



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

A. D. López Solís<sup>1</sup>, C. H. Zepeda Fernández<sup>2</sup>

[1]alejandra.lopezso@alumno\_buap.mx, [2]hzepeda@cfm.buap.mx

LINK: [https://teams.microsoft.com/l/meetup-join/19%3ameeting\\_MWE1ZTBmZDIhZTZmOC00NDQ5LWUyYVYtZWVhNml4OTU4NTVh%40thread.v2/0?context=%7b%22Tid%22%3a%223ae3c7c7-ea51-46de-9540-12f4110eac9b%22%2c%22Oid%22%3a%227e224b68-29e3-40b7-9cb5-bc9eba4e7797%22%7d](https://teams.microsoft.com/l/meetup-join/19%3ameeting_MWE1ZTBmZDIhZTZmOC00NDQ5LWUyYVYtZWVhNml4OTU4NTVh%40thread.v2/0?context=%7b%22Tid%22%3a%223ae3c7c7-ea51-46de-9540-12f4110eac9b%22%2c%22Oid%22%3a%227e224b68-29e3-40b7-9cb5-bc9eba4e7797%22%7d)

## 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 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 diferentes 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 disponibles en la base de datos de libre acceso para la investigación científica 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, conducirá 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.

## INTRODUCCIÓN

El cáncer de mama representa un importante problema social. 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 la detección precoz sigue siendo la mejor oportunidad de quienes luchan contra esta enfermedad.

Actualmente, la mamografía es por excelencia la prueba de diagnóstico mamario, ya que permite detectar lesiones en la mama hasta dos años antes de que sean palpables. A fin de mejorar la capacidad de diagnóstico en este medio, se han propuesto diferentes métodos y algoritmos computarizados que permiten llevar a cabo una fácil manipulación de la imagen para su estudio, propiciando de forma considerable la rápida detección y caracterización de lesiones mamarias que en ocasiones se ve limitada por la densidad de las mamas y la gran cantidad de tonos de gris representados en las imágenes de diagnóstico, tal es el caso del procesamiento de imágenes mamográficas mediante análisis de datos.

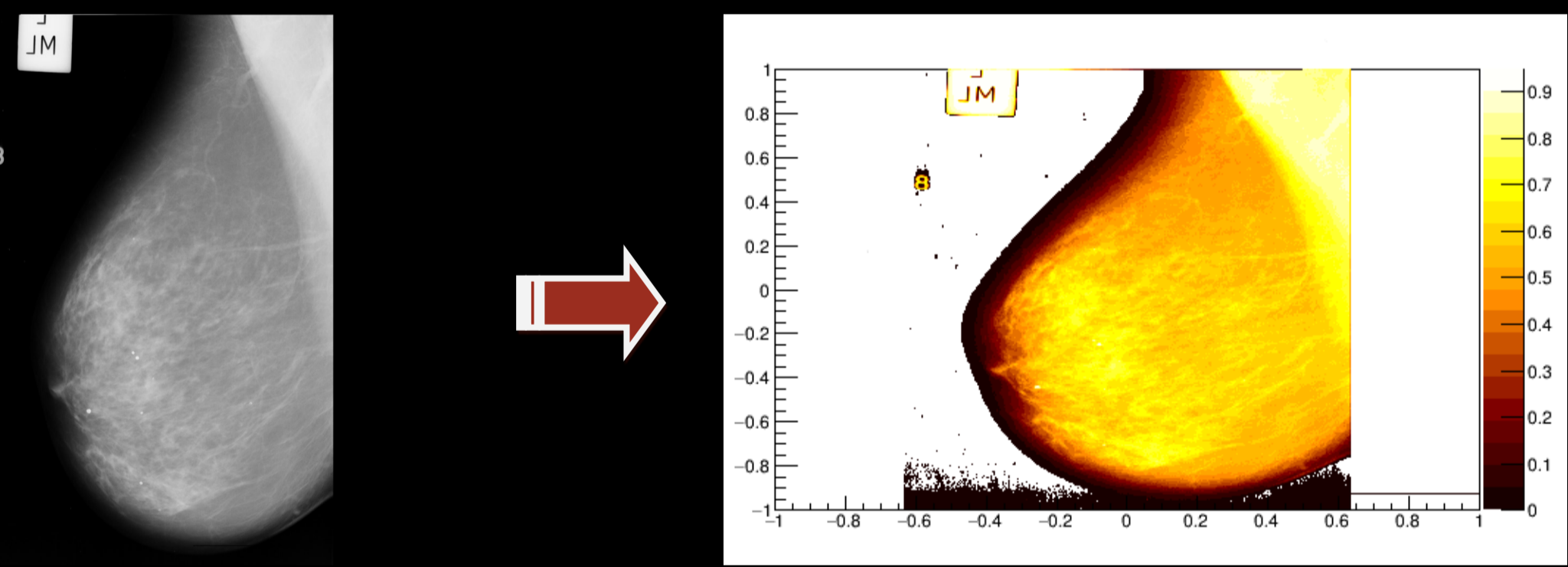
## OBJETIVO

Se busca mostrar que es posible caracterizar el tejido sano presente en la mama haciendo uso de los parámetros media y desviación estándar de la distribución de intensidades para áreas representativas de dicho tejido en mamografías, esto con la finalidad de poder 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), facilitando así su localización. Lo anterior, utilizando un método computacional de análisis de datos como 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.

## METODOLOGÍA

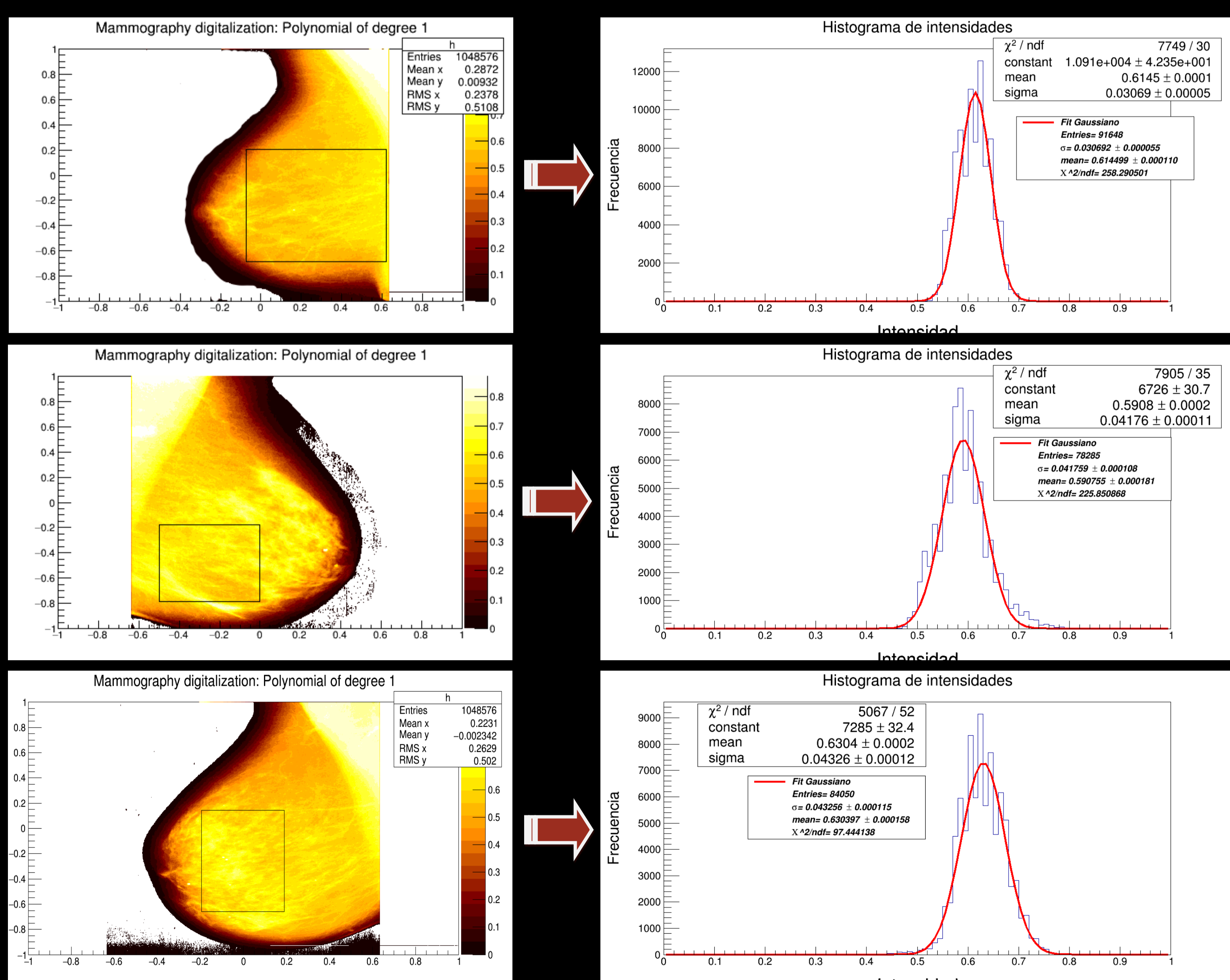
Se recopiló una muestra de 90 mamografías disponibles en la base de datos de libre acceso para la investigación científica mini-MIAS, perteneciente a PEIPA: el archivo piloto europeo de procesamiento de imágenes. Cada imagen cuenta con marcadores descriptivos según su tipo de tejido, el tipo de lesión que presentan y la gravedad de la misma, además, se proporcionan datos referentes al diagnóstico como las coordenadas del centro de la anomalía y su radio en píxeles.

Se trata de imágenes digitalizadas en formato .pgm de un tamaño de  $1024 \times 1024$  píxeles, las cuales se convirtieron a formato .png para ser sometidas a procesamiento, pues el código C++ utilizado para el análisis requiere de una imagen en este formato como datos de entrada. Dichas películas se transformaron a mapas de intensidades, cuyos valores de píxeles corresponden a una escala de grises, de modo que, la información de las imágenes originales se representa como matrices 2D, donde la ubicación del bin queda determinada por cada entrada y el valor de la misma presenta la intensidad, que se normalizó para este caso (0=negro hasta 1=blanco). A estas nuevas imágenes las denominaremos *redigitalizadas*.



Posteriormente, se creó un archivo de texto que ordena en tres columnas toda la información correspondiente a las imágenes redigitalizadas: Coordenada  $x$ , Coordenada  $y$  y valor de intensidad. Esto con el fin de convertir dicha información en histogramas de frecuencias utilizando nuevamente el sistema ROOT V. 6.23/01. Con esta información, fue posible localizar el bin de mayor intensidad que coincide con el centro de la anomalía en mamografías anormales.

La muestra original se dividió en 3 clasificaciones según el tipo de tejido de fondo: graso, glandular y denso (30 mamografías de cada uno), de las cuales se seleccionaron áreas representativas de tejido sano y se obtuvo su respectivo histograma de intensidades, el cual nos proporciona información sobre la densidad del tejido, se encontraron distribuciones de tipo Gaussiano y se les realizó el ajuste correspondiente para obtener los parámetros que caracterizan a esa distribución ( $\mu$  y  $\sigma$ ).



## RESULTADOS

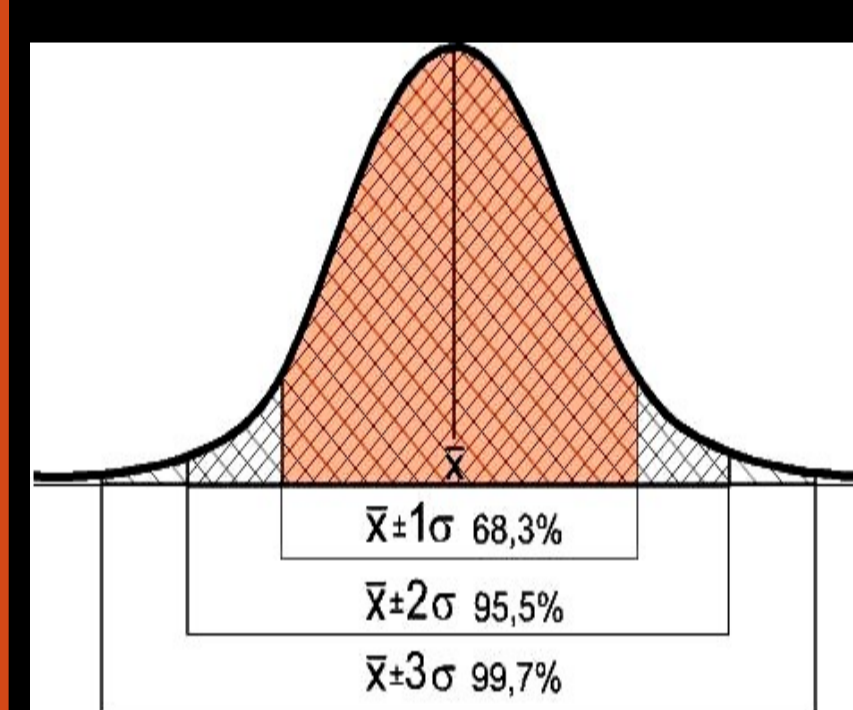
Para cada tipo de tejido se obtuvieron distribuciones similares. Se encontró:  
 $\mu_F < \mu_G < \mu_D$

Además, de la estadística descriptiva, de las medias de las distribuciones de intensidad, se encontró que el tejido D es el más irregular de los 3 tipos, pues su  $\sigma$  es mayor.

	N Total	$\mu$	$\sigma$	Suma	Mínimo	Mediana	Máximo
$\mu$ Tejido F	30	0.55536	0.07642	16.66082	0.39379	0.55515	0.66121
$\mu$ Tejido G	30	0.6384	0.08634	19.15203	0.45636	0.63125	0.81866
$\mu$ Tejido D	30	0.70773	0.10237	21.23203	0.47375	0.70894	0.90059

No se tiene un valor de intensidad definido para separar cuantitativamente cada una de las muestras según su tipo de tejido.

El valor del bin de intensidad máxima, que coincide con el centro de la anomalía para las mamografías anormales es siempre mayor que el valor  $\mu + 3\sigma$  de los parámetros de su distribución, es decir, los valores de intensidad e los píxeles que conforman el tejido sano no distan más de  $3\sigma$  de la media obtenida para cada muestra. Esto coincide con la teoría conocida para el tipo de distribuciones encontradas.



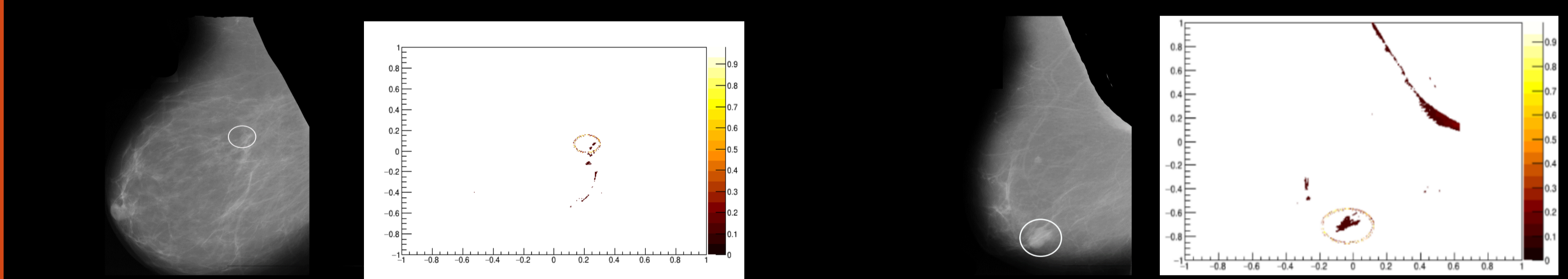
Por lo tanto, para cada tipo de tejido es posible definir un valor "umbral" a partir del cual se reconstruirá la imagen únicamente con aquellos valores que lo sobrepasen, aislando de esta forma las áreas de interés para cualquier mamografía.

Haciendo uso de la estadística se calcularon:

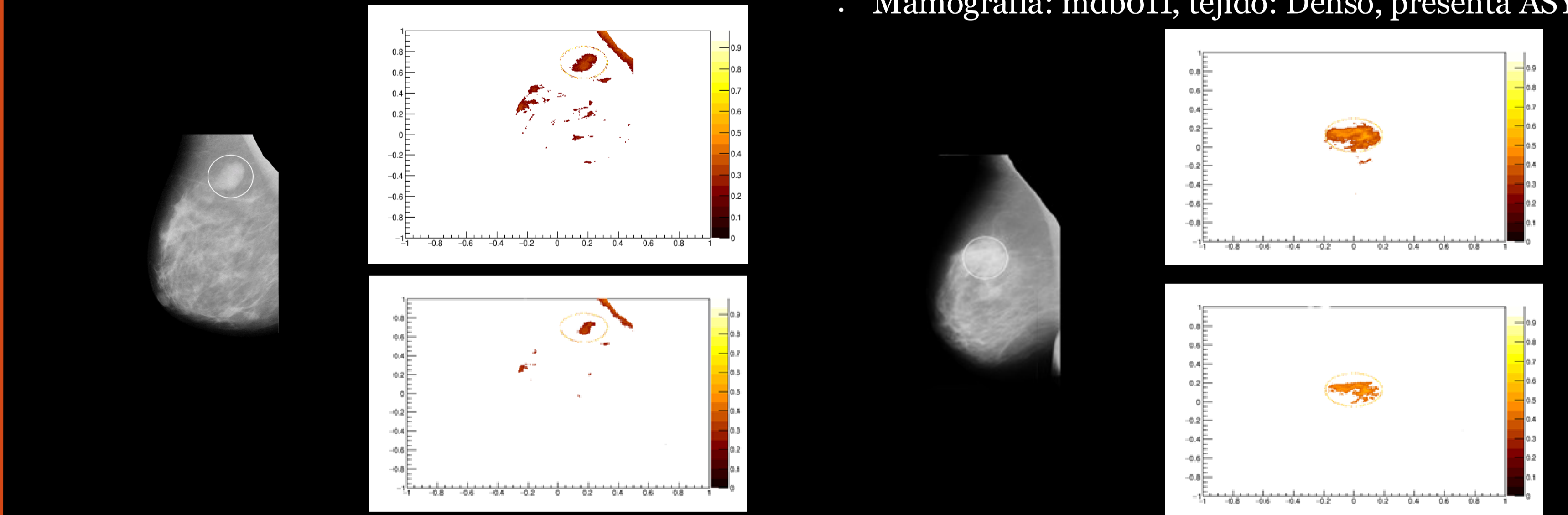
Tipo de tejido	Valor umbral 1 $\mu + 3\sigma$	Valor umbral 2 $\mu + 4\sigma$
Graso	0.6434629	
Glandular	0.7379555	0.771140367
Denso	0.8026723	0.83431828

Algunas reconstrucciones se muestran en las siguientes imágenes. Para las muestras de tejidos de mayor densidad se calcularon 2 candidatos de valores umbral, se muestran ambos resultados. El segundo valor umbral discrimina mayor cantidad de bins. Se aprovechó la normalización de intensidades para resaltar las regiones persistentes.

• Mamografía: mdb231, tejido: Graso, presenta CALC M • Mamografía: mdb005, tejido: Graso, presenta CIRC B



• Mamografía: mdb015, tejido: Glandular, presenta CIRC • Mamografía: mdb011, tejido: Denso, presenta ASYM M



## CONCLUSIONES

No se obtuvo un método cuantitativo eficaz para distinguir entre los tres diferentes tipos de parénquima mamario sano: F, G y D, pues se tiene que no es posible definir un valor de intensidad de píxel límite que separe cada modelo. Para una sola mamografía, se tienen píxeles que corresponden a cada tipo de tejido. Así, la mejor forma de clasificar las mamografías según su densidad es cualitativamente 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 al darnos una idea de las regiones que presentan valores de intensidad mucho mayores que los esperados según las diferentes clasificaciones.

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.

## REFERENCIAS

[1] Organización mundial de la salud. <https://www.who.int/es/news-room/fact-sheets/detail/breast-cancer>  
 [2] The mini-MIAS database of mammograms: MAMMOGRAPHIC IMAGE ANALYSIS SOCIETY MiniMammographic Database <http://peipa.essex.ac.uk/info/mias.html>  
 [3] C.H. Zepeda Fernández, M.G. Vázquez Domínguez, E. Moreno Barbosa, B. De Celis Alonso, K. Herrera, and M. Rodríguez Cahuantzi. (2021). Tumor and microcalcification characterization using Entropy, Fractal Dimension and intensity values statistical analysis in a mammography. 05/09/2021, de ResearchGate Sitio web: [https://www.researchgate.net/publication/348832943\\_Tumor\\_and\\_microcalcification\\_characterization\\_using\\_Entropy\\_Fractal\\_Dimension\\_and\\_intensity\\_values\\_statistical\\_analysis\\_in\\_mammography](https://www.researchgate.net/publication/348832943_Tumor_and_microcalcification_characterization_using_Entropy_Fractal_Dimension_and_intensity_values_statistical_analysis_in_mammography)