral, que no tiene ningún valor a la hora de clasificar la mama y puede producir errores en el diagnóstico y aumentar el tiempo de procesamiento. Por este motivo se optó por aislar estas áreas de la imagen, manteniendo únicamente la región de los senos para luego encontrar las coordenadas del valor de intensidad máxima (Figura 3.4). También es importante mencionar que, dado que las intensidades están normalizadas, nos es posible elevar los valores de intensidad a ciertas potencias con el fin de resaltar las áreas que podrían ser de interés, es decir las que presentan mayor intensidad (Figura 3.5). Esta herramienta se utilizará posteriormente para una mejor visualización de regiones.

Una vez removidas las áreas que no son de interés para el diagnóstico, se recopilieron los datos del valor de intensidad máxima para cada una de las mamografías, se obtuvo que, para las mamografías que presentan anomalías el bin de mayor intensidad se encuentra localizado dentro del área de interés (ROI). Este primer resultado coincide con el obtenido en el antecedente principal de este

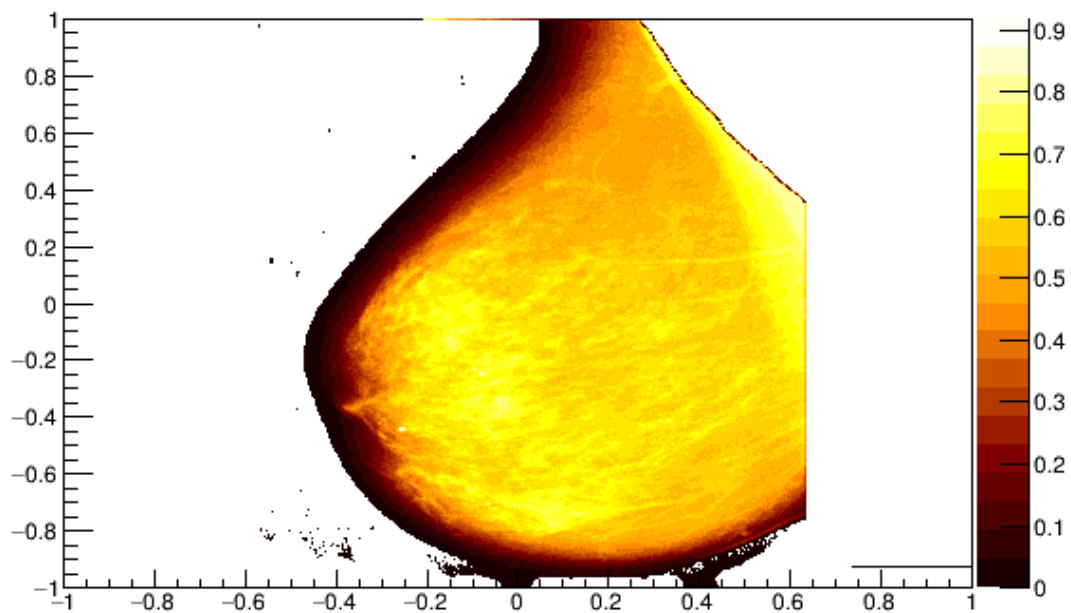


Figura 3.4: Área del seno de la imagen mdb 057 aislada

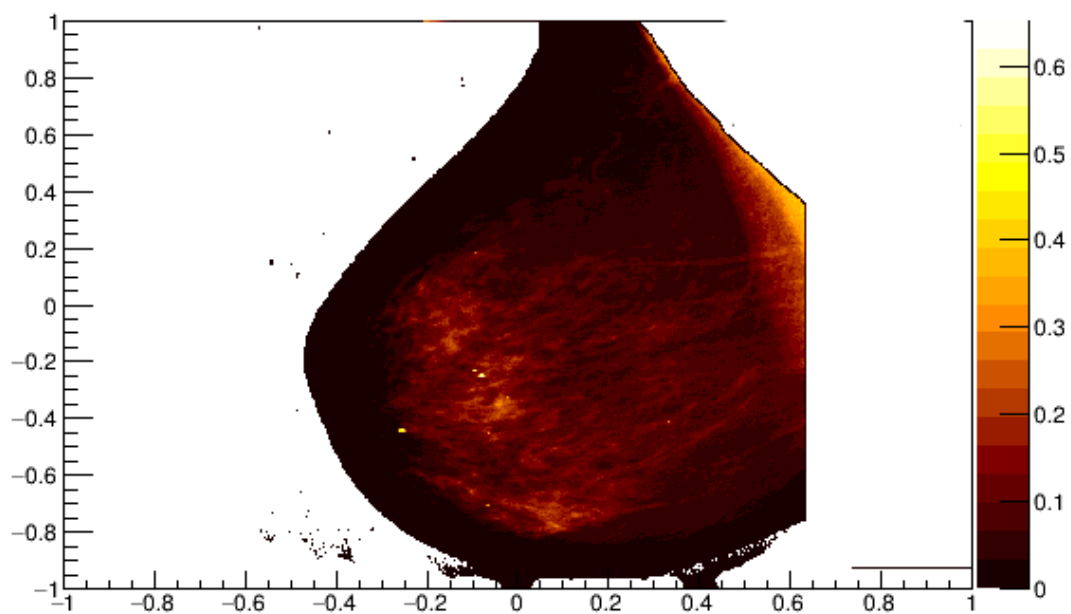


Figura 3.5: Imagen redigitalizada de la MI mdb 057 cuyos valores de intensidad han sido elevados a la quinta potencia

**CAPÍTULO 3. METODOLOGÍA**  
**3.2. PROCESAMIENTO DE IMÁGENES**

---

análisis [10], en el que se realizó el respectivo cálculo de errores de coincidencia de las coordenadas del bin de mayor intensidad y las coordenadas presentadas por el diagnóstico como centro de la anomalía, dicho error fue menor al 2%. Las tablas 3.4, 3.5, 3.6 presentan el valor de intensidad máxima para cada mamografía; es importante mencionar que este se encuentra normalizado.

Muestra de tejido graso (F)			
Imagen	Número de identificación	Valor de intensidad máximo	Presenta anomalía
1	mdb 005	0.761719	SI
2	mdb 009	0.90625	NO
3	mdb 010	0.8125	SI
4	mdb 011	0.894531	NO
5	mdb 012	0.777344	SI
6	mdb 025	0.804688	SI
7	mdb 026	0.707031	NO
8	mdb 027	0.722656	NO
9	mdb 028	0.851563	SI
10	mdb 059	0.671875	SI
11	mdb 060	0.671875	NO
12	mdb 069	0.878906	SI
13	mdb 070	0.824219	NO
14	mdb 075	0.941406	SI
15	mdb 076	0.808594	NO
16	mdb 077	0.738281	NO
17	mdb 078	0.792969	NO
18	mdb 079	0.851563	NO
19	mdb 080	0.796875	SI
20	mdb 087	0.777344	NO
21	mdb 092	0.746094	SI
22	mdb 095	0.859375	SI
23	mdb 134	0.820313	SI
24	mdb 141	0.796875	SI
25	mdb 231	0.683594	SI
26	mdb 238	0.792969	SI
27	mdb 248	0.882813	SI
28	mdb 252	0.828125	SI
29	mdb 256	0.757813	SI
30	mdb 267	0.828125	SI

Tabla 3.4: Valor del bin de intensidad máxima para cada una de las mamografías utilizadas en el estudio que presentan tejido graso

**CAPÍTULO 3. METODOLOGÍA**  
**3.2. PROCESAMIENTO DE IMÁGENES**

Muestra de tejido glandular (G)			
Imagen	Número de identificación	Valor de intensidad máximo	Presenta anomalía
1	mdb 007	0.855469	NO
2	mdb 008	0.855469	NO
3	mdb 013	0.878906	SI
4	mdb 014	0.878906	NO
5	mdb 015	0.851563	SI
6	mdb 016	0.839844	NO
7	mdb 017	0.929688	SI
8	mdb 018	0.847656	NO
9	mdb 019	0.828125	SI
10	mdb 020	0.878906	NO
11	mdb 021	0.90625	SI
12	mdb 022	0.863281	NO
13	mdb 023	0.875	SI
14	mdb 024	0.832031	NO
15	mdb 029	0.925781	NO
16	mdb 030	0.898438	SI
17	mdb 031	0.871094	NO
18	mdb 032	0.875	SI
19	mdb 041	0.871094	NO
20	mdb 042	0.863281	NO
21	mdb 072	0.898438	SI
22	mdb 090	0.890625	SI
23	mdb 209	0.890625	SI
24	mdb 211	0.886719	SI
25	mdb 213	0.78125	SI
26	mdb 218	0.894531	SI
27	mdb 219	0.867188	SI
28	mdb 264	0.78125	SI
29	mdb 265	0.886719	SI
30	mdb 270	0.941406	SI

Tabla 3.5: Valor del bin de intensidad máxima para cada una de las mamografías utilizadas en el estudio que presentan tejido glandular

**CAPÍTULO 3. METODOLOGÍA**  
**3.2. PROCESAMIENTO DE IMÁGENES**

Muestra de tejido denso (D)			
Imagen	Número de identificación	Valor de intensidad máximo	Presenta anomalía
1	mdb 003	0.917969	NO
2	mdb 004	0.90625	NO
3	mdb 033	0.832031	NO
4	mdb 034	0.855469	NO
5	mdb 035	0.820313	NO
6	mdb 036	0.816406	NO
7	mdb 037	0.855469	NO
8	mdb 038	0.78125	NO
9	mdb 039	0.878906	NO
10	mdb 040	0.882813	NO
11	mdb 053	0.894531	NO
12	mdb 054	0.882813	NO
13	mdb 057	0.917969	NO
14	mdb 058	0.886719	SI
15	mdb 061	0.839844	NO
16	mdb 062	0.851563	NO
17	mdb 063	0.886719	SI
18	mdb 064	0.945313	NO
19	mdb 065	0.882813	NO
20	mdb 066	0.769531	NO
21	mdb 102	0.875	SI
22	mdb 105	0.945313	SI
23	mdb 110	0.878906	SI
24	mdb 111	0.886719	SI
25	mdb 170	0.921875	SI
26	mdb 236	0.88283	SI
27	mdb 239	0.917969	SI
28	mdb 241	0.863281	SI
29	mdb 249	0.851563	SI
30	mdb 253	0.882813	SI

Tabla 3.6: Valor del bin de intensidad máxima para cada una de las mamografías utilizadas en el estudio que presentan tejido denso

Asimismo, con el objetivo de poder llevar a cabo la manipulación de los datos de cada RMI se creó, para cada elemento de muestra, un archivo de texto que ordena en tres columnas toda la información necesaria: Coordenada  $x$ , Coordenada  $y$  y valor de intensidad, pues dada la ubicación e intensidad de cada bin se puede llevar a cabo la reconstrucción total de la imagen. Esto con el fin de convertir dicha información en histogramas de frecuencias utilizando nuevamente el sistema ROOT V. 6.23/01, para el posterior análisis de la distribución de valores de intensidades.

### 3.2.2. Conversión de imágenes a Histogramas 2D

Una característica importante de las imágenes mamográficas (y en general, de cualquier imagen radiológica) es que los contrastes que se presentan en los valores de intensidad se deben a que la permeabilidad de los rayos X es diferente en cada tipo de estructura, para el caso del seno se tiene que el tejido del mismo, su respectivo fondo y otras formas que se encuentran presentes como el músculo pectoral, tumores o anomalías y la glándula mamaria presentan, aunque mínimas, diferencias en densidad y es este hecho el que hace posible la obtención de la imagen radiológica

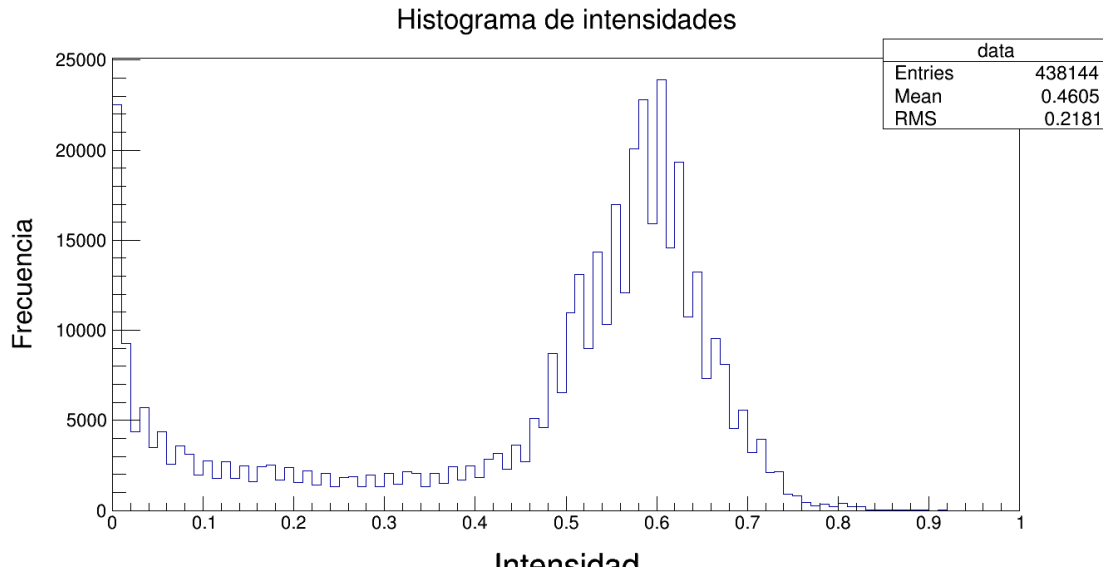


Figura 3.6: Distribución de valores de intensidad de la Figura 3.4

de interés, se trata del principio de la radiografía. Dicho lo anterior, se tiene que la información de la intensidad de una imagen mamográfica, contiene también la información de densidad de tejido de la mama estudiada. Dado este hecho y sabiendo que el valor de densidad del área b es más bajo que el correspondiente a las zonas con anomalías (ROI), el histograma de frecuencias de intensidad permite arribar conclusiones importantes.

Utilizando el mismo software [21], se llevó a cabo la creación de un programa (Véase Anexo B.2.) capaz de realizar el histograma de frecuencias de intensidades con los datos importados del archivo de texto que contiene toda la información de los RMI.

El histograma representa gráficamente el número de píxeles asociado a cada tono de gris de una MI en particular. En esta nueva representación se pierde por completo la información respecto a la posición de los píxeles en la matriz virtual o el sistema de coordenadas definido. Únicamente presenta la distribución de grises que se encuentra correlacionada con las diferentes densidades de los objetos presentes en la imagen que es lo que nos interesa para el estudio. Se tiene que: la presencia de tejidos densos se visualiza por valores significativos (en cantidad de píxeles) a la derecha de la escala horizontal de intensidad. Tonos claros de gris corresponden a tejido de alta densidad que absorben una mayor cantidad de radiación y por ende están representados por valores altos en la escala, mientras que, tonos oscuros pertenecen a tejidos de baja densidad cuyos valores de intensidad son bajos. Al respecto conviene recordar que la escala de intensidad se normalizó para este análisis: negro es 0, blanco es 1. De lo anterior, el análisis de esta nueva representación de la información puede proporcionar los parámetros cualitativos para caracterizar el tejido mamario.

Las figura 3.6 muestra la distribución de intensidades obtenida para la imagen mdb 057, cuya MRI se muestra en la figura 3.4. Para esta mamografía, que presenta un tejido denso uniforme y carece de anomalías se observa que los píxeles de la imagen se aglomeran en dos grandes grupos según su nivel de gris, representados como dos altas crestas (mayor cantidad de datos) separadas que corresponden (la primera) a los píxeles del fondo, esto es, zonas donde no hay tejido y por ende los rayos x no son absorbidos así que se visualizan como zonas oscuras y (la segunda, con apariencia de tipo gaussina) a los píxeles que conforman la mama y que presentan diferentes niveles de brillo. Una comparación entre ambas agrupaciones hace evidente que el número de píxeles pertenecientes al tejido es mucho mayor que el del fondo para este caso, lo que nos habla del tamaño de la mama estudiada.

Para ejemplificar el caso en que se tiene una anomalía, la figura 3.8 presenta la distribución

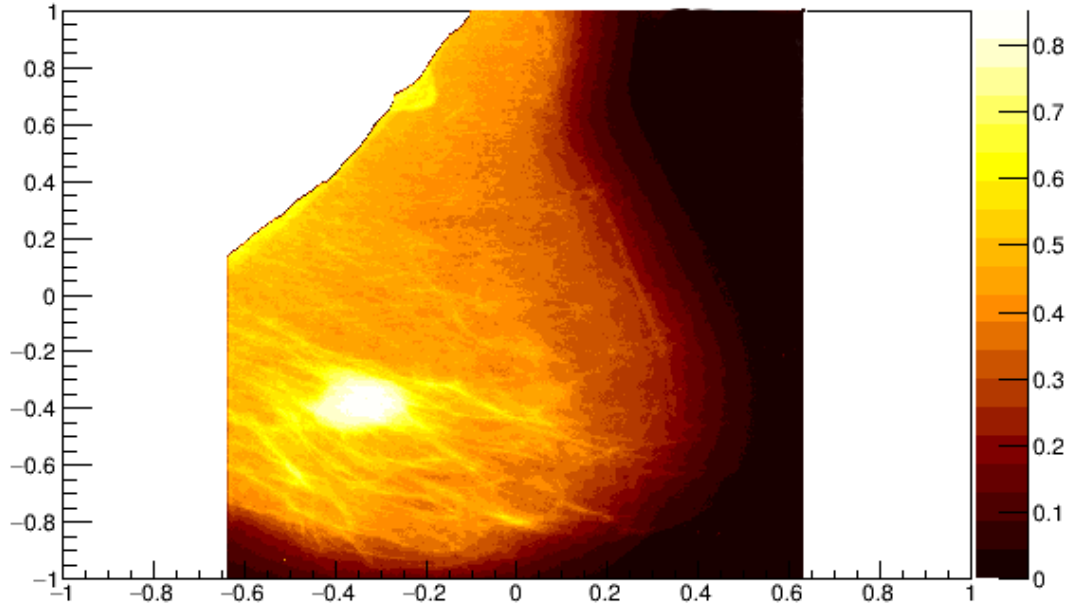


Figura 3.7: MRI de la mamografía mdb 028, clasificada como tejido F con una masa bien definida

de intensidad de la figura 3.7. En esta distribución se aprecian las dos grandes agrupaciones de píxeles de menor y mayor intensidad que pertenecen al fondo y a la zona de tejido (que una vez más presenta apariencia gaussiana), respectivamente, pero a diferencia de la distribución obtenida para la mamografía de tejido denso sano en la figura 3.6, en esta segunda distribución se puede observar tercera agrupación con mucho menor cantidad de píxeles que las otras dos pero de valores de intensidad máximos, cuyos datos corresponden a la ROI que presenta una alta densidad a comparación del resto del tejido. Puesto que las crestas que representan al tejido b y a las ROI (zonas que presentan una alta densidad) se encuentran notoriamente separadas, se propone que es posible definir un valor de intensidad umbral que marque la separación entre el área sana del seno y las posibles ROI.

Teniendo en cuenta que se tienen tres clasificaciones de tejido sano según sus diferentes densidades y que el valor de intensidad umbral a definir corresponde también a un valor de densidad umbral, se obtuvo un valor umbral para cada tipo. Adviértase que, a mayor densidad del tejido, la separación entre la cresta de los píxeles de tejido sano y la de las posibles ROI será cada vez menor, es decir, se tendrá un histograma muy parejo. En consecuencia, la elección correcta de un umbral se dificultará, siendo altamente probable que existan valores de intensidad pertenecientes a tejido sano que rebasen al valor definido, entorpeciendo la correcta delimitación de las ROI, no obstante, se tendrá una buena aproximación de regiones candidatas a presentar anomalías.

El proceso de determinación del valor de intensidad umbral ( $\rho$ ) se llevó a cabo mediante la caracterización de zonas sanas representativas de los tres tipos de tejido de fondo haciendo uso de las herramientas de la estadística, pues una inspección visual de los histogramas obtenidos, basada en las características mencionadas al inicio de esta sección no basta para proporcionar resultados aceptables. Se realizaron pruebas con distintos valores de umbral y se arribó a las conclusiones presentadas en capítulos posteriores.

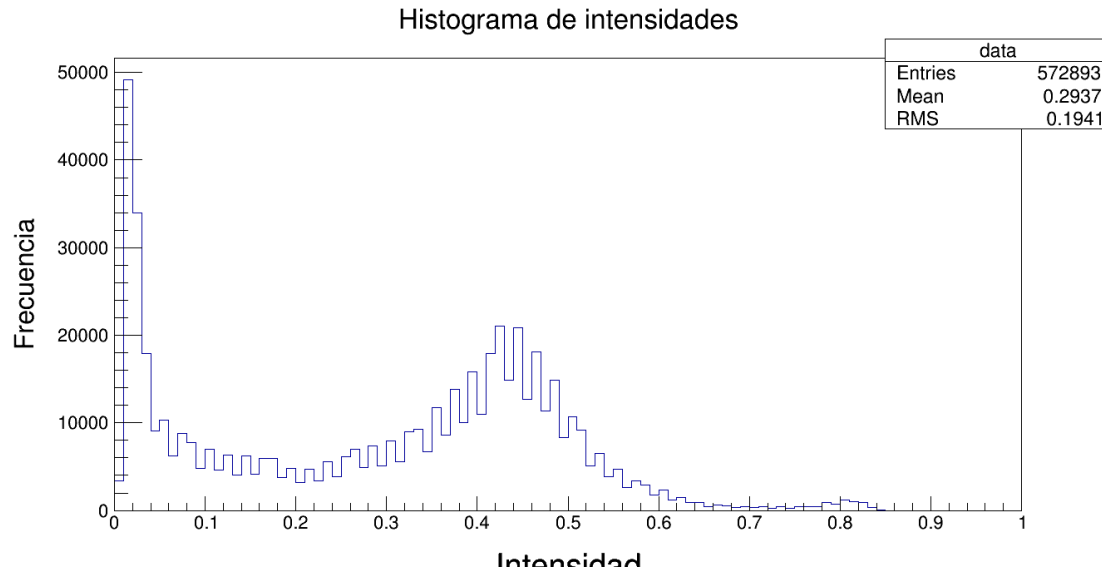


Figura 3.8: Distribución de valores de intensidad de la Figura 3.7

### 3.3. Caracterización de tejido

#### 3.3.1. Delimitación de tejido y creación de histogramas

Al programa para la creación de histogramas se le implementaron las líneas de código necesarias para delimitar los datos que se presentan en el mismo (Véase Anexo B.2.). Se busca aislar las zonas de mayor intensidad que pueden ser relevantes para el diagnóstico mediante la definición de un valor de intensidad umbral para cada uno de los tres tipos de tejido según los cuales se clasifican las mamografías en la base de datos MINI-mias [20]. Para ello, se llevará a cabo la caracterización de los mismos mediante el análisis estadístico de las distribuciones de intensidad obtenidas para áreas representativas de tejido sano seleccionadas de los 90 elementos para estudio.

Para cada una de las 90 mamografías se seleccionó un área rectangular representativa de tejido sano, es decir, dichas áreas se delimitaron de tal manera que, para el caso de mamografías con lesiones, estas no se incluyeran en la región seleccionada. Se garantizó que el área elegida no contemplara zonas afectadas basándose en la posición y el tamaño de las ROI indicadas por los médicos en el diagnóstico clínico, esto fue hecho manualmente por un operador (no se realizó ninguna segmentación automática). Se definió el tejido de fondo (b) como las regiones de la mama en las que no se tienen lesiones localizados (área b).

La delimitación del área se llevó a cabo estableciendo un valor de  $x$  y  $y$  máximo y mínimo en el sistema de coordenadas implementado en los MRI (Figura 3.9). Estas zonas corresponden a las muestras de tejido sano a analizar y posteriormente se obtuvieron sus respectivos histogramas de frecuencias de intensidades (Figura 3.10). Con respecto a la selección de estas áreas es importante decir que, de acuerdo con el estudio previo que es antecedente principal de este trabajo [10], se tiene que para los fines de este análisis independientemente de la selección de estudio, los resultados son consistentes. Esto se explicará con mayor detalle más adelante en este trabajo.

En las figuras 3.11, 3.12, 3.13, 3.14, 3.15 y 3.16 se muestran los resultados de la transformación de la información de secciones representativas de tejido sano de los MRI en histogramas de frecuencia de intensidades para diferentes mamografías de acuerdo a la siguiente información:

- **TEJIDO GRASO**

**Figura 3.11.** Ejemplo de la transformación de la información de un área de tejido sano perteneciente a una mamografía normal en histograma de frecuencia de intensidades.

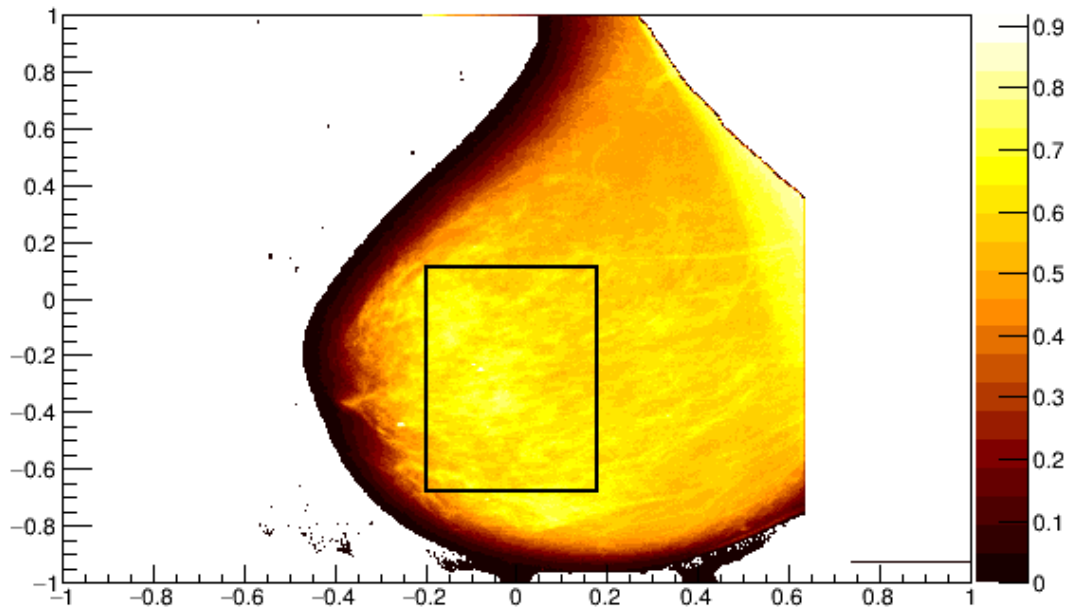


Figura 3.9: Muestra de tejido sano seleccionada para una mamografía

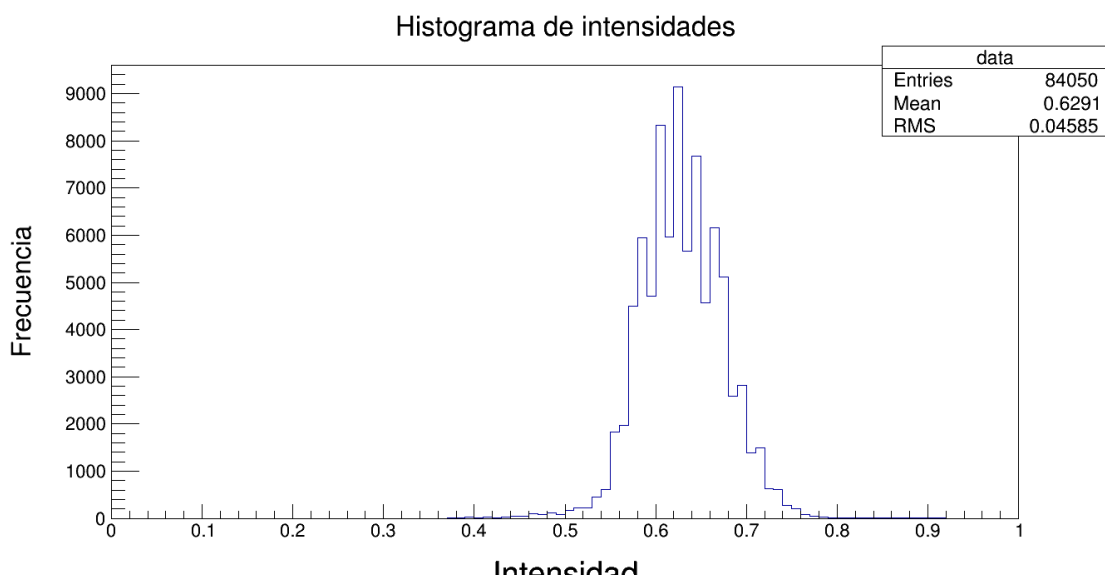
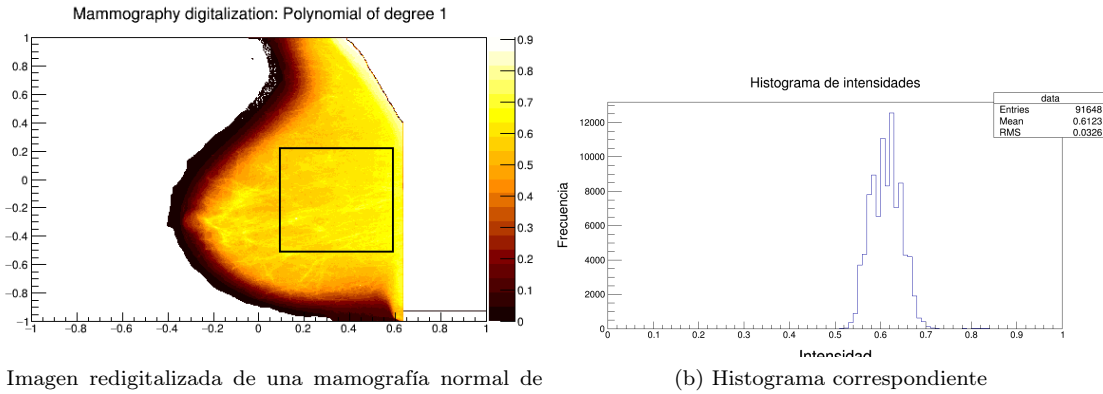


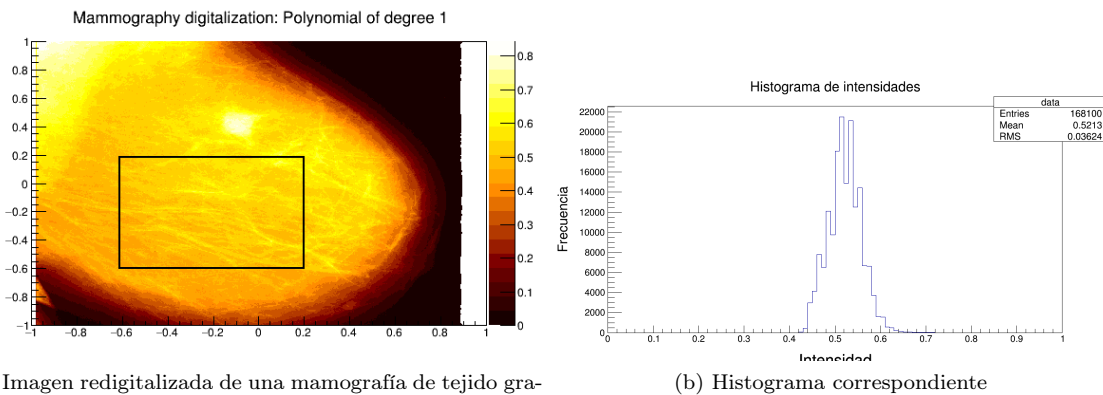
Figura 3.10: Muestra de tejido sano seleccionada para una mamografía



(a) Imagen redigitalizada de una mamografía normal de tejido graso

(b) Histograma correspondiente

Figura 3.11: Transformación a histograma de la información correspondiente a un área de tejido sano en una mamografía de tejido graso normal (mdb 009)



(a) Imagen redigitalizada de una mamografía de tejido graso que presenta anomalía

(b) Histograma correspondiente

Figura 3.12: Transformación a histograma de la información correspondiente a un área de tejido sano en una mamografía de tejido graso que presenta una masa mal definida maligna (mdb 134)

**Figura 3.12.** Ejemplo de la transformación de la información de un área de tejido sano perteneciente a una mamografía que presenta una masa mal definida maligna en histograma de frecuencia de intensidades.

■ **TEJIDO GLANDULAR**

**Figura 3.13.** Ejemplo de la transformación de la información de un área de tejido sano perteneciente a una mamografía normal en histograma de frecuencia de intensidades.

**Figura 3.14.** Ejemplo de la transformación de la información de un área de tejido sano perteneciente a una mamografía que presenta una masa circunscrita benigna en histograma de frecuencia de intensidades.

**TEJIDO DENSO**

**Figura 3.15.** Ejemplo de la transformación de la información de un área de tejido sano perteneciente a una mamografía normal en histograma de frecuencia de intensidades.

**Figura 3.16.** Ejemplo de la transformación de la información de un área de tejido sano perteneciente a una mamografía que presenta una masa mal definida maligna en histograma de frecuencia de intensidades.

**CAPÍTULO 3. METODOLOGÍA**  
**3.3. CARACTERIZACIÓN DE TEJIDO**

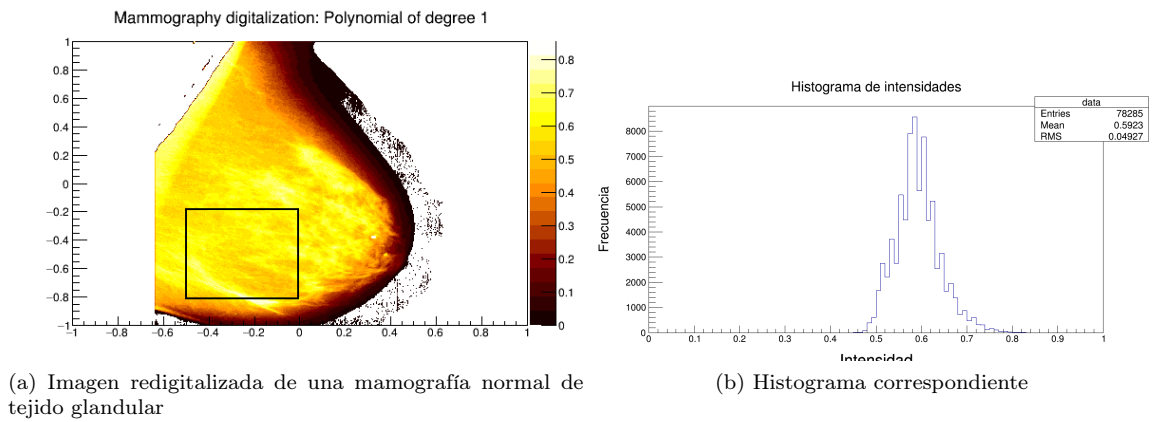


Figura 3.13: Transformación a histograma de la información correspondiente a un área de tejido sano en una mamografía de tejido glandular normal (mdb 008)

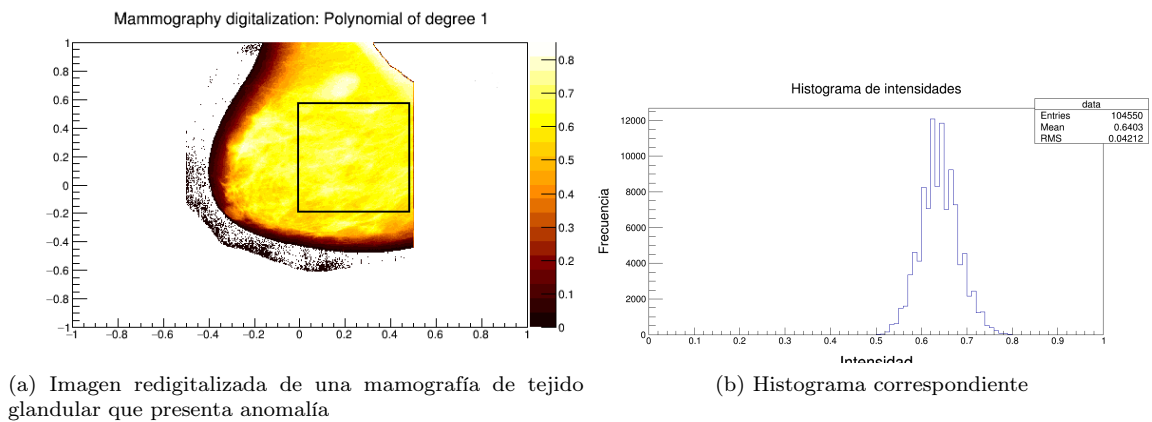


Figura 3.14: Transformación a histograma de la información correspondiente a un área de tejido sano en una mamografía de tejido glandular que presenta una masa circunscrita benigna (mdb 015)

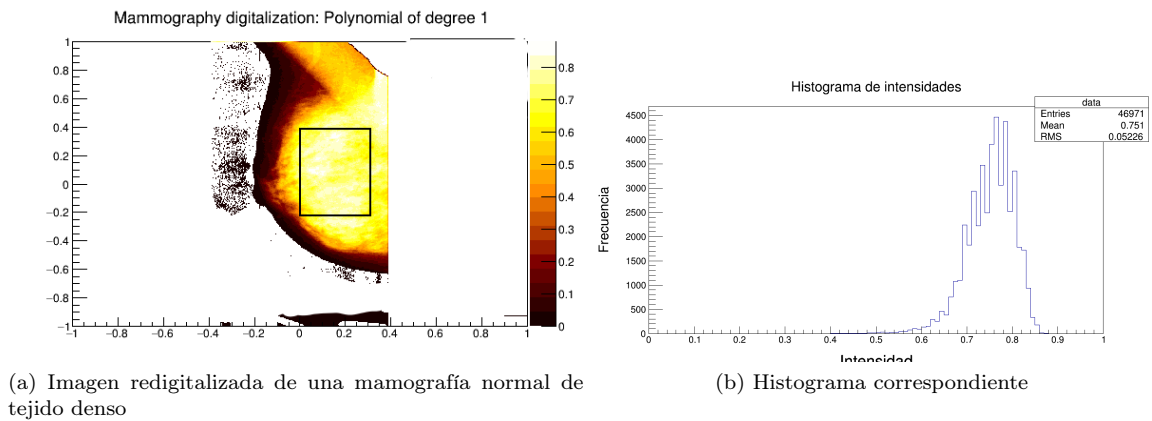
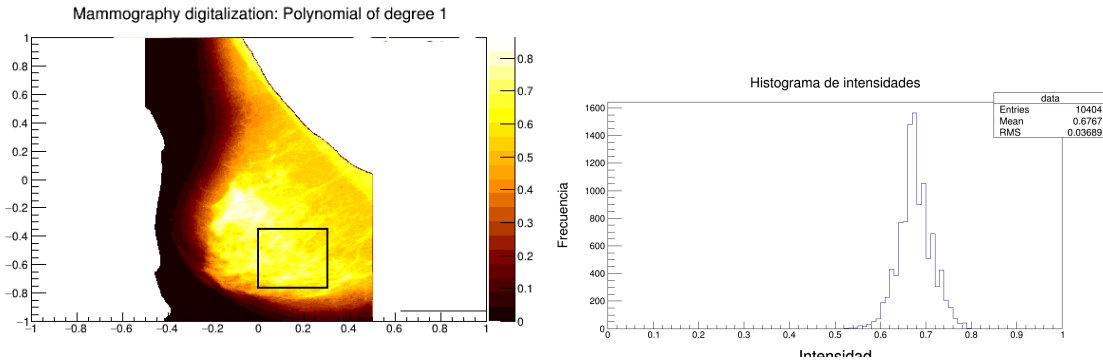


Figura 3.15: Transformación a histograma de la información correspondiente a un área de tejido sano en una mamografía normal de tejido denso (mdb 039)



(a) Imagen redigitalizada de una mamografía de tejido denso que presenta anomalía

(b) Histograma correspondiente

Figura 3.16: Transformación a histograma de la información correspondiente a un área de tejido sano en una mamografía de tejido denso que presenta calcificaciones malignas (mdb 249)

ROOT genera automáticamente la leyenda correspondiente para cada histograma creado, en esta presenta tres valores del mismo: Entries, Mean y RMS; el primero se refiere al total de las entradas del histograma, el segundo al promedio de estas y el tercero corresponde al término "Root-mean-square" que, en general, los físicos usan como sinónimo de desviación estándar para un conjunto finito de datos.

Se encontraron distribuciones similares para cada uno de los tres tipos de tejido sano, la mayoría presenta una distribución que parece ser simétrica o casi simétrica respecto a su valor medio, en apariencia, todas las distribuciones obtenidas se aproximan a una de tipo gaussiano. Además, se observó que, dependiendo del tipo de tejido dentro del cuál se encuentra clasificada la mamografía, el valor entorno al cual se distribuyen los valores varía, a mayor densidad del tejido, la distribución se recorre a la derecha, esto era de esperarse dada la relación intensidad de pixel- densidad de tejido.

### 3.3.2. Ajuste

Se realizó el ajuste de los datos presentados en los histogramas agregando las líneas de código necesarias al programa utilizado para la creación de los mismos, dada la apariencia de las distribuciones obtenidas, se optó por un ajuste de tipo gaussiano.

Este tipo de curva queda completamente definida por dos parámetros: la media poblacional  $\mu$  y la desviación estándar de la población  $\sigma$ . Una vez más, root genera automáticamente la leyenda, que esta vez muestra los valores del ajuste. Se observa que los valores del ajuste, Mean ( $\mu$ ) y Sigma ( $\sigma$ ) son muy cercanos a los valores Mean y RMS del histograma, además se presenta el valor de la constante, que se refiere a la amplitud de la función gaussiana, así como los datos obtenidos del ajuste para  $\chi^2$ , que se refiere a la suma de los cuadrados de los residuos normalizados y el valor *ndf* que corresponde a las siglas de *number of degrees of freedom*, es decir, el número de grados de libertad en el ajuste, que para este caso están dados por el número de bins usados en el cálculo de  $\chi^2$ — el número de parámetros del ajuste, el cociente de estas dos cantidades nos da una idea de la bondad del ajuste utilizado, aquí se obtuvieron valores grandes para este resultado, dado que  $\chi^2$  se calcula de la siguiente manera:

$$\sum \left[ \frac{(\text{Contenido del bin} - \text{valor de la función})}{\text{Error del bin}} \right]^2 \quad (3.1)$$

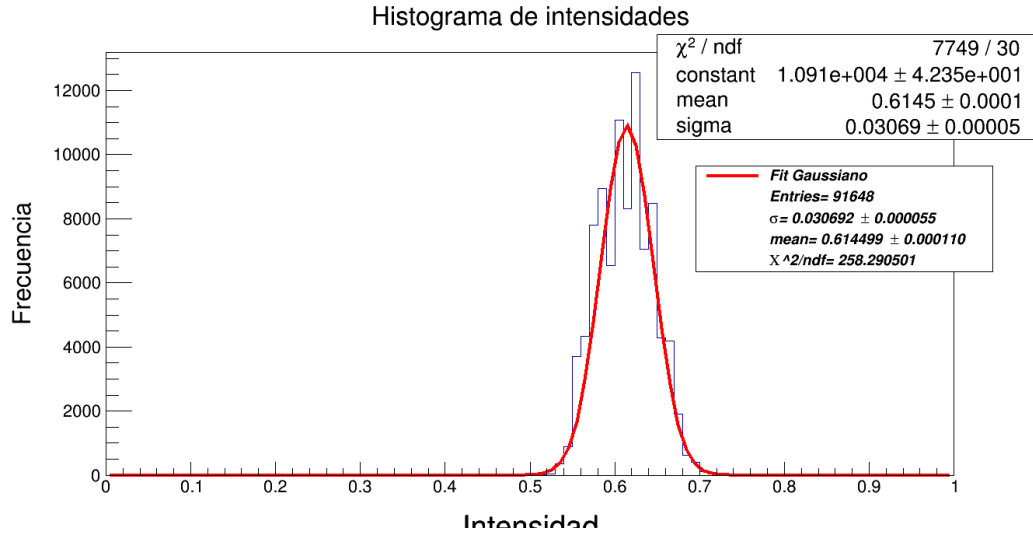


Figura 3.17: Ajuste gaussiano de la distribución de intensidades obtenida para una región de tejido sano de la mamografía mdb 009

En analogía con la definición del estadístico  $\chi^2$ :

$$\chi^2 = \sum_{i=1}^n \frac{(X_i - \mu_i)^2}{\sigma_i^2} \quad (3.2)$$

Si el ajuste es bueno, se espera que cada uno de estos cocientes se acerque a 1, por tanto la suma sería cercana a  $n$  y  $\frac{\chi^2}{ndf}$  da un número cercano a 1. No obstante, el programa trata de ajustar esta información sobre toda la ventana, incluso donde no hay datos, de ahí los resultados que se tienen. Pero dado que el error asociado a los parámetros de interés es tres órdenes de magnitud menor que estos, se dice que el ajuste es bueno. Las figuras 3.17, 3.18, 3.19, 3.20, 3.21 y 3.22 presentan los ajustes obtenidos para las distribuciones de las mamografías presentadas en el apartado anterior.

Este procedimiento se realizó para las 90 distribuciones de las mamografías estudiadas, las tablas 3.7, 3.8 y 3.9 muestran los datos obtenidos para ambos parámetros de cada una de las muestras de tejido.

Como se observa en los datos presentados, las distribuciones se recorren a la derecha conforme aumenta la densidad del tejido, es decir, los datos correspondientes a tejido graso (F) presentan menor  $\mu$  que los obtenidos para tejido glandular (G) y estos a su vez tienen menor  $\mu$  que los de tejido denso (D). Para promedio de las medias poblaciones de los datos previos, se espera que:

$$\mu_F < \mu_G < \mu_D \quad (3.3)$$

Además, respecto a las desviaciones estándar, se puede decir que un conjunto de píxeles que tiene una desviación estándar pequeña en cuanto a intensidad implica que el tejido de la mama a la que pertenece la mamografía en cuestión, es más uniforme que el tejido de una cuya desviación estándar de la intensidad de los píxeles sea grande.

En la primera sección del capítulo 4 se presenta la estadística descriptiva de los resultados anteriores.

### 3.3.3. Análisis de datos

Comparando los datos correspondiente al bin de intensidad máxima para mamografías que presentan anomalías de las tablas 3.4, 3.5 y 3.6 con los obtenidos para los valores de  $\sigma$  y  $\mu$  en los ajustes de las distribuciones del área b de las tablas 3.7, 3.8 y 3.9, se encuentra lo siguiente:

CAPÍTULO 3. METODOLOGÍA  
3.3. CARACTERIZACIÓN DE TEJIDO

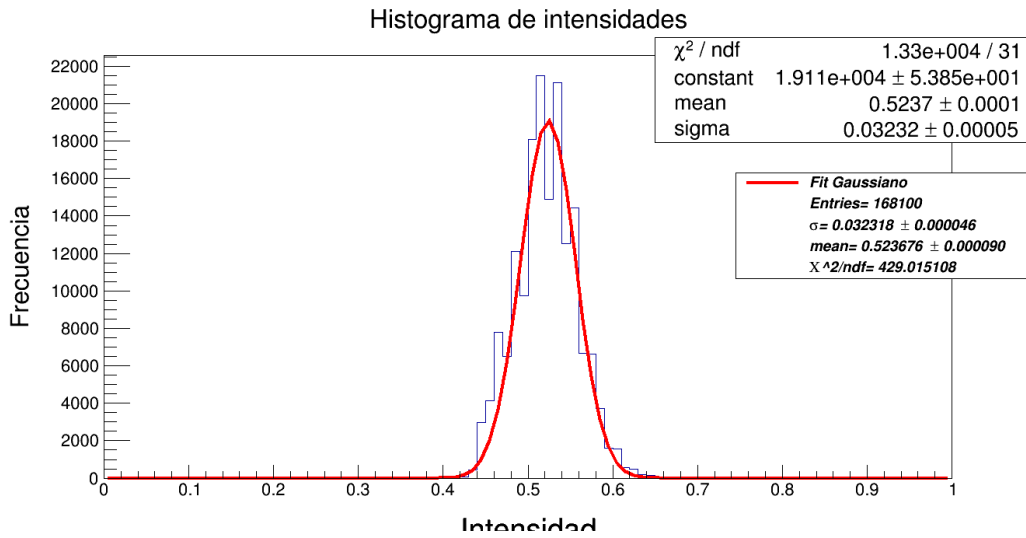


Figura 3.18: Ajuste gaussiano de la distribución de intensidades obtenida para una región de tejido sano de la mamografía mdb 134

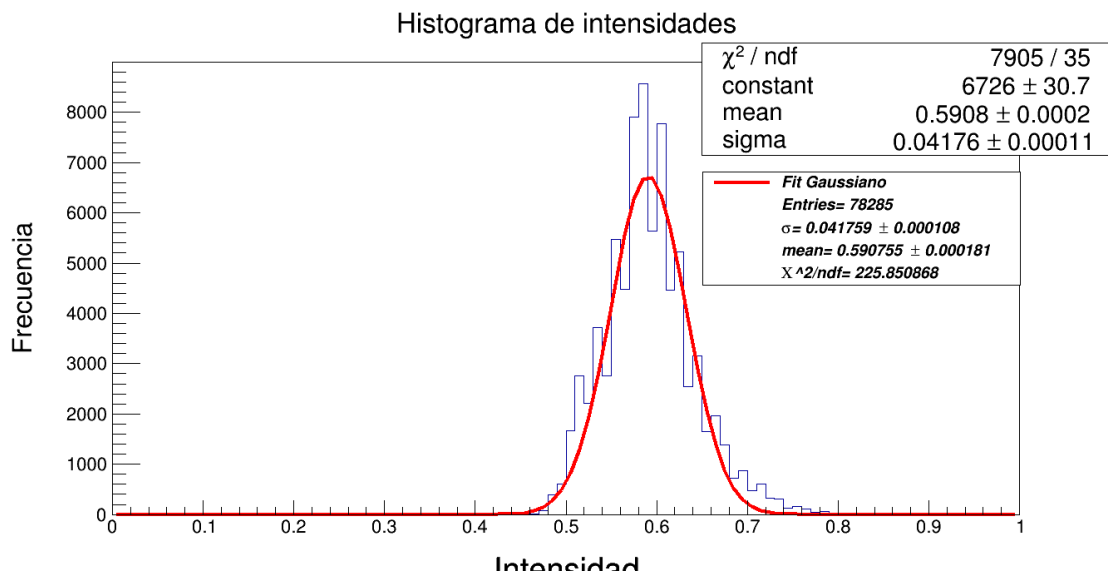


Figura 3.19: Ajuste gaussiano de la distribución de intensidades obtenida para una región de tejido sano de la mamografía mdb 008

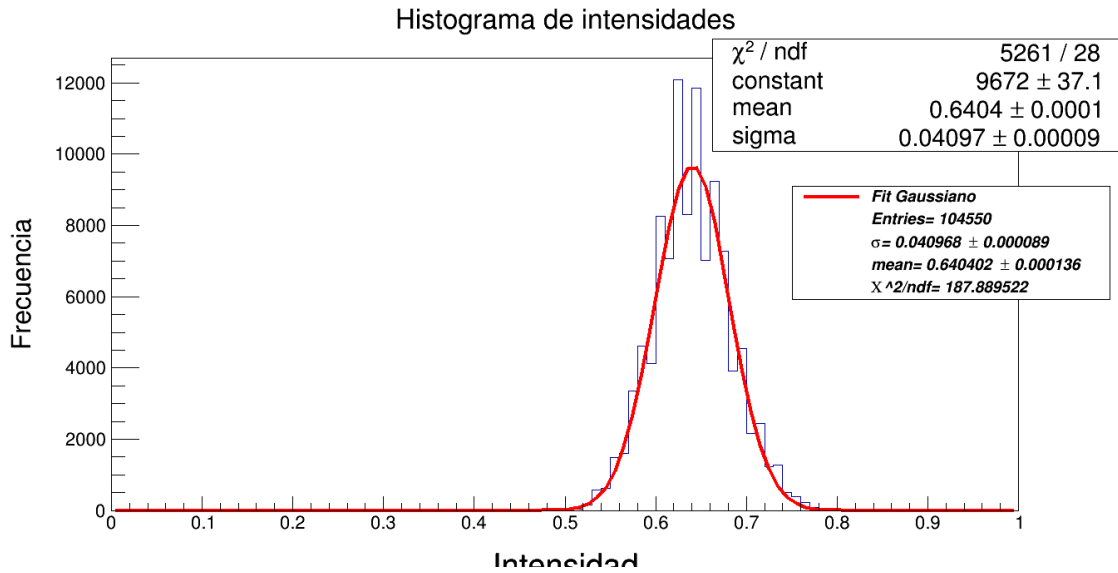


Figura 3.20: Ajuste gaussiano de la distribución de intensidades obtenida para una región de tejido sano de la mamografía mdb 015

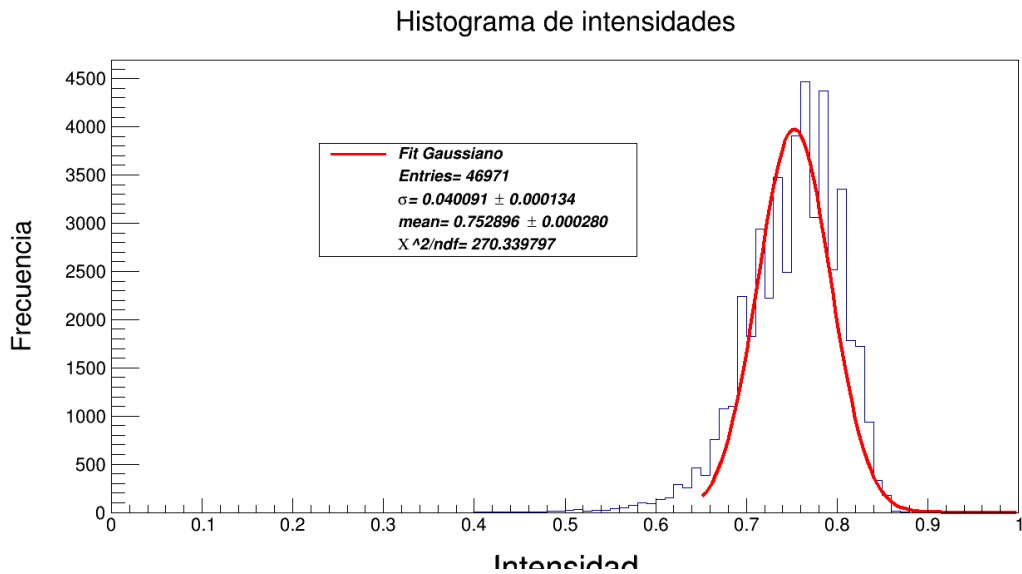


Figura 3.21: Ajuste gaussiano de la distribución de intensidades obtenida para una región de tejido sano de la mamografía mdb 039

**CAPÍTULO 3. METODOLOGÍA**  
**3.3. CARACTERIZACIÓN DE TEJIDO**

---

Tejido Graso			
Mamografía	ID	Valores de ajuste obtenidos	
		$\mu$	$\sigma$
1	mdb 005	0.552111 ± 0.0001	0.027591 ± 0.000066
2	mdb 009	0.614499 ± 0.00110	0.030692 ± 0.000055
3	mdb 010	0.661207 ± 0.000117	0.021141 ± 0.000066
4	mdb 011	0.648367 ± 0.000084	0.026663 ± 0.000066
5	mdb 012	0.618670 ± 0.000116	0.027570 ± 0.000061
6	mdb 025	0.489498 ± 0.000380	0.036539 ± 0.000200
7	mdb 026	0.531307 ± 0.000183	0.044620 ± 0.000161
8	mdb 027	0.417201 ± 0.000095	0.018682 ± 0.000079
9	mdb 028	0.449732 ± 0.000170	0.029107 ± 0.000111
10	mdb 059	0.393786 ± 0.000137	0.023533 ± 0.000091
11	mdb 060	0.398775 ± 0.000211	0.020152 ± 0.000096
12	mdb 069	0.551476 ± 0.000133	0.037497 ± 0.000113
13	mdb 070	0.646524 ± 0.000107	0.033139 ± 0.000067
14	mdb 075	0.543285 ± 0.000111	0.016687 ± 0.000066
15	mdb 076	0.562066 ± 0.000083	0.017295 ± 0.000058
16	mdb 077	0.548384 ± 0.000080	0.026709 ± 0.000068
17	mdb 078	0.579101 ± 0.000087	0.031282 ± 0.000103
18	mdb 079	0.607011 ± 0.000326	0.074230 ± 0.000211
19	mdb 080	0.566214 ± 0.000297	0.026211 ± 0.000263
20	mdb 087	0.521488 ± 0.000199	0.035768 ± 0.000124
21	mdb 092	0.545910 ± 0.000340	0.038899 ± 0.000231
22	mdb 095	0.558176 ± 0.000239	0.040327 ± 0.000123
23	mdb 134	0.523676 ± 0.000090	0.032318 ± 0.000046
24	mdb 141	0.598441 ± 0.000096	0.024547 ± 0.000074
25	mdb 231	0.446882 ± 0.000123	0.023949 ± 0.000084
26	mdb 238	0.635932 ± 0.000198	0.034507 ± 0.000124
27	mdb 248	0.624266 ± 0.000119	0.027164 ± 0.000063
28	mdb 252	0.534901 ± 0.000149	0.026394 ± 0.000149
29	mdb 256	0.632852 ± 0.000099	0.018774 ± 0.000064
30	mdb 267	0.659072 ± 0.000079	0.018039 ± 0.000043

Tabla 3.7: Datos obtenidos para la muestra de tejido graso (F). Primera columna: mamografía. Siguiendo columna: Número de identificación en la base de datos. Dos siguientes columnas: media y desviación estándar del ajuste de la distribución de intensidad en el área de tejido sano seleccionada.

**CAPÍTULO 3. METODOLOGÍA**  
**3.3. CARACTERIZACIÓN DE TEJIDO**

---

Tejido Glandular			
Muestra	ID	Valores de ajuste obtenidos	
		$\mu$	$\sigma$
1	mdb 007	0.611471 $\pm$ 0.000241	0.043508 $\pm$ 0.000115
2	mdb 008	0.590755 $\pm$ 0.000181	0.041759 $\pm$ 0.000108
3	mdb 013	0.635956 $\pm$ 0.000168	0.027847 $\pm$ 0.000110
4	mdb 014	0.692132 $\pm$ 0.000193	0.024614 $\pm$ 0.000096
5	mdb 015	0.640402 $\pm$ 0.000136	0.040968 $\pm$ 0.000089
6	mdb 016	0.621089 $\pm$ 0.000292	0.050615 $\pm$ 0.000162
7	mdb 017	0.630461 $\pm$ 0.000282	0.032500 $\pm$ 0.000161
8	mdb 018	0.724731 $\pm$ 0.000240	0.024931 $\pm$ 0.000118
9	mdb 019	0.630389 $\pm$ 0.000143	0.036151 $\pm$ 0.000105
10	mdb 020	0.595217 $\pm$ 0.000177	0.022706 $\pm$ 0.000090
11	mdb 021	0.622349 $\pm$ 0.000144	0.024836 $\pm$ 0.000095
12	mdb 022	0.616400 $\pm$ 0.000226	0.030220 $\pm$ 0.000138
13	mdb 023	0.683781 $\pm$ 0.000090	0.028891 $\pm$ 0.000058
14	mdb 024	0.668996 $\pm$ 0.000103	0.022095 $\pm$ 0.000067
15	mdb 029	0.757422 $\pm$ 0.000222	0.036485 $\pm$ 0.000142
16	mdb 030	0.808219 $\pm$ 0.000104	0.021675 $\pm$ 0.000059
17	mdb 031	0.704396 $\pm$ 0.000214	0.046351 $\pm$ 0.000098
18	mdb 032	0.740251 $\pm$ 0.000355	0.046055 $\pm$ 0.000206
19	mdb 041	0.501223 $\pm$ 0.000113	0.026859 $\pm$ 0.000082
20	mdb 042	0.456362 $\pm$ 0.000167	0.027729 $\pm$ 0.000165
21	mdb 072	0.574021 $\pm$ 0.000310	0.020997 $\pm$ 0.000187
22	mdb 090	0.665206 $\pm$ 0.000519	0.031096 $\pm$ 0.000325
23	mdb 209	0.505826 $\pm$ 0.000325	0.031221 $\pm$ 0.000182
24	mdb 211	0.818661 $\pm$ 0.000435	0.026258 $\pm$ 0.000339
25	mdb 213	0.632037 $\pm$ 0.000580	0.037303 $\pm$ 0.000403
26	mdb 218	0.628012 $\pm$ 0.000203	0.025472 $\pm$ 0.000099
27	mdb 219	0.627752 $\pm$ 0.000472	0.042733 $\pm$ 0.000269
28	mdb 264	0.462124 $\pm$ 0.000380	0.053902 $\pm$ 0.000213
29	mdb 265	0.639033 $\pm$ 0.000431	0.024974 $\pm$ 0.000208
30	mdb 270	0.667353 $\pm$ 0.000156	0.044795 $\pm$ 0.000106

Tabla 3.8: Datos obtenidos para la muestra de tejido glandular (G). Primera columna: mamografía. Siguiendo columna: Número de identificación en la base de datos. Dos siguientes columnas: media y desviación estándar del ajuste de la distribución de intensidad en el área de tejido sano seleccionada.

**CAPÍTULO 3. METODOLOGÍA**  
**3.3. CARACTERIZACIÓN DE TEJIDO**

---

Tejido Denso			
Muestra	ID	Valores de ajuste obtenidos	
		$\mu$	$\sigma$
1	mdb 003	0.819689 ± 0.000276	0.054269 ± 0.000280
2	mdb 004	0.822126 ± 0.000148	0.024983 ± 0.000091
3	mdb 033	0.654287 ± 0.000314	0.05203 ± 0.000184
4	mdb 034	0.661765 ± 0.000420	0.036345 ± 0.000350
5	mdb 035	0.722555 ± 0.000378	0.031832 ± 0.000320
6	mdb 036	0.706817 ± 0.000418	0.041564 ± 0.000222
7	mdb 037	0.758089 ± 0.000326	0.028369 ± 0.000140
8	mdb 038	0.704709 ± 0.000226	0.024236 ± 0.000099
9	mdb 039	0.752896 ± 0.000280	0.040091 ± 0.000134
10	mdb 040	0.775538 ± 0.000238	0.031213 ± 0.000138
11	mdb 053	0.786402 ± 0.000218	0.027541 ± 0.000100
12	mdb 054	0.752060 ± 0.000224	0.018199 ± 0.000115
13	mdb 057	0.630347 ± 0.000158	0.043256 ± 0.000115
14	mdb 058	0.580435 ± 0.000218	0.021914 ± 0.000152
15	mdb 061	0.673925 ± 0.000388	0.023089 ± 0.000183
16	mdb 062	0.570651 ± 0.000285	0.031089 ± 0.000165
17	mdb 063	0.587727 ± 0.0003	0.024796 ± 0.000229
18	mdb 064	0.749332 ± 0.000313	0.035494 ± 0.000171
19	mdb 065	0.813314 ± 0.000339	0.025234 ± 0.000170
20	mdb 066	0.546191 ± 0.000451	0.067199 ± 0.000328
21	mdb 102	0.568146 ± 0.00032	0.021921 ± 0.000228
22	mdb 105	0.900590 ± 0.000130	0.019276 ± 0.000075
23	mdb 110	0.688347 ± 0.000339	0.025498 ± 0.000220
24	mdb 111	0.665251 ± 0.000295	0.039784 ± 0.000264
25	mdb 170	0.473752 ± 0.000231	0.025352 ± 0.000152
26	mdb 236	0.782870 ± 0.000337	0.030524 ± 0.000232
27	mdb 239	0.863344 ± 0.000358	0.021615 ± 0.000287
28	mdb 241	0.711070 ± 0.000345	0.030822 ± 0.000252
29	mdb 249	0.676427 ± 0.000376	0.036537 ± 0.000323
30	mdb 253	0.833378 ± 0.000138	0.015356 ± 0.000065

Tabla 3.9: Datos obtenidos para la muestra de tejido denso (D). Primera columna: mamografía. Siguiente columna: Número de identificación en la base de datos. Dos siguientes columnas: media y desviación estándar del ajuste de la distribución de intensidad en el área de tejido sano seleccionada.

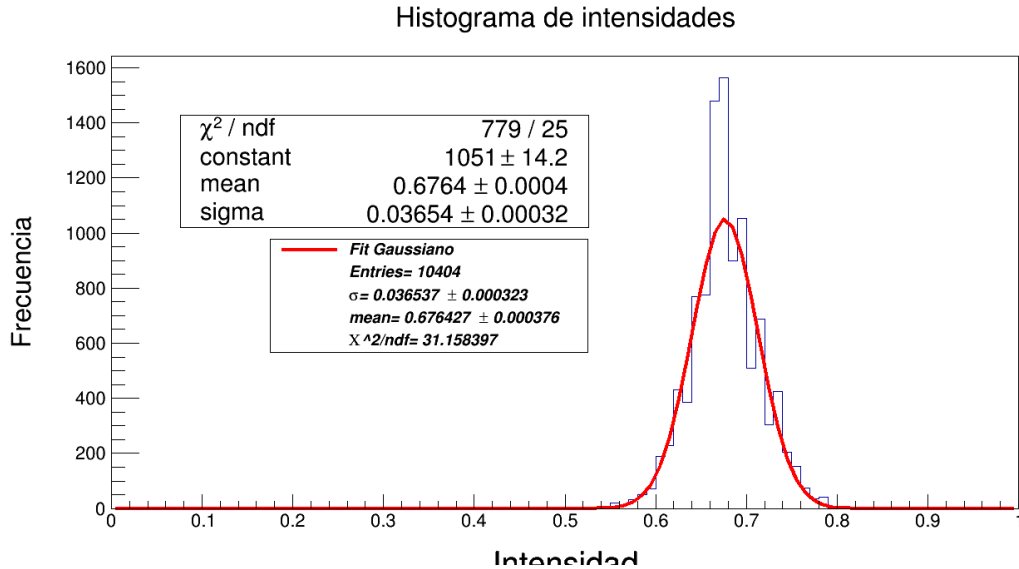


Figura 3.22: Ajuste gaussiano de la distribución de intensidades obtenida para una región de tejido sano de la mamografía mdb 249

El valor máximo de intensidad, ubicado en el área de interés (ROI), para cada muestra es mayor al valor de la media poblacional  $\mu$  obtenida para su respectiva área b y dado que la desviación estándar de la población es una medida de la anchura de la curva de Gauss, es posible extraer conclusiones sobre la población utilizando los parámetros que describen dicha curva. Siendo que, para este caso la curva gaussiana caracteriza al tejido sano, los datos correspondientes al mismo se encuentran considerados dentro del área bajo esta curva, mientras que, el valor de intensidad más alto no se encuentra contemplado en la misma, pues no pertenece a tejido sano y sabiendo que el 99.6% de los valores medidos para la intensidad en el área b, se encuentran contenidos a tres intervalos de confianza de  $\mu$ , esto de acuerdo con la teoría conocida para este tipo de distribución que se presenta en el Anexo C; se espera que, un valor de intensidad dado mayor que el  $\mu + 3\sigma$  obtenido para el tejido sano (área b), se encuentre en una ROI y sugiera muy probablemente la existencia de una anomalía. Por lo anterior, se procedió a calcular los valores  $\mu + 3\sigma$  para todos los MRI que presentan alguna lesión, a fin de compararlos con los valores máximos de intensidad conocidos. Se obtuvieron los resultados de las tablas 3.10, 3.11 y 3.12.

En vista de que los datos obtenidos para ambos parámetros van acompañados de errores, esto debido a que por tratarse de una serie finita de datos esta se desviará algo de la curva de Gauss, se empleó la ecuación 2.2 para calcular la incertidumbre de la suma requerida para el cálculo del valor de interés.

Dados los resultados anteriores se verifica que es posible distinguir entre las ROI y el tejido de fondo definiendo un valor umbral de intensidad de pixel a partir del cual se reconstruirá la imagen, es decir, únicamente se mostrarán aquellos píxeles que superen el valor de intensidad umbral definido. Esto utilizando como primera aproximación la caracterización del área b, para cada tipo de tejido de acuerdo a los parámetros obtenidos para las distribuciones de cada una de las muestras (Tablas 3.7, 3.8 y 3.9) y la teoría del área comprendida bajo la curva de Gauss, pues se espera que, casi la totalidad de los datos que pertenezcan a tejido sano no disten más de  $3\sigma$  de la media obtenida para cada muestra. De este modo, los siguientes parámetros del histograma: media promedio y desviación estándar pueden acusar cuantitativamente la presencia de tejidos de alta densidad, pues aquellos valores tan altos como para superar un valor umbral definido por los parámetros de caracterización para cada tipo de tejido sano, podrían corresponder zonas con anomalías. En este punto, debe estar claro que, este resultado es independiente de la selección

**CAPÍTULO 3. METODOLOGÍA**  
**3.3. CARACTERIZACIÓN DE TEJIDO**

Muestra de tejido graso (F)		
Número de identificación	Valor $\mu + 3\sigma$	Valor de intensidad máximo
mdb 005	$0.634884 \pm 0.00015188$	0.761719
mdb 010	$0.72463 \pm 0.00016358$	0.8125
mdb 012	$0.70138 \pm 0.0001569$	0.777344
mdb 025	$0.599115 \pm 0.0005142$	0.804688
mdb 028	$0.510053 \pm 0.00025664$	0.851563
mdb 059	$0.464385 \pm 0.00020883$	0.671875
mdb 069	$0.663967 \pm 0.00023663$	0.878906
mdb 075	$0.593346 \pm 0.00114853$	0.941406
mdb 080	$0.644847 \pm 0.0005438$	0.796875
mdb 092	$0.662607 \pm 0.00052506$	0.746094
mdb 095	$0.679157 \pm 0.00032017$	0.859375
mdb 134	$0.62063 \pm 0.0001202$	0.820313
mdb 141	$0.672082 \pm 0.00016014$	0.796875
mdb 231	$0.518729 \pm 0.00019052$	0.683594
mdb 238	$0.739453 \pm 0.00029212$	0.792969
mdb 248	$0.705758 \pm 0.00109766$	0.882813
mdb 252	$0.614083 \pm 0.000298$	0.828125
mdb 256	$0.689174 \pm 0.00014862$	0.757813
mdb 267	$0.713189 \pm 0.00010857$	0.828125

Tabla 3.10: Valores de intensidad máxima ubicados en el ROI son mayores que el valor medio de la distribución de fondo más  $3\sigma$  para la muestra de tejido graso

Muestra de tejido glandular (G)		
Número de identificación	Valor $\mu + 3\sigma$	Valor de intensidad máximo
mdb 013	$0.719497 \pm 0.000254$	0.878906
mdb 015	$0.763306 \pm 0.0002056$	0.851563
mdb 017	$0.727961 \pm 0.0003966$	0.929688
mdb 019	$0.738842 \pm 0.0002314$	0.828125
mdb 021	$0.696857 \pm 0.0002187$	0.90625
mdb 023	$0.770454 \pm 0.0001349$	0.875
mdb 030	$0.873244 \pm 0.0001458$	0.898438
mdb 032	$0.878416 \pm 0.0005033$	0.875
mdb 072	$0.637012 \pm 0.0004483$	0.898438
mdb 090	$0.758494 \pm 0.0007657$	0.890625
mdb 209	$0.599489 \pm 0.0004528$	0.890625
mdb 211	$0.897435 \pm 0.0007307$	0.886719
mdb 213	$0.743946 \pm 0.0009075$	0.78125
mdb 218	$0.704428 \pm 0.0002657$	0.894531
mdb 219	$0.755951 \pm 0.0006632$	0.867188
mdb 264	$0.62383 \pm 0.0005296$	0.78125
mdb 265	$0.713955 \pm 0.0005617$	0.886719
mdb 270	$0.801738 \pm 0.0002409$	0.941406

Tabla 3.11: Valores de intensidad máxima ubicados en el ROI son mayores que el valor medio de la distribución de fondo más  $3\sigma$  para la muestra de tejido glandular

**CAPÍTULO 3. METODOLOGÍA**  
**3.3. CARACTERIZACIÓN DE TEJIDO**

Muestra de tejido denso (D)		
Número de identificación	Valor $\mu + 3\sigma$	Valor de intensidad máximo
mdb 058	$0.646177 \pm 0.0003418$	0.886719
mdb 063	$0.662115 \pm 0.0004973$	0.886719
mdb 102	$0.633909 \pm 0.0005083$	0.875
mdb 105	$0.9582728 \pm 0.0001838$	0.945313
mdb 110	$0.764841 \pm 0.0005517$	0.878906
mdb 111	$0.784603 \pm 0.0005442$	0.886719
mdb 170	$0.549808 \pm 0.0003502$	0.921875
mdb 236	$0.874442 \pm 0.0005244$	0.88283
mdb 239	$0.928189 \pm 0.0006126$	0.917969
mdb 241	$0.803536 \pm 0.0005564$	0.863281
mdb 249	$0.786038 \pm 0.0006741$	0.851563
mdb 253	$0.879446 \pm 0.0011343$	0.882813

Tabla 3.12: Valores de intensidad máxima ubicados en el ROI son mayores que el valor medio de la distribución de fondo más  $3\sigma$  para la muestra de tejido denso

de tejido sano para estudio, es decir, los resultados mostrados en las tablas 3.4, 3.5 y 3.6 son consistentes independientemente de la región de tejido sano presentada en el histograma, pues las distribuciones se superponen en los mismos valores de intensidad [10]. Una selección más suave de la región de estudio únicamente otorgará una definición más exacta del valor de intensidad a definir para la separación entre las regiones b y ROI.

Para obtener el valor umbral para cada tipo de tejido se calculó el valor  $\mu + 3\sigma$  para las mamografías restantes de la muestra original, es decir, también se obtuvo este valor para aquellas mamografías que no presentan anomalías. Los resultados se presentan en la tabla 3.13. Posteriormente, basándose en las herramientas de la estadística, aceptando la conclusión de que para un conjunto de datos se considera que el promedio de los mismos corresponde a la mejor estimación del valor real, es decir, aquel que tiene la más alta probabilidad de ser verdadero, se procedió a obtener la media  $\mu \pm 3\sigma$  de estos valores para cada tipo de muestra utilizando la ecuación 2.1 y estos tres resultados serán designados como los valores umbral ( $\rho$ ) para excluir la sección del área b y distinguir la ROI. Los resultados se presentan en la sección correspondiente.

Como conforme aumenta la densidad de tejido, la intensidad de los píxeles también lo hace, para el caso de tejidos de mayor densidad (G y D) se calculó también el valor  $\mu + 4\sigma$  que se propone como segundo candidato de valor de intensidad umbral y se muestra en las tablas 3.14 y 3.15 . Comparando este valor con los valores de intensidad máximos de píxel para las mamografías que presentan lesiones (tablas 3.5 y 3.6 ) se encuentra que estos son aún mayores que el nuevo parámetro calculado, no obstante, este nuevo valor ya no está considerado dentro de las distribuciones características de tejido sano, por lo que proporciona un dato más certero para realizar el corte entre la región b y las ROI, de modo que, este se define como el "valor umbral" para los tejidos glandular y denso.

Los resultados (valores umbrales obtenidos) y las reconstrucciones se presentan en el siguiente apartado. Finalmente, para lograr una mejor apreciación de los resultados, se hizo uso de la normalización de la intensidad de los píxeles del mapa de intensidades, al elevar a ciertas potencias dichos valores, se destacan los píxeles persistentes.

**CAPÍTULO 3. METODOLOGÍA**  
**3.3. CARACTERIZACIÓN DE TEJIDO**

---

Mamografías normales		
Tipo de tejido	Número de identificación	Valor $\mu + 3\sigma$
F	mdb 009	0.706575 $\pm$ 0.00110412
F	mdb 011	0.728356 $\pm$ 0.00014186
F	mdb 026	0.665167 $\pm$ 0.00033354
F	mdb 027	0.473246 $\pm$ 0.00016658
F	mdb 060	0.459231 $\pm$ 0.00026864
F	mdb 070	0.745941 $\pm$ 0.00015785
F	mdb 076	0.613851 $\pm$ 0.00013031
F	mdb 077	0.628511 $\pm$ 0.00014238
F	mdb 078	0.672947 $\pm$ 0.00019848
F	mdb 079	0.829701 $\pm$ 0.00366914
F	mdb 087	0.628792 $\pm$ 0.0002928
G	mdb 007	0.741995 $\pm$ 0.0003127
G	mdb 008	0.716032 $\pm$ 0.0002603
G	mdb 014	0.765974 $\pm$ 0.0016739
G	mdb 016	0.772934 $\pm$ 0.000405
G	mdb 018	0.799524 $\pm$ 0.0003152
G	mdb 020	0.663335 $\pm$ 0.0002359
G	mdb 022	0.70706 $\pm$ 0.0003289
G	mdb 024	0.73281 $\pm$ 0.0001552
G	mdb 029	0.866877 $\pm$ 0.0003313
G	mdb 031	0.843449 $\pm$ 0.0002731
G	mdb 041	0.5818 $\pm$ 0.0001815
G	mdb 042	0.539549 $\pm$ 0.000331
D	mdb 003	0.982496 $\pm$ 0.000558
D	mdb 004	0.897075 $\pm$ 0.0002162
D	mdb 033	0.810377 $\pm$ 0.0003457
D	mdb 034	0.7708 $\pm$ 0.0007375
D	mdb 035	0.818051 $\pm$ 0.0006709
D	mdb 036	0.831509 $\pm$ 0.000568
D	mdb 037	0.843196 $\pm$ 0.0004063
D	mdb 038	0.777417 $\pm$ 0.0002837
D	mdb 039	0.873169 $\pm$ 0.0003637
D	mdb 040	0.869177 $\pm$ 0.0003373
D	mdb 053	0.869025 $\pm$ 0.0002784
D	mdb 054	0.806657 $\pm$ 0.0002998
D	mdb 057	0.760115 $\pm$ 0.0002542
D	mdb 061	0.743192 $\pm$ 0.000501
D	mdb 062	0.663918 $\pm$ 0.0004036
D	mdb 064	0.855814 $\pm$ 0.0004309
D	mdb 065	0.889016 $\pm$ 0.000449
D	mdb 066	0.747788 $\pm$ 0.0007254

Tabla 3.13: Valor medio de la distribución de fondo más  $3\sigma$  para las mamografías que no presentan anomalía de la muestra original

**CAPÍTULO 3. METODOLOGÍA**  
**3.3. CARACTERIZACIÓN DE TEJIDO**

---

Muestra de tejido glandular (G)	
Número de identificación	Valor $\mu + 4\sigma$
mdb 007	0.785503 $\pm$ 0.0003331
mdb 008	0.757791 $\pm$ 0.0002818
mdb 013	0.747344 $\pm$ 0.0002768
mdb 014	0.790588 $\pm$ 0.0019297
mdb 015	0.804274 $\pm$ 0.000224
mdb 016	0.823549 $\pm$ 0.0004362
mdb 017	0.760461 $\pm$ 0.000428
mdb 018	0.824455 $\pm$ 0.0003366
mdb 019	0.774993 $\pm$ 0.0002541
mdb 020	0.686041 $\pm$ 0.0002524
mdb 021	0.721693 $\pm$ 0.0002384
mdb 022	0.73728 $\pm$ 0.0003567
mdb 023	0.799345 $\pm$ 0.0001468
mdb 024	0.757376 $\pm$ 0.000169
mdb 029	0.903362 $\pm$ 0.0003605
mdb 030	0.894919 $\pm$ 0.0001573
mdb 031	0.8898 $\pm$ 0.0002902
mdb 032	0.924471 $\pm$ 0.0005438
mdb 041	0.608659 $\pm$ 0.0001992
mdb 042	0.567278 $\pm$ 0.0003698
mdb 072	0.658009 $\pm$ 0.0004858
mdb 090	0.78959 $\pm$ 0.0008318
mdb 209	0.63071 $\pm$ 0.000488
mdb 211	0.923693 $\pm$ 0.0008055
mdb 213	0.781249 $\pm$ 0.000993
mdb 218	0.7299 $\pm$ 0.0002836
mdb 219	0.98684 $\pm$ 0.0007157
mdb 264	0.677732 $\pm$ 0.0005709
mdb 265	0.738929 $\pm$ 0.000599
mdb 270	0.846533 $\pm$ 0.0002632

Tabla 3.14: Valores  $\mu + 4\sigma$  calculados la muestra de tejido glandular

**CAPÍTULO 3. METODOLOGÍA**  
**3.3. CARACTERIZACIÓN DE TEJIDO**

---

Muestra de tejido denso (D)	
Número de identificación	Valor $\mu + 4\sigma$
mdb 003	1.036765 $\pm$ 0.0006243
mdb 004	0.922058 $\pm$ 0.0002346
mdb 033	0.862407 $\pm$ 0.0003916
mdb 034	0.807145 $\pm$ 0.0008163
mdb 035	0.849883 $\pm$ 0.0007433
mdb 036	0.873073 $\pm$ 0.0006098
mdb 037	0.871565 $\pm$ 0.0004297
mdb 038	0.801653 $\pm$ 0.0003005
mdb 039	0.91326 $\pm$ 0.0003876
mdb 040	0.90039 $\pm$ 0.0003644
mdb 053	0.896566 $\pm$ 0.0002958
mdb 054	0.824856 $\pm$ 0.0003211
mdb 057	0.803371 $\pm$ 0.000279
mdb 058	0.668091 $\pm$ 0.0003741
mdb 061	0.766281 $\pm$ 0.0005334
mdb 062	0.695007 $\pm$ 0.000436
mdb 063	0.686911 $\pm$ 0.0005475
mdb 064	0.891308 $\pm$ 0.0004636
mdb 065	0.91425 $\pm$ 0.0004801
mdb 066	0.814987 $\pm$ 0.0007961
mdb 102	0.65583 $\pm$ 0.0005571
mdb 105	0.9775004 $\pm$ 0.0001985
mdb 110	0.790339 $\pm$ 0.000594
mdb 111	0.824387 $\pm$ 0.0006048
mdb 170	0.57516 $\pm$ 0.0003818
mdb 236	0.904966 $\pm$ 0.0005735
mdb 239	0.949804 $\pm$ 0.0006765
mdb 241	0.834358 $\pm$ 0.0006108
mdb 249	0.822575 $\pm$ 0.0007475
mdb 253	0.894802 $\pm$ 0.0013073

Tabla 3.15: Valores  $\mu + 4\sigma$  calculados la muestra de tejido denso

## Capítulo 4

# Resultados y Discusión

En este capítulo se presentan los resultados más relevantes de este análisis, la separación entre la información que se considera resultado y metodología en este análisis se hizo considerando las hipótesis de esta tesis. Primero se muestran los resultados estadísticos para las medias muestrales de todas las distribuciones separados según el tipo de tejido, esto con el fin de determinar la posibilidad de distinguir entre los diferentes tipos de tejido sano cuantitativamente haciendo uso de los parámetros que caracterizan a sus distribuciones, pues este hecho es uno de los objetivos planteados en esta tesis. Posteriormente se exponen los valores candidatos a intensidades umbrales  $\rho$  obtenidos para cada tipo de tejido y la consiguiente separación de regiones de alta densidad y el tejido de fondo en mamografías basada en los mismos. Las herramientas necesarias para la realización de las reconstrucciones ya han sido expuestas en los capítulos anteriores. Aquí, sólo queda mostrar los resultados y su confiabilidad basándose en los diagnósticos médicos que acompañan a cada imagen.

### 4.1. Estadística descriptiva de las medias de las distribuciones de intensidad

En la primera muestra de 30 mamografías oblicuas laterales medias (OML), clasificadas como tejido graso (F), para la media se obtuvieron los datos estadísticos presentados en la tabla 4.1 .

	N Total	$\bar{\mu}$	$\bar{\sigma}$	Suma	Mínimo	Mediana	Máximo
$\mu$ Tejido F	30	0.55536	0.07642	16.66082	0.39379	0.55515	0.66121

Tabla 4.1: Datos estadísticos de los valores promedio obtenidos de las distribuciones de intensidades para la muestra de tejido graso (F)

Para la segunda muestra de 30 mamografías OML de tejido glandular (G), obtuvieron los datos estadísticos de la media presentados en la tabla 4.2 .

	N Total	$\bar{\mu}$	$\bar{\sigma}$	Suma	Mínimo	Mediana	Máximo
$\mu$ Tejido G	30	0.6384	0.08634	19.15203	0.45636	0.63125	0.81866

Tabla 4.2: Datos estadísticos de los valores promedio obtenidos de las distribuciones de intensidades para la muestra de tejido glandular (G)

**CAPÍTULO 4. RESULTADOS Y DISCUSIÓN**  
**4.2. CALCULO DEL "VALOR UMBRAL" PARA CADA TIPO DE TEJIDO**

---

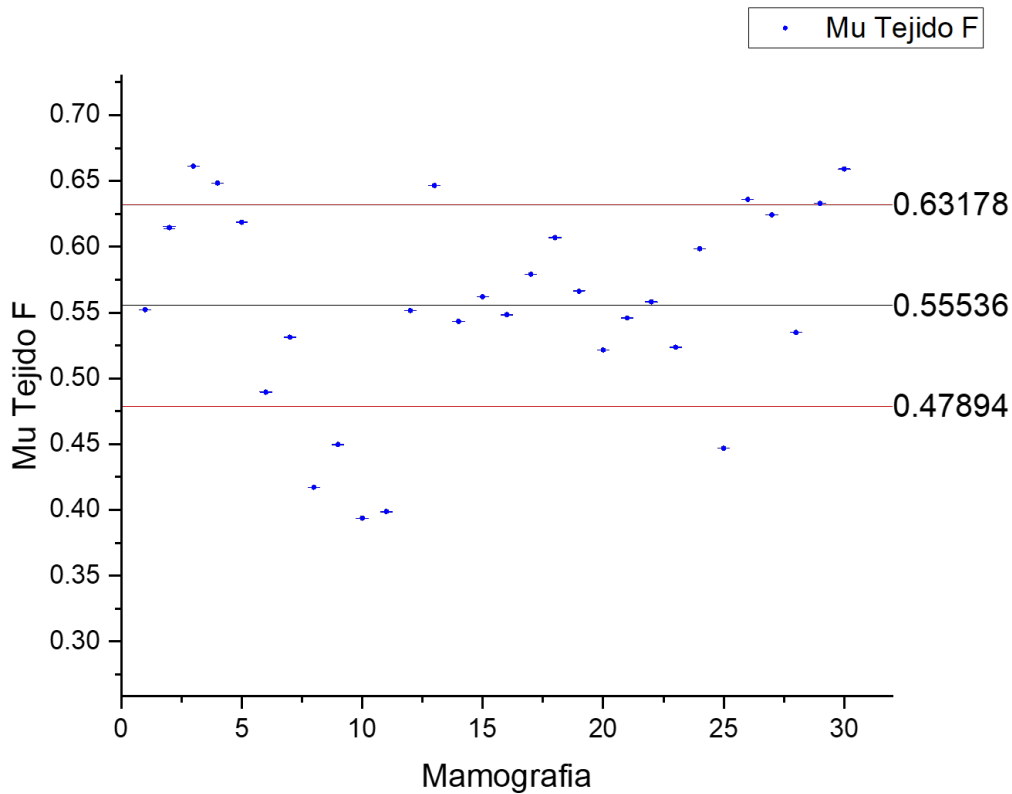


Figura 4.1: Se presentan los valores de la media obtenidos para cada una de las mamografías de la muestra F con sus respectivas barras de error. La línea horizontal negra corresponde a la media de estos datos y las líneas horizontales rojas presentan el primer intervalo de confianza de los mismos.

Finalmente, para la tercera muestra de 30 mamografías OML, catalogadas como de tejido denso (D), se obtuvieron los datos estadísticos de la media presentados en la tabla 4.3 .

	N Total	$\bar{\mu}$	$\bar{\sigma}$	Suma	Mínimo	Mediana	Máximo
$\mu$ Tejido D	30	0.70773	0.10237	21.23203	0.47375	0.70894	0.90059

Tabla 4.3: Datos estadísticos de los valores promedio obtenidos de las distribuciones de intensidades para la muestra de tejido denso (D)

Las representaciones gráficas de estos datos para las muestras F, G y D se presentan en las figuras 4.1, 4.2 y 4.3 respectivamente.

Como otra forma de representación de la información de las medias muestrales para cada tipo de tejido se realizaron sus gráficas de cajas y bigotes. Estas se muestran en la figura 4.4.

## 4.2. Calculo del "valor umbral" para cada tipo de tejido

Se calculó la media muestral  $\bar{x}$  de los valores presentados en las tablas 3.10, 3.5 , 3.12 y 3.13. Los errores calculados en este cálculo no son relevantes para los resultados pues tienen un efecto mínimo

**CAPÍTULO 4. RESULTADOS Y DISCUSIÓN**  
**4.2. CALCULO DEL "VALOR UMBRAL" PARA CADA TIPO DE TEJIDO**

---

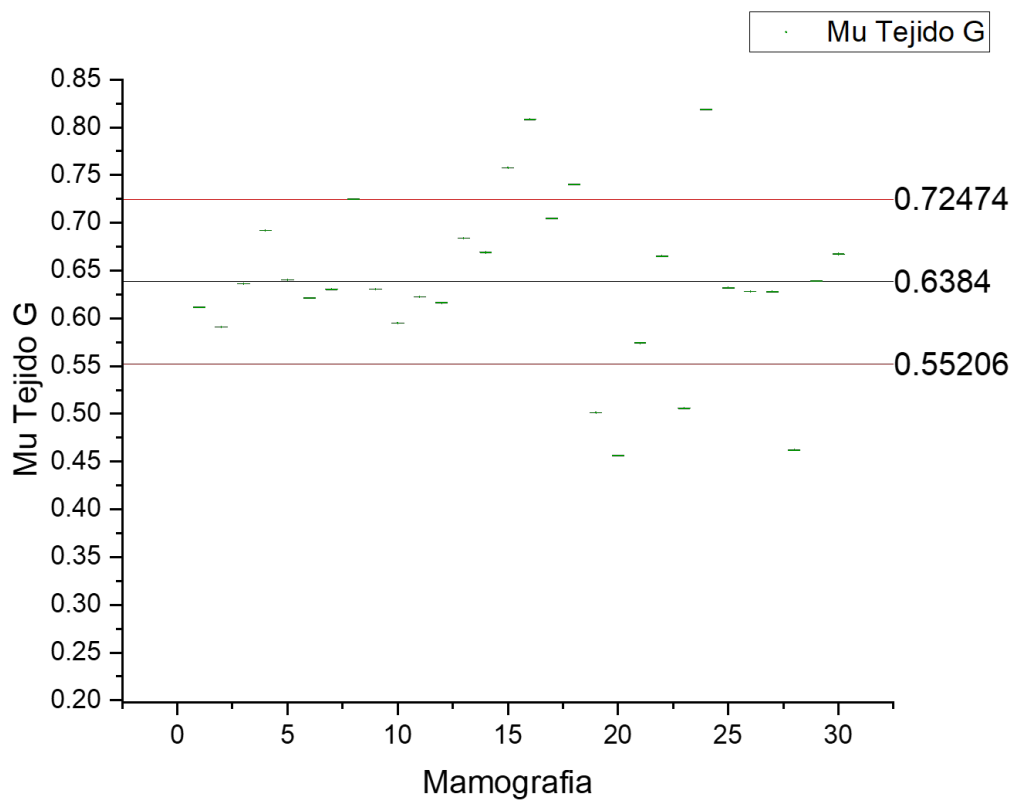


Figura 4.2: Se presentan los valores de la media obtenidos para cada una de las mamografías de la muestra G con sus respectivas barras de error. La línea horizontal negra corresponde a la media de estos datos y las líneas horizontales rojas presentan el primer intervalo de confianza de los mismos.

**CAPÍTULO 4. RESULTADOS Y DISCUSIÓN**  
**4.2. CALCULO DEL "VALOR UMBRAL" PARA CADA TIPO DE TEJIDO**

---

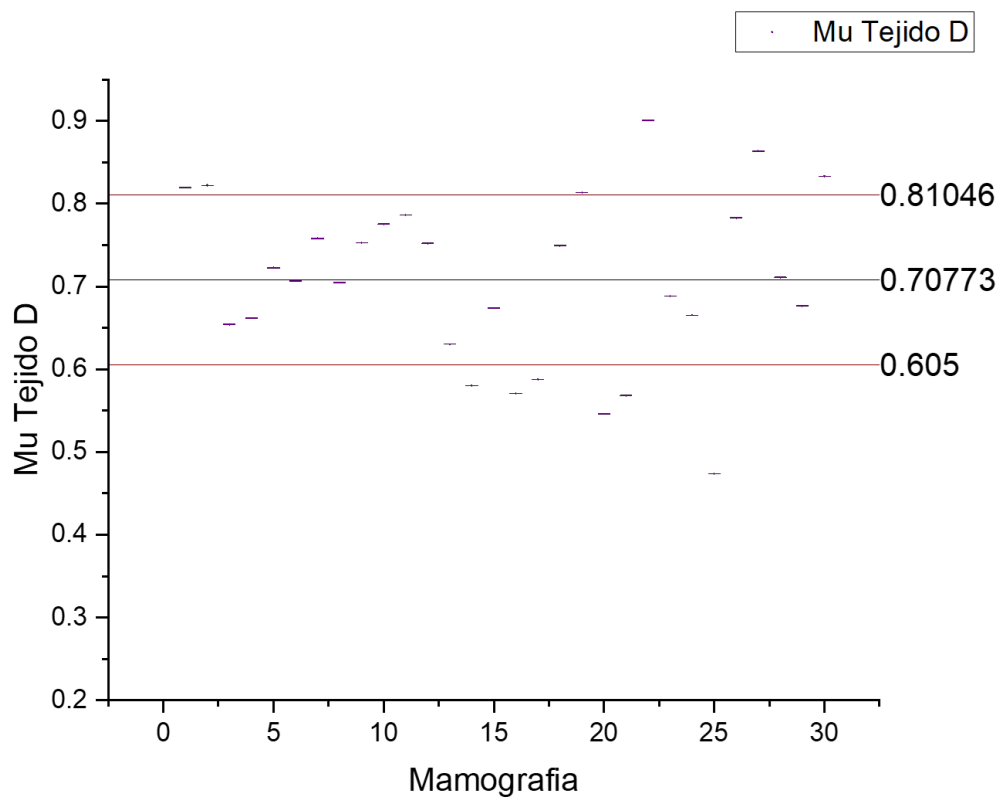


Figura 4.3: Se presentan los valores de la media obtenidos para cada una de las mamografías de la muestra D con sus respectivas barras de error. La línea horizontal negra corresponde a la media de estos datos y las líneas horizontales rojas presentan el primer intervalo de confianza de los mismos.

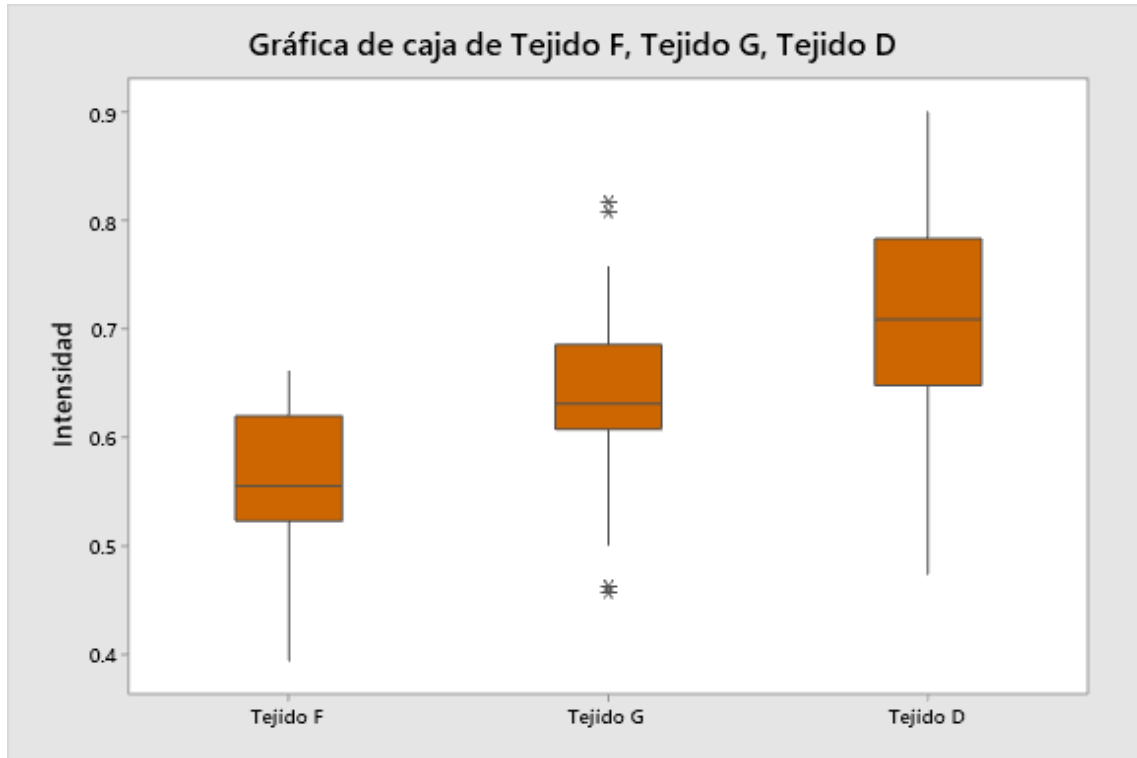


Figura 4.4: Distribución de caja para la densidad media de las tres muestras

sobre el valor de la media que no hacen diferencia en los píxeles mostrados en las redigitalizaciones finales debido a la cantidad de niveles de gris de las imágenes estudiadas, es por ello que ya no se consideran en este apartado, lo mismo para el apartado anterior.

#### 4.2.1. Tejido F

Utilizando la ecuación 2.1 para los valores calculados  $\mu + 3\sigma$  se obtiene:

$$\bar{x}_F = 0,6434629$$

Este valor se define como el valor umbral para este tipo de tejido:

$$\rho_F = 0,6434629$$

#### 4.2.2. Tejido G

El cálculo de la media de los valores  $\mu + 3\sigma$  para este tipo de tejido da como resultado:

$$\bar{x}_G = 0,7379555$$

El segundo valor umbral, obtenido del promedio de los valores  $\mu + 4\sigma$  para las 30 mamografías de esta clasificación que se muestran en la tabla 3.14 es el siguiente:

$$\bar{x}'_G = 0,771140367$$

Como se indicó en la sección final del capítulo anterior, este segundo valor umbral se calculó debido a que, este tipo de tejido, al ser de mayor densidad, presenta también valores de intensidad de píxel

más altos, por ende, la dificultad de definir correctamente el valor de interés incrementa, se busca el mejor candidato a valor umbral. Dado este hecho, se tiene que  $\bar{x}'_G$  excluye casi en su totalidad los píxeles correspondientes a tejido sano, pues se encuentra fuera del área bajo la curva que caracteriza al mismo. De este modo, definiendo como  $\rho_{G1}$  y  $\rho_{G2}$  los valores  $\bar{x}_G$  y  $\bar{x}'_G$  respectivamente; para tejido glandular se determinaron dos valores umbrales:

$$\rho_{G1} = 0,7379555$$

$$\rho_{G2} = 0,7711403667$$

Los resultados de las regiones persistentes para ambos valores umbrales se presentan en el siguiente apartado.

### 4.2.3. Tejido D

Finalmente, los valores umbral calculados para el tejido de mayor densidad son los siguientes:

$$\rho_{D1} = 0,8026723$$

$$\rho_{D2} = 0,8343183$$

## 4.3. ROI aislada

A continuación se muestran los mejores resultados para los MRI obtenidos de la redigitalización de las MI. Para observar todos los resultados de las mamografías que presentan una anomalía y fueron utilizadas para este análisis véase el Apéndice D. Como el objetivo de este trabajo era la localización de zonas con posibles anomalías no se incluyen resultados con falsos positivos, es decir, no se examinaron mamografías de tipo normal con los valores umbrales definidos, pero se esperaba que de analizarse las mismas, sus respectivas redigitalizaciones prácticamente no mostrarán la presencia de píxeles persistentes. Este análisis se propone como trabajo futuro.

Las siguientes listas presenta los datos de interés de las MI que se muestran en esta sección de resultados, clasificadas según su tipo de tejido.

### TEJIDO GRASO (F)

- Mamografía: mdb 005. Presenta CIRC Benigna. (Figura 4.5)
- Mamografía: mdb 012. Presenta CIRC Benigna. (Figura 4.6)
- Mamografía: mdb 025. Presenta CIRC Benigna. (Figura 4.7)
- Mamografía: mdb 028. Presenta CIRC Maligna. (Figura 4.8)
- Mamografía: mdb 069. Presenta CIRC Benigna. (Figura 4.9)
- Mamografía: mdb 134. Presenta MISC Maligna. (Figura 4.10)
- Mamografía: mdb 231. Presenta CALC Maligna. (Figura 4.11)

### TEJIDO GLANDULAR (G)

- Mamografía: mdb 015. Presenta CIRC Benigna. (Figura 4.12)
- Mamografía: mdb 211. Presenta CALC Maligna. (Figura 4.13)
- Mamografía: mdb 213. Presenta CALC Maligna. (Figura 4.14)
- Mamografía: mdb 218. Presenta CALC Maligna. (Figura 4.15)

- Mamografía: mdb 219. Presenta CALC Maligna. (Figura 4.16)

#### TEJIDO DENSO (D)

- Mamografía: mdb 063. Presenta MISC Maligna. (Figura 4.17)
- Mamografía: mdb 102. Presenta ASYM Maligna. (Figura 4.18)
- Mamografía: mdb 111. Presenta ASYM Maligna. (Figura 4.19)
- Mamografía: mdb 170. Presenta ARCH Maligna. (Figura 4.20)
- Mamografía: mdb 249. Presenta CALC Benigna. (Figura 4.21)

Cada una de las anomalías fue localizada en la MI original haciendo uso de los datos de diagnóstico proporcionados. Las ROI se encerraron en un perímetro color blanco, el inciso (a) de cada figura presenta este hecho, posteriormente, se llevó a cabo la redigitalización de estas nuevas imágenes. Dado el color elegido para señalar las ROI, este perímetro persiste en las imágenes redigitalizadas, señalando así las anomalías en los MRI. Los incisos (b) y (c) de cada figura muestran los MRI obtenidos para el primer valor umbral definido, es decir, el correspondiente a la media muestral de los valores  $\mu + 3\sigma$  con los valores de intensidad de los píxeles a la primera y quinta potencia respectivamente. Esto con el fin de aumentar el contraste de los píxeles persistentes y que estos puedan visualizarse más fácilmente para el caso de lesiones pequeñas como es el caso de las microcalcificaciones. Finalmente, las figuras que presentan incisos (d) y (e) corresponden a MI de tejido glandular o denso, para los cuales se calculó un segundo valor umbral correspondiente a la media muestral de los valores  $\mu + 4\sigma$  y de igual manera se presentan los valores de intensidad de los píxeles a la primera y quinta potencia.

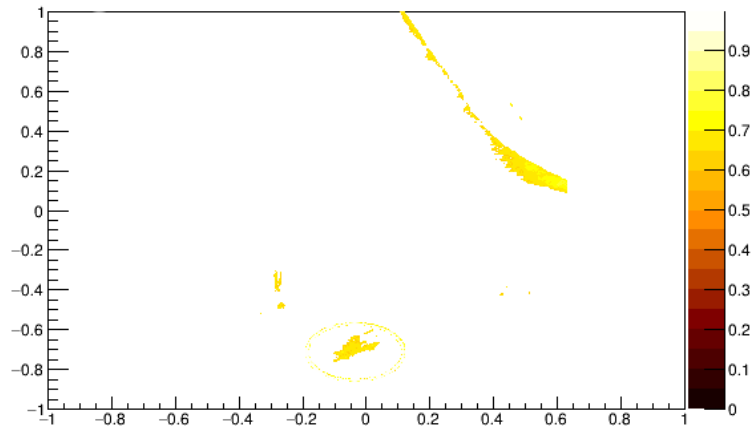
## 4.4. Discusión

Con respecto a los datos presentados en las tablas 4.1, 4.2 y 4.3 se puede decir lo siguiente:

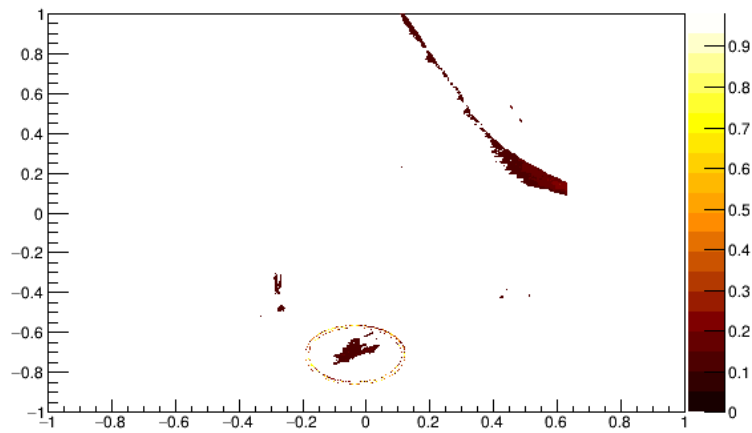
- Se observa que el valor promedio de las medias de las distribuciones incrementa conforme la densidad del tejido es mayor.
- La desviación estándar de los datos para las dos primeras muestras es pequeña y de un orden de magnitud menor que la de tejido D, lo que nos dice que el tejido de tipo F y G de las mamas es más uniforme que el de tipo denso, lo que se ve reflejado como una variación menor entre las intensidades de los píxeles que conforman la imagen, este es un resultado conocido dada la limitación que se tiene para detectar anomalías en las mamografías conformadas por tejido de mayor densidad.
- Para los tres tipos de tejido la media aritmética presenta un valor muy cercano al de la mediana, lo que habla de que los datos se distribuyen casi simétricamente entorno a la media lo que hace pensar que estos siguen una distribución gaussiana.
- Las gráficas presentadas para cada tipo de muestra muestran la distribución de los datos obtenidos, las barras de error que acompañan a los valores, generadas durante el ajuste a la curva de Gauss, son casi imperceptibles a simple vista.
- La distribución de los datos tiende a una distribución de Gauss, aproximadamente  $\frac{2}{3}$  de los datos de las medias no distan de la media en más de una desviación estándar, es decir, se encuentran dentro del primer intervalo de confianza. Este comportamiento es característico de la mayoría de los procesos analíticos cuantitativos [19] por ello, una mayor cantidad de datos a analizar se espera se distribuyan de forma normal, para este caso no se tienen conclusiones al respecto dado que la cantidad de datos obtenidos (30) es el límite inferior de información para realizar estadística según la bibliografía para análisis cuantitativo utilizada.



(a) Localización de la ROI en la mamografía original



(b) ROI aislada para el valor umbral  $\mu + 3\sigma$

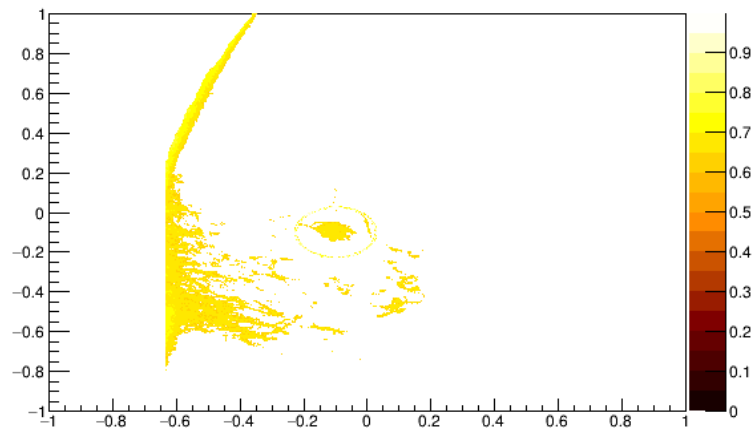


(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

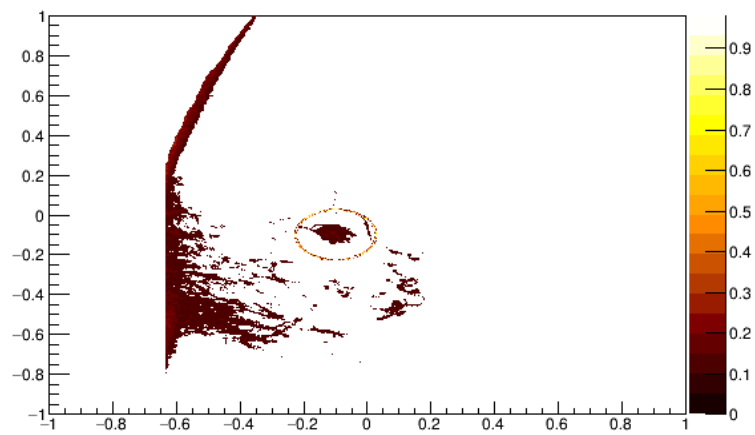
Figura 4.5: Resultados para la mamografía con ID mdb 005



(a) Localización de la ROI en la mamografía original

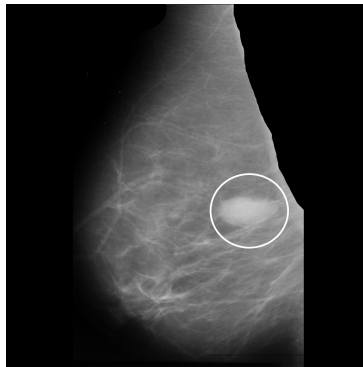


(b) ROI aislada para el valor umbral  $\mu + 3\sigma$

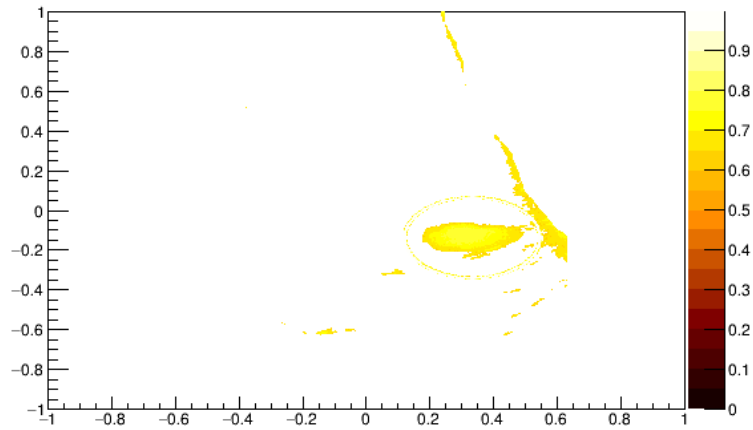


(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

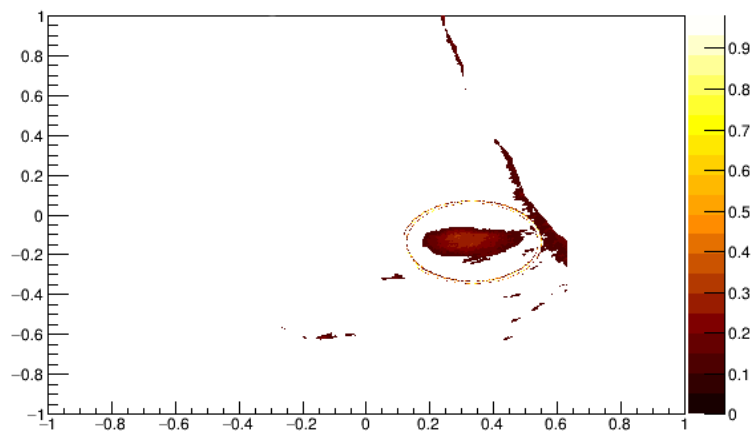
Figura 4.6: Resultados para la mamografía con ID mdb 012



(a) Localización de la ROI en la mamografía original

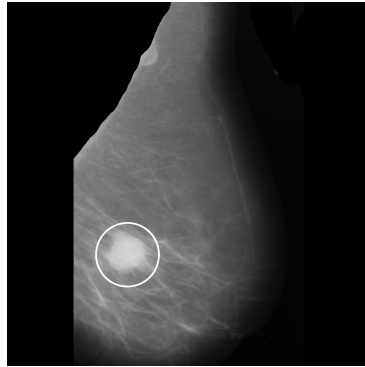


(b) ROI aislada para el valor umbral  $\mu + 3\sigma$

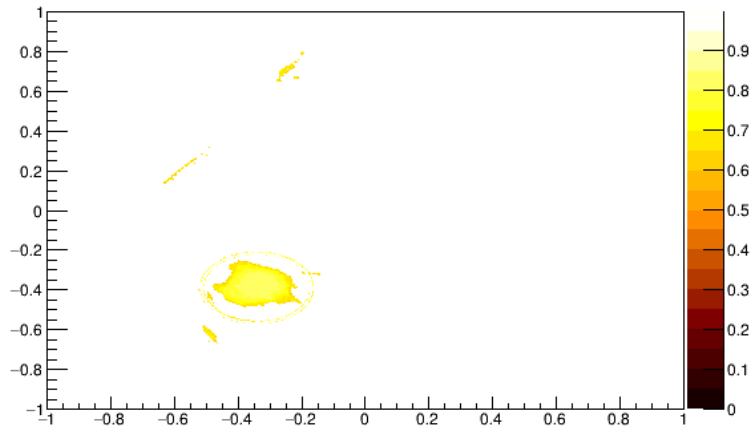


(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

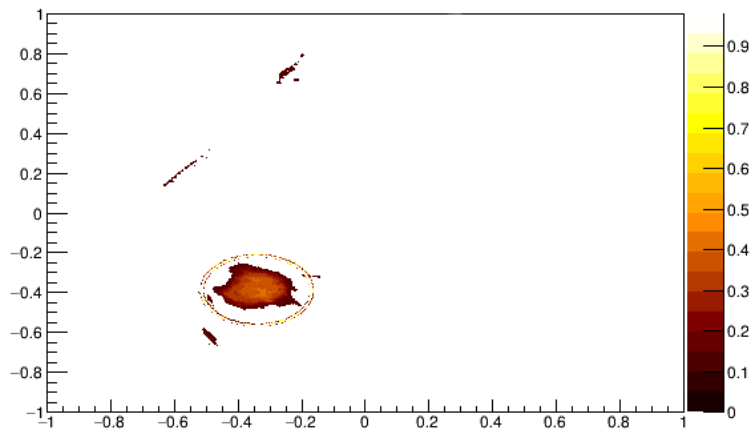
Figura 4.7: Resultados para la mamografía con ID mdb 025



(a) Localización de la ROI en la mamografía original

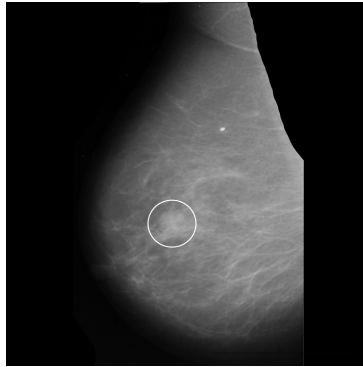


(b) ROI aislada para el valor umbral  $\mu + 3\sigma$

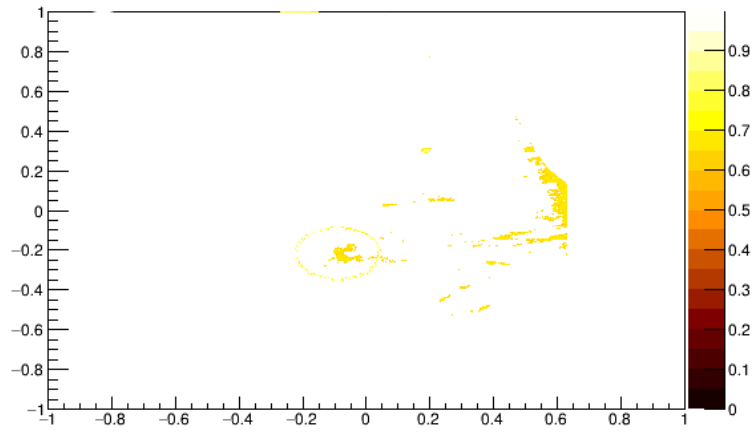


(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

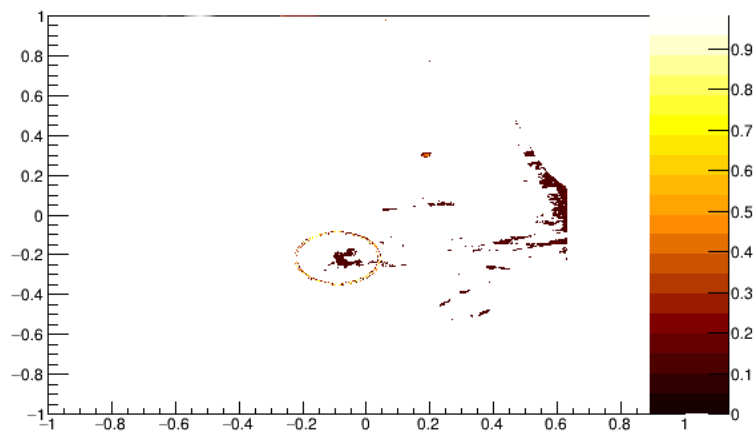
Figura 4.8: Resultados para la mamografía con ID mdb 028



(a) Localización de la ROI en la mamografía original

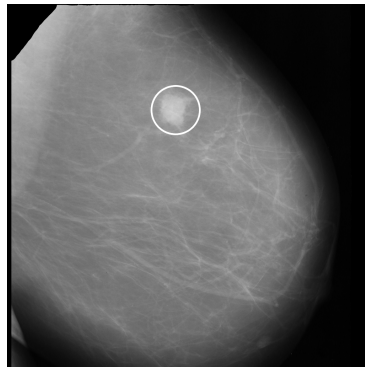


(b) ROI aislada para el valor umbral  $\mu + 3\sigma$

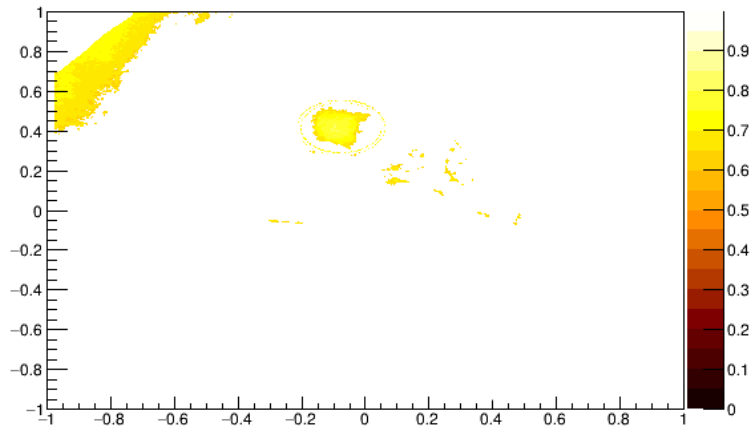


(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

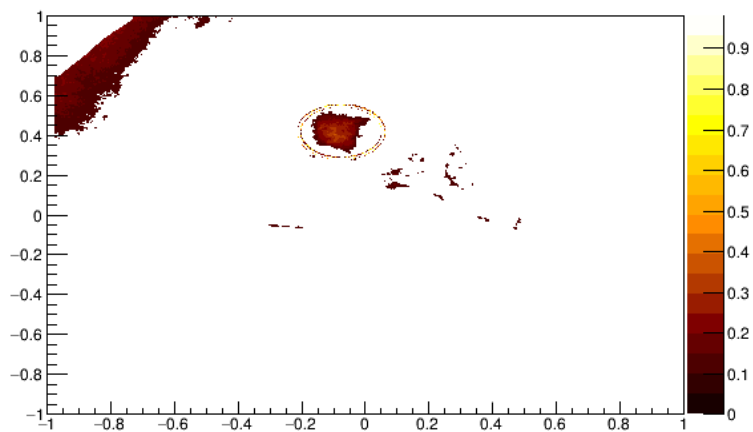
Figura 4.9: Resultados para la mamografía con ID mdb 069



(a) Localización de la ROI en la mamografía original

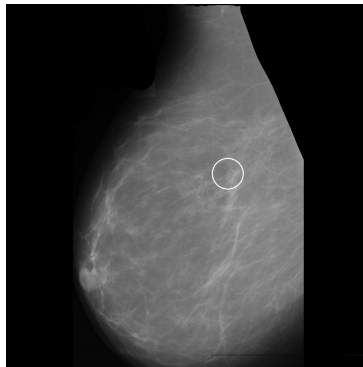


(b) ROI aislada para el valor umbral  $\mu + 3\sigma$

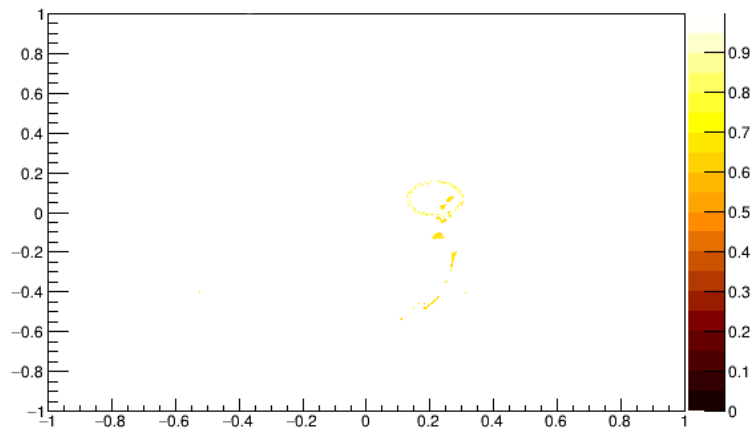


(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

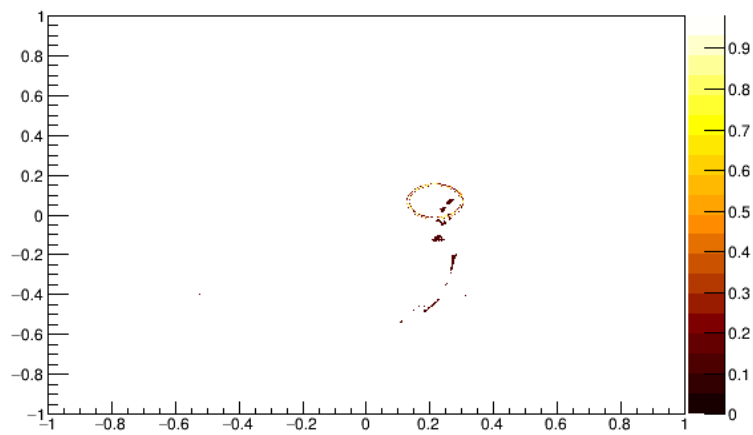
Figura 4.10: Resultados para la mamografía con ID mdb 134



(a) Localización de la ROI en la mamografía original

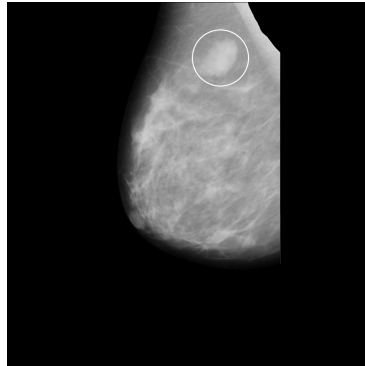


(b) ROI aislada para el valor umbral  $\mu + 3\sigma$

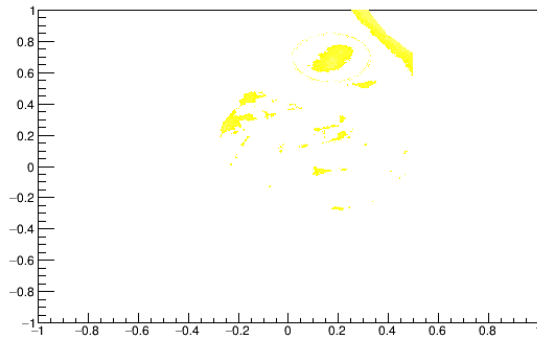


(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

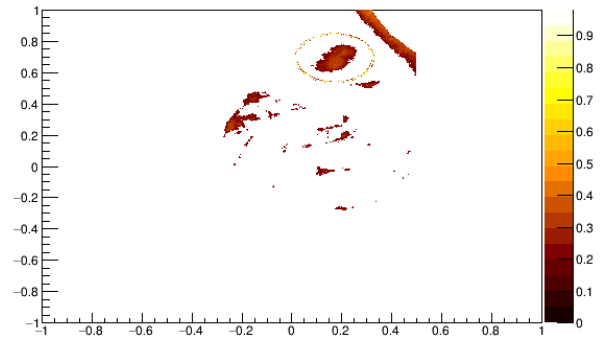
Figura 4.11: Resultados para la mamografía con ID mdb 231



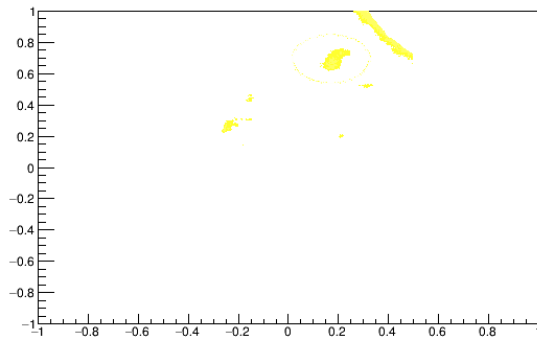
(a) Localización de la ROI en la mamografía original



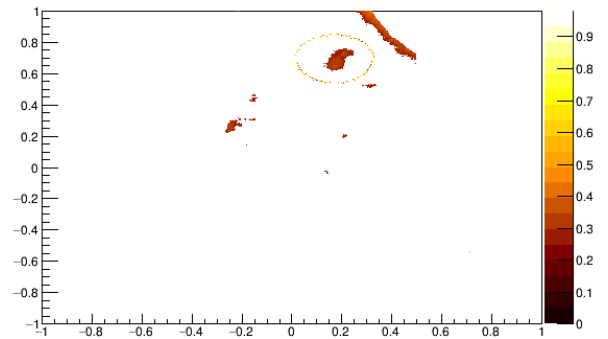
(b) ROI aislada para el valor umbral  $\mu + 3\sigma$



(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

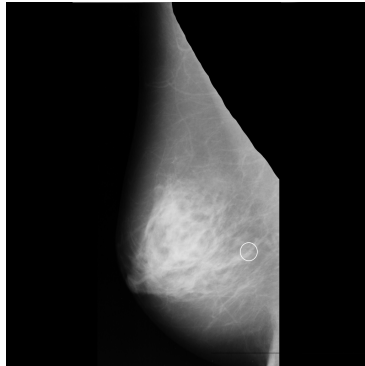


(d) ROI aislada para el valor umbral  $\mu + 4\sigma$

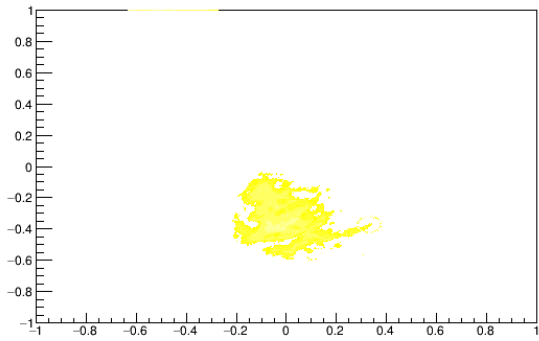


(e) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (d)

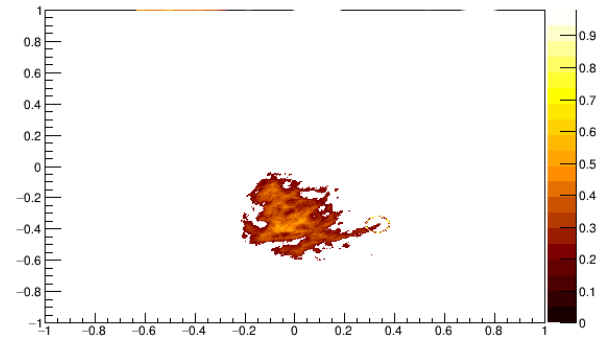
Figura 4.12: Resultados para la mamografía con ID mdb 015



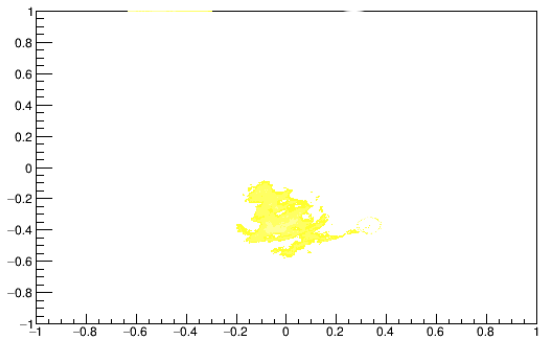
(a) Localización de la ROI en la mamografía original



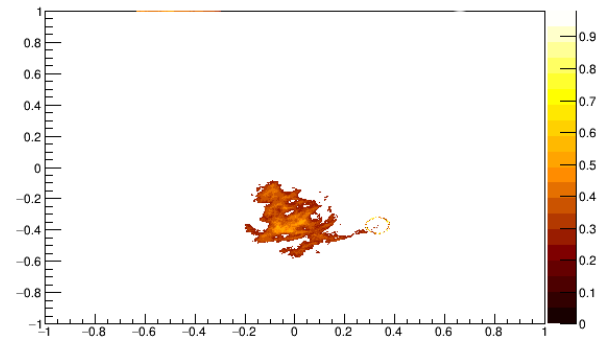
(b) ROI aislada para el valor umbral  $\mu + 3\sigma$



(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

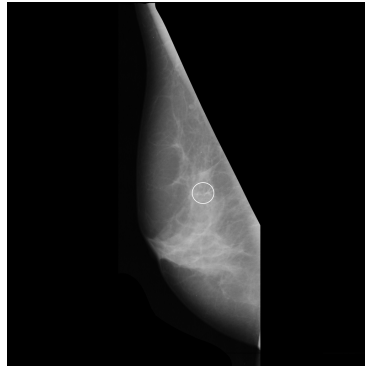


(d) ROI aislada para el valor umbral  $\mu + 4\sigma$

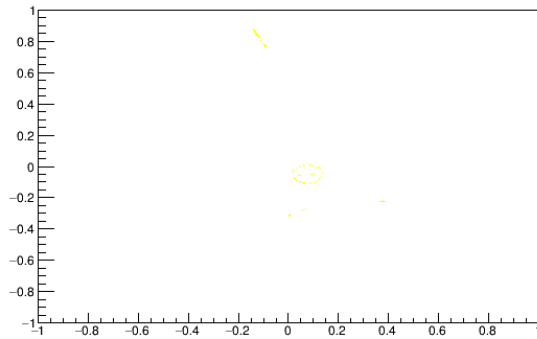


(e) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (d)

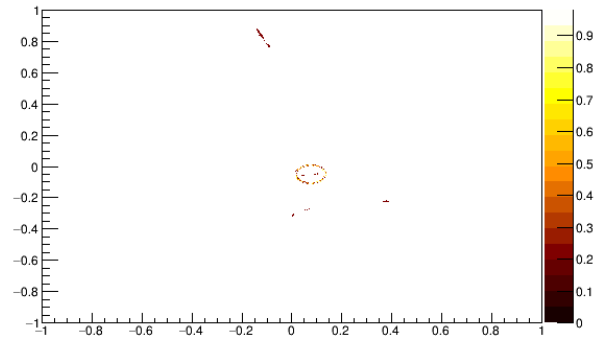
Figura 4.13: Resultados para la mamografía con ID mdb 211



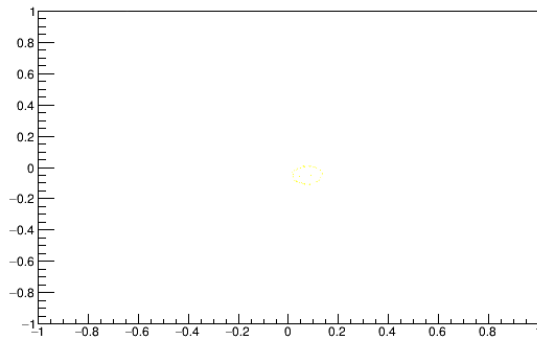
(a) Localización de la ROI en la mamografía original



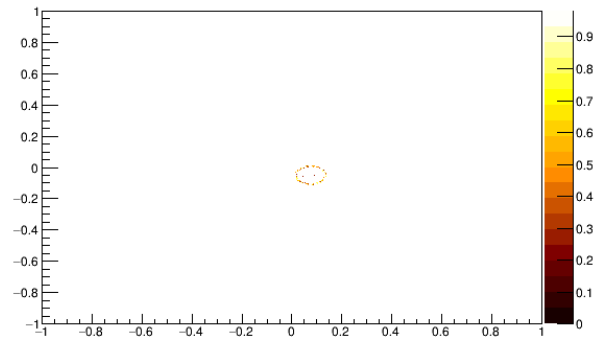
(b) ROI aislada para el valor umbral  $\mu + 3\sigma$



(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

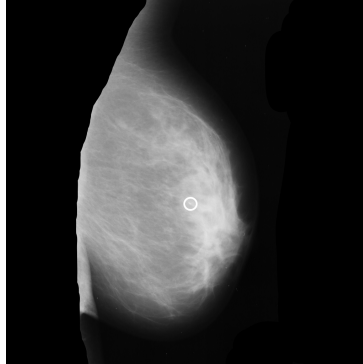


(d) ROI aislada para el valor umbral  $\mu + 4\sigma$

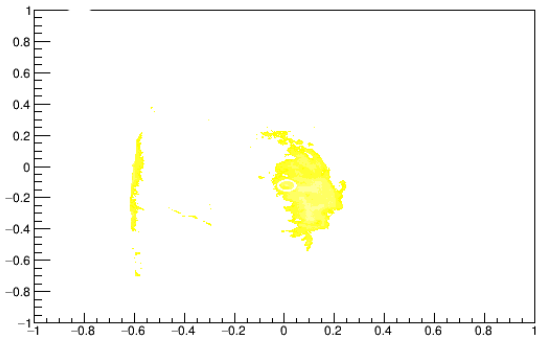


(e) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (d)

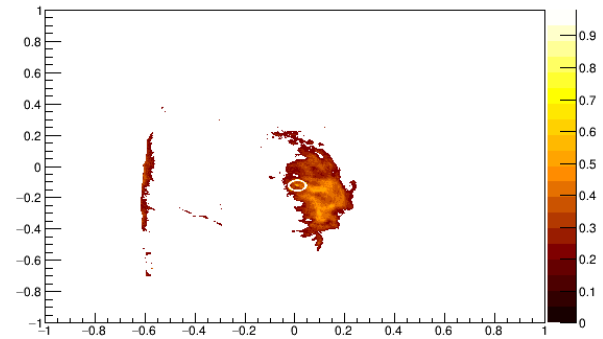
Figura 4.14: Resultados para la mamografía con ID mdb 213



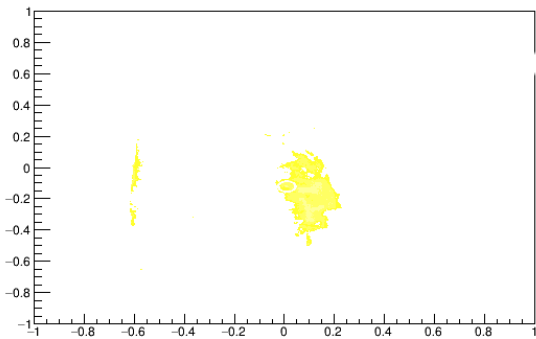
(a) Localización de la ROI en la mamografía original



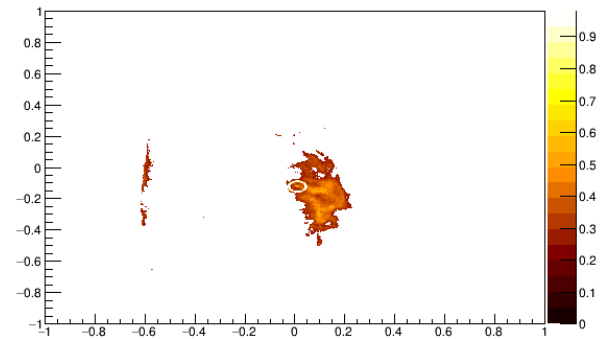
(b) ROI aislada para el valor umbral  $\mu + 3\sigma$



(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

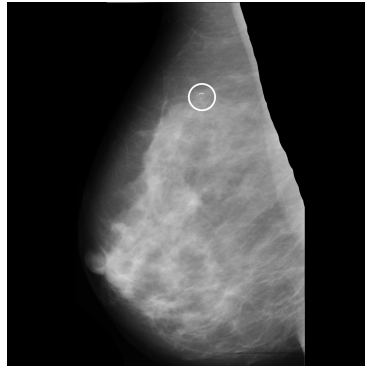


(d) ROI aislada para el valor umbral  $\mu + 4\sigma$

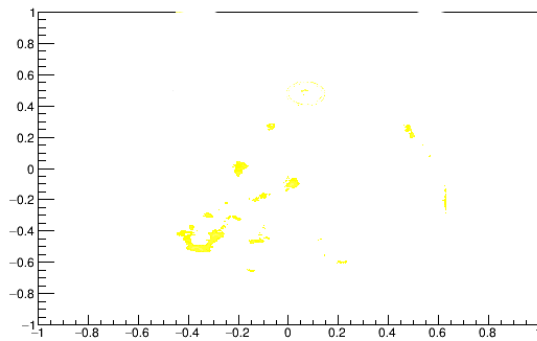


(e) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (d)

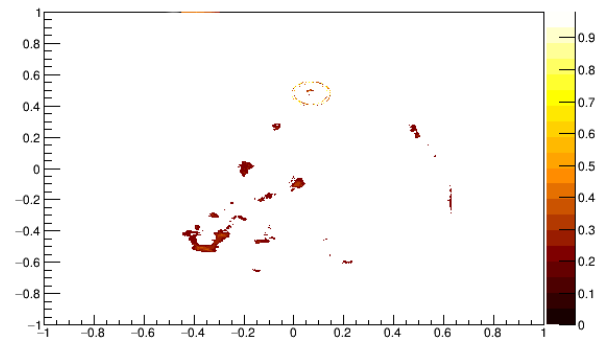
Figura 4.15: Resultados para la mamografía con ID mdb 218



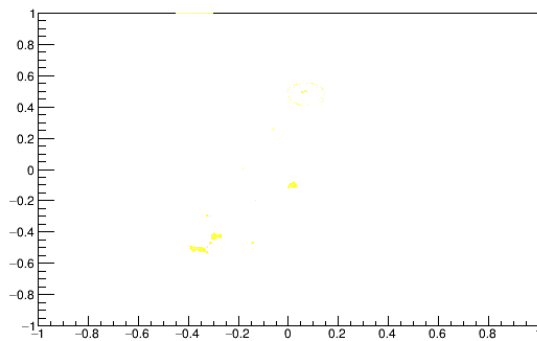
(a) Localización de la ROI en la mamografía original



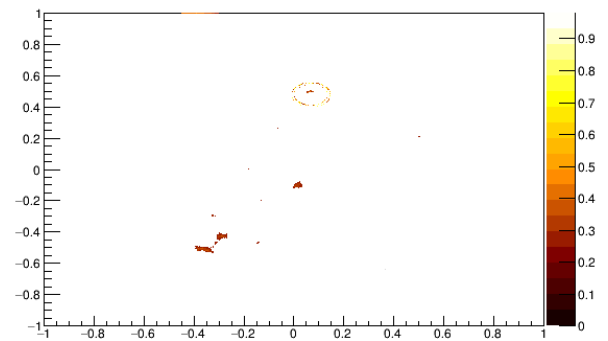
(b) ROI aislada para el valor umbral  $\mu + 3\sigma$



(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

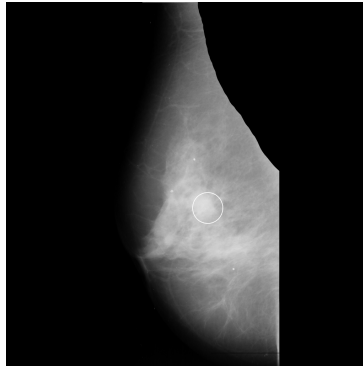


(d) ROI aislada para el valor umbral  $\mu + 4\sigma$

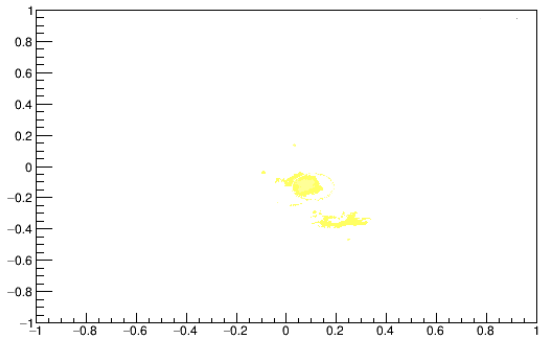


(e) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (d)

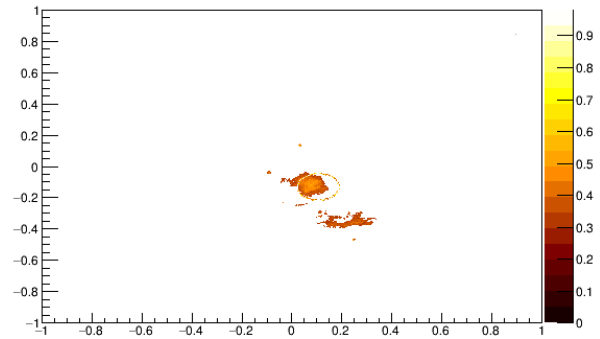
Figura 4.16: Resultados para la mamografía con ID mdb 219



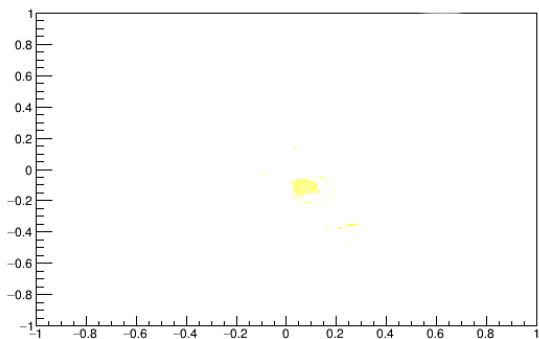
(a) Localización de la ROI en la mamografía original



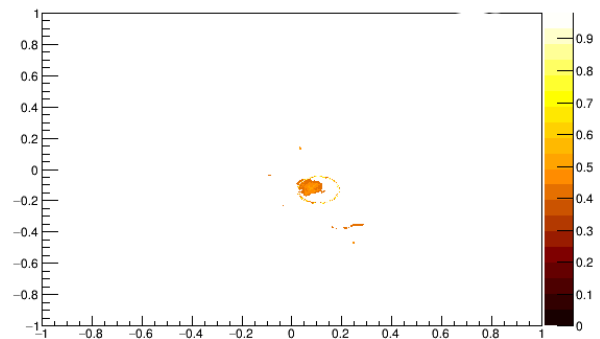
(b) ROI aislada para el valor umbral  $\mu + 3\sigma$



(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

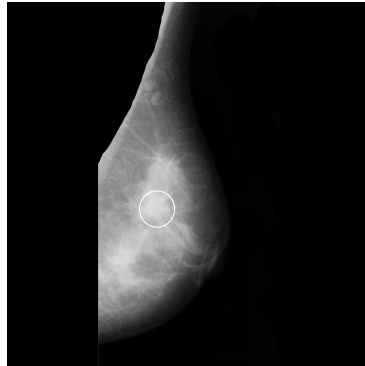


(d) ROI aislada para el valor umbral  $\mu + 4\sigma$

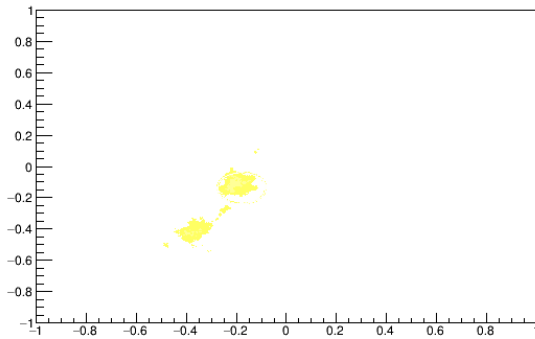


(e) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (d)

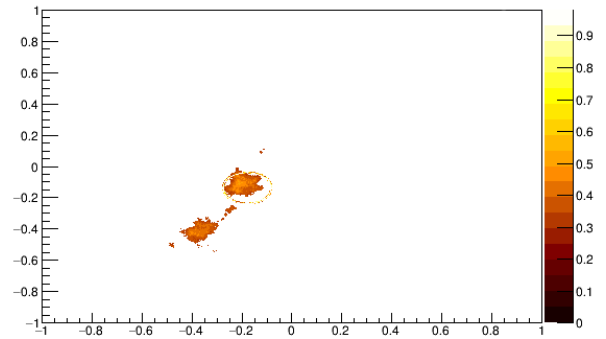
Figura 4.17: Resultados para la mamografía con ID mdb 063



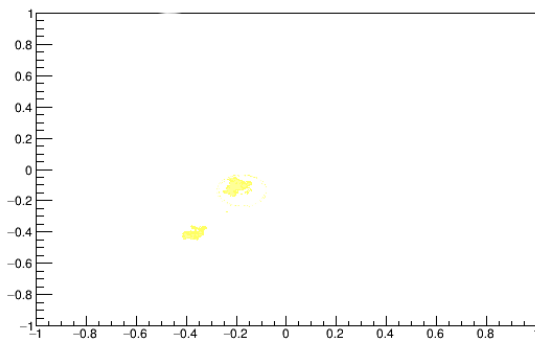
(a) Localización de la ROI en la mamografía original



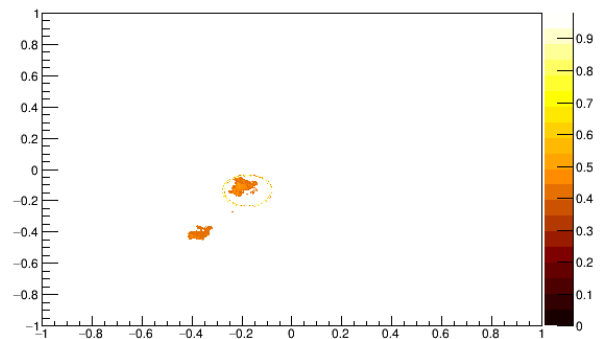
(b) ROI aislada para el valor umbral  $\mu + 3\sigma$



(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

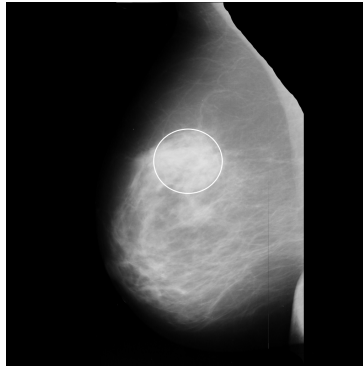


(d) ROI aislada para el valor umbral  $\mu + 4\sigma$

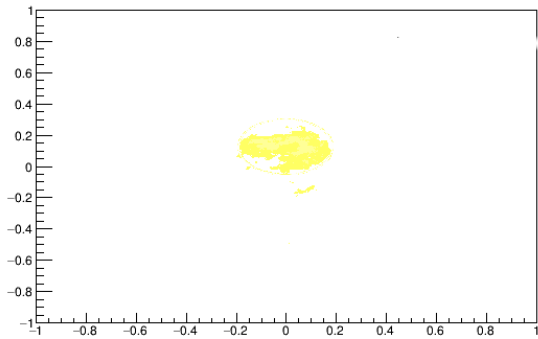


(e) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (d)

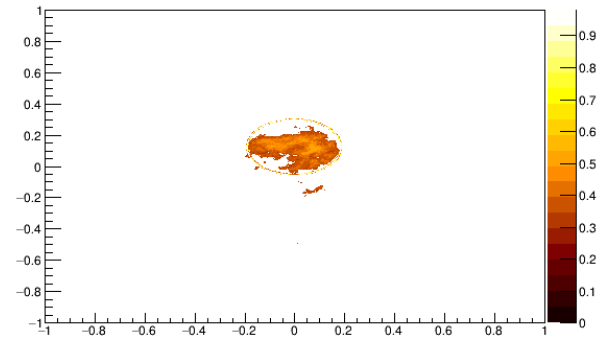
Figura 4.18: Resultados para la mamografía con ID mdb 102



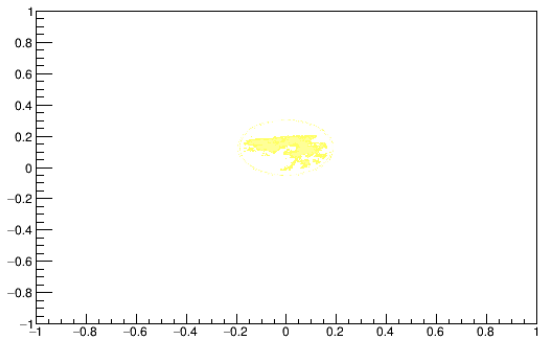
(a) Localización de la ROI en la mamografía original



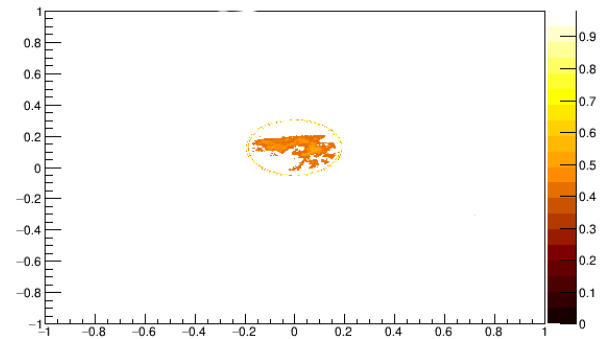
(b) ROI aislada para el valor umbral  $\mu + 3\sigma$



(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)



(d) ROI aislada para el valor umbral  $\mu + 4\sigma$

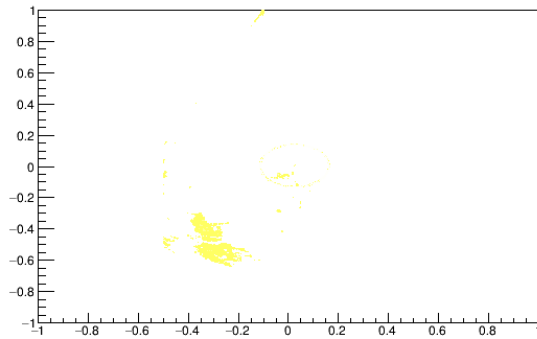


(e) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (d)

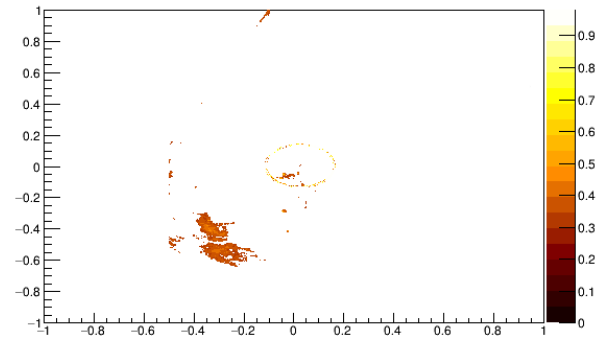
Figura 4.19: Resultados para la mamografía con ID mdb 111



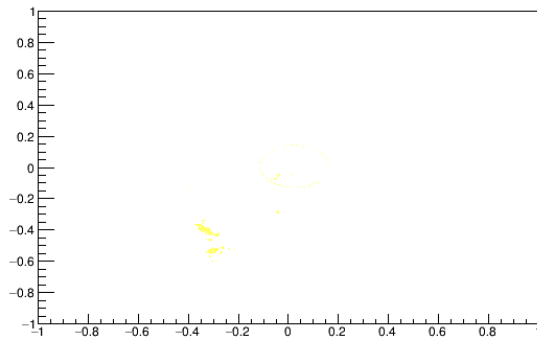
(a) Localización de la ROI en la mamografía original



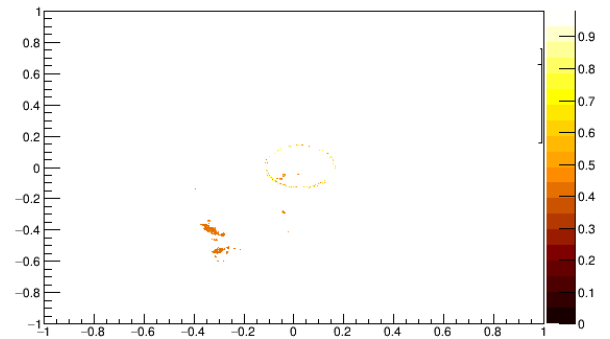
(b) ROI aislada para el valor umbral  $\mu + 3\sigma$



(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

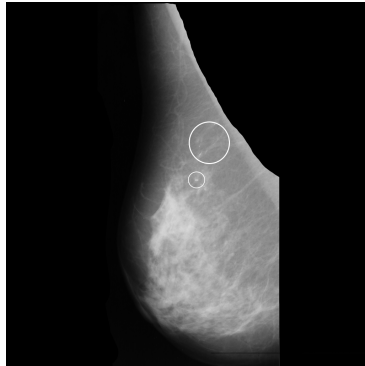


(d) ROI aislada para el valor umbral  $\mu + 4\sigma$

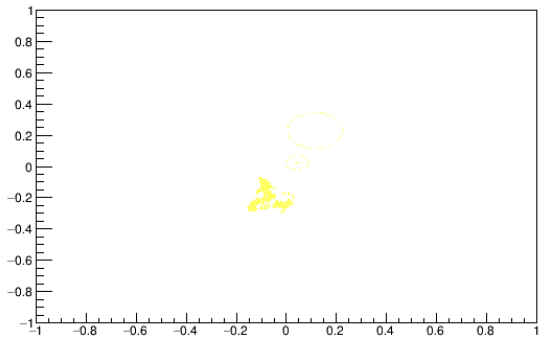


(e) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (d)

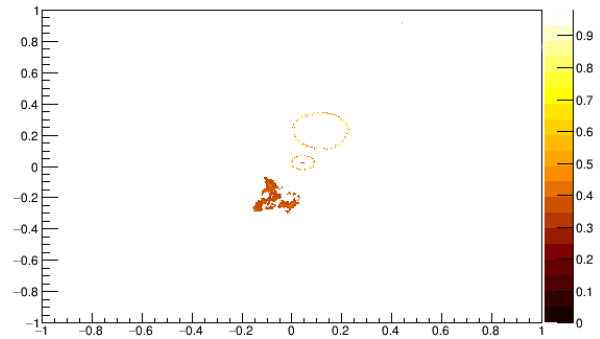
Figura 4.20: Resultados para la mamografía con ID mdb 170



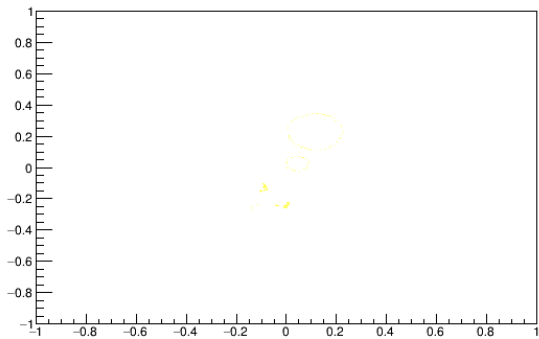
(a) Localización de la ROI en la mamografía original



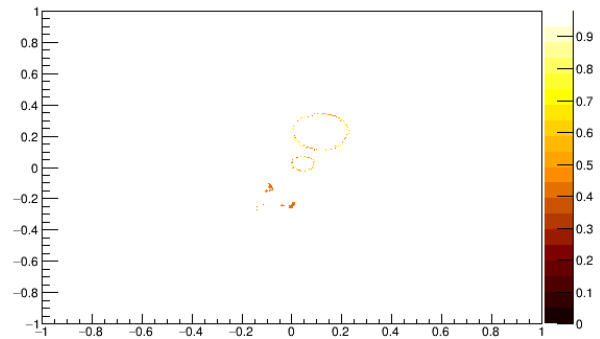
(b) ROI aislada para el valor umbral  $\mu + 3\sigma$



(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)



(d) ROI aislada para el valor umbral  $\mu + 4\sigma$



(e) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (d)

Figura 4.21: Resultados para la mamografía con ID mdb 249

- Los valores mínimos y máximos obtenidos muestran que para el tejido sano de los 3 tipos de muestra hay valores de intensidad que coinciden.

Finalmente, estos resultados muestran que la media y la desviación estándar del histograma de una región de tejido sano de la imagen mamográfica es un índice confiable de la densidad del tejido mamario al estudiar las distribuciones de intensidades de los MIH, no obstante, la figura 4.4 hace evidente que no se tiene un valor definido para separar cuantitativamente cada una de las muestras. Es por ello que según BIRADS, los tipos de tejido se clasifican de manera cualitativa teniendo categorías en las que se combinan las denominaciones graso-glandular y glandular-denso.

Respecto a los valores umbrales obtenidos, se tiene que para el tejido F, este puede no llegar a separar correctamente las zonas que presentan anomalías del tejido normal, teniendo que en algunos casos persisten regiones que no presentan lesiones luego de reconstruir la imagen, véase figura 4.6. No obstante, en la mayoría de los casos, la reconstrucción muestra correctamente los píxeles de la imagen que corresponden a anomalías. Para el caso de masas, estas se aprecian bien definidas y debido a su tamaño no es difícil identificarlas, no obstante, cuando se trata de microcalcificaciones, se tiene que posterior al corte en el valor umbral, en la RMI se aprecian claramente los píxeles que persisten y pertenecen a estas anomalías que eran prácticamente indistinguibles en la MI original. Para la posible obtención de mejores resultados podría calcularse un segundo candidato de valor umbral con el promedio de los valores  $\mu + 4\sigma$  para esta clasificación de tejido, este valor probablemente descartará mejor los píxeles de zonas sanas.

Para el tejido G, se tiene que el corte en el valor  $\rho_{G2}$  proporciona RMI con las zonas que presentan anomalías mejor delimitadas. En todos los casos el área de la lesión sigue presente en las reconstrucciones y al igual que para el tejido F, los píxeles que corresponden a microcalcificaciones son evidenciados en el resultado, especialmente en el que presenta mayor contraste. Sin embargo, no son las regiones con problema las únicas que se presentan luego del análisis, también se pueden visualizar regiones que según el diagnóstico son tejido normal.

Finalmente, respecto al tejido D se observó que no siempre las regiones de anomalías persisten luego de la redigitalización con valores que superan al valor umbral, pues para algunos casos, este supera las intensidades de los píxeles que corresponden a las lesiones, incluso para el valor  $\rho_{D1}$  cuando se trata de microcalcificaciones; las regiones que competen a masas y asimetrías de gran tamaño en general se aprecian en el resultado, junto con tejido sano de alta densidad.



## Capítulo 5

# Conclusiones

En este trabajo se buscó caracterizar cuantitativamente los diferentes tipos de tejido mamario sano analizando las distribuciones de intensidades obtenidas del procesamiento de una muestra de 90 mamografías a fin de posteriormente separarlo de las regiones que presentaran anomalías. Lo anterior haciendo uso del análisis de valores de intensidad que conforman la imagen.

El primer resultado importante del análisis, que es el que da las bases para llevar a cabo el mismo, se presenta en la sección de metodología pues no se trata del objetivo principal de estudio, y es que al transformar la MI original a un mapa de intensidades o histograma 2D, donde dos ejes proporcionan el sistema de coordenadas para la ubicación de los píxeles y para cada uno se tiene un valor de intensidad, se encontró que el píxel de mayor valor de intensidad en la RMI se ubica en el la región que el diagnóstico establece como anomalía, incluso para el caso de mamografías clasificadas como de tejido denso-glandular, en las que la ubicación de las lesiones puede ser complicada, es este hecho, presentado en el trabajo que es principal antecedente de este estudio [10], es el que nos lleva a pensar que la ROI se puede separar del tejido sano, pues la primera se puede ubicar en la mamografía encontrando el bin de mayor intensidad una vez que la región del músculo pectoral y otras regiones que entorpecen el análisis es eliminada.

Pensando en realizar un análisis estadístico de intensidad como método de discriminación entre el área b y las ROI se buscó caracterizar a la primera para luego excluirla manteniendo únicamente los valores de intensidad que no correspondieran a tejido sano. Esto fundamentado en el histograma de intensidades obtenido para las mamografías en el que fue posible distinguir entre el fondo de la imagen, el área de tejido de fondo y los píxeles de intensidad máxima. Siendo que la agrupación de píxeles correspondientes al tejido de fondo se apreciaba como una distribución gaussina se procedió a seleccionar regiones significativas del mismo para analizarlas por separado, conociendo la teoría de dicho tipo de distribuciones. Una vez obtenidos los parámetros ( $\mu$  y  $\sigma$ ) que describen los ajustes se encontró que los valores de intensidad máximo para las mamografías que presentan anomalías se ubicaron a más de  $3\sigma$  de las medias obtenidas para b (Tablas 3.10, 3.11 y 3.12) y este resultado es consistente sin importar la selección de tejido sano realizada. Por lo que se sugirió llevar a cabo el análisis de datos necesario para establecer un valor umbral de intensidad para mantener la ROI, lo que traería consigo la caracterización de cada tipo de tejido pues este procedimiento se llevo a cabo por separado en 30 mamografías de cada clasificación.

Al respecto de los resultados podemos decir que no se obtuvo un método cuantitativo eficaz para distinguir entre los tres diferentes tipos de parénquima mamario sano que en los que se encontraban clasificadas las mamografías de este estudio, pues se tiene que no es posible definir un valor de intensidad de píxel límite que separe cada modelo. En general, para una sola mamografía, se tienen píxeles que corresponden a cada tipo de tejido: F, G y D; pues las mamas presentan los tres tipos, siendo que se trata de estructuras complejas donde los tejidos fibroso y glandular estructuran los conductos, glándulas y tejidos conjuntivos y que estos se encuentran rodeados por una capa de grasa. Así, la mejor forma de clasificar las mamografías según su densidad es cualitati-

vamente utilizando el método BIRADS. No obstante, también se encontró que los parámetros que caracterizan a la distribución de intensidades obtenida de la mamografía, son una buena medida de la densidad del tejido estudiado. Si bien no nos ayudan a distinguir entre los distintos tipos, sí pueden darnos una idea de las regiones que presentan valores de intensidad mucho mayores que los esperados según las diferentes clasificaciones. Lo que lleva a las conclusiones de la siguiente parte del análisis: la separación de las ROI y el tejido de fondo.

Cada uno de los valores umbrales definidos presenta cierta exactitud, pues las redigitalizaciones que los toman en cuenta proporcionan buenos resultados respecto a la permanencia de los píxeles que corresponden a las lesiones, no obstante, no se logra aislar por completo la región de la anomalía, es decir, se tienen píxeles persistentes que pertenecen a tejido normal. Aun así, para el caso de lesiones que son difíciles de identificar a simple vista, como lo son las microcalcificaciones, se tiene que este método presenta los píxeles que las identifican, pudiendo ser de ayuda como herramienta para localizar estos bins de interés realizando la redigitalización considerando el valor umbral y posteriormente elevando los valores de intensidad a alguna potencia. Como tal, el objetivo del estudio era la localización de ROI para el diagnóstico, no la definición correcta de su forma, por ejemplo, por lo que se considera que la propuesta realizada cumple con el objetivo.

A fin de mejorar los resultados de este análisis, un trabajo futuro podría consistir en examinar un mayor número de mamografías, pues un número pequeño de nuestras no conlleva a resultados estadísticos significativos. Además debe buscarse eliminar la limitación que representa el hecho de que el área de tejido a estudiar se seleccione con una figura tan elemental como un rectángulo, pues una definición más suave del área del estudio daría pie a menor pérdida de información, para ello podrían utilizarse métodos de procesamiento digital de imágenes a fin de llevar a cabo la extracción del área total del seno y que esta, sin anomalías, se defina como la región de estudio, es decir que se analice la intensidad de todos los píxeles que conforman la imagen de la mama, de esta manera se tendrá mayor exactitud para definir los valores umbral y por ende la separación entre ROI y tejido sano será más confiable. También hace falta llevar a cabo la verificación de los valores umbrales obtenidos aplicándolos a mamografías de tipo normal con el objetivo de analizar los falsos positivos.

Por último, es importante señalar que para la utilización de este método, el formato de imagen de la mamografía original no es tan relevante pues al transformar la misma a un mapa de escala de grises, sólo se trabaja con la información de la intensidad, no obstante, sería interesante realizar este procedimiento a una diferente muestra de estudio, es decir, llevar a cabo una recopilación de mamografías con una mayor cantidad de datos descriptivos y especificaciones e información sobre el diagnóstico médico de las pacientes, tomando en cuenta factores importantes como la edad y la etapa de desarrollo de las anomalías asegurando también que la obtención de estos exámenes sea llevada a cabo bajo las mismas condiciones y parámetros, para lo anterior se requiere de una colaboración con algún tipo de institución de salud, lo que no fue posible en este trabajo debido a la situación de la pandemia mundial de Covid-19 que se vivió durante su realización, es por ello que se utilizaron los recursos con los que se contaban para la conclusión de la misma, siendo que este análisis no habría sido posible sin las mamografías disponibles para libre acceso en MINIMIAS por lo que se menciona a dicha base de datos en los agradecimientos.

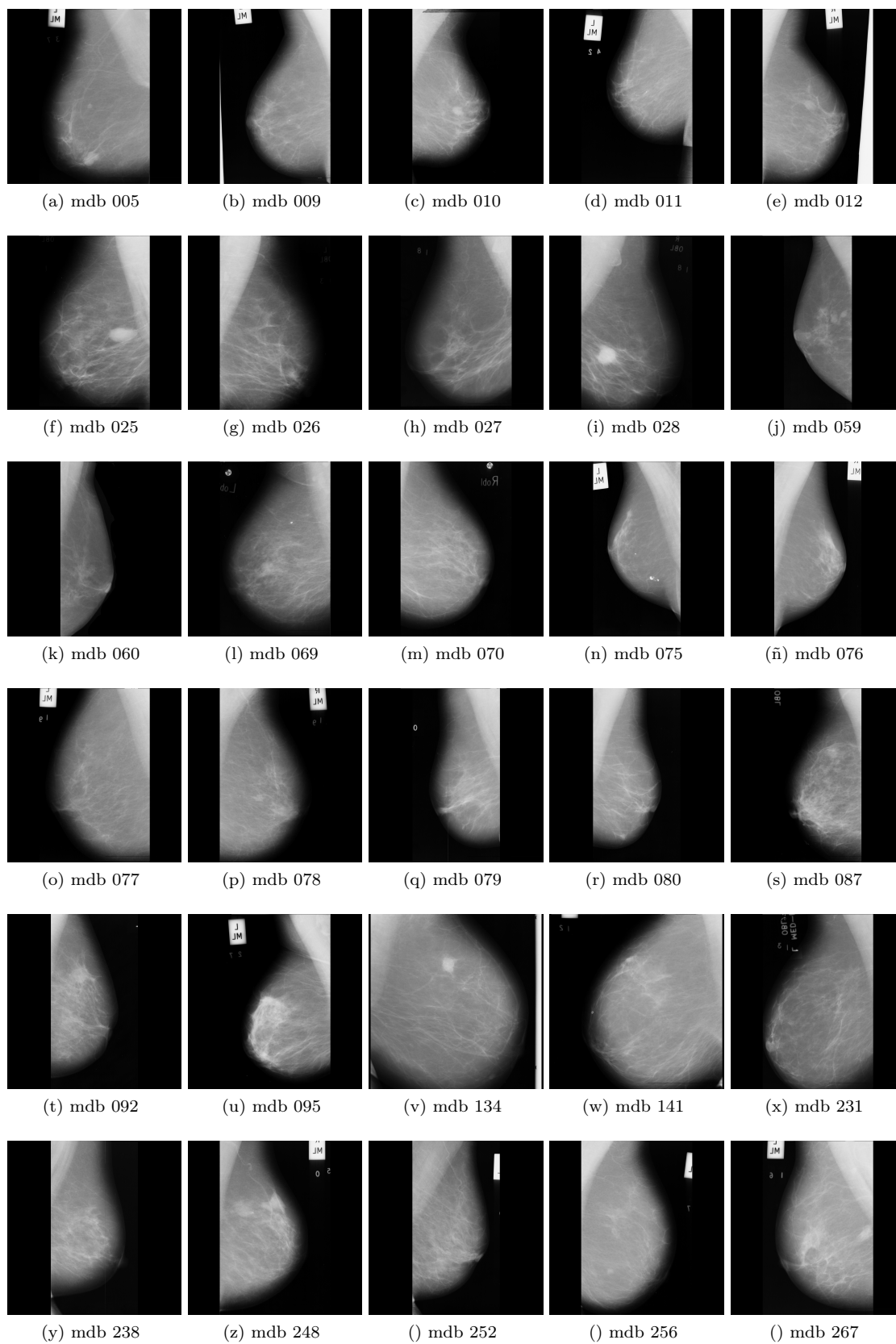
## Apéndice A

### Imágenes obtenidas de la base de datos mini-MIAS empleadas para el estudio

- A.1. Muestra de tejido F
- A.2. Muestra de tejido G
- A.3. Muestra de tejido D

APÉNDICE A. IMÁGENES OBTENIDAS DE LA BASE DE DATOS MINI-MIAS  
EMPLEADAS PARA EL ESTUDIO  
A.3. MUESTRA DE TEJIDO D

---



---

Figura A.1: Imágenes recopiladas para la muestra de tejido F

APÉNDICE A. IMÁGENES OBTENIDAS DE LA BASE DE DATOS MINI-MIAS  
 EMPLEADAS PARA EL ESTUDIO  
 A.3. MUESTRA DE TEJIDO D

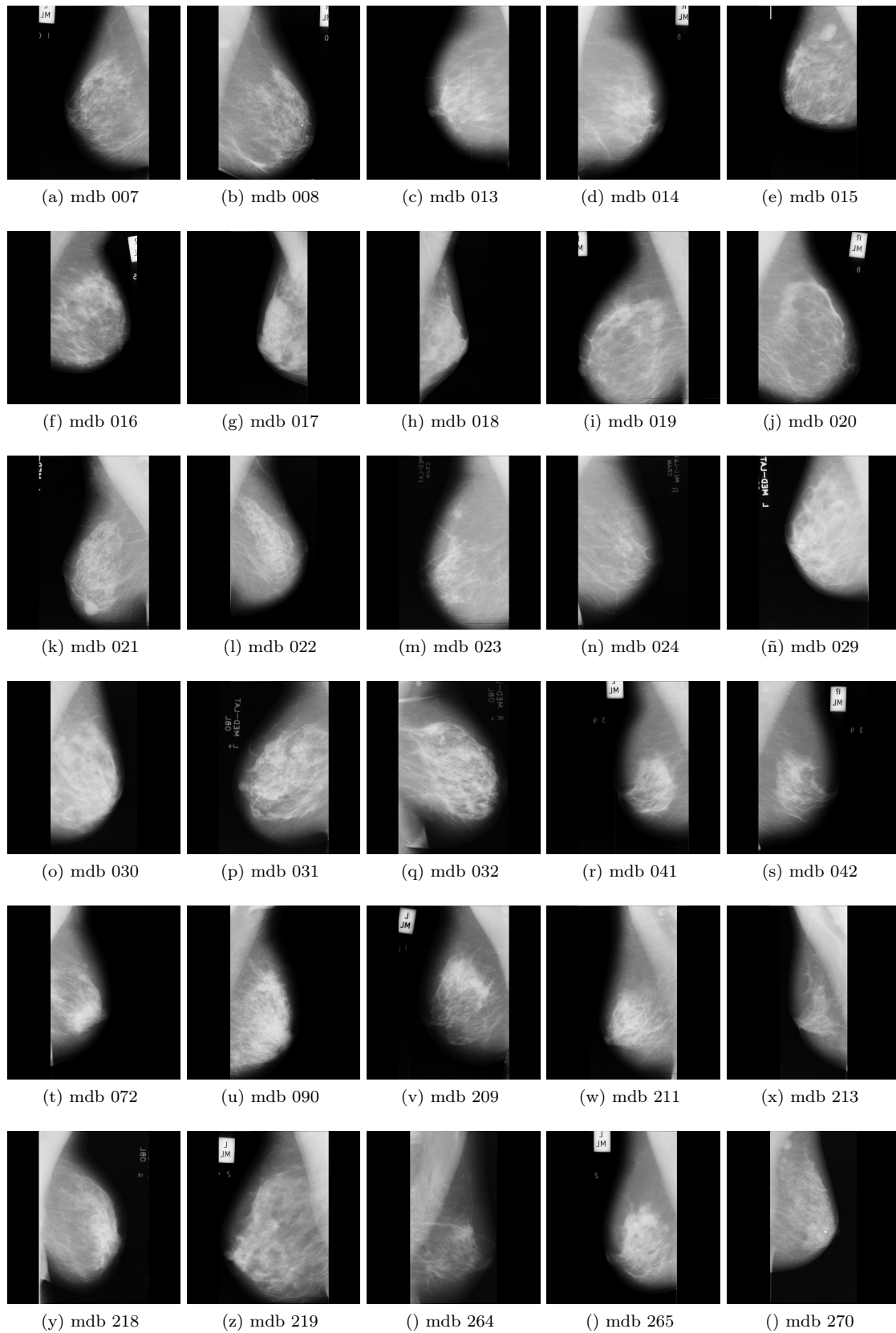


Figura A.2: Imágenes recopiladas para la muestra de tejido G

APÉNDICE A. IMÁGENES OBTENIDAS DE LA BASE DE DATOS MINI-MIAS  
EMPLEADAS PARA EL ESTUDIO  
A.3. MUESTRA DE TEJIDO D

---

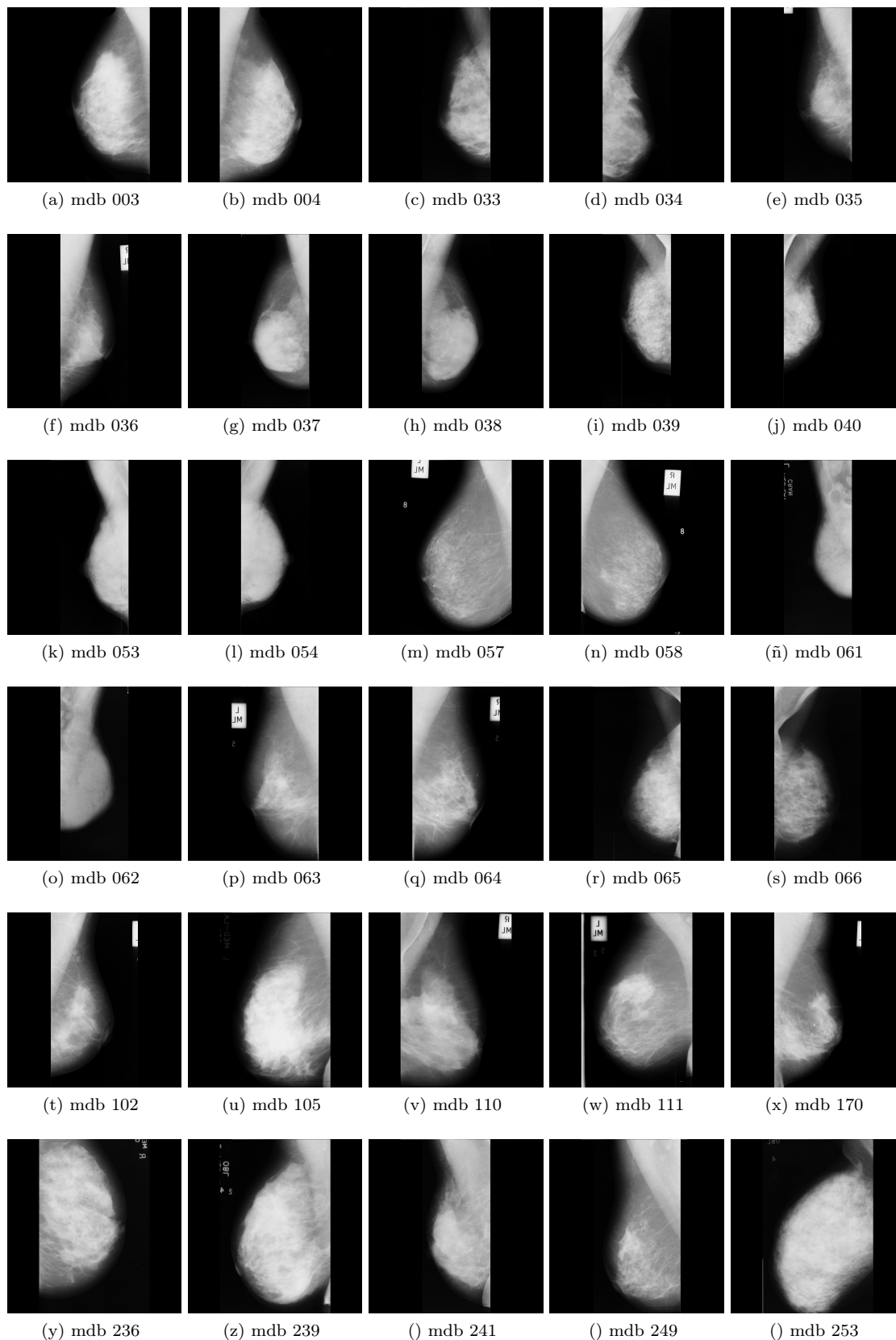


Figura A.3: Imágenes recopiladas para la muestra de tejido D

## Apéndice B

# Códigos en ROOT

### B.1. Código para la conversión de imágenes a mapas de colores

//PROGRAMA PARA CONVERTIR LAS IMAGENES EN FORMATO .PNG A MAPAS DE NIVELES DE GRISES CON LAS INTENSIDADES NORMALIZADAS, REESCALADOS A UNA MATRIZ CUADRADA A LA QUE SE LE IMPLEMENTO UN SISTEMA COORDENADA PARA FACILITAR LA UBICACIÓN DE LOS PÍXELES, CON EL FIN DE PRESENTAR LA INFORMACIÓN DE UBICACIÓN E INTENSIDAD DEL BIN DE MAYOR BRILLO Y FINALMENTE GENERAR UN ARCHIVO DE TEXTO CON TODA LA INFORMACIÓN DE LA IMAGEN ORIGINAL.

```
#include "TCanvas.h"
#include "TH1.h"
#include "TCanvas.h"
#include "TAxis.h"
#include "TStyle.h"
#include "TSystem.h"
#include <string.h>
#include <iostream>
#include <ostream>
#include <cstdio>
#include <fstream>
using namespace std;
void ImageToHisto2()

{

////////////////////
//////// Initial parameters //////////
////////////////////

//Number of boxes in X and Y axis
const Int_t boxes=3;
// Input Image
TASImage image("mdb-0XX.png");
```

## APÉNDICE B. CÓDIGOS EN ROOT

### B.1. CÓDIGO PARA LA CONVERSIÓN DE IMÁGENES A MAPAS DE COLORES

---

```
////////////////////////////////////
//////// Image digitalization //////////
////////////////////////////////////

const UInt_t yPixels = image.GetHeight();
const UInt_t xPixels = image.GetWidth();
UInt_t *argb = image.GetArgbArray();

Double_t X[xPixels];
Double_t Y[yPixels];

const Int_t n = yPixels*xPixels;
cout<<xPixels<<"_"<<yPixels<<endl;
Float_t Grey[xPixels][yPixels];

X[0]=-1;
Y[0]=1;

Double_t XX,YY;

Double_t incx=2.0/xPixels;
Double_t incy=2.0/yPixels;

TH2D* h = new TH2D( "h","Mammography_digitalization", xPixels, -1, 1, yPixels, -1, 1 );
Double_t maxA=0.;
Double_t minA=50.;

//Creation of the text file

std::ofstream outFile("Info_mdb0XX.dat",std::ios::app);

////////////////////////////////////
//////// Construction of grid //////////
////////////////////////////////////
for (int j=1;j<xPixels; j++)
{
    X[j]=X[0]+j*incx;
    Y[j]=Y[0]-j*incy;
}
Double_t cordx;
Double_t cordy;
Double_t x,y;

for (int row=0; row<yPixels; row++)
{
    for (int col=0; col<yPixels; col++)
    {
        int index = col*xPixels+row;
        float grey = float (argb[index]&0xff)/256;
        if (grey>0)
        {
```

```

h->SetBinContent(row, yPixels-col, pow(grey,1));
    }
    Grey[row][col]=grey;

    if(Grey[row][col]>maxA)
    {
        maxA=Grey[row][col];
        cordx=row;
        cordy=col;
        x=X[row];
        y=Y[col];
    }
    if(Grey[row][col]<minA)
    {
        minA=Grey[row][col];
    }
    //Highest Intensity bin results
    outFile<<X[row]<<" "<<Y[col]<<" "<<Grey[row][col]<<endl;
}
}
cout<<"Total_Segment_:_"<<cordx<<" "<<cordy<<" "<<maxA<<" "<<x<<" "<<y<<endl;

//Histogram Creation

TCanvas *c1 = new TCanvas("c1", "reconstruction", 0, 0, 600, 400);

gStyle->SetPalette(53);
h->Draw("COLZ");
Double_t binmax = h->GetMaximum();

return 0;
}

```

Este código me fue proporcionado por el Dr. Cristian Heber Zepeda Fernández

## B.2. Código para la creación de histogramas de frecuencia de intensidades

//SE LLEVA A CABO LA CREACIÓN DEL HISTOGRAMA DE INTENSIDADES CON LOS DATOS EXPORTADOS DEL ARCHIVO DE TEXTO GENERADO PARA CADA MAMOGRAFÍA, PARA LLEVAR A CABO EL ANÁLISIS DE TEJIDO SANO SE DELIMITA LA REGIÓN A ESTUDIAR MEDIANTE UN CUADRADO, DEFINIENDO XMAX, XMIN, YMAX Y YMIN, ÚNICAMENTE LA INFORMACIÓN DE LA REGIÓN CORRESPONDIENTE SE PRESENTARÁ EN EL HISTOGRAMA. ADEMÁS SE AGREGAN LAS LÍNEAS REQUERIDAS PARA QUE ROOT REALICE EL AJUSTE GAUSSIANO.

```
{
```

**APÉNDICE B. CÓDIGOS EN ROOT**  
B.2. CÓDIGO PARA LA CREACIÓN DE HISTOGRAMAS DE FRECUENCIA DE  
INTENSIDADES

---

```
TH1F*h = new TH1F("data","Histograma de intensidades",100,0,1);

//Se exportan los datos del archivo de texto de cada mamografía

ifstream File("Info_mdb0XX.dat");
TCanvas*c1 = new TCanvas("Histograma de intensidades","Histograma",600,600);

//Se definen las 3 variables de interes: posicion en X, en Y e intensidad I

Double_t x,y,I;

//Se delimita el rea de tejido del cual se obtendr el
histograma

Double_t xmin=XMIN;
Double_t xmax=XMAX;
Double_t ymin=YMIN;
Double_t ymax=YMAX;
Double_t Inffs=0;
Double_t Maxfs=1;

//Se lleva a cabo el llenado del histograma

while(File>>x && File>>y && File>>I)
{
    if (I>0 && xmin<x && x<xmax && ymin<y && y<ymax){

        h->Fill(I);
    }
}

//Presentaci n del histograma

h->GetXaxis()->SetTitle("Intensidad");
h->GetXaxis()->SetTitleSize(0.06);
h->GetXaxis()->CenterTitle();

h->SetTitle("Histograma de intensidades");
h->GetYaxis()->SetTitle("Frecuencia");
h->GetYaxis()->SetTitleSize(0.05);
h->GetYaxis()->CenterTitle();

//Se realiza el ajuste de tipo gaussiano al histograma

Double_t par[3];
TF1 *g1= new TF1("g1","gaus", Inffs , Maxfs);
TF1 *total= new TF1("total", "gaus(0)",Inffs , Maxfs);
g1->SetLineWidth(4);
g1->SetParName(0,"constant");
g1->SetParName(1,"mean");
g1->SetParName(2,"sigma");
gStyle->SetOptFit(111);
```

```

total->SetParameters(par);
total->SetLineColor(kRed);
total->SetLineWidth(3);
h->SetStats(0);
h->Fit(total,"R+");
h->Fit("total");
h->Fit(g1,"R");
h->Draw("");

g1->GetParameters(par);

g1->GetParameters(&par[0]);

Float_t sigma = par[2];
Float_t mean = par[1];
Double_t Chi2 = g1->GetChisquare();
Int_t Entries = h->GetEntries();
Double_t Errorsigma = g1->GetParError(2);
Double_t Errormean = g1->GetParError(1);
Double_t NDF = g1->GetNDF();
Double_t Chi2NDF = Chi2/NDF;

cout<<Chi2<<" " <<NDF<<" " <<Chisquare/ndf: " <<Chi2NDF<<endl;

//Se presenta la leyenda del ajuste

TLegend *leg=new TLegend(0.6,0.65,0.88,0.85,"");
leg->SetTextFont(72);
leg->SetTextSize(0.03);

leg->AddEntry(total,"Fit Gaussiano","l");
leg->AddEntry((TObject*)0, Form("Entries= %a ",Entries), "");
leg->AddEntry((TObject*)0, Form("#sigma= %f #pm %f ",sigma,Errorsigma), "");
leg->AddEntry((TObject*)0, Form("mean= %f #pm %f ",mean,Errormean), "");
leg->AddEntry((TObject*)0, Form("#Chi^2/ndf= %f ",Chi2NDF), "");
leg->Draw();
}

```



## Apéndice C

# Determinación del área debajo de la curva de Gauss

Para fines estadísticos, es importante determinar el área bajo la curva de Gauss dentro de los límites que se definan [19]. Esto dado que, el área bajo la curva entre un par de límites establece la probabilidad de un valor medido de ser encontrado en ese intervalo. La ecuación 2.3 describe la curva de Gauss en términos de la media poblacional  $\mu$  y de la desviación estándar  $\sigma$ . En términos de la variable  $z$  se tiene:

$$y = \frac{e^{-\frac{z^2}{2}}}{\sigma\sqrt{2\pi}} \quad (\text{C.1})$$

Para conocer el área bajo la curva a un intervalo de confianza de la media, es decir, entre  $-1\sigma$  y  $+1\sigma$  de  $\mu$  ( $\mu - \sigma$  y  $\mu + \sigma$ ) se procede a utilizar el cálculo porque se sabe que la integración de una ecuación determina el área bajo la curva descrita por la ecuación misma. Así, se tiene:

$$A = \int_{-\sigma}^{+\sigma} \frac{e^{(-x-\mu)^2/2\sigma^2}}{\sigma\sqrt{2\pi}} dx \quad (\text{C.2})$$

Utilizando la ecuación anterior en términos de la variable  $z$  se tiene:

$$A = \int_{-1}^{1} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz \quad (\text{C.3})$$

Puesto que no existe una forma cerrada de la solución, la integral debe ser evaluada numéricamente. El resultado es

$$A = \int_{-1}^{1} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz = 0,683 \quad (\text{C.4})$$

Ahora bien, si se busca conocer el área bajo la curva de Gauss a  $2\sigma$  a cada lado de la media, se debe evaluar la siguiente integral

$$A = \int_{-2}^{2} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz = 0,954 \quad (\text{C.5})$$

Y para  $\pm 3\sigma$ :

$$A = \int_{-3}^{3} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz = 0,997 \quad (\text{C.6})$$

## APÉNDICE C. DETERMINACIÓN DEL ÁREA DEBAJO DE LA CURVA DE GAUSS

---

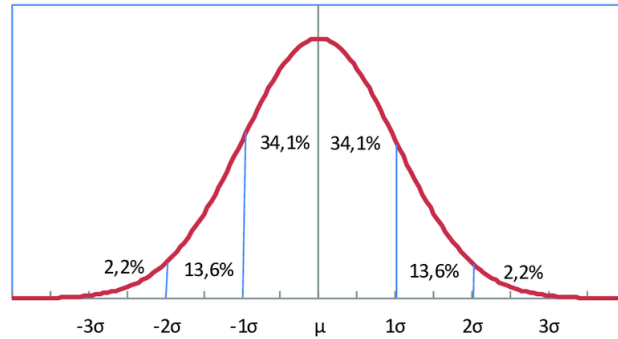


Figura C.1: Área bajo la curva de Gauss.

De ResearchGate , por Alfredo Baronio ([https://www.researchgate.net/figure/Figura-213-Areas-bajo-la-curva-normal\\_fig8\\_317800983](https://www.researchgate.net/figure/Figura-213-Areas-bajo-la-curva-normal_fig8_317800983))

A partir de estas integrales se concluye que las áreas bajo la curva de Gauss de una, dos y tres desviaciones estándar de la media son 68.3, 95.4 y 99.7 %, respectivamente, del área total bajo la curva. La figura muestra las diferentes áreas bajo la curva calculadas.

# Apéndice D

## Resultados

### D.1. Tejido F

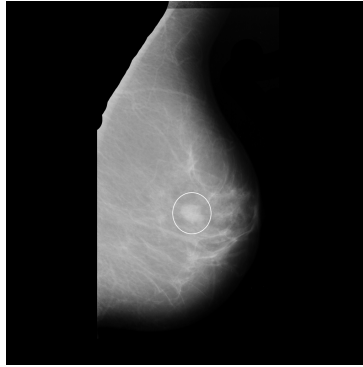
Se presentan los resultados de la siguiente lista de MI.

- Mamografía: mdb 010. Presenta CIRC Benigna. (Figura D.1)
- Mamografía: mdb 075. Presenta ASYM Maligna. (Figura D.2)
- Mamografía: mdb 080. Presenta CIRC Benigna. (Figura D.3)
- Mamografía: mdb 092. Presenta ASYM Maligna. (Figura D.4)
- Mamografía: mdb 095. Presenta ASYM Maligna. (Figura D.5)
- Mamografía: mdb 141. Presenta CIRC Maligna. (Figura D.6)
- Mamografía: mdb 238. Presenta CALC Maligna. (Figura D.7)
- Mamografía: mdb 248. Presenta CALC Benigna. (Figura D.8)
- Mamografía: mdb 252. Presenta CALC Benigna. (Figura D.9)
- Mamografía: mdb 256. Presenta CALC Maligna. (Figura D.10)
- Mamografía: mdb 267. Presenta MISC Maligna. (Figura D.11)

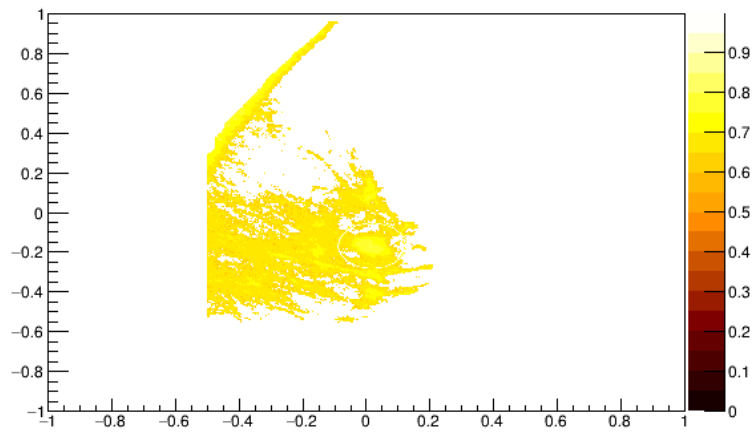
### D.2. Tejido G

Se presentan los resultados de la siguiente lista de MI.

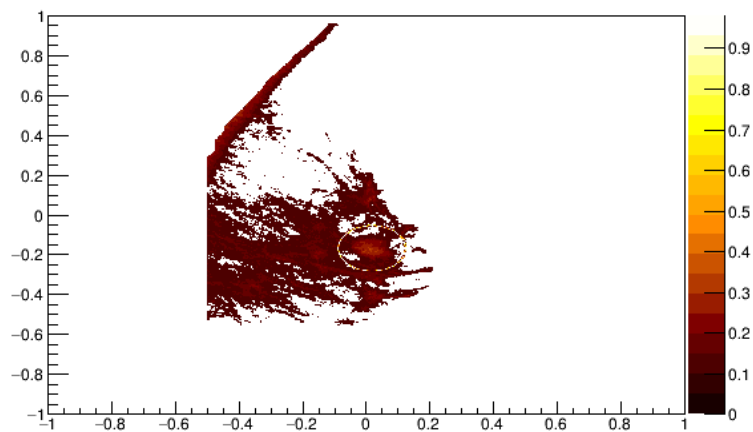
- Mamografía: mdb 013. Presenta MISC Benigna. (Figura D.12)
- Mamografía: mdb 017. Presenta CIRC Benigna. (Figura D.13)
- Mamografía: mdb 019. Presenta CIRC Benigna. (Figura D.14)
- Mamografía: mdb 021. Presenta CIRC Benigna. (Figura D.15)
- Mamografía: mdb 023. Presenta CIRC Maligna. (Figura D.16)
- Mamografía: mdb 030. Presenta MISC Benigna. (Figura D.17)



(a) Localización de la ROI en la mamografía original

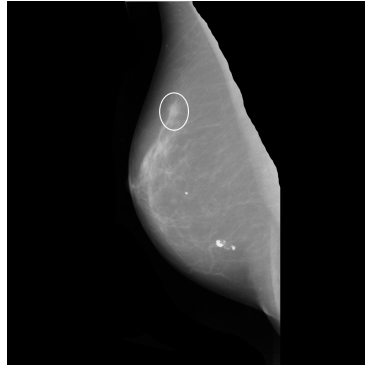


(b) ROI aislada para el valor umbral  $\mu + 3\sigma$

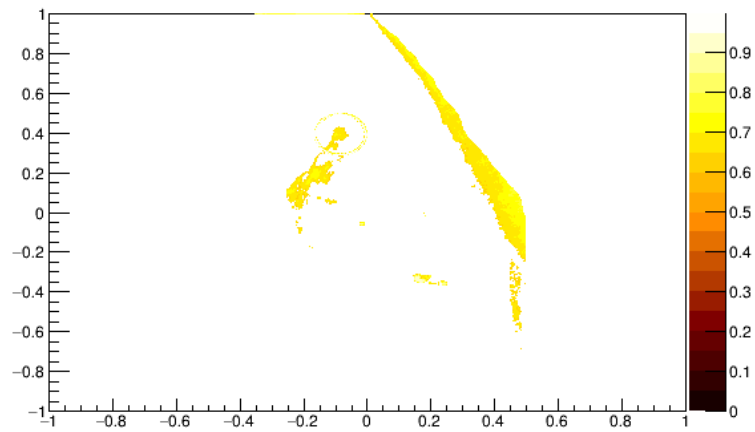


(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

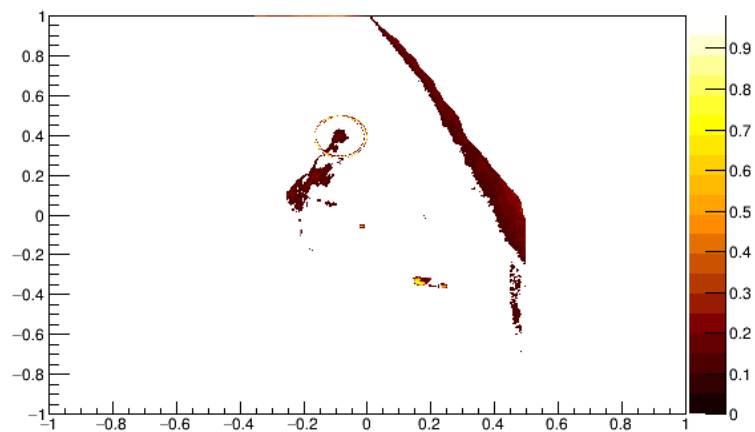
Figura D.1: Resultados para la mamografía con ID mdb 010



(a) Localización de la ROI en la mamografía original

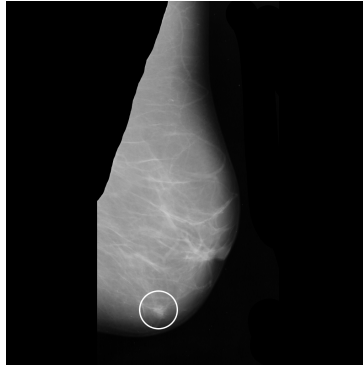


(b) ROI aislada para el valor umbral  $\mu + 3\sigma$

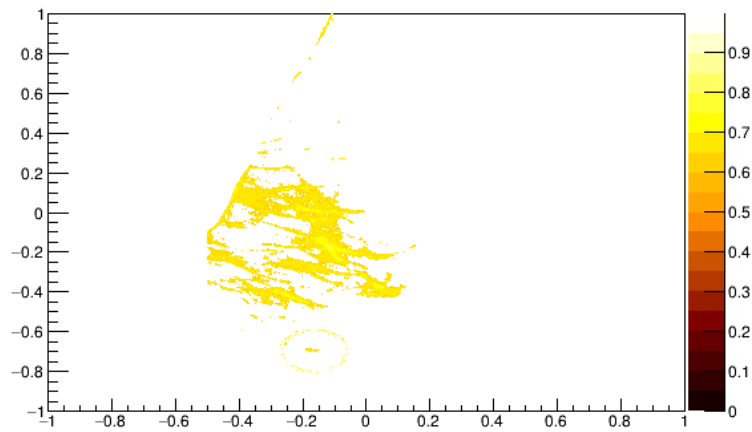


(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

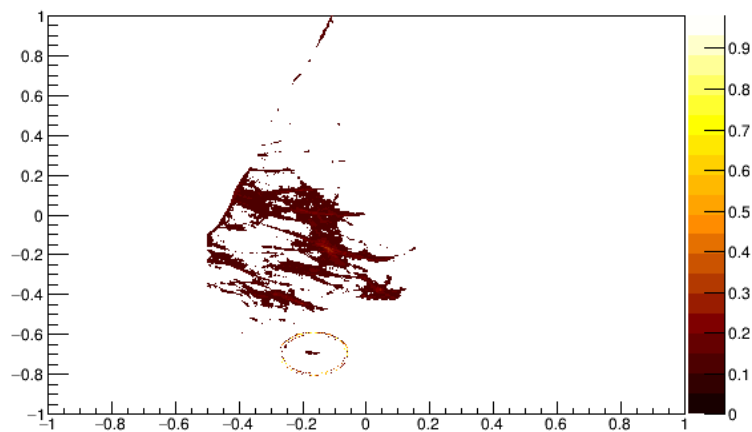
Figura D.2: Resultados para la mamografía con ID mdb 075



(a) Localización de la ROI en la mamografía original

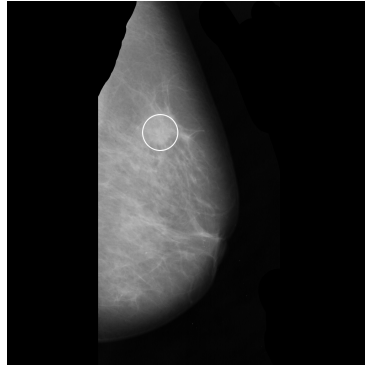


(b) ROI aislada para el valor umbral  $\mu + 3\sigma$

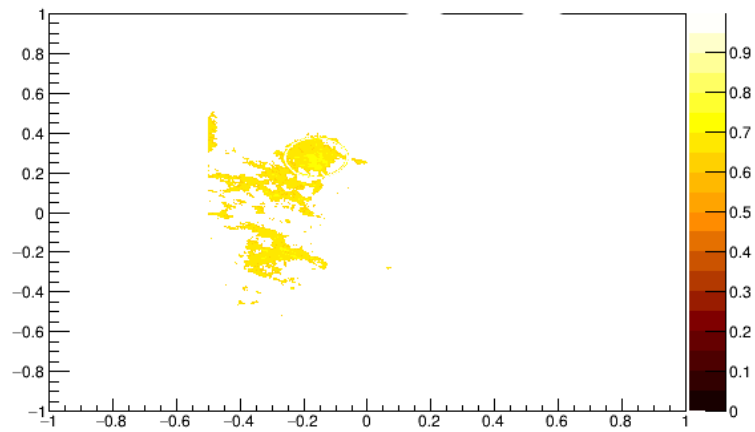


(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

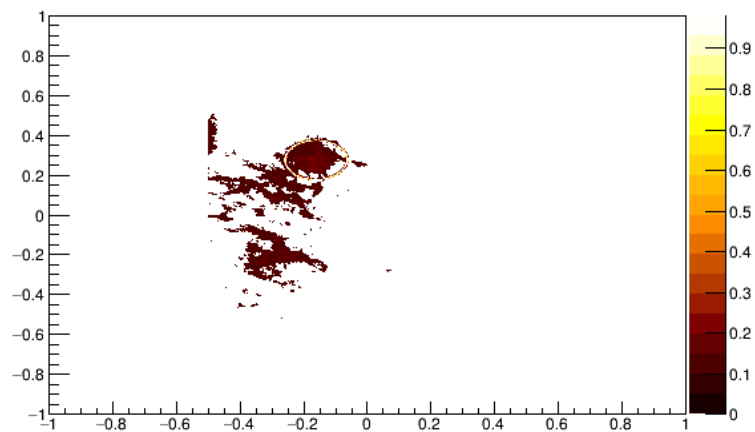
Figura D.3: Resultados para la mamografía con ID mdb 080



(a) Localización de la ROI en la mamografía original



(b) ROI aislada para el valor umbral  $\mu + 3\sigma$

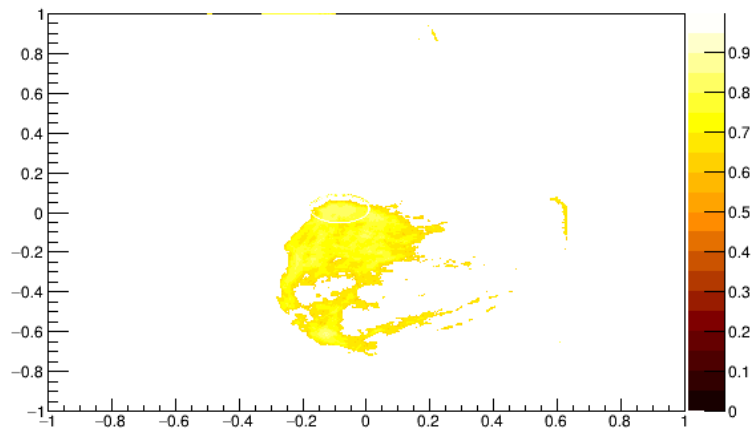


(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

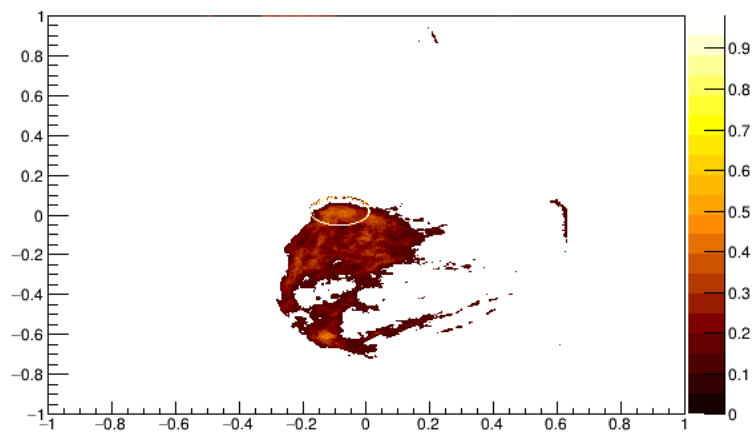
Figura D.4: Resultados para la mamografía con ID mdb 092



(a) Localización de la ROI en la mamografía original

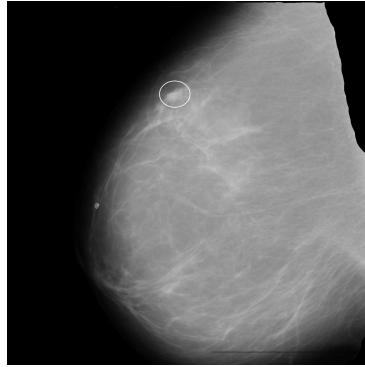


(b) ROI aislada para el valor umbral  $\mu + 3\sigma$

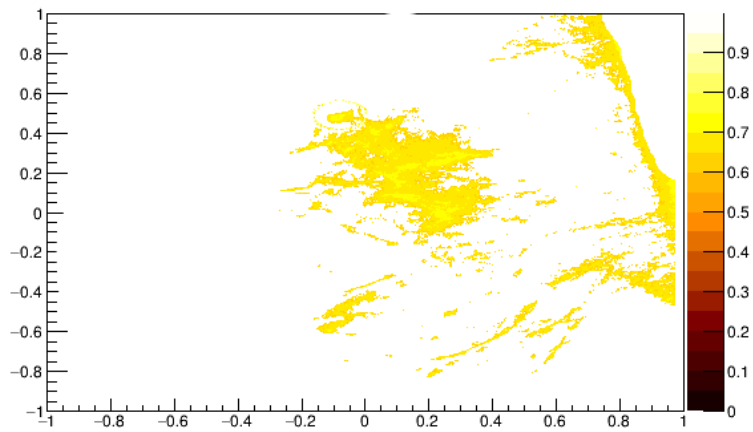


(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

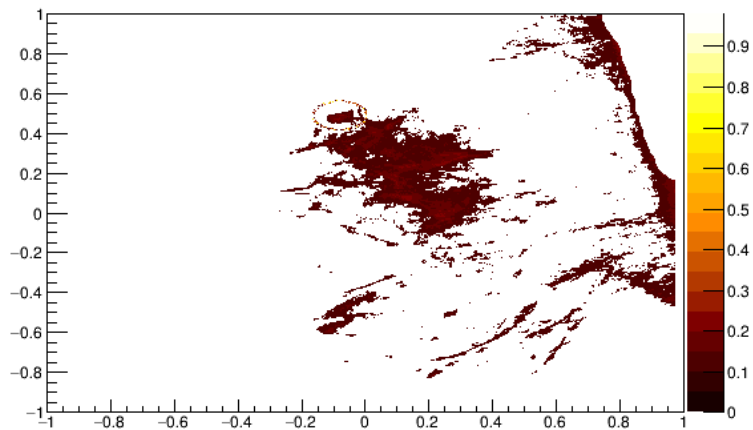
Figura D.5: Resultados para la mamografía con ID mdb 095



(a) Localización de la ROI en la mamografía original

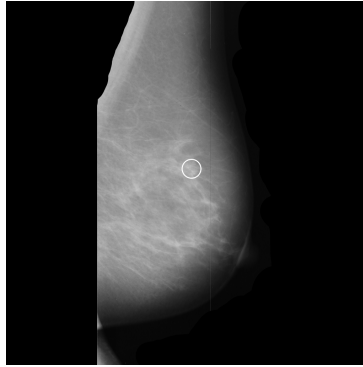


(b) ROI aislada para el valor umbral  $\mu + 3\sigma$

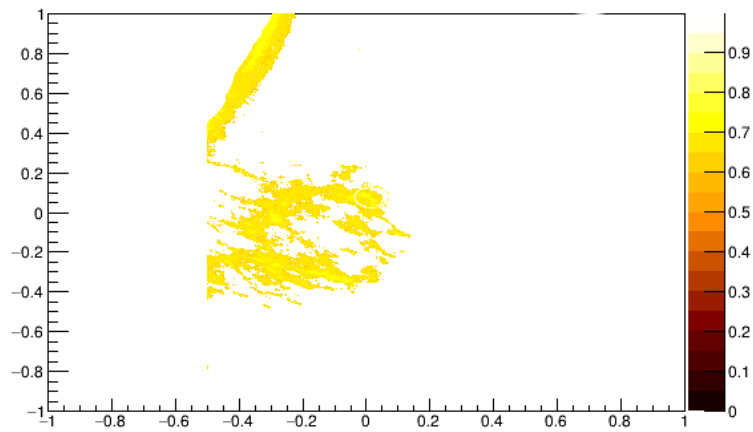


(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

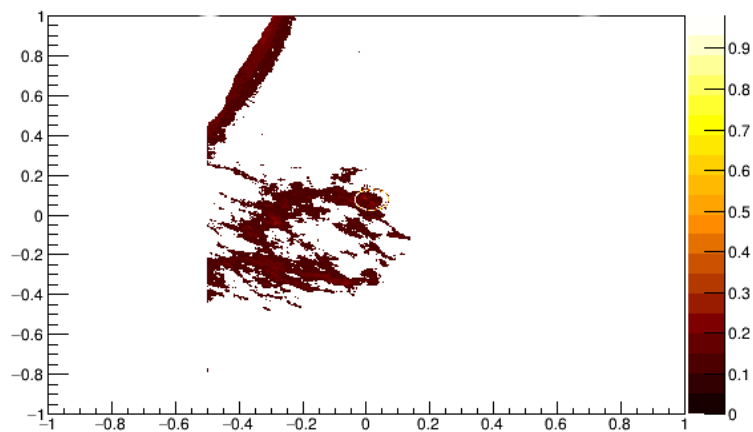
Figura D.6: Resultados para la mamografía con ID mdb 141



(a) Localización de la ROI en la mamografía original

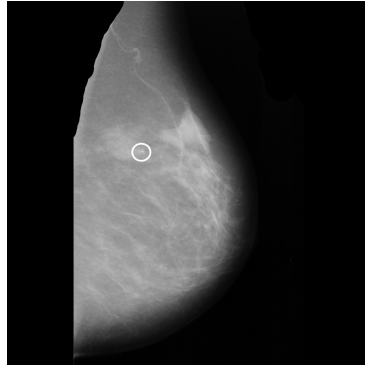


(b) ROI aislada para el valor umbral  $\mu + 3\sigma$

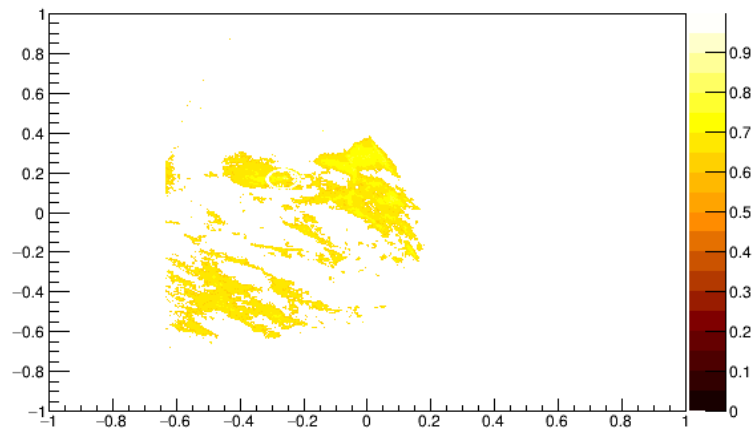


(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

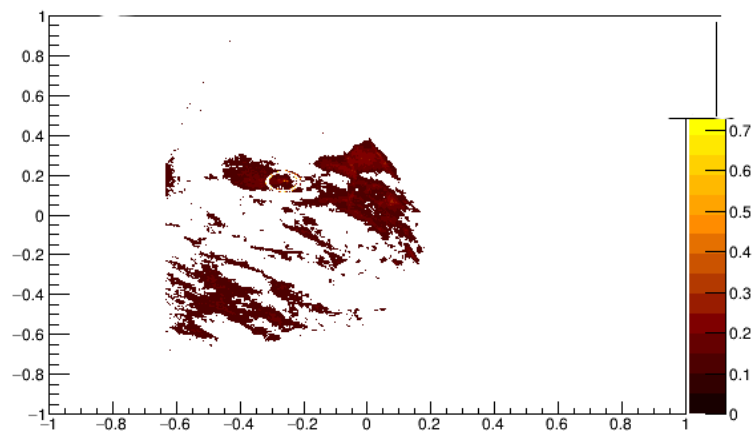
Figura D.7: Resultados para la mamografía con ID mdb 238



(a) Localización de la ROI en la mamografía original



(b) ROI aislada para el valor umbral  $\mu + 3\sigma$

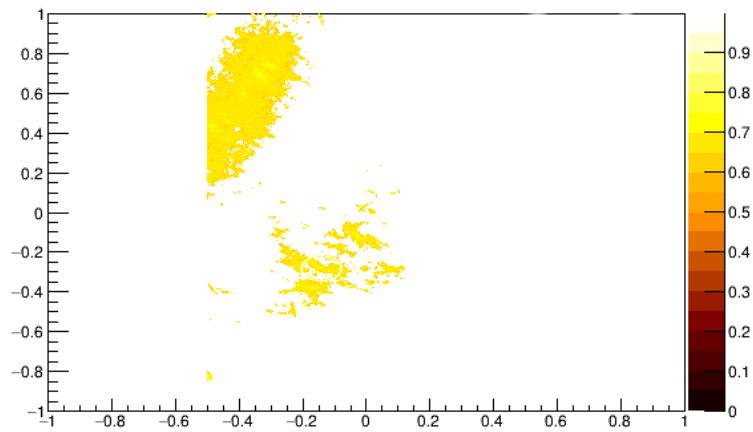


(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

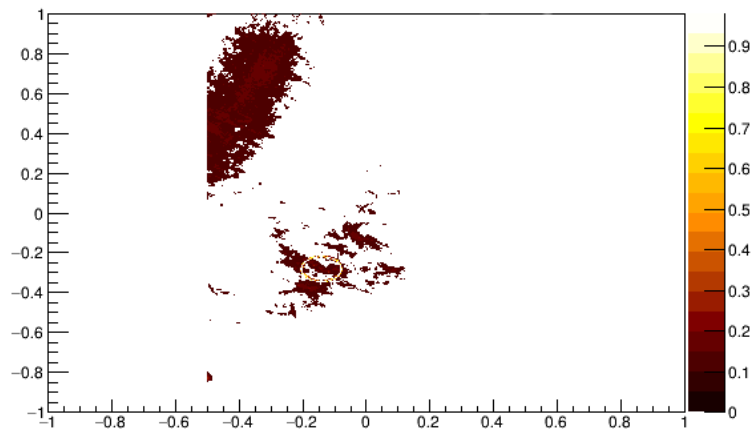
Figura D.8: Resultados para la mamografía con ID mdb 248



(a) Localización de la ROI en la mamografía original



(b) ROI aislada para el valor umbral  $\mu + 3\sigma$

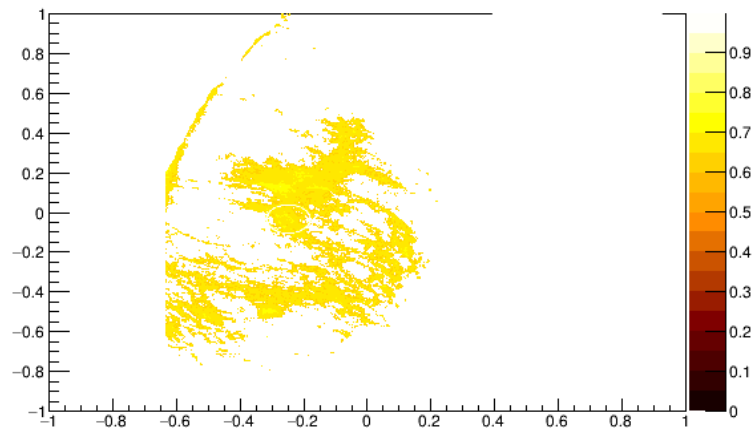


(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

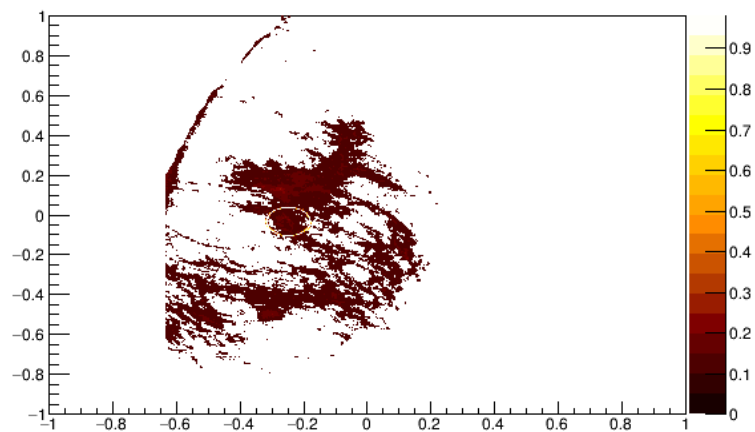
Figura D.9: Resultados para la mamografía con ID mdb 252



(a) Localización de la ROI en la mamografía original

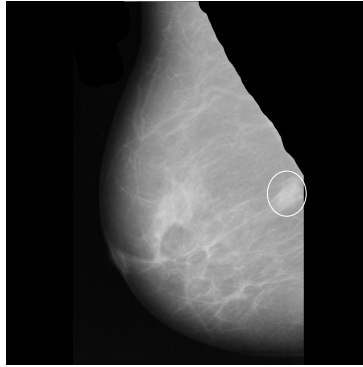


(b) ROI aislada para el valor umbral  $\mu + 3\sigma$

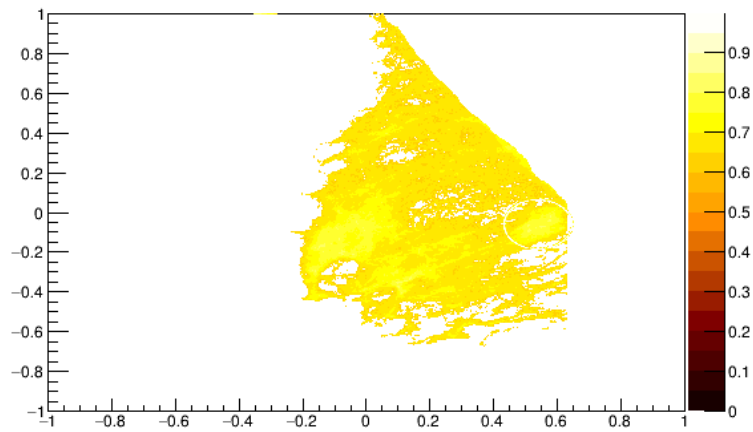


(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

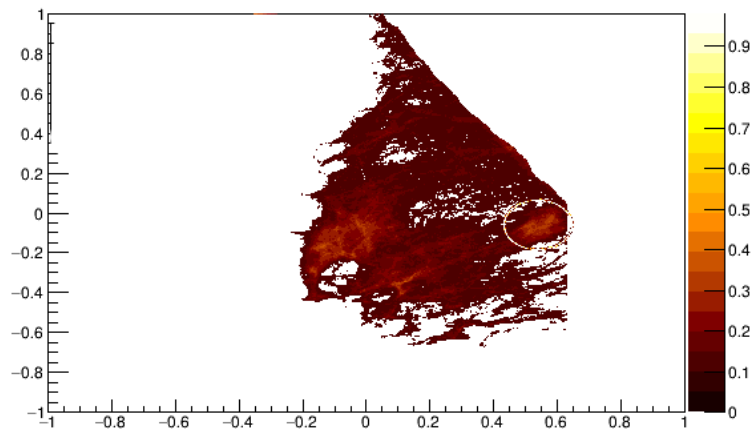
Figura D.10: Resultados para la mamografía con ID mdb 256



(a) Localización de la ROI en la mamografía original



(b) ROI aislada para el valor umbral  $\mu + 3\sigma$



(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

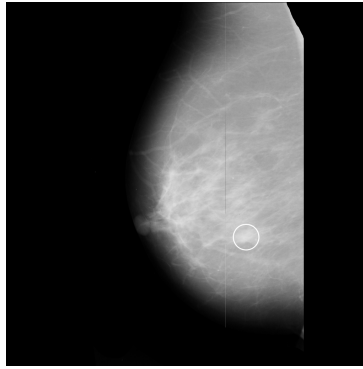
Figura D.11: Resultados para la mamografía con ID mdb 267

- Mamografía: mdb 032. Presenta MISC Benigna. (Figura D.18)
- Mamografía: mdb 090. Presenta ASYM Maligna. (Figura D.19)
- Mamografía: mdb 209. Presenta CALC Maligna. (Figura D.20)
- Mamografía: mdb 264. Presenta MISC Maligna. (Figura D.21)
- Mamografía: mdb 265. Presenta MISC Maligna. (Figura D.22)
- Mamografía: mdb 270. Presenta CIRC Maligna. (Figura D.23)

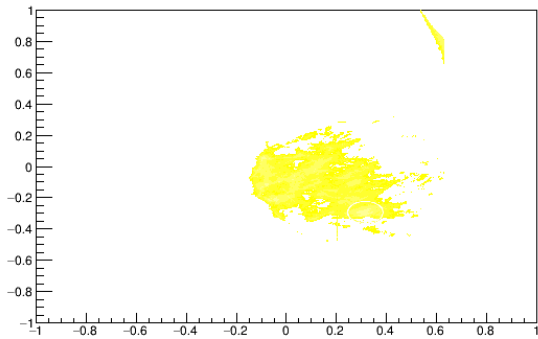
### D.3. Tejido D

Se presentan los resultados de la siguiente lista de MI.

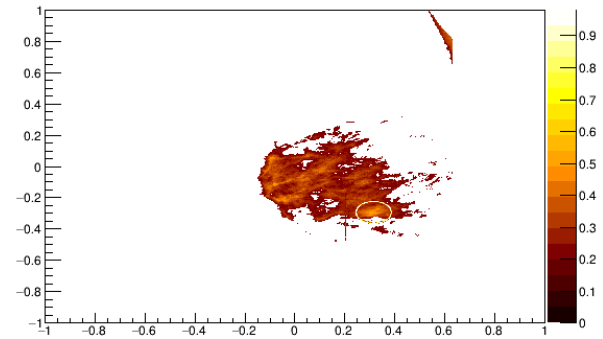
- Mamografía: mdb 058. Presenta MISC Maligna. (Figura D.24)
- Mamografía: mdb 105. Presenta ASYM Maligna. (Figura D.25)
- Mamografía: mdb 110. Presenta ASYM Maligna. (Figura D.26)
- Mamografía: mdb 236. Presenta CALC Benigna. (Figura D.27)
- Mamografía: mdb 239. Presenta CALC Maligna. (Figura D.28)
- Mamografía: mdb 241. Presenta CALC Maligna. (Figura D.29)
- Mamografía: mdb 253. Presenta CALC Maligna. (Figura D.30)



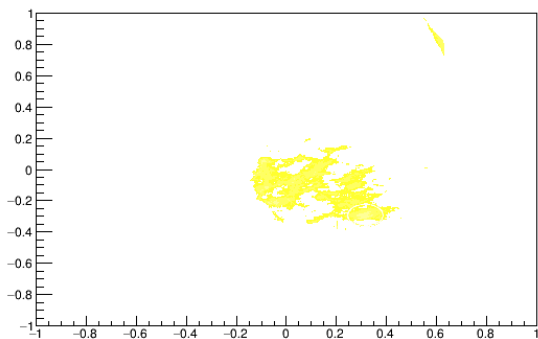
(a) Localización de la ROI en la mamografía original



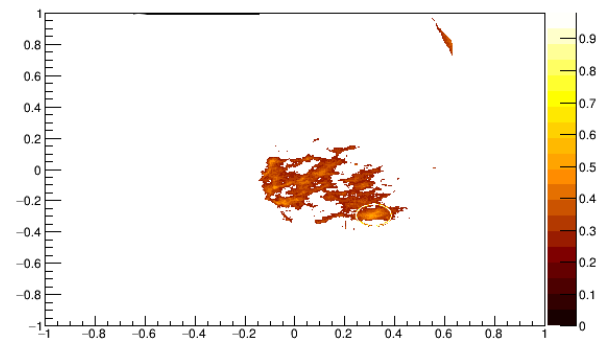
(b) ROI aislada para el valor umbral  $\mu + 3\sigma$



(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

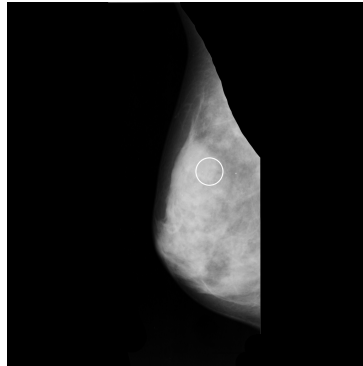


(d) ROI aislada para el valor umbral  $\mu + 4\sigma$

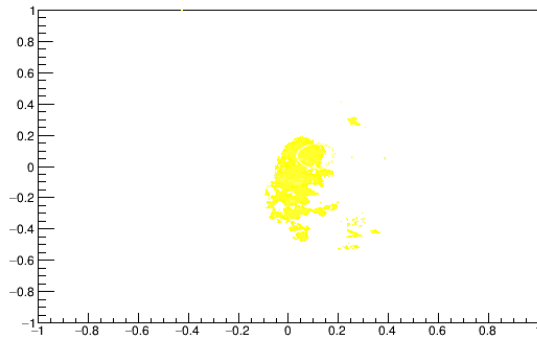


(e) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (d)

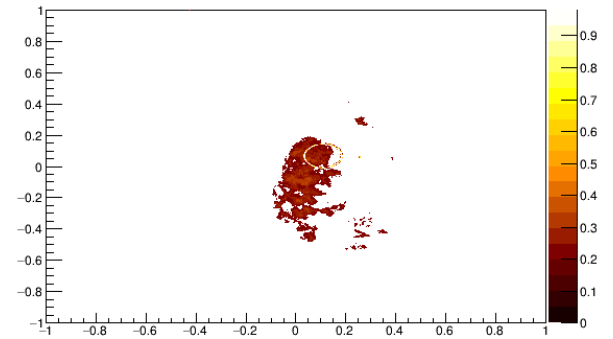
Figura D.12: Resultados para la mamografía con ID mdb 013



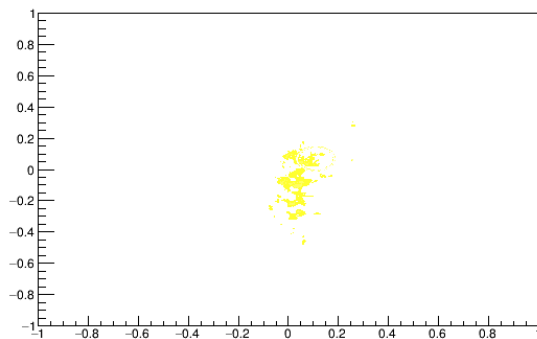
(a) Localización de la ROI en la mamografía original



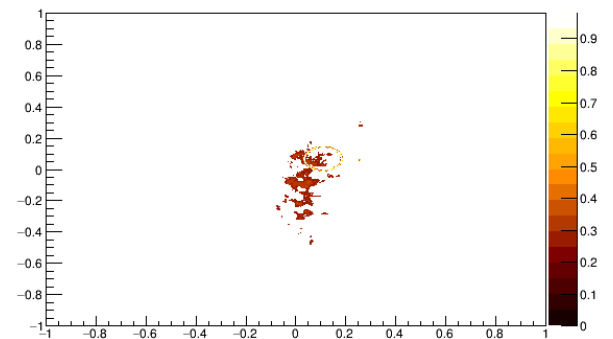
(b) ROI aislada para el valor umbral  $\mu + 3\sigma$



(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

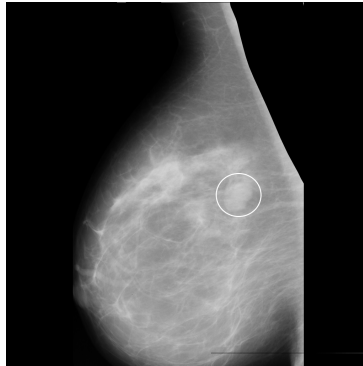


(d) ROI aislada para el valor umbral  $\mu + 4\sigma$

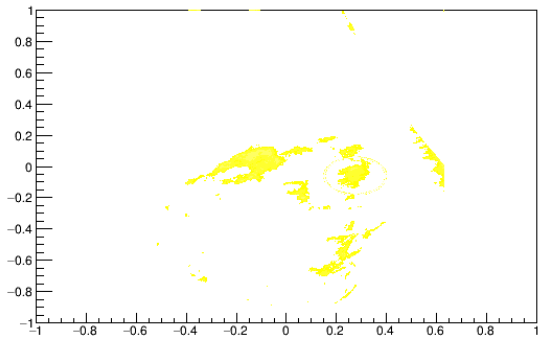


(e) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (d)

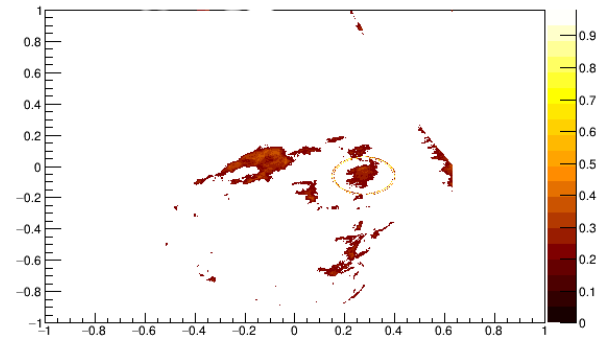
Figura D.13: Resultados para la mamografía con ID mdb 017



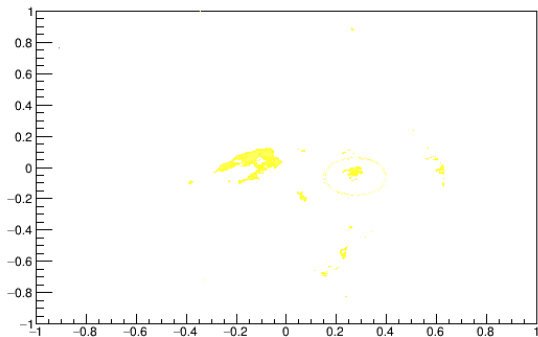
(a) Localización de la ROI en la mamografía original



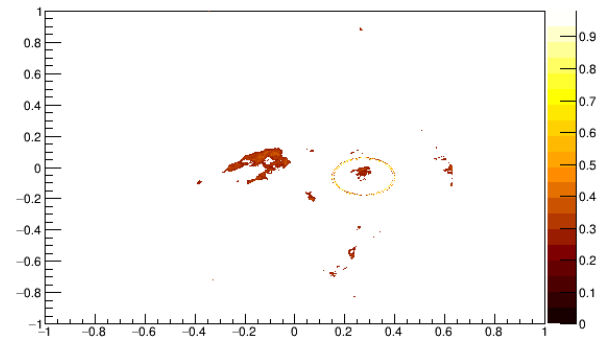
(b) ROI aislada para el valor umbral  $\mu + 3\sigma$



(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)



(d) ROI aislada para el valor umbral  $\mu + 4\sigma$

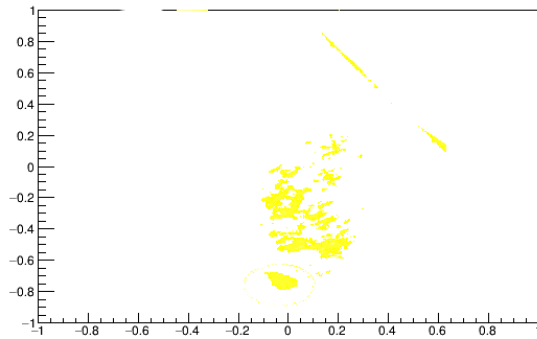


(e) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (d)

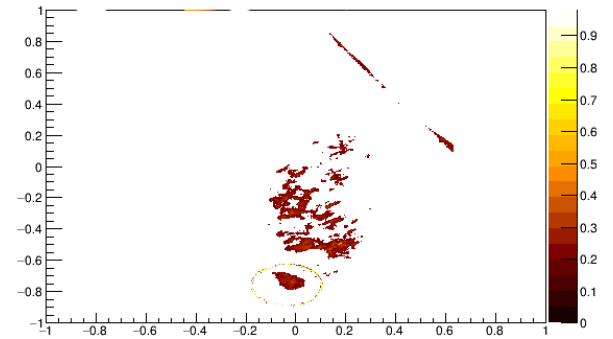
Figura D.14: Resultados para la mamografía con ID mdb 019



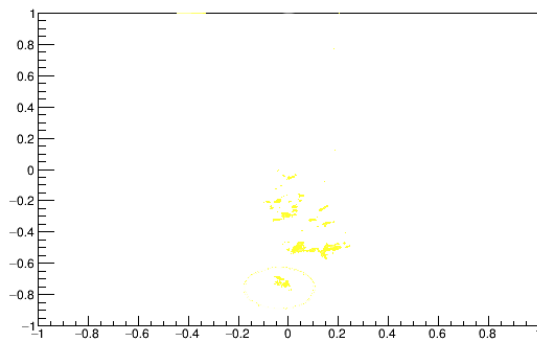
(a) Localización de la ROI en la mamografía original



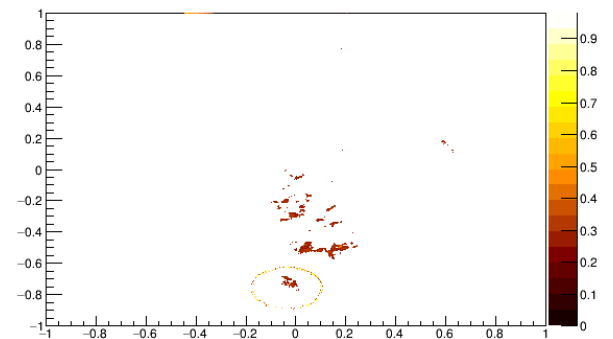
(b) ROI aislada para el valor umbral  $\mu + 3\sigma$



(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

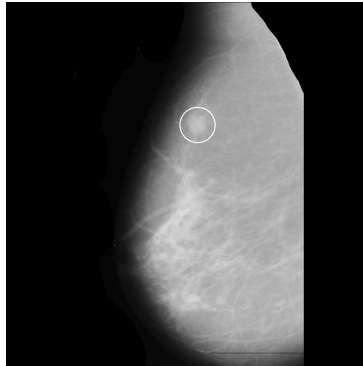


(d) ROI aislada para el valor umbral  $\mu + 4\sigma$

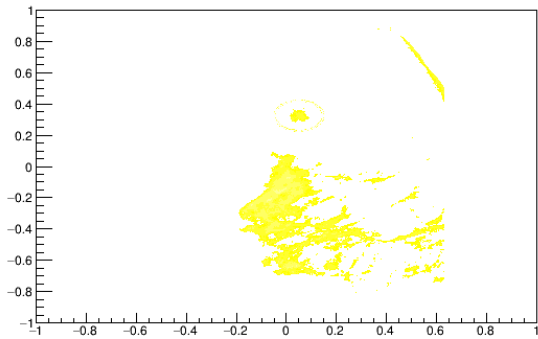


(e) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (d)

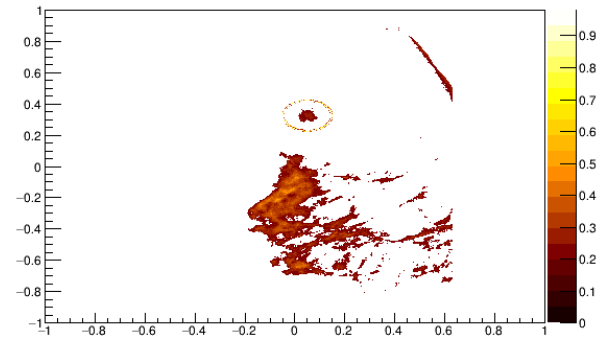
Figura D.15: Resultados para la mamografía con ID mdb 021



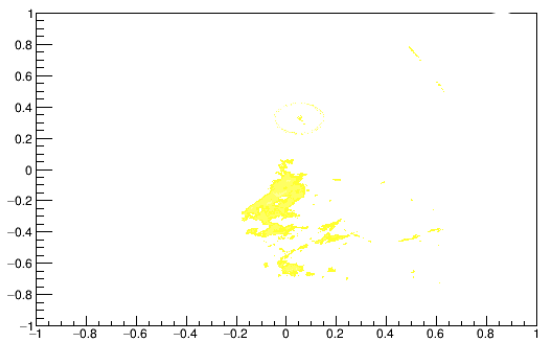
(a) Localización de la ROI en la mamografía original



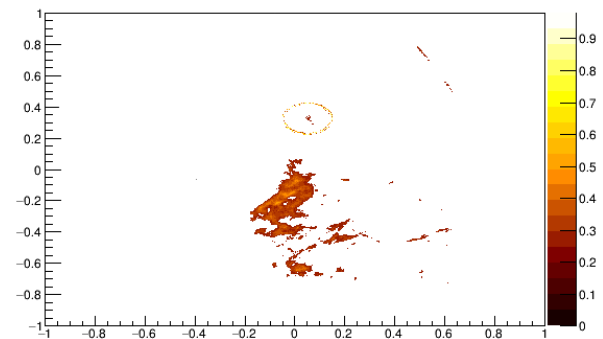
(b) ROI aislada para el valor umbral  $\mu + 3\sigma$



(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

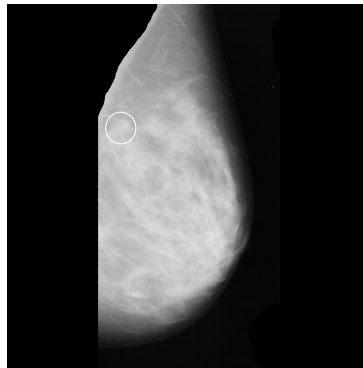


(d) ROI aislada para el valor umbral  $\mu + 4\sigma$

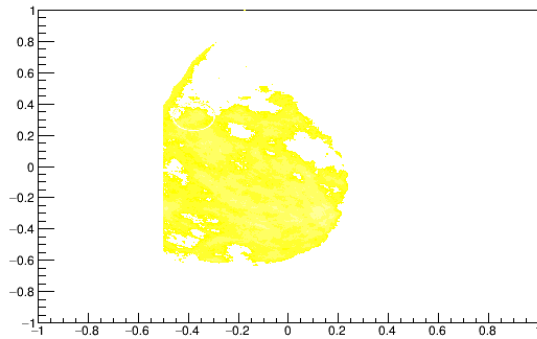


(e) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (d)

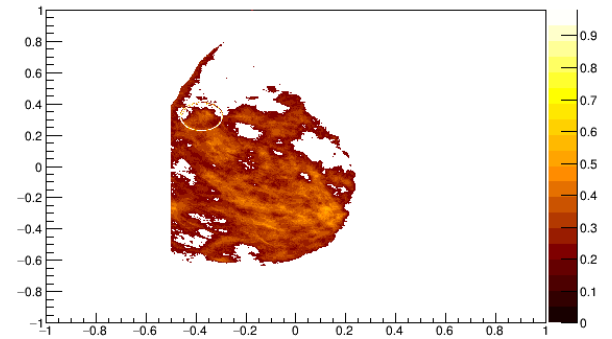
Figura D.16: Resultados para la mamografía con ID mdb 023



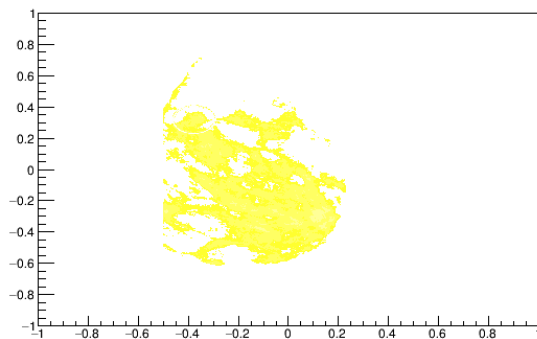
(a) Localización de la ROI en la mamografía original



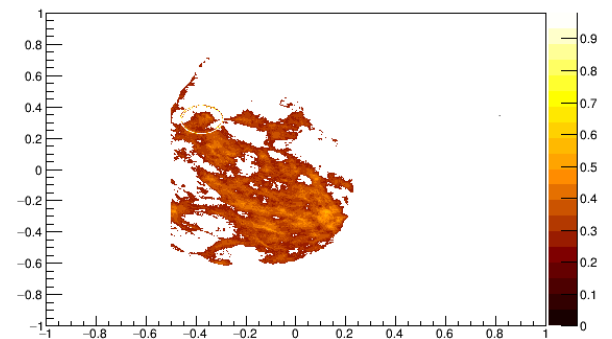
(b) ROI aislada para el valor umbral  $\mu + 3\sigma$



(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

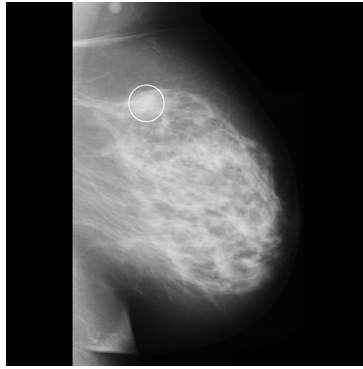


(d) ROI aislada para el valor umbral  $\mu + 4\sigma$

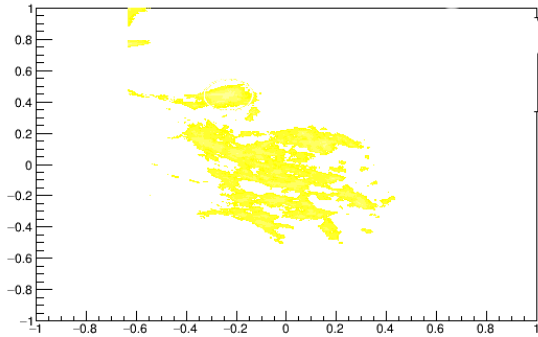


(e) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (d)

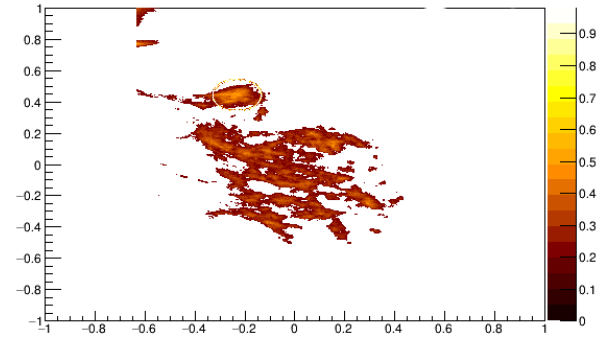
Figura D.17: Resultados para la mamografía con ID mdb 030



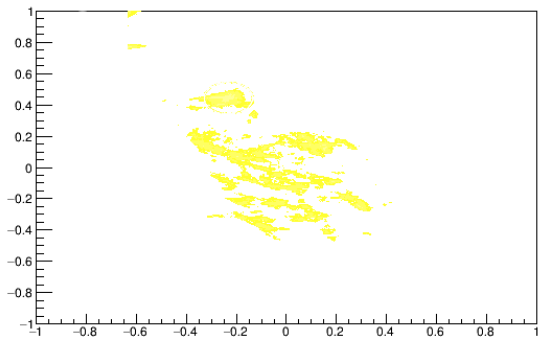
(a) Localización de la ROI en la mamografía original



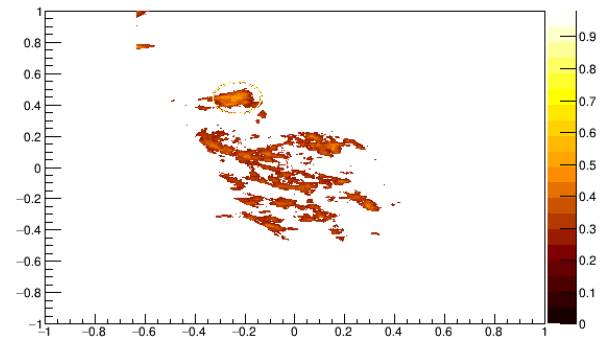
(b) ROI aislada para el valor umbral  $\mu + 3\sigma$



(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)



(d) ROI aislada para el valor umbral  $\mu + 4\sigma$

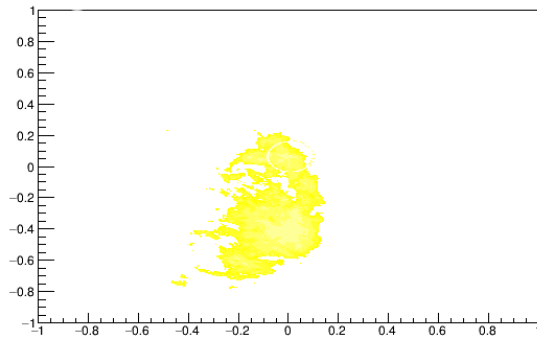


(e) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (d)

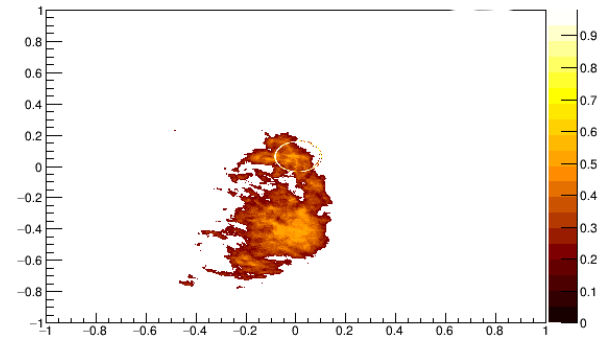
Figura D.18: Resultados para la mamografía con ID mdb 032



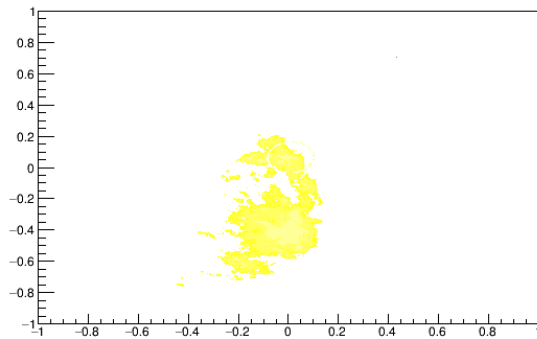
(a) Localización de la ROI en la mamografía original



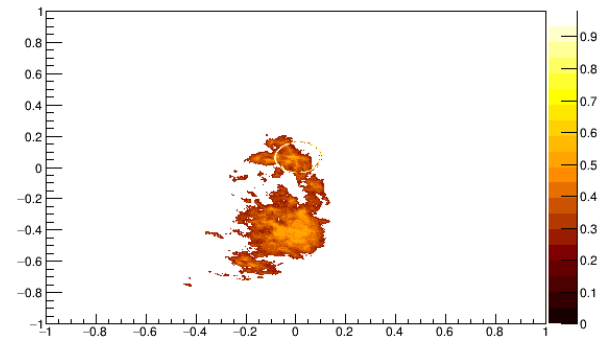
(b) ROI aislada para el valor umbral  $\mu + 3\sigma$



(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

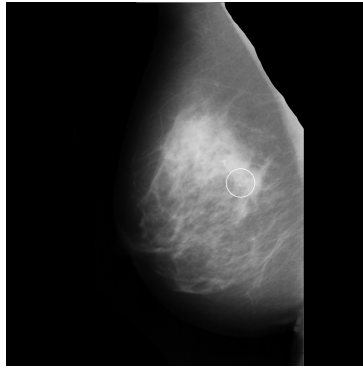


(d) ROI aislada para el valor umbral  $\mu + 4\sigma$

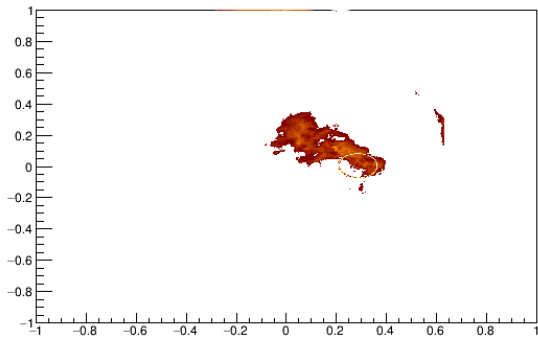


(e) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (d)

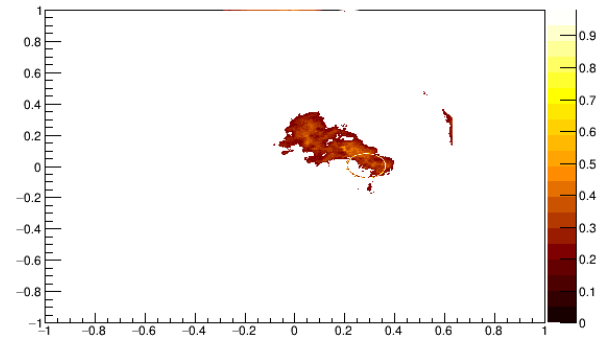
Figura D.19: Resultados para la mamografía con ID mdb 090



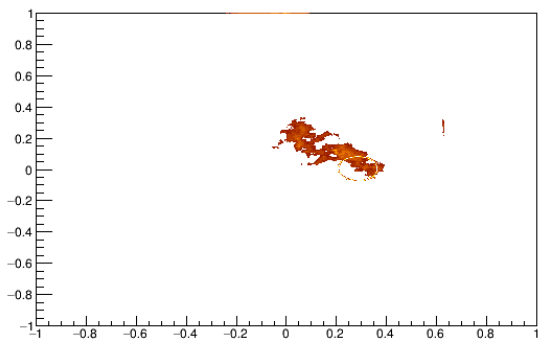
(a) Localización de la ROI en la mamografía original



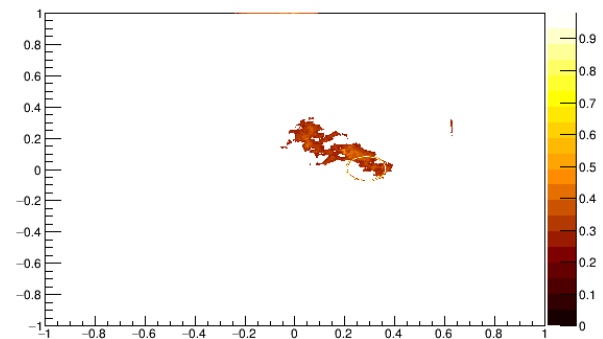
(b) ROI aislada para el valor umbral  $\mu + 3\sigma$



(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

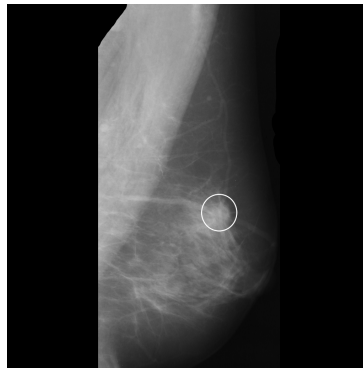


(d) ROI aislada para el valor umbral  $\mu + 4\sigma$

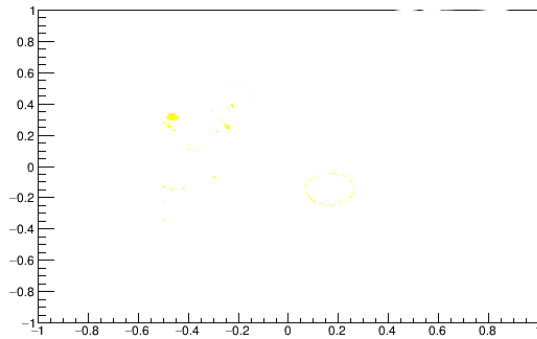


(e) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (d)

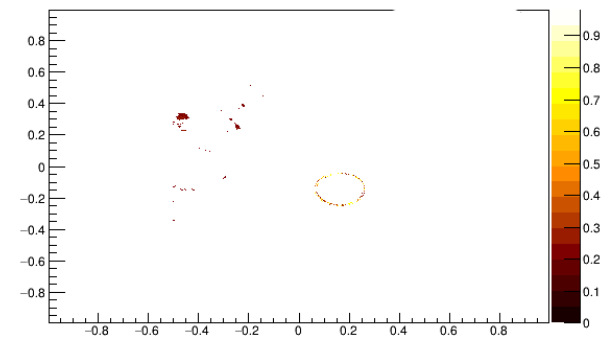
Figura D.20: Resultados para la mamografía con ID mdb 209



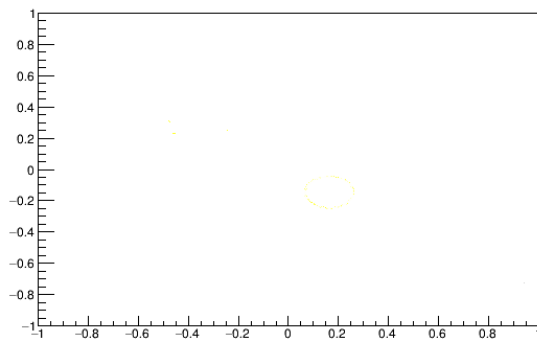
(a) Localización de la ROI en la mamografía original



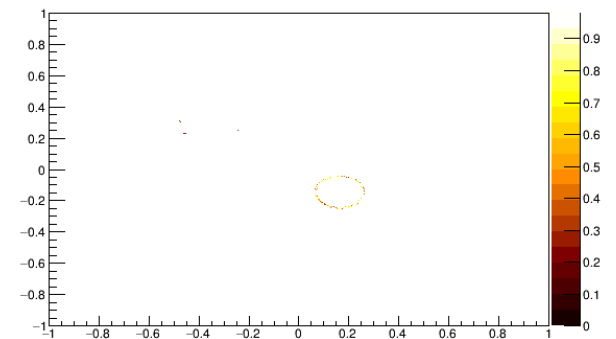
(b) ROI aislada para el valor umbral  $\mu + 3\sigma$



(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

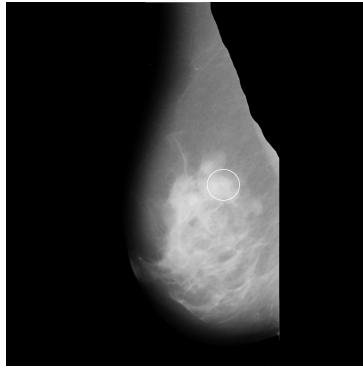


(d) ROI aislada para el valor umbral  $\mu + 4\sigma$

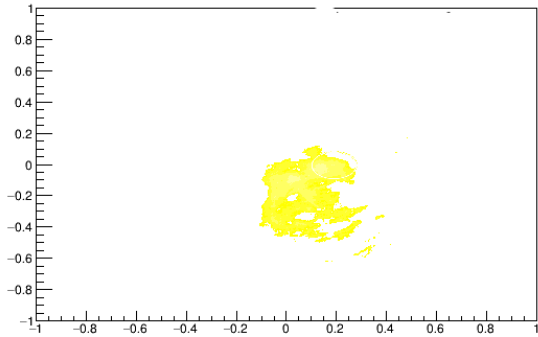


(e) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (d)

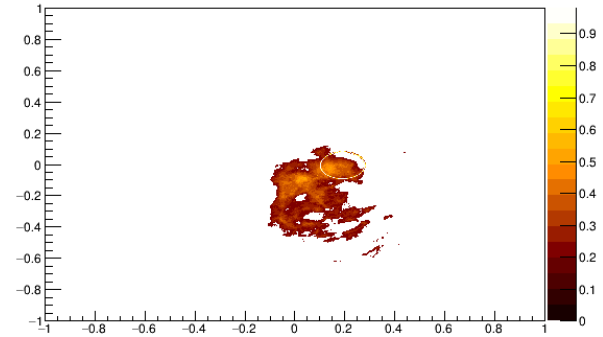
Figura D.21: Resultados para la mamografía con ID mdb 264



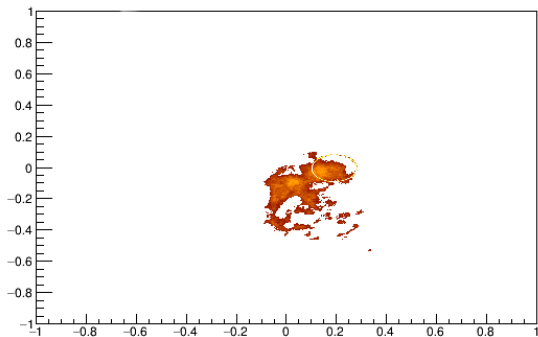
(a) Localización de la ROI en la mamografía original



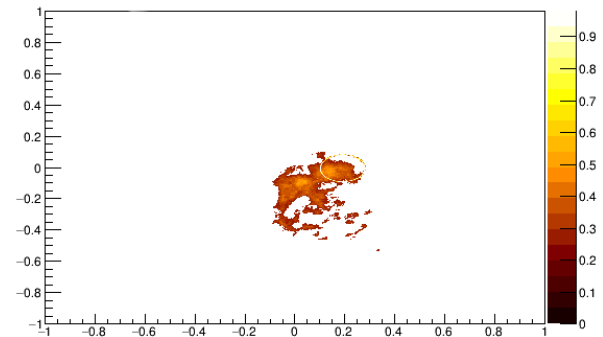
(b) ROI aislada para el valor umbral  $\mu + 3\sigma$



(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

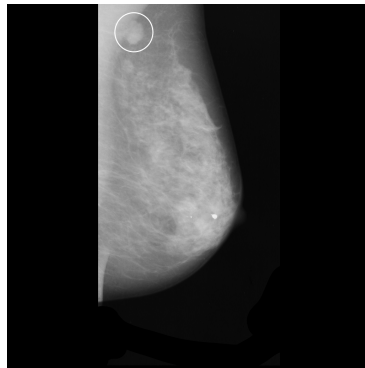


(d) ROI aislada para el valor umbral  $\mu + 4\sigma$

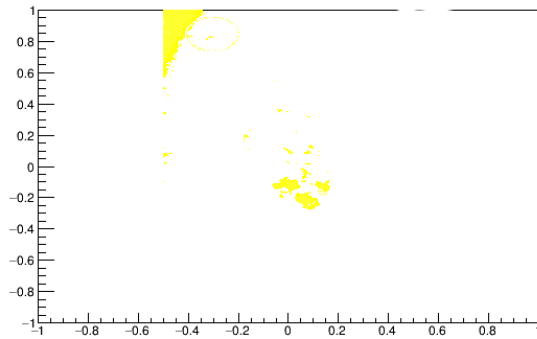


(e) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (d)

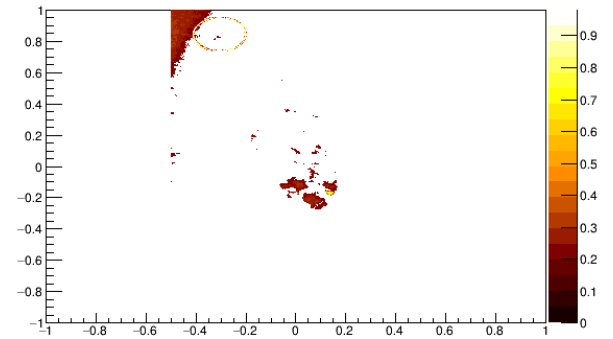
Figura D.22: Resultados para la mamografía con ID mdb 265



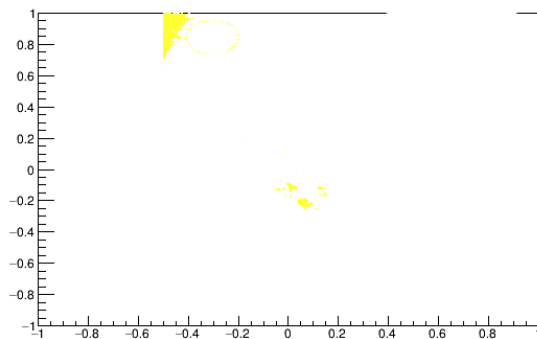
(a) Localización de la ROI en la mamografía original



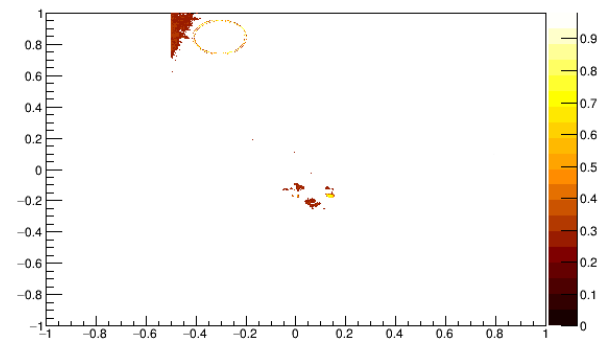
(b) ROI aislada para el valor umbral  $\mu + 3\sigma$



(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

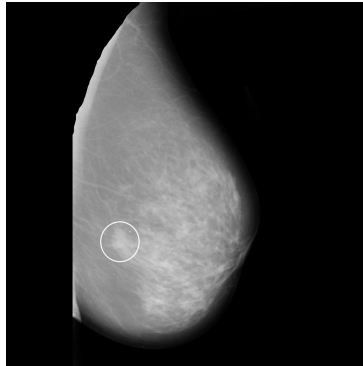


(d) ROI aislada para el valor umbral  $\mu + 4\sigma$

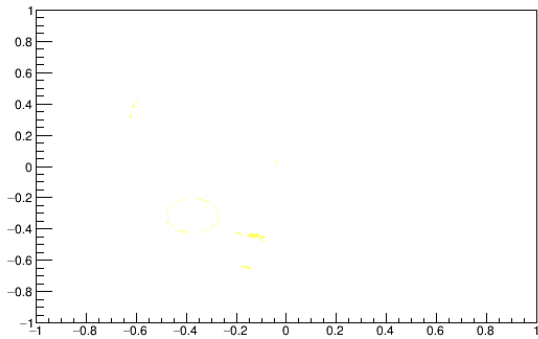


(e) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (d)

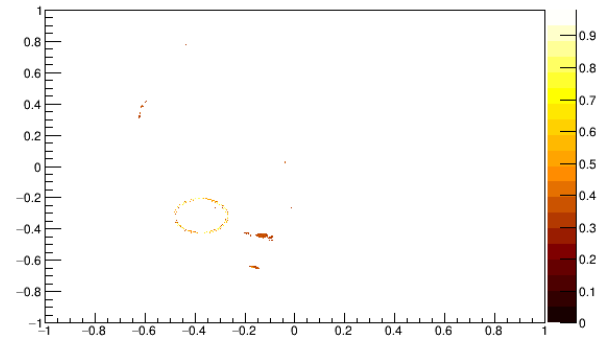
Figura D.23: Resultados para la mamografía con ID mdb 270



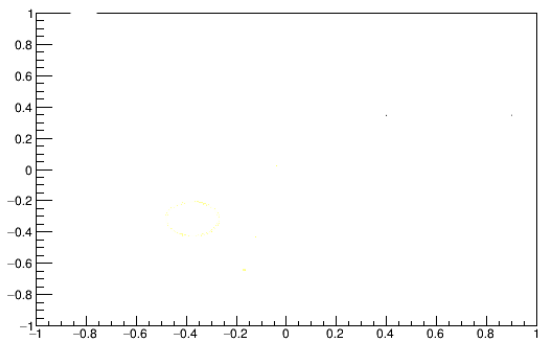
(a) Localización de la ROI en la mamografía original



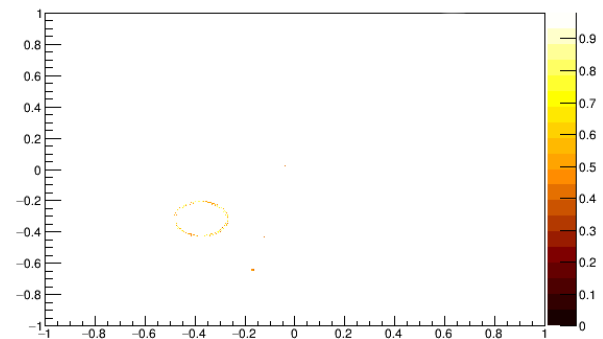
(b) ROI aislada para el valor umbral  $\mu + 3\sigma$



(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

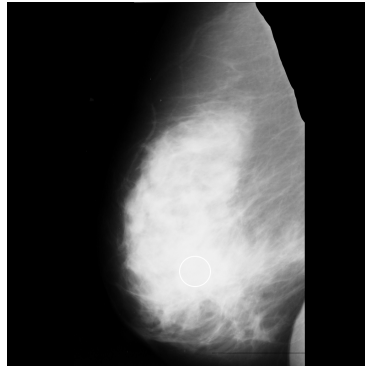


(d) ROI aislada para el valor umbral  $\mu + 4\sigma$

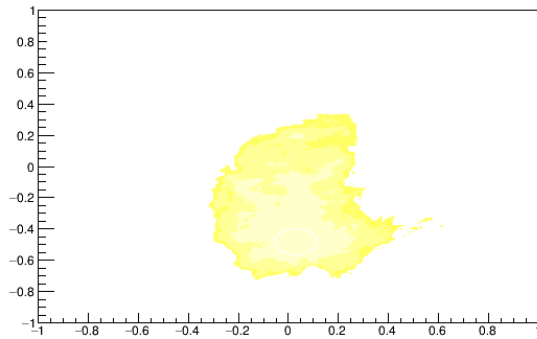


(e) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (d)

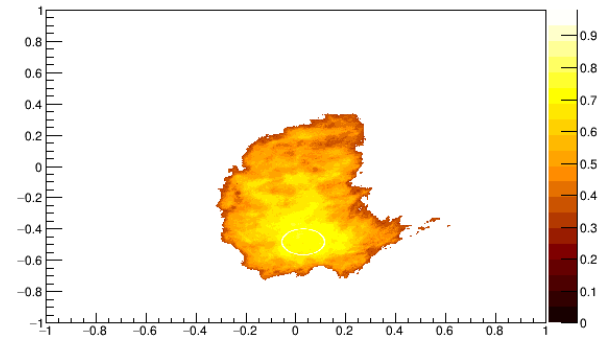
Figura D.24: Resultados para la mamografía con ID mdb 058



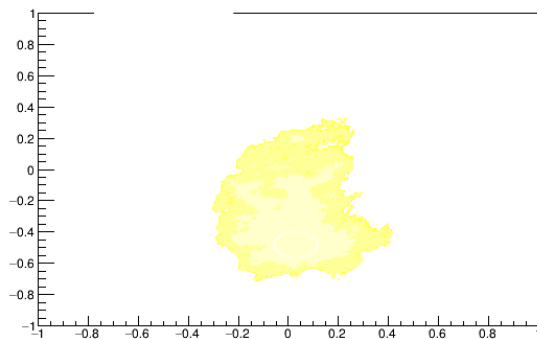
(a) Localización de la ROI en la mamografía original



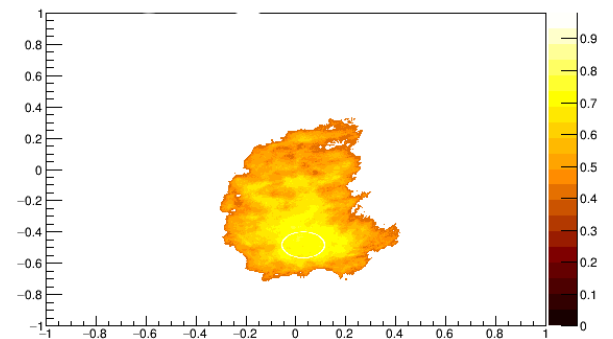
(b) ROI aislada para el valor umbral  $\mu + 3\sigma$



(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

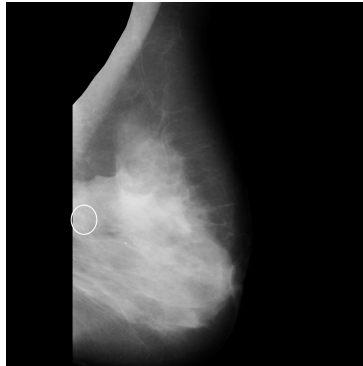


(d) ROI aislada para el valor umbral  $\mu + 4\sigma$

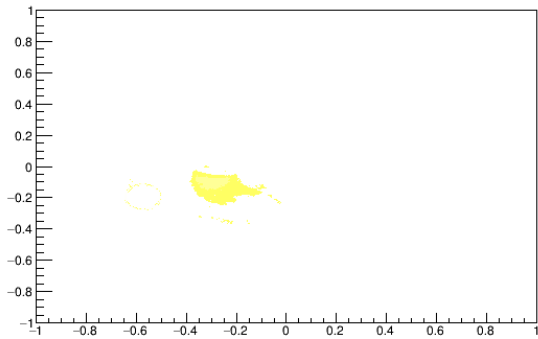


(e) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (d)

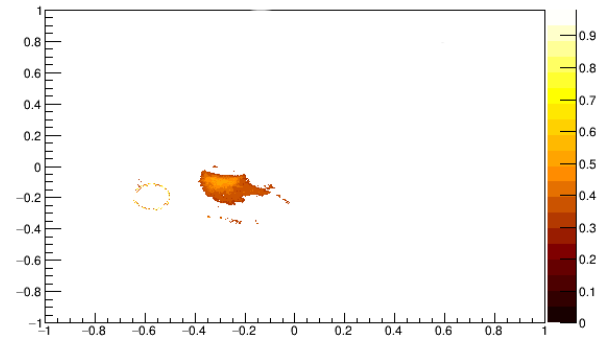
Figura D.25: Resultados para la mamografía con ID mdb 105



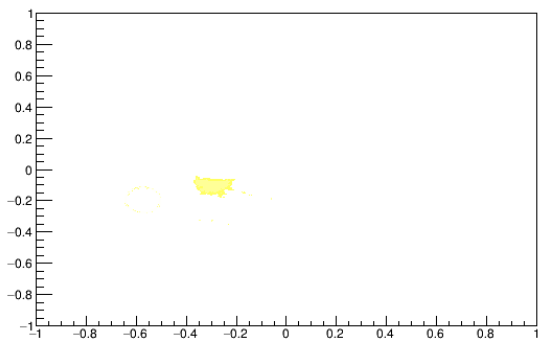
(a) Localización de la ROI en la mamografía original



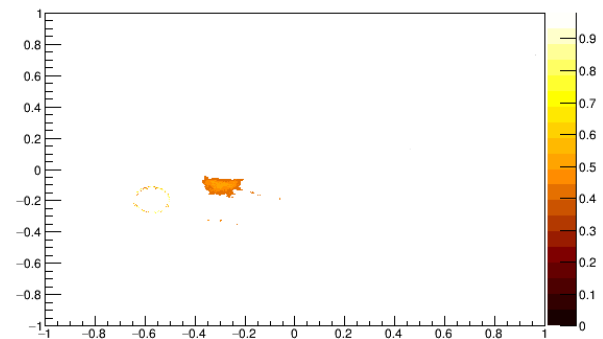
(b) ROI aislada para el valor umbral  $\mu + 3\sigma$



(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

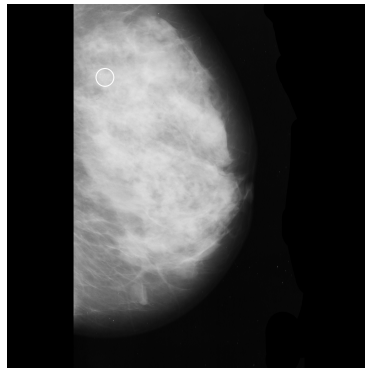


(d) ROI aislada para el valor umbral  $\mu + 4\sigma$

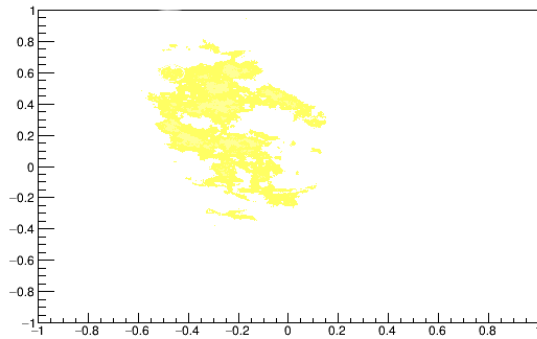


(e) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (d)

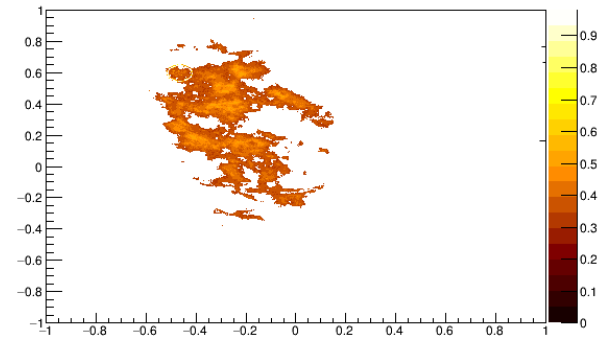
Figura D.26: Resultados para la mamografía con ID mdb 110



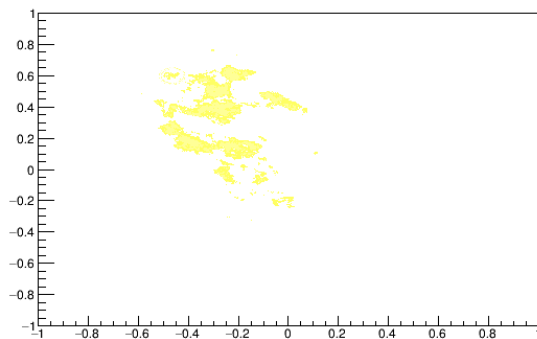
(a) Localización de la ROI en la mamografía original



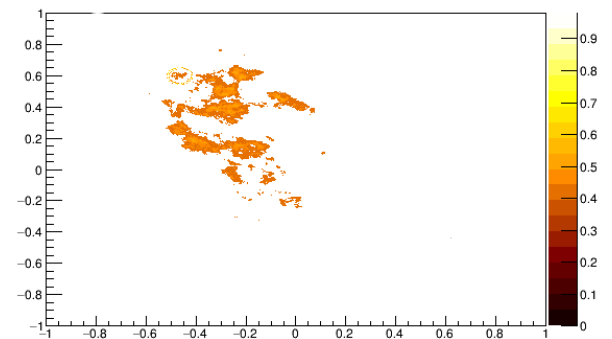
(b) ROI aislada para el valor umbral  $\mu + 3\sigma$



(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

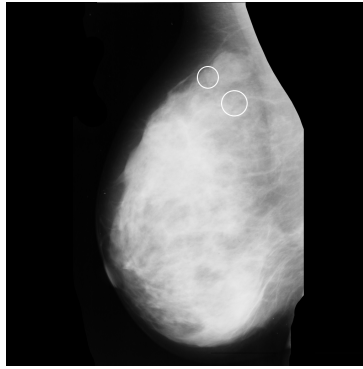


(d) ROI aislada para el valor umbral  $\mu + 4\sigma$

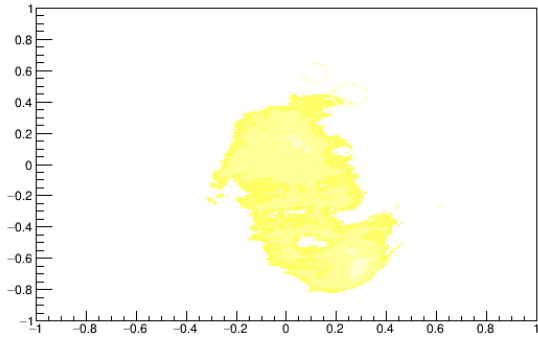


(e) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (d)

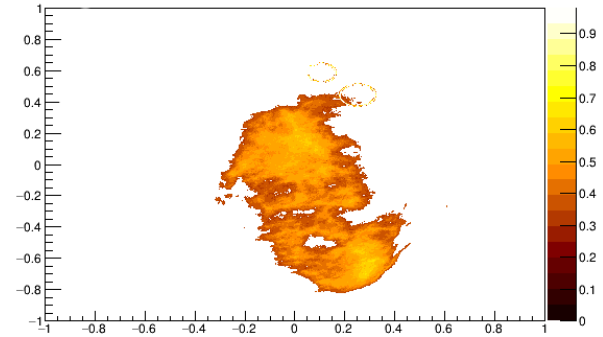
Figura D.27: Resultados para la mamografía con ID mdb 236



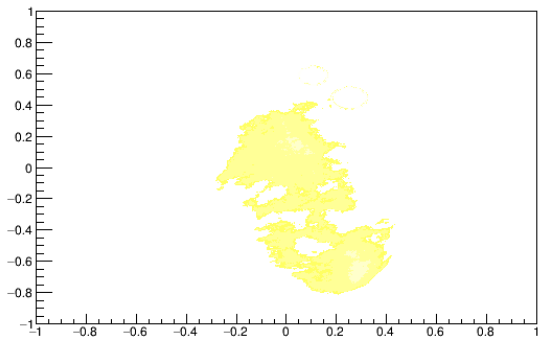
(a) Localización de la ROI en la mamografía original



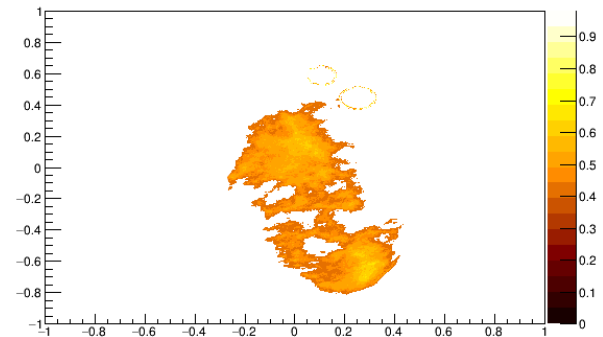
(b) ROI aislada para el valor umbral  $\mu + 3\sigma$



(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)



(d) ROI aislada para el valor umbral  $\mu + 4\sigma$

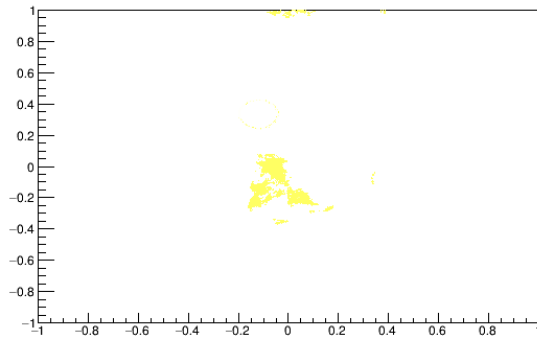


(e) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (d)

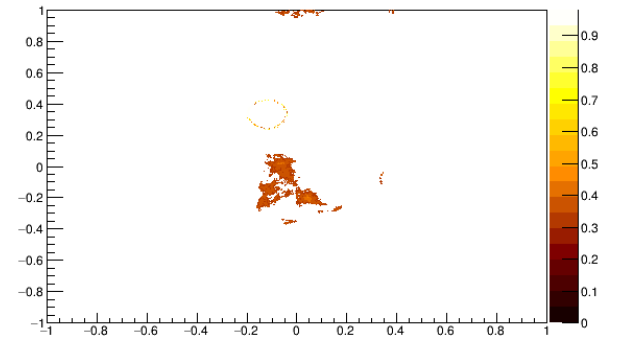
Figura D.28: Resultados para la mamografía con ID mdb 239



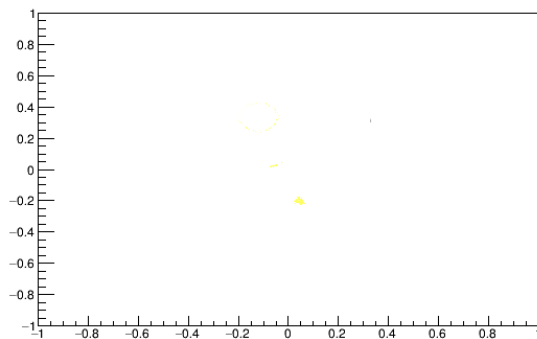
(a) Localización de la ROI en la mamografía original



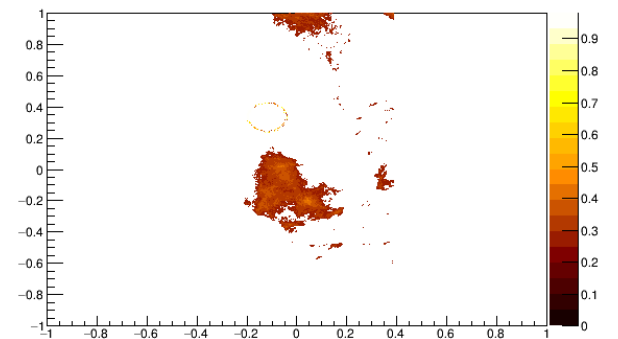
(b) ROI aislada para el valor umbral  $\mu + 3\sigma$



(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)

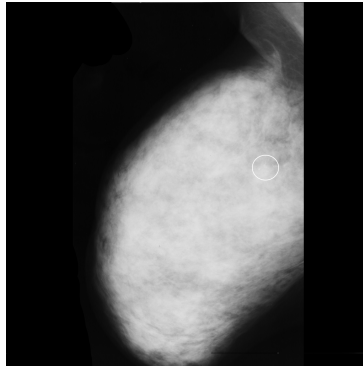


(d) ROI aislada para el valor umbral  $\mu + 4\sigma$

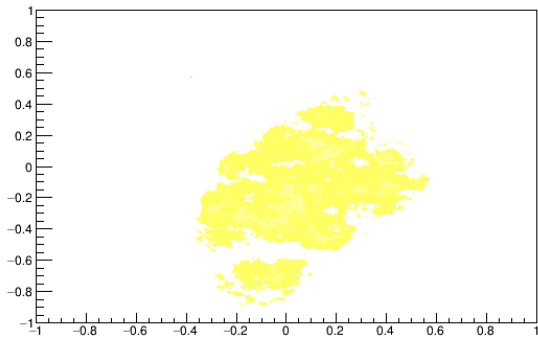


(e) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (d)

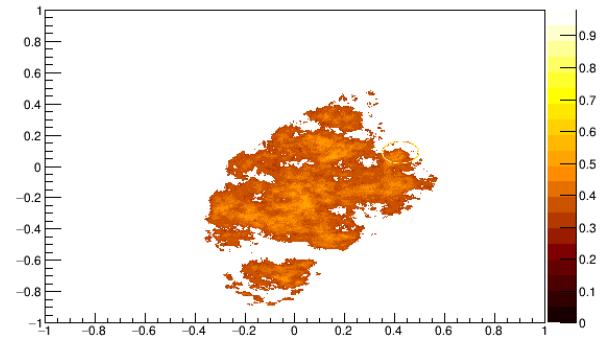
Figura D.29: Resultados para la mamografía con ID mdb 241



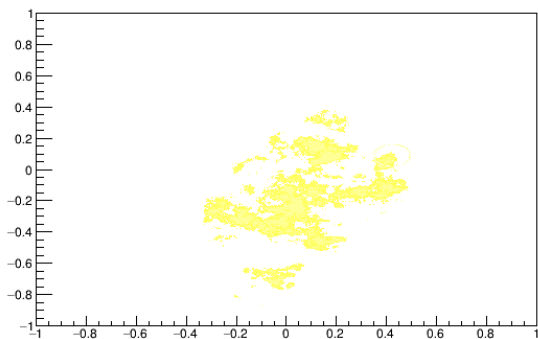
(a) Localización de la ROI en la mamografía original



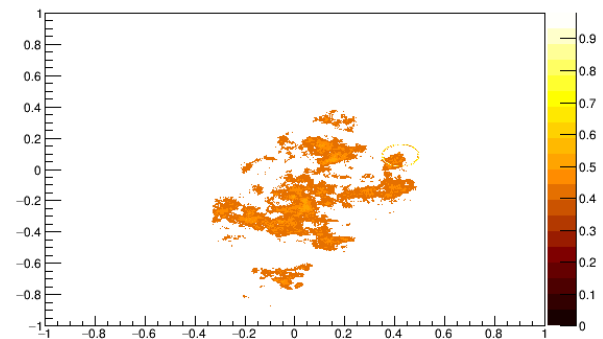
(b) ROI aislada para el valor umbral  $\mu + 3\sigma$



(c) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (b)



(d) ROI aislada para el valor umbral  $\mu + 4\sigma$



(e) Resultado de elevar a la quinta potencia la imagen presentada en el inciso (d)

Figura D.30: Resultados para la mamografía con ID mdb 253

# Bibliografía

- [1] ORGANIZACIÓN MUNDIAL DE SALUD (2020A). OCTUBRE: MES DE SENSIBILIZACIÓN SOBRE EL CÁNCER DE MAMA. [https://www.who.int/cancer/events/breast\\_cancer\\_month/es/](https://www.who.int/cancer/events/breast_cancer_month/es/)
- [2] BECERRA ALCÁNTARA G. I. , DORADO GASPAS M. P. BASAVILVAZO RODRÍGUEZ M. A. , TORRES ARREOLA L. P. , JUÁREZ DE LA LUZ E. (2012). *Diagnóstico y Tratamiento de la Patología Mamaria Benigna en primer y segundo nivel de atención*. México: Instituto Mexicano del Seguro Social.
- [3] AMERICAN CANCER SOCIETY. CANCER: DETECCIÓN TEMPRANA Y DIAGNÓSTICO DEL CÁNCER DE SENO <https://www.cancer.org/es/cancer/cancer-de-seno/pruebas-de-deteccion-y-deteccion-temprana-del-cancer-de-seno/mamogramas/la-densidad-de-los-senos-y-el-informe-de-su-mamograma.html>
- [4] ARANCIBIA H, PATRICIA, TAUB E, TERESA, DE GRAZIA K, JOSÉ A, DÍAZ C, M. LORENA, LÓPEZ P, ALEJANDRA, SÁEZ T, CARLA, CERMENATI B, TOMÁS. (2013). *Microcalcificaciones mamarias: revisión de los descriptores y categorías BI-RADS*. Revista chilena de obstetricia y ginecología, 78(5), 383-394. <https://dx.doi.org/10.4067/S0717-75262013000500010>
- [5] AMERICAN CANCER SOCIETY. CANCER: LIMITACIONES DE LOS MAMOGRAMAS <https://www.cancer.org/es/cancer/cancer-de-seno/pruebas-de-deteccion-y-deteccion-temprana-del-cancer-de-seno/mamogramas/limitaciones-de-los-mamogramas.html>
- [6] AMERICAN CANCER SOCIETY. CANCER: DENSIDAD DE LOS SENOS E INFORME DE SU MAMOGRAMA <https://www.cancer.org/es/cancer/cancer-de-seno/pruebas-de-deteccion-y-deteccion-temprana-del-cancer-de-seno/mamogramas/la-densidad-de-los-senos-y-el-informe-de-su-mamograma.html>
- [7] WOLFE, J. (1976). BREAST PATTERNS AS AN INDEX OF RISK FOR DEVELOPING BREAST CANCER. AMERICAN JOURNAL OF ROENTGENOLOGY, 126(6), 1130–1137. DOI:10.2214/AJR.126.6.1130
- [8] MAURICIO VERGARA, MSc; JOSÉ F. ROBLEDO, MD; SANDRA GÓMEZ, MD; ROXANA AYCARDI, MD; NICOLÁS JIMÉNEZ, MD. (2013). *La desviación estándar del histograma mamográfico como índice de densidad del tejido mamario*. de revistamedicacountry Sitio web: <https://revistamedicacountry.com/app/default/files-module/local/documents/La-desviacion-estandar-del-histograma-mamografico.pdf>
- [9] ROXANA AYCARDI, MD; MAURICIO VERGARA, MSc(2012) *Densidad media de la escala de grises y su correlación con la escala Birads en mamografía* de revista-medicaacountry Sitio web: <https://revistamedicacountry.com/posts/densidad-media-de-la-escala-de-grises-y-su-correlacion-con-la-birads-en-mamografia>
- [10] C.H. ZEPEDA FERNÁNDEZ , M.G. VÁZQUEZ DOMÍNGUEZ , E. MORENO BARBOSA , B. DE CELIS ALONSO, AND M. RODRÍGUEZ CAHUANTZI. (2021). *Tumor and microcalcification*

*characterization using Entropy, Fractal Dimension and intensity values statistical analysis in a mammography.* de arxiv.org Sitio web: <https://arxiv.org/pdf/2101.11090.pdf>

- [11] CARLYLE BUSHONG S. (2010). *Manual de radiología para técnicos.* Barcelona, España: ELSEVIER.
- [12] WENDIE A. BERG, JESSICA W. T. LEUNG . (2019). *Diagnostic Imaging Breast.* Manitoba, Canada: ELSEVIER.
- [13] ORGANIZACIÓN MUNDIAL DE LA SALUD. (S.F.-A). CÁNCER DE MAMA: PREVENCIÓN Y CONTROL. CONTROL DEL CÁNCER DE MAMA. <https://www.who.int/topics/cancer/breastcancer/es/index3.html>
- [14] BOYD, N. F., BYNG, J. W., JONG, R. A., FISHELL, E. K., LITTLE, L. E., MILLER, A. B., YAFFE, M. J. (1995). *Quantitative Classification of Mammographic Densities and Breast Cancer Risk: Results From the Canadian National Breast Screening Study.* JNCI Journal of the National Cancer Institute, 87(9), 670–675. doi:10.1093/jnci/87.9.670
- [15] LÁSZLÓ TABÁR, PETER B. DEAN. (2012). *Teaching Atlas of Mammography.* New York, USA: Thieme Publishing Group.
- [16] LASZLO TABAR. *Imaging of the Breast – Technical Aspects and Clinical Implication .* Rijeka, Croatia: InTech, 2012.
- [17] MURRAY R. SPIEGEL (2010) *Probabilidad y Estadística.* Mc Graw Hill, México.
- [18] DANIEL C. HARRIS (2003) *Análisis químico cuantitativo.* Editorial Reverté, Barcelona, España.
- [19] DOUGLAS A. SKOOG, DONALD M. WEST(2014)*Fundamentos de química analítica.* CEN-CEGE Learning,
- [20] THE MINI-MIAS DATABASE OF MAMMOGRAMS: MAMMOGRAPHIC IMAGE ANALYSIS SOCIETY MINIMAMMOGRAPHIC DATABASE. <http://peipa.essex.ac.uk/info/mias.html>
- [21] R. Brun, ROOT - An object oriented data analysis framework, NUCL INSTRUM METH A 389 (1997) 1-2 pg. 81-86.
- [22] TASImage Class Referencet <https://root.cern.ch/doc/master/classTASImage.html>.
- [23] SetBinContent Class Referencet <https://root.cern.ch/doc/master/classTHn.html>.