

ANÁLISIS ESTADÍSTICO DEL FACTOR DE RIESGO BASADO EN IMÁGENES DE
ULTRASONIDO DEL TEJIDO GLANDULAR MAMARIO.

SANDRA CAROLINA GÓMEZ ESPINAL
VICTOR MANUEL RODRIGUEZ DE LA HOZ

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICOMECÁNICAS
ESCUELA DE INGENIERÍAS ELÉCTRICA, ELECTRÓNICA Y DE
TELECOMUNICACIONES
BUCARAMANGA

2022

ANÁLISIS ESTADÍSTICO DEL FACTOR DE RIESGO BASADO EN IMÁGENES DE
ULTRASONIDO DEL TEJIDO GLANDULAR MAMARIO.

SANDRA CAROLINA GÓMEZ ESPINAL
VICTOR MANUEL RODRIGUEZ DE LA HOZ

Trabajo de Grado para optar al título de
Ingeniero Electrónico

Director
ANA BEATRIZ RAMIREZ SILVA
Ph.D. en Ingeniería Eléctrica

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICOMECÁNICAS
ESCUELA DE INGENIERÍAS ELÉCTRICA, ELECTRÓNICA Y DE
TELECOMUNICACIONES
BUCARAMANGA
2022

Dedicado

A mis padres Jorge y Julieta por su amor, su esfuerzo y por darme todo el apoyo que necesito. A mi hermana Sofía que me acompañó durante todo el proceso. A mi novio Andrés por su amor y motivación, a mi amigo Víctor con quien empecé y pude concluir esta etapa y para finalizar a mis amigos y amigas que hicieron parte de mi vida estos últimos años y me dieron el apoyo que necesitaba dentro y fuera de las aulas.

SANDRA CAROLINA GÓMEZ ESPINAL

Dedicado a todas aquellas personas que me brindaron apoyo en esta etapa de mi vida que finaliza llena de conocimientos y experiencias. Dedicado a mi mamá Adriana De La Hoz por apoyarme, brindarme sus consejos y darme toda la motivación necesaria, dedicado a mi padre Víctor Manuel Rodríguez Rojas por estar en todo momento, a Natalia Peña por entenderme y darme apoyo durante parte del proceso , a mi abuela Zorayda Rojas por apoyarme en cada momento de la carrera , a mi mejor amiga Sandra Gomez por su dedicación en el proyecto y cada ayuda dentro de este proceso, a mi abuelo Victor Rodriguez Martinez por creer en todo el proceso y apoyarme .

VÍCTOR MANUEL RÓDRIGUEZ DE LA HOZ

CONTENIDO

	pág.
INTRODUCCIÓN	10
JUSTIFICACIÓN	12
1. OBJETIVOS	13
2. MARCO TEÓRICO	14
2.1. ULTRASONIDO COMO TÉCNICA PARA ANÁLISIS DEL SENO	14
2.2. DENSIDAD MAMARIA COMO FACTOR DE RIESGO	16
2.3. BASES DE DATOS Y GENERACIÓN DE LAS IMÁGENES	18
3. DESARROLLO DEL PROYECTO	22
3.1. CREACIÓN DE LA BASE DE DATOS	22
3.2. SOFTWARE	23
3.3. HERRAMIENTA PARA EL ANÁLISIS DE FACTOR DE RIESGO	27
3.4. MODELO DE TYRER-CUZICK	29
4. RESULTADOS	31
5. CONCLUSIONES Y RECOMENDACIONES	37
BIBLIOGRAFÍA	39

LISTA DE FIGURAS

	pág.
Figura 1. Sistema de ultrasonido de mama completo en 3D	15
Figura 2. Posición del seno en el transductor con forma de anillo	16
Figura 3. Modelo de mama sintética construido a partir de una resonancia magnética de una mama cancerosa 33 con diferentes tejidos y matriz circular que contiene 157 transductores (puntos blancos).	19
Figura 4. Modelo de mama sintética en corte transversal generado para el desarrollo del proyecto indicando en el Index las diferentes velocidades del sonido (m/s) de cada tejido.	21
Figura 5. Seno predominantemente graso en copa B	24
Figura 6. Seno predominantemente graso en copa C	24
Figura 7. Seno predominantemente graso en copa D	24
Figura 8. Seno con densidades fibrogandulares en copa B	25
Figura 9. Seno con densidades fibrogandulares en copa C	25
Figura 10. Seno con densidades fibrogandulares en copa C	25
Figura 11. Seno heterogéneamente denso en copa B	26
Figura 12. Seno heterogéneamente denso en copa C	26
Figura 13. Seno heterogéneamente denso en copa D	26
Figura 14. Seno extremadamente denso en copa B	27
Figura 15. Seno extremadamente denso en copa C	27
Figura 16. Seno extremadamente denso en copa C	27
Figura 17. Factor de riesgo a largo plazo según la densidad del seno en copa B.	34

Figura 18. Factor de riesgo a largo plazo según la densidad del seno tamaño seno en copa C	34
Figura 19. Factor de riesgo a largo plazo según la densidad del seno en copa D	35
Figura 20. Factor de riesgo a 10 años según la densidad del seno en copa B.	35
Figura 21. Factor de riesgo a 10 años según la densidad del seno tamaño seno en copa C	36
Figura 22. Factor de riesgo a 10 años según la densidad del seno en copa D	36

LISTA DE TABLAS

	pág.
Tabla 1. Propiedades físicas de diferentes tejidos del seno y agua.	21
Tabla 2. Muestra de datos clasificados en la escala BI-RADS.	32
Tabla 3. Muestra de valores de entrada y resultados de la calculadora de evaluación de riesgos de Tyrer-Cuzick	33

RESUMEN

TÍTULO: ANÁLISIS ESTADÍSTICO DEL FACTOR DE RIESGO BASADO EN IMÁGENES DE ULTRASONIDO DEL TEJIDO GLANDULAR MAMARIO. *

AUTOR: SANDRA CAROLINA GÓMEZ ESPINAL, VÍCTOR MANUEL RODRIGUEZ DE LA HOZ **

PALABRAS CLAVE: BASES DE DATOS, ANÁLISIS DE IMÁGENES, ULTRASONIDO, CÁNCER DE SENO.

DESCRIPCIÓN:

El cáncer de mama es uno de los cinco tipos de cáncer con mayor tasa de mortalidad en el mundo, para poder reducir estas cifras es importante tener programas de prevención y detección temprana. Aunque hay varias técnicas de detección del cáncer, se enfocó en el ultrasonido ya que además de ser más eficiente es apta para senos densos. El objetivo del proyecto era identificar si la densidad del seno está relacionada con la probabilidad de tener cáncer de seno a futuro, entonces se analizó este factor y otros como la edad y su relación con la posibilidad de contraer la enfermedad. En este proyecto se generó una base de datos en Matlab de 150 imágenes sintéticas de senos 3D con diferentes densidades y volúmenes, se generó un código también en Matlab para clasificar cada seno en densidades según el método BI-RADS, para finalmente analizar el factor de riesgo de cada una de las imágenes a los próximos 10 años y a largo plazo con una herramienta online basada en el método de Tyrer-Cuzick. Con los resultados obtenidos se realiza un análisis de datos para calcular las correlaciones de las diferentes características del seno con el factor de riesgo obtenido.

* Trabajo de grado

** Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingenierías Eléctrica, Electrónica y telecomunicaciones. Director: Ana Beatriz Ramirez Silva Ph.D. en Ingeniería Eléctrica

ABSTRACT

TITLE: STATISTICAL ANALYSIS OF THE RISK FACTOR BASED ON IMAGES OF ULTRASOUND OF BREAST GLANDULAR TISSUE *

AUTHOR: SANDRA CAROLINA GÓMEZ ESPINAL, VICTOR MANUEL RODRIGUEZ DE LA HOZ **

KEYWORDS: IMAGE PROCESSING, DATABASES, ULTRASOUND, BREAST CANCER.

DESCRIPTION:

Breast cancer is one of the five types of cancer with the highest mortality rate in the world, in order to reduce these figures, it is important to have prevention and early detection programs. Although there are several cancer detection techniques, it focuses on ultrasound since, in addition to being more efficient, it is suitable for dense breasts. The objective of the project was to identify if the density of the breast is related to the probability of having breast cancer in the future, then this factor and others such as age and its relationship with the possibility of contracting the disease were analyzed. In this project, a database in Matlab of 150 synthetic images of 3D breasts with different densities and volumes will be presented, a code will also be closed in Matlab to classify each breast in densities according to the BI-RADS method, to finally analyze the factor of risk of each of the images in the next 10 years and in the long term with an online tool based on the Tyrer-Cuzick method. With the results obtained, a data analysis is carried out to calculate the correlations of the different characteristics of the breast with the risk factor obtained.

* Bachelor Thesis

** Faculty of Physical-Mechanical Engineering; School of Electrical, Electronic and Telecommunications Engineering. Director: Ana Beatriz Ramirez Silva Ph.D. en Ingeniería Eléctrica

INTRODUCCIÓN

Hay una gran variedad de enfermedades oncológicas que se pueden desarrollar en el cuerpo humano. Según la organización mundial de la salud (OMS), el cáncer es la principal causa de muerte en todo el mundo y el cáncer de mama está dentro de los 5 tipos de cáncer que causan el mayor número de fallecimientos en el mundo. Y según el Instituto Nacional de la Salud, en Colombia se detectaron 13376 casos nuevos en 2018 con una incidencia de 63.9 casos por cada 100000 mujeres mayores de 15 años y su tasa de muerte fue de 17.2 muertes por cada 100000 habitantes[15]. Aunque estas cifras parecen alarmantes, la detección temprana es la estrategia principal para reducir la mortalidad de mujeres, por cáncer de mama. Para esto, es fundamental que se genere más información sobre evidencia científica de los factores de riesgos, para que esta enfermedad pueda ser detectada en etapas tempranas y pueda ser tratada adecuadamente.

La identificación de los diferentes factores de riesgos es de vital importancia porque brinda la ventaja de conocer la población que podría ser más vulnerable a sufrir esta enfermedad. De esta manera se pueden crear estadísticas, cuantificar la información y generar métodos de prevención para mitigar el riesgo de que se desarrolle el cáncer. [14]

La mamografía es la prueba que más se utiliza actualmente para la búsqueda de tumores malignos que no se pueden detectar por medio del tacto. Consiste en generar una imagen de la mama por medio de rayos X, suponiendo que si hay un tejido maligno será más denso considerablemente que el tejido circundante. Sin embargo, esta técnica tiene ciertas desventajas que pueden tener repercusiones en el paciente como por ejemplo el sobre diagnóstico, falsos positivos, falsos negativos y el

cáncer inducido por radiación, además de que no funciona de la misma forma para todas las pacientes, ya que este método tiene menor efectividad para las mujeres jóvenes y para mujeres con senos densos.

Por esta razón, se ha optado por estudiar otros métodos, entre ellos el ultrasonido, que es un método no invasivo que ha demostrado alta efectividad, gracias a que permite un diagnóstico temprano, y además permite determinar las condiciones de una paciente con riesgo de cáncer de mama. Por medio del ultrasonido se pueden generar imágenes de reflectividad, velocidad del sonido del tejido, atenuación del sonido y densidad del tejido, entre otros. En este proyecto se plantea generar una base de datos de imágenes sintéticas de velocidad de sonido para senos con diferentes densidades y tamaños. A partir de esta base de datos, se propone un algoritmo que permita la estimación del tejido glandular mamario y su clasificación de acuerdo a la proporción de tejido denso y no denso en uno de los 4 niveles BI-RADS, para finalmente analizar el factor de riesgo.

JUSTIFICACIÓN

El cáncer de mama es el cáncer más común, de mayor prevalencia y el que genera más muertes en el mundo [5]. En Colombia, el cáncer de seno se ubica en los primeros dos lugares de causa de muertes en las mujeres[11]. La Organización Panamericana de la Salud afirmó que, según estadísticas en el año 2020, de 210.000 mujeres que fueron diagnosticadas en América latina y el Caribe, y fallecieron casi 68.000 [12], la cual es una cifra preocupante para la salud pública ya que cada vez va en aumento. Estas cifras tan elevadas son alertas para la comunidad científica enfocada en la salud y por tal motivo es un área de interés.

Esta enfermedad no distingue de estratos socio-económicos, sin embargo, en los países con mayores ingresos han logrado disminuir la mortalidad ya que se han enfocado en programas de detección temprana y protocolos de tratamientos estandarizados, a diferencia de los países de ingresos medios y bajos, ya que esta implementación de programas aún son un reto. El cáncer es una enfermedad que no se puede erradicar por completo por el momento, pero si se pueden establecer factores de riesgo que puedan disminuir la mortalidad. Por ejemplo, ya se puede afirmar que hay características propias del seno y de su anatomía que tienen mayor probabilidad de desarrollar el cáncer, esta y otras condiciones se van encontrando mediante el estudio de las imágenes para la prevención y esta capacidad de predecir si hay mayores probabilidades de que aparezca la enfermedad permite mejorar las estrategias preventivas y una mejor toma de decisiones clínicas para mejorar la calidad de vida del paciente.

Para la detección de cáncer de seno, se han estudiado diferentes técnicas, entre ellas el ultrasonido. Esta técnica ha demostrado no ser invasiva y además ha sido eficiente a la hora de generar un diagnóstico en pacientes jóvenes o con senos densos, a diferencia de los rayos X que son más costosos y utilizan radiación.

1. OBJETIVOS

Objetivo general

- Determinar el factor de riesgo de cáncer de seno, a partir de una base de datos de imágenes sintéticas de ultrasonido del tejido glandular mamario

Objetivos específicos

- Crear una base de datos de imágenes sintéticas de seno que contenga diferentes tipos y porcentajes de tejidos.
- Implementar un algoritmo que permita determinar el porcentaje de tejido glandular mamario presente en las imágenes sintéticas.
- Desarrollar un algoritmo para determinar el factor de riesgo de cáncer de seno a partir de los porcentajes de tejido glandular mamario presente en las imágenes sintéticas.

2. MARCO TEÓRICO

2.1. ULTRASONIDO COMO TÉCNICA PARA ANÁLISIS DEL SENO

El ultrasonido como examen médico de detección de cáncer de seno es una técnica de imágenes utilizada para el diagnóstico de bultos, anormalidades o lesiones mamarias. Es un método seguro, no invasivo y no utiliza radiación. El elemento principal de este examen es el transductor, este dispositivo es el encargado de enviar las ondas sonoras de altas frecuencias a través del seno que son las hacen posible la visualización de la imagen. Después de ser enviadas, el mismo elemento recibe las ondas que rebotan debido a los órganos internos, fluidos y diferentes tejidos en el interior, determinando así formas, tamaños y diferentes densidades. Simultáneamente esas ondas son convertidas en impulsos eléctricos, lo cual hace posible que se creen imágenes que se visualizan a tiempo real en una computadora [10]. Cuanto mayor sea la frecuencia (MHz) de la sonda, mayor va a ser la resolución de la imagen, pero menor va a ser la distancia que alcanzan las ondas.

Hay varios tipos de transductores. Están los transductores lineales para diagnósticos poco profundos, es decir, con alta resolución, sus frecuencias se emplean desde los 7.5 hasta los 20MHz. Los transductores convex son ligeramente curvos, se emplean para mayores profundidades por lo que trabajan a bajas frecuencias generando imágenes trapezoidales de menor resolución, las frecuencias que utilizan están entre los 3.5 y 5 MHz. Los transductores sectoriales emplean frecuencias entre 3.5 y 5 MHz, el origen del haz de ultrasonido es un punto único, por lo que permite mejor accesibilidad. Este proyecto se basó en un diseño de transductor en configuración de anillo. Este puede generar una imagen coronal tomográfica de 360 grados con el rebote de las ondas sonoras y además cuantifica las señales de velocidad y

atenuación del sonido transmitidas a través de la mama mostrando diferencias entre los tipos de tejidos. Como resultado se crea un conjunto de imágenes de sección transversal coronal de toda la mama.

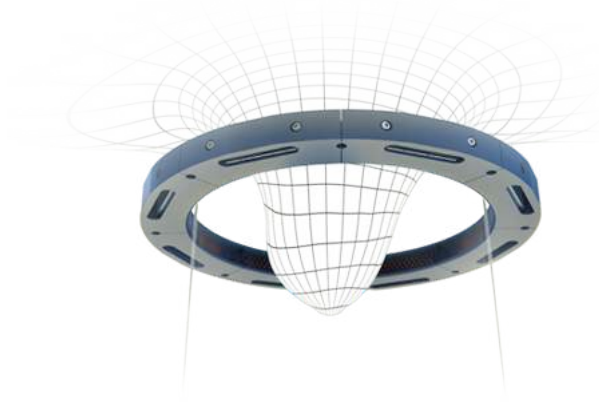
Para realizar el examen de ultrasonido, la paciente se acuesta en decúbito prono (posición corporal en donde la paciente se tiende boca abajo con la cabeza de lado) sobre una mesa como se muestra en la **Figura 1** con la mama sumergida dentro de un baño de agua tibia. En la **Figura 2** se observa como los sensores de ultrasonido en forma de anillo rodean el seno y son los que recopilan información gracias a los transductores.

Figura 1. Sistema de ultrasonido de mama completo en 3D



Fuente: Página web Delphinus medical technologies [Consultado: Noviembre de 2021] Disponible en: <https://www.delphinusmt.com/technology/>

Figura 2. Posición del seno en el transductor con forma de anillo



Fuente: Página web Delphinus medical technologies [Consultado: Noviembre de 2021] Disponible en: <https://www.delphinusmt.com/technology/triadtm-technology/>

2.2. DENSIDAD MAMARIA COMO FACTOR DE RIESGO

El tejido mamario está compuesto por tres componentes: El tejido fibroso, el tejido glandular y el tejido adiposo. El tejido fibroso (conectivo) desempeña principalmente las funciones mecánicas de relleno, armazón y sostén del seno. El tejido glandular es el componente funcional y corresponde a los lóbulos y conductos, el conjunto de estos dos tejidos es llamado tejido fibroglandular. Por último el tejido adiposo o tejido graso que le da el tamaño y forma al seno y es el rellena el espacio entre los otros tejidos, lóbulos y conductos. [6]

La densidad mamaria es un término que compara la cantidad del tejido fibroglandular con la cantidad del tejido graso del seno. Este concepto es relevante ya que según gran cantidad de estudios se puede afirmar que tener mamas densas aumenta las probabilidades de presentar cáncer de mama. McCormack y dos Santos Silva hicieron estudios utilizando más de 14000 mujeres con cáncer de mama y 226000 mujeres sin la enfermedad. Al final descubrieron que la densidad mamaria es un

factor fuertemente relacionado con el riesgo de cáncer de mama, es independiente a otros factores de riesgo y tiene un gradiente de riesgo mayor que los otros factores [2]. Otros estudios [3] demuestran que mujeres con densidades superiores al 75 % tienen mayor riesgo de cáncer de mama en comparación de las mujeres con densidades menores al 10 %. Se puede añadir que por algunos métodos, como por tomografía computarizada, es más difícil la detección temprana debido a las altas densidades, ya que no permite diferenciar fácilmente las células cancerosas de los otros tejidos. También se debe tener en cuenta que no es el único factor que puede influir: la edad, el peso o índice de masa corporal (IMC), hormonas, herencia, algunos factores epidemiológicos pueden afectar para que una persona desarrolle la enfermedad.

Hay 4 métodos principales para medir la densidad mamaria: Wolfe, Estimación visual de la proporción de mama ocupada por tejido mamario radiológicamente denso (Boyd), métodos de medición asistidos por computador basados en un umbral ubicado por un observador, entre la grasa y el tejido fibroglandular y por último el método BI-RADS.

BI-RADS (Breast Imaging Report and Database System), traduce como Sistema de Informes y Registro de Datos de Imagen de la Mama. Es una escala que representa cuantitativamente y estandariza el diagnóstico mamográfico para facilitar la comunicación entre los médicos, especialistas, radiólogos y pacientes. Este sistema categoriza la densidad en cuatro grados [4]:

1. Predominantemente grasa: se encuentra en aproximadamente 10 % de la población y la densidad mamaria está por debajo del 25 %.
2. Densidades fibroglandulares dispersas: se encuentra en aproximadamente el 40 % de la población y la densidad está entre 25 % y 50 %.

3. Heterogéneamente densa: se encuentra en aproximadamente 40 % de la población y la densidad está entre 50 % y 75 %.
4. Extremadamente densa: se encuentra en aproximadamente 10 % de la población y la densidad es mayor al 75 %.

La densidad mamaria es una característica que disminuye con la edad, lo cual es paradójico ya que el riesgo de cáncer también aumenta a mayor edad. Esto se puede explicar con el modelo de Pike que se basa en más "la exposición al tejido mamario" que en la edad de la mujer, esto se refiere al efecto de las variaciones hormonales que sufre una mujer a lo largo de su vida: la menarquía, embarazos, uso de anticonceptivos hormonales, menopausia y como esta exposición hace susceptible el tejido mamario a los carcinógenos. Esta predicción dice que la exposición acumulativa del tejido mamario hace que entre mayor edad haya mas probabilidad de incidencia para el cáncer de mama [13].

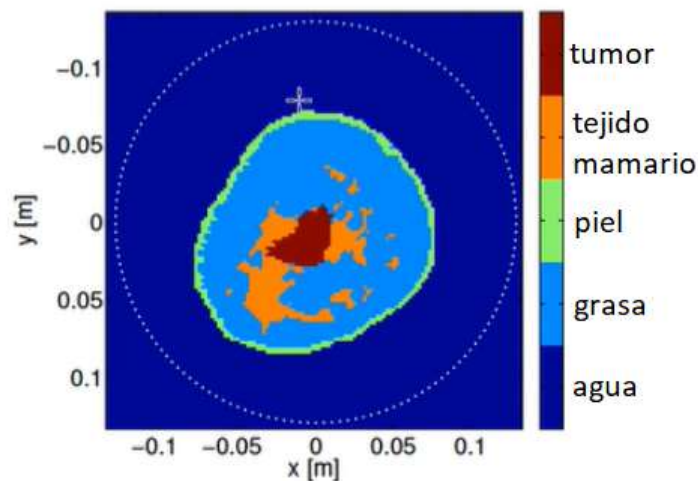
2.3. BASES DE DATOS Y GENERACIÓN DE LAS IMÁGENES

Para este proyecto se requieren bases de datos de imágenes de ultrasonido del seno de las cuales se van a generar los análisis y conclusiones. Esta tecnología es relativamente nueva para esta aplicación por lo cual no se obtuvieron bases de datos con información real a disposición del público en general, por esta razón se generó una base de datos de imágenes propias, con datos sintéticos que incluyen diferentes características del seno que representan una población pequeña y aproximada a la realidad según estadísticas.

Utilizar ondas sonoras de altas frecuencias es una modalidad para generar imágenes médicas y poder detectar y caracterizar lesiones mamarias. Hay métodos que consisten en obtener imágenes transversales de la mama, como se observa en

la **Figura 3** resolviendo la ecuación de onda para la reconstrucción, aprovechando información adjunta como los efectos de dispersión y difracción, estos métodos se denominan métodos de inversión de onda completa (FWI), entre ellos existe el método de inversión en dominio de tiempo (TDI) y el método de inversión en dominio de frecuencia (CSI). Una de las ventajas de las imágenes generadas por los métodos de FWI son es que proporcionan imágenes de mejor calidad gracias a los parámetros del medio acústico que tienen los diferentes tejidos.

Figura 3. Modelo de mama sintética construido a partir de una resonancia magnética de una mama cancerosa 33 con diferentes tejidos y matriz circular que contiene 157 transductores (puntos blancos).



Para reconstruir el perfil de sonido de una mama sintética fue utilizado el método de la aproximación en el dominio del tiempo en diferencias finitas (FDTD). El método numérico que se utiliza para reconstruir la velocidad 2-D del perfil de sonido de la mama sintética lleva consigo la resolución de ciertas ecuaciones integrales, uso de la función de Green 2-D y con esto se descartan las singularidades presentes en esta función. Los datos que se consideran sintéticos y hacen parte del dominio de la frecuencia se calculan resolviendo el modelo de problema directo con un método de

onda completa y este hace uso del esquema de gradiente conjugado. Se puede decir que los métodos TDI y CSI producen imágenes bastante precisas de la velocidad del sonido de la mama en escenarios sin ruido teniendo una configuración completa de fuentes y receptores.

Con el fin de que la información de la base de datos fuera una representación de la mama de la población femenina se tuvieron en cuenta los porcentajes mencionados en el método BI-RADS y además se tuvieron en cuenta 3 diferentes tamaños posibles del seno. Se generaron 150 imágenes en total, 50 de senos talla B, 50 de senos talla C y 50 de senos talla D. De cada talla se dividieron por densidades: un 10% con senos que cumplen la característica de predominantemente grasa (5 imágenes), un 40% con densidades fibroglandulares dispersas (20 imágenes), un 40% con imágenes de senos que son heterogéneamente densos (20 imágenes) y el último 10% con senos extremadamente densos (5 imágenes).

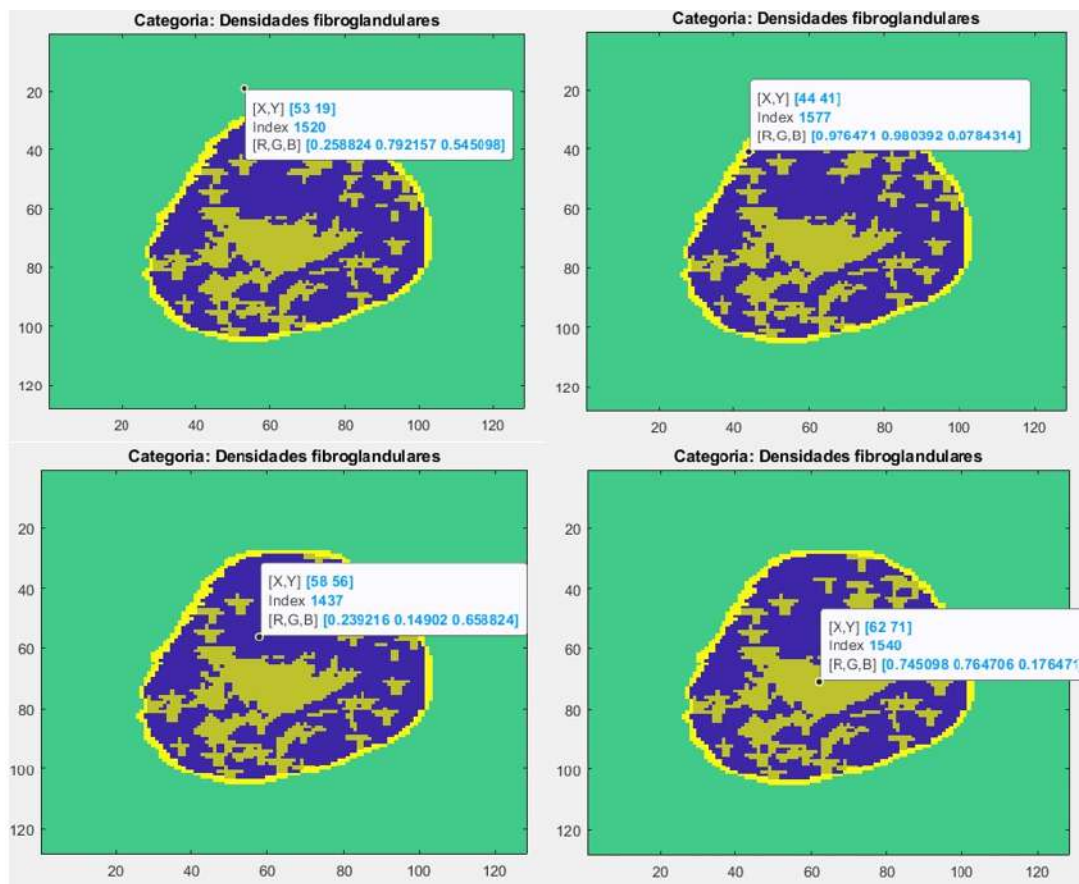
En la **Figura 4** se muestra un ejemplo de las imágenes sintéticas utilizadas y en cada una se señala un punto en específico. En el campo index se evidencia la velocidad del sonido que le pertenece a cada punto, es decir, que con ayuda de los valores de la Tabla 1 se puede determinar que tipo de tejido esta señalando.

El dominio espacial de estas imágenes contienen 128x128 elementos cuadrados de 2.2mm de resolución espacial y el tiempo de duración de cada scan en el transductor es de 552 μs . Cada transductor dentro del anillo actúa como una fuente puntual y genera un campo de ondas modulado gaussiano con una frecuencia central de $f_0 \approx 90kHz$.

	Densidad [kg/m^3]	Velocidad del Sonido [m/s]
Tejido mamario	1020	1540
Piel	1200	1577
Grasa	950	1437
Músculo	1040	1580
Tumor	1000	1572
Agua	993	1520

Tabla 1. Propiedades físicas de diferentes tejidos del seno y agua.

Figura 4. Modelo de mama sintética en corte transversal generado para el desarrollo del proyecto indicando en el Index las diferentes velocidades del sonido (m/s) de cada tejido.



3. DESARROLLO DEL PROYECTO

3.1. CREACIÓN DE LA BASE DE DATOS

Se utilizó una herramienta del grupo de investigación CPS, la cual consiste en generar imágenes de modelos de senos sintéticos en la que se le pueden variar las características como la densidad mamaria, el factor de escala y volumen en general, esta herramienta se basa en un tipo de escaneo de imagen el cual es imagen de resonancia magnética (MRI Scan), se generaron 150 imágenes con 3 tipos de tamaños diferentes en partes iguales y cada tamaño se subdivide en los cuatro tipos de densidades mamarias que establece el método BI-RADS, con el fin de hacer un proceso estadístico efectivo y sin sesgos con el objetivo de determinar los principales factores de riesgos. Se debe tener en cuenta que estas son imágenes para predecir el riesgo de cáncer, es decir, ninguna imagen contiene la enfermedad.

Este código se basa a partir de la teoría de un artículo de ultrasonido para el seno [1] y se construyen dos modelos 3D heterogéneos que son de la anatomía mamaria. Para esto se hizo uso de una toolbox de segmentación para delinear los MRI o valores grises en los tejidos con sus valores de velocidad de sonido correspondientes para el agua, músculo, grasa, piel y tejido mamario. El código recibe un archivo de voxels que tiene todas las coordenadas y para cada coordenada tiene unas propiedades que indica si es grasa, musculo, piel y esa genera las diferentes proporciones entre la grasa y el tejido glandular.

Este código nos permite determinar el tamaño del seno que se quiere generar, para este caso se utilizaron 3 tamaños diferentes del seno los cuales clasificamos en copa B, copa C y copa D. Los otros parámetros que determinan los tejidos se van generando dentro del código siendo valores variantes.

3.2. SOFTWARE

Teniendo la información del seno sintético generado en formato de matriz .mat, se procede a cargar los datos a una nueva variable en la herramienta MATLAB para descomponerlos en las 3 dimensiones de la matriz.

Todas las imágenes contienen 128x128 píxeles y dentro de la matriz se almacenan los datos de 88 imágenes. Estas imágenes son los cortes transversales de una imagen de ultrasonido del seno que al ser sumados forman un volumen (el de un solo seno). El código implementado funciona para el análisis de una imagen 3D. Inicialmente se hace un recorrido por las 88 imágenes con el objetivo de hacer un barrido de todos los píxeles que la componen para identificar la velocidad del sonido, lo cual permite determinar a que parte del tejido pertenece cada píxel. Cada valor de velocidad de cada píxel se va almacenando para que cuando se hayan recorrido el total de píxeles de las 88 imágenes, se tenga la información exacta de la cantidad de píxeles que pertenecen a agua, tejido mamario, grasa y piel.

Con la información de los píxeles también se obtuvo el área del seno, excluyendo los que contenían agua. Para esto se hizo binaria la imagen y se apartaron los píxeles que contenían agua a los que tenían algún tipo de tejido y calculamos el área con la librería Bwarea, la cual permite determinar el área de objetos en imagen binaria [9]. Finalmente se hace la suma de todas las áreas y obtener el volumen total del seno en términos de píxeles y teniendo estos resultados se obtiene el porcentaje de tejido glandular mamario. El porcentaje de tejido glandular mamario corresponde a la relación entre la suma de tejido mamario y la suma total del tejidos presentes en el seno.

Por último, se clasifica la imagen completa en una de las 4 categorías del método BI-RADS que depende del porcentaje del tejido glandular: 1. Predominantemente grasa, 2. densidades fibroglandulares, 3. heterogeneamente densa o 4. extremadamente densa

En las **Figuras 5, 6 y 7** se evidencia un ejemplo de las imágenes transversales que pertenecen a la categoría predominantemente grasa en tres tamaños diferentes: copa B, copa C y copa D.

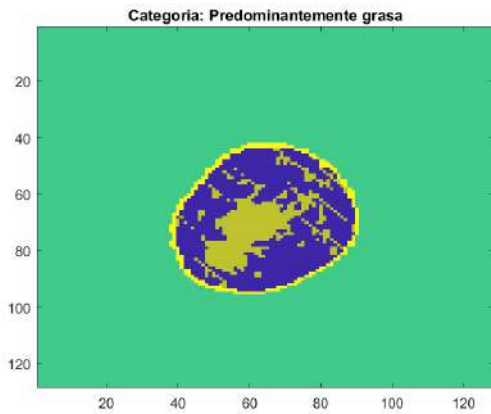


Figura 5. Seno predominantemente graso en copa B

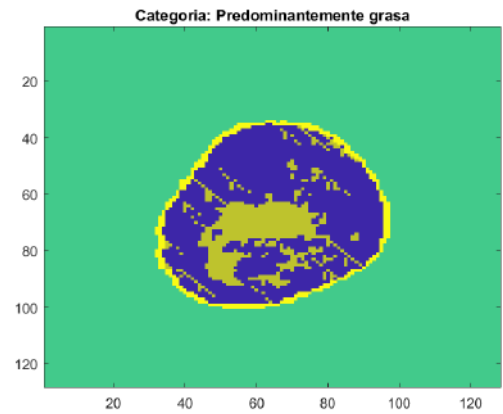


Figura 6. Seno predominantemente graso en copa C

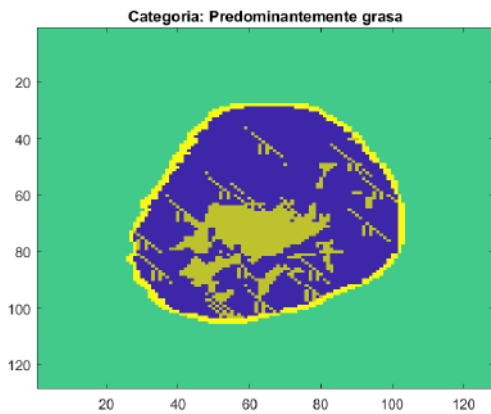


Figura 7. Seno predominantemente graso en copa D

En las **Figuras 8, 9 y 10** se evidencia un ejemplo de las imágenes transversales que pertenecen a la categoría densidades fibroglandulares dispersas en los tres tamaños diferentes.

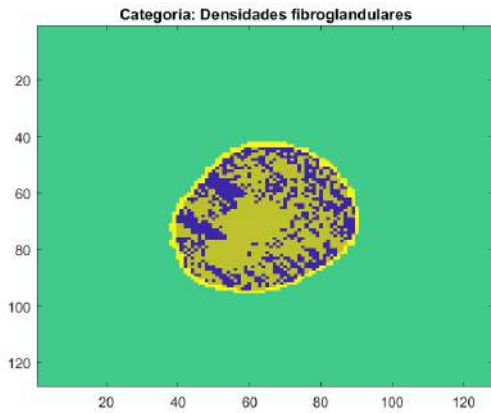


Figura 8. Seno con densidades fibroglandulares en copa B

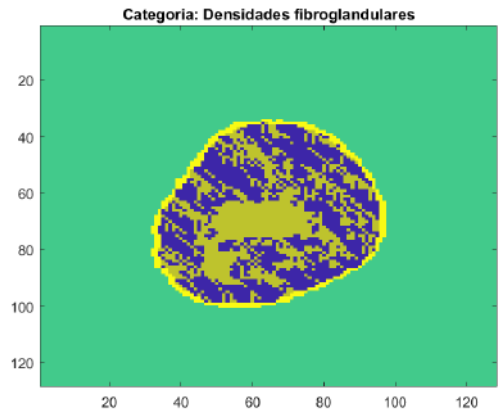


Figura 9. Seno con densidades fibroglandulares en copa C

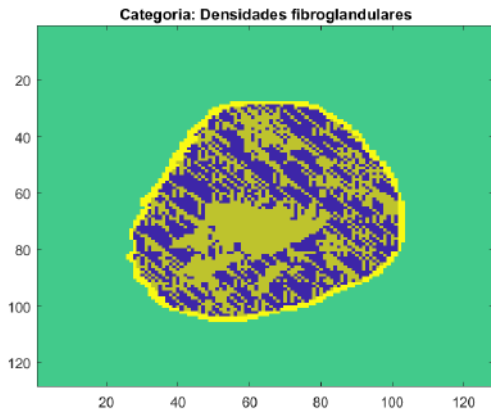


Figura 10. Seno con densidades fibroglandulares en copa C

En las **Figuras 11, 12 y 13** se evidencia un ejemplo de las imágenes transversales que pertenecen a la categoría heterogeneamente densas en los tres tamaños diferentes.

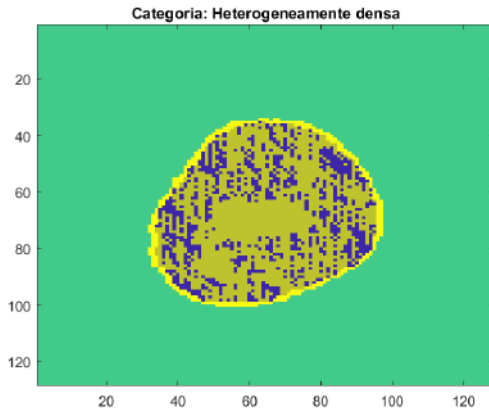


Figura 11. Seno heterogéneamente denso en copa B

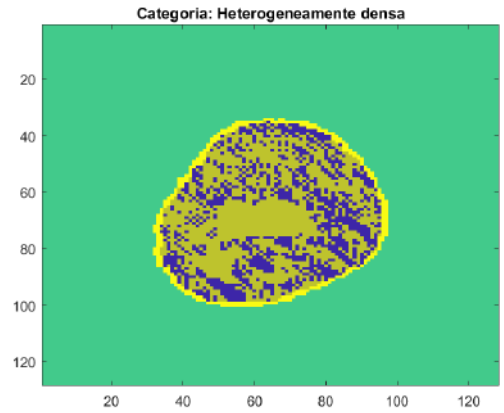


Figura 12. Seno heterogéneamente denso en copa C

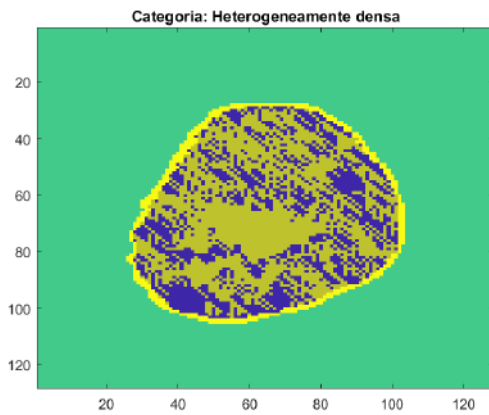


Figura 13. Seno heterogéneamente denso en copa D

En las **Figuras 14, 15 y 16** se evidencia un ejemplo de las imágenes transversales que pertenecen a la categoría extremadamente densas en los tres tamaños diferentes.

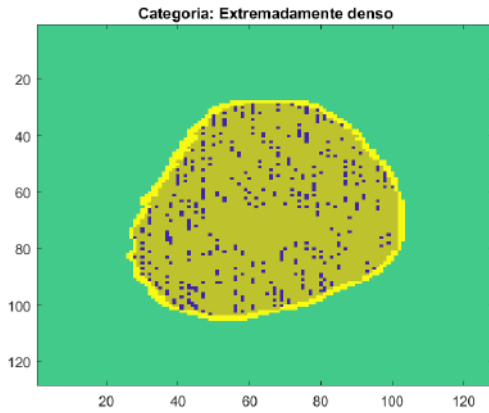


Figura 14. Seno extremadamente denso en copa B

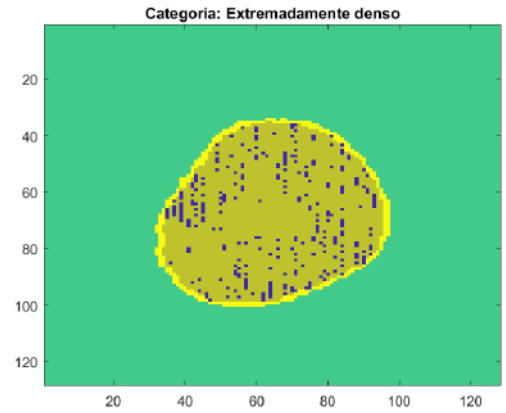


Figura 15. Seno extremadamente denso en copa C

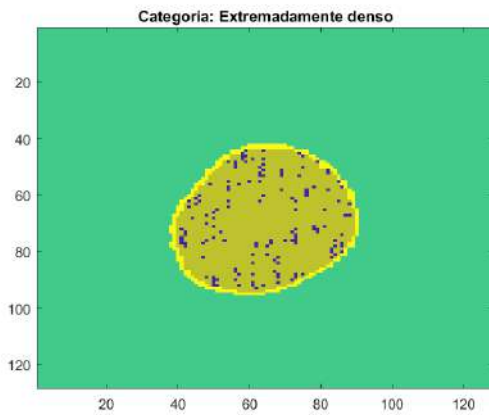


Figura 16. Seno extremadamente denso en copa C

3.3. HERRAMIENTA PARA EL ANÁLISIS DE FACTOR DE RIESGO

Existen herramientas basadas en diferentes modelos que dan como resultado el factor de riesgo de que una persona pueda desarrollar la enfermedad a futuro. Se analizaron tres modelos para al final escoger uno con el cual haríamos el procedimiento:

El modelo Claus calcula el riesgo en una persona sana pero es basado en la historia

familiar, incluye antecedentes de primer y segundo grado[7].

El modelo Gail estima el factor de riesgo a cinco años con factores personales como la edad, edad de menarquía, de primer embarazo, raza, variables hormonales, cantidad de biopsias de mama, solo tiene en cuenta familiares de primer grado con cáncer de mama y a diferencia del Claus y Tyrer-Cuzick no tiene en cuenta el riesgo de BRCA1 y BRCA2 que son genes heredados que pueden traer variantes dañinas los cuales aumentan significativamente el riesgo de cáncer de seno y de ovario. [8]

El modelo mas completo es el Tyrer-Cuzick, es un modelo inglés que puede predecir el riesgo e incluye factores reproductivos, mutación de gene BRCA1-2, incluye historia familiar de hasta tercer grado y factores personales como edad, menarquía, raza, edad de primer parto, menopausia, índice de masa corporal y ascendencia.

Como la necesidad más importante era poder incluir el porcentaje de densidad mamaria, el modelo Claus no incluye esa información de entrada entonces fue descartado de primero, el modelo Gail se le incluyó esta característica después de un tiempo pero a términos generales Tyrer-Cuzick, que es un modelo relativamente reciente, incluye los factores que se trabajaron, además es el más completo de los tres.

Se determinó una herramienta hecha por MagView, una empresa que lidera la industria en soluciones de software de informes y seguimiento de mamografías. Esta calculadora consiste en introducir unos valores de entrada y dependiendo de estos da dos tipos de resultados: el factor de riesgo a 10 años y el factor de riesgo a largo plazo. Esta calculadora tiene diferentes campos de entrada determinados por el modelo de Tyrer-Cuzick, pero para este caso solo se utilizaron los campos del porcentaje de tejido glandular mamario, edad, altura, peso, edad en la que tuvo su el

primer hijo y la menarquía.

Para hacer una representación poblacional se les dieron características de mujeres a cada imagen de cada seno generada de manera sintética, ya teniendo los porcentajes de tejido glandular mamario se les estableció los otros aspectos que necesitaban para poder introducir en la calculadora. Para la edad de escogieron valores entre 20 y 70 años, como eran 150 imágenes en total hay 3 personas de cada edad para tener un balance en las edades. La estatura y el peso fueron números generados aleatoriamente con ayuda de excel dentro de los rangos 1.50 - 1.75 metros y 50 - 80 kilogramos respectivamente. Se estableció que la mitad de los volúmenes provenían de mujeres sin hijos y la otra mitad de mujeres con al menos un hijo y que todas habían tenido su primer periodo menstrual a los 12 años. De esta manera ya había información para hacer los cálculos de probabilidad en la herramienta.

3.4. MODELO DE TYRER-CUZICK

Este modelo tiene como función estimar el factor de riesgo de que una persona desarrolle cáncer de mama en los siguientes 10 años y a lo largo de la vida. El modelo de Tyrer-Cuzick tiene como resultado una puntuación que determina el riesgo que una persona pueda desarrollar mutaciones genéticas específicas relacionadas con el cáncer de mama. Esta estimación ha sido utilizada por los médicos para determinar el nivel de riesgo de una persona de desarrollar la enfermedad. Estas pruebas ayudan a identificar a las personas con un riesgo de cáncer de mama superior al promedio, después de esto el médico puede recomendar pruebas de detección ó pruebas genéticas adicionales a esas personas.

Para poder realizar el cálculo de esta puntuación para una persona, el modelo va a considerar el historial de salud familiar y otros factores de riesgo que son poten-

ciales. Varios de los factores que pueden influir en la puntuación de Tyrer-Cuzick en un individuo son la edad, la altura, el peso ,resultados de biopsias mamarias y antecedentes personales de cáncer de mama o de ovario, raza, la menarquía, edad en la que persona dio a luz por primera vez (si ya ha tenido hijos) y por ultimo la densidad mamaria la cual es la intención de este proyecto.

Normalmente se considera que una persona que en este examen indica una puntuación menor de 15% tiene un riesgo promedio de desarrollar cáncer de mama. Esta puntuación indica que la persona no necesita necesariamente pruebas adicionales, además de las generalmente recomendadas. El Ministerio de Salud y Protección Social de la República de Colombia recomienda que las mujeres mayores de 20 años se realicen un examen clínico de la mama por lo menos cada tres años y después de los 40 años, que el examen sea una vez por año.

4. RESULTADOS

Los primeros resultados se generan de la toma de las 150 imágenes del banco las cuales pasaron por un software que determina el porcentaje de la densidad mamaria de cada una de ellas. Con este porcentaje se logró clasificar cada imagen en una de los cuatro tipos de seno en la escala BI-RADS. En la Tabla 2 se presenta una muestra de los resultados para evidenciar los datos obtenidos.

Al tener cada imagen clasificada, se utilizó la calculadora de evaluación de riesgos de Tyrer-Cuzick ya mencionada que tiene como valores de entrada el porcentaje de tejido glandular mamario y otras características de una persona determinada como edad, peso, altura, edad con la que tuvieron el primer hijo y menarquía, para generar los resultados del porcentaje de la probabilidad de que en 10 años esa persona pueda desarrollar cáncer de seno y del porcentaje de la probabilidad de que esa misma persona desarrolle cáncer en el resto de su vida. En la Tabla 3 se pone una muestra de los valores de entrada con los resultados para evidenciar los datos obtenidos. Teniendo ya los resultados, se hicieron una serie de análisis de datos para poder entender mejor el comportamiento.

Se quiere saber si hay una relación simple entre los valores de entrada con los resultados, en la estadística las correlaciones evidencian como la relación lineal puede ser positiva: es cuando ambos valores tienden a incrementar juntos, negativa: cuando uno de los valores incrementa el otro descende y nula: cuando esta relación lineal es débil y significa que no están asociadas entre ellas.

Nombre	Porcentaje Tejido Glandular Mamario	Número de Píxeles (Tamaño del seno)	Categoría BI-RADS
Volumen 1	23.03	112660	1.Predominantemente Grasa
Volumen 2	34.65	112660	2.Densidades Fibroglandulares
Volumen 3	66.95	112660	3.Heterogeneamente Densa
Volumen 4	75.25	112660	4.Extremadamente Densa
Volumen 5	17.64	179186	1.Predominantemente Grasa
Volumen 6	41.30	179186	2.Densidades Fibroglandulares
Volumen 7	52.83	179186	3.Heterogeneamente Densa
Volumen 8	75.88	179186	4.Extremadamente Densa
Volumen 9	10.59	269191	1.Predominantemente Grasa
Volumen 10	29.41	269191	2.Densidades Fibroglandulares
Volumen 11	65.72	269191	3.Heterogeneamente Densa
Volumen 12	76.24	269191	4.Extremadamente Densa

Tabla 2. Muestra de datos clasificados en la escala BI-RADS.

- La correlación entre el porcentaje de tejido glandular mamario y la probabilidad de padecer cáncer fue de 0.683, es decir, es positiva e indica que los valores de estas dos variables tienen una relación directa. Esto afirma que entre más denso es el seno, mayor probabilidad hay de desarrollar el cáncer a futuro.
- La correlación entre la edad y la probabilidad de padecer cáncer fue de 0.7911, es decir, al ser positiva e indica que a mayor edad, mayor es la probabilidad de padecer la enfermedad.
- La correlación entre peso y porcentaje de tejido glandular mamario fue de 0.145, como es un valor cercano a cero podemos decir que hay poca relación entre estas dos variables, lo mismo para el caso de la relación de la altura con el porcentaje de tejido glandular mamario cuyo resultado fue 0.124.

De la información que nos generó la calculadora de evaluación de riesgos de Tyrer-Cuzick también se obtuvieron los resultados del factor de riesgo de los 3 tamaños diferentes dependiendo del tejido glandular mamario, los cuales se pueden observar en las Figuras **Figuras 17, 18, 19.**

Nombre	Edad	Peso [kg]	Estatura [m]	Categoría BI-RADS	Edad Primer Hijo	Menarquía	Riesgo 10 años [%]	Riesgo a largo plazo [%]
Volumen1	66	60	1.68	1	28	12	1.4	3.0
Volumen2	69	79	1.58	2	-	12	3.2	4.8
Volumen3	47	64	1.56	3	-	12	2.5	10.6
Volumen4	27	71	1.69	4	28	12	0.3	12.1
Volumen5	65	68	1.68	1	-	12	1.9	3.5
Volumen6	43	72	1.52	2	-	12	1.5	8.1
Volumen7	51	76	1.63	3	28	12	3.4	11.6
Volumen8	46	64	1.59	4	-	12	3.3	14.0
Volumen9	64	65	1.50	1	-	12	2.1	3.9
Volumen10	40	80	1.74	2	-	12	1.3	9.1
Volumen11	35	80	1.74	3	-	12	1.3	14.2
Volumen12	21	58	1.61	4	18	12	0.1	8.8

Tabla 3. Muestra de valores de entrada y resultados de la calculadora de evaluación de riesgos de Tyrer-Cuzick

Gracias a la regresión lineal en color naranja que se visualiza en las figuras, podemos evidenciar el mismo patrón: cuando el porcentaje de tejido glandular mamario es bajo, el factor de riesgo de desarrollar la enfermedad a largo plazo es menor que cuando el tejido tiende a ser más denso. Se intuye de estos resultados que uno de los parámetros más predominantes es el tejido glandular mamario.

Con estos resultados se puede afirmar que a mayor densidad del tejido glandular mamario, el factor de riesgo es más alto para desarrollar el cáncer de seno.

La calculadora también genera el resultado de el factor de riesgo de en el intervalo de los próximos diez años. Estos resultados no dejan en evidencia la afirmación anterior debido a que hay otros factores que también afectan los resultados como la edad, el índice de masa corporal, genética. Por ejemplo las mujeres más jóvenes no habrán pasado por la menopausia lo cual es una alteración hormonal que podría afectar negativamente, entre otros factores, todo esto limita un poco el resultado a

diferencia de las gráficas a largo plazo. Estas imágenes del factor de riesgo a 10 años en los tres tamaños diferentes las podemos visualizar en las **Figuras 20, 21, 22.**

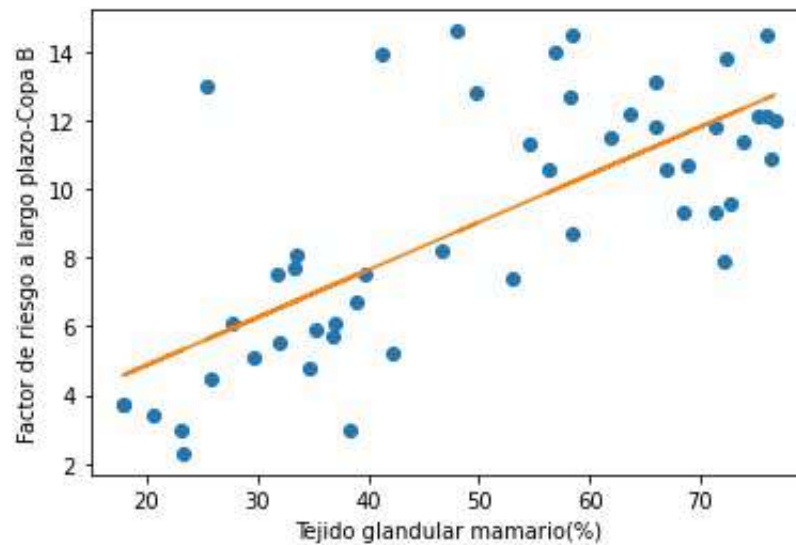


Figura 17. Factor de riesgo a largo plazo según la densidad del seno en copa B.

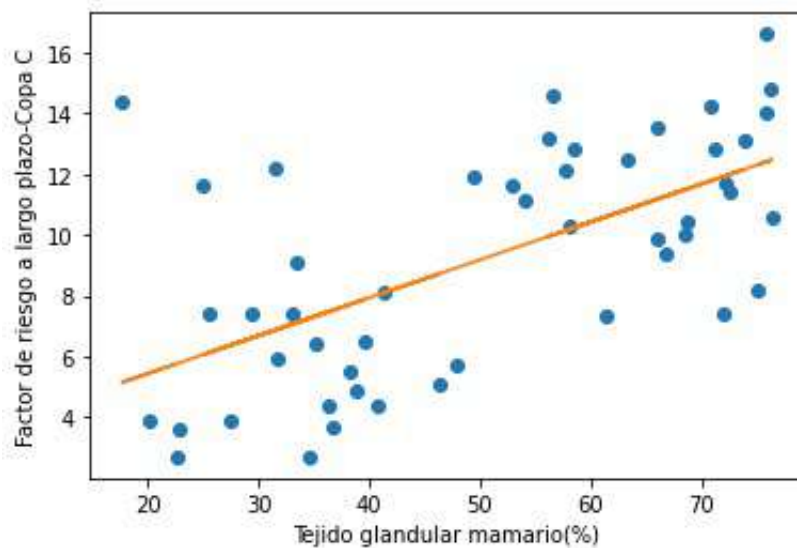


Figura 18. Factor de riesgo a largo plazo según la densidad del seno tamaño seno en copa C

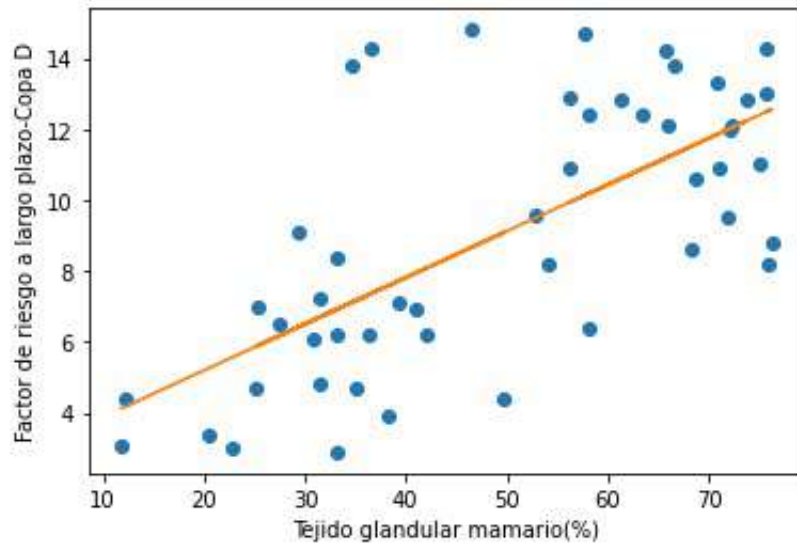


Figura 19. Factor de riesgo a largo plazo según la densidad del seno en copa D

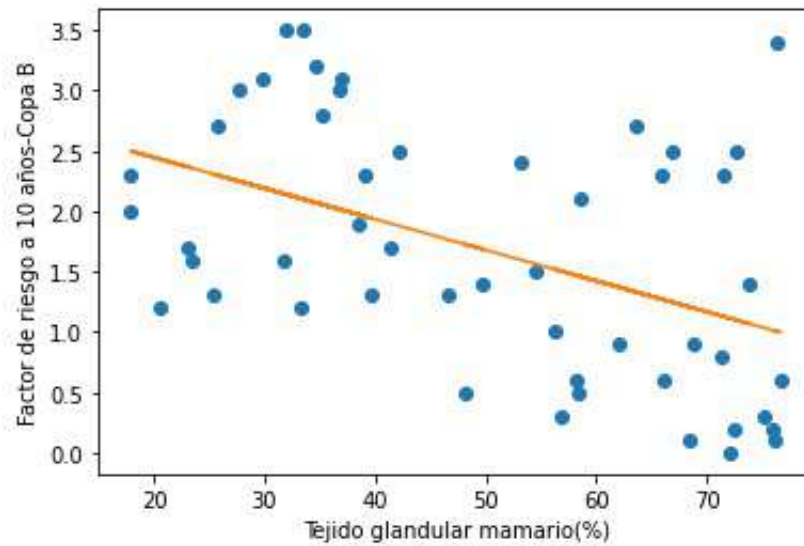


Figura 20. Factor de riesgo a 10 años según la densidad del seno en copa B.

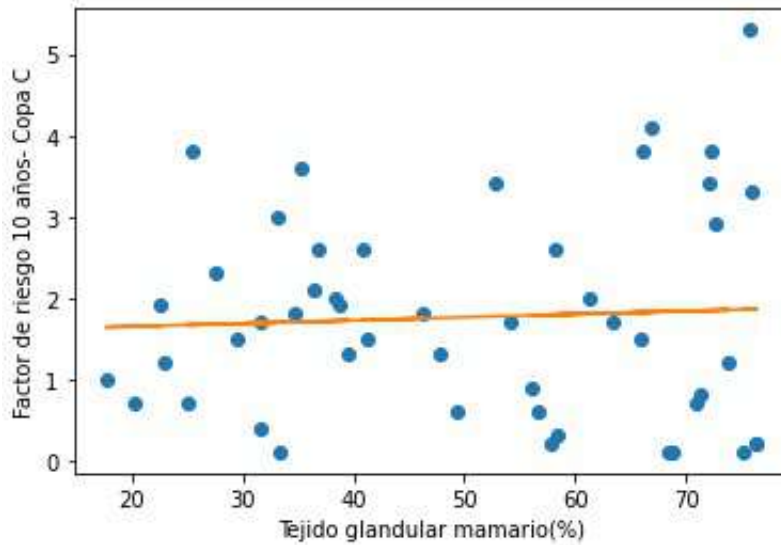


Figura 21. Factor de riesgo a 10 años según la densidad del seno tamaño seno en copa C

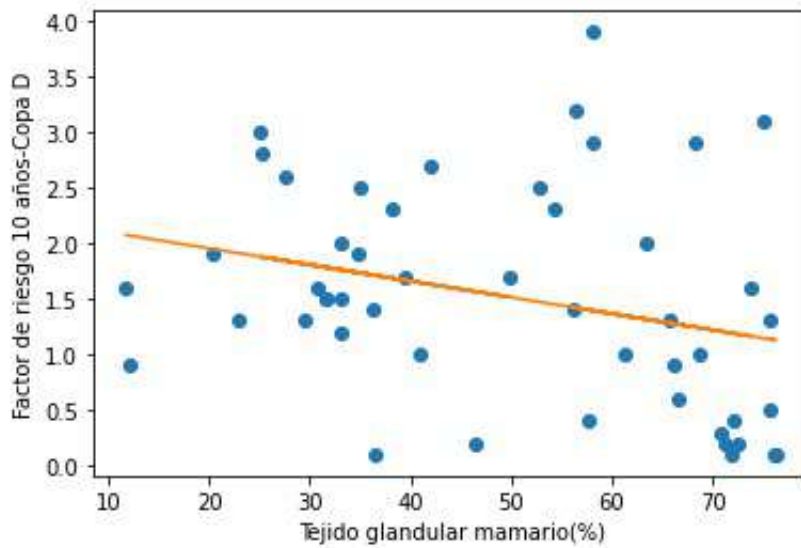


Figura 22. Factor de riesgo a 10 años según la densidad del seno en copa D

5. CONCLUSIONES Y RECOMENDACIONES

Utilizar herramientas como la generación de imágenes de los modelos sintéticos de senos nos permitió manipular las imágenes, determinar los tipos de tejidos, hacer variaciones de tamaños del seno todo esto para al final tener una base de datos completa y variada.

Una de los procesos más importantes para realizar el cálculo de la densidad mamaria fue poder hacer volúmenes con la sumatoria de las áreas, y con ayuda de un programa convertir las imágenes a binarias para diferenciar los tejidos del agua y poder hacer la sumatoria de píxeles necesaria para obtener el porcentaje de tejido glandular mamario.

Se comprobó que el tejido glandular mamario representa un factor de riesgo para desarrollar cáncer de seno debido a las altas correlaciones que se dieron entre este valor y los resultados que arroja el software de cálculo de factor de riesgo.

Las mujeres con senos más densos son más propensas a desarrollar cáncer de seno.

Se debe tener en cuenta que la densidad no es el único factor, hay otras variables que también influyen fuertemente en el resultado como la edad de las pacientes ya que al estar más expuestas a cambios hormonales como menarquía, embarazos y menopausia genera alteraciones y aumenta el factor de riesgo.

El peso y la altura no están tan relacionados con el riesgo de desarrollar cáncer como si lo tiene el porcentaje de tejido glandular mamario y la edad. Esto debido a

que los resultados finales no tienen una correlación alta con respecto a estos dos parámetros.

Se comprobó que hay una gran diferencia entre el factor de riesgo que tiene una mujer para desarrollar cáncer de mama calculado para el tiempo actual, a calcular el factor de riesgo de desarrollar cáncer de mama en 10 años. Esto se debe a que en el riesgo a largo plazo se incluyen más cambios y alteraciones hormonales y se ve más reflejada la edad que en el resultado de los siguientes 10 años.

Se dejó en evidencia con los dos resultados positivos de las dos correlaciones (porcentaje de tejido glandular mamario con factor de riesgo y edad con factor de riesgo), que la probabilidad de desarrollar la enfermedad aumenta a con la edad inclusive si el tejido glandular mamario va disminuyendo con el tiempo

BIBLIOGRAFÍA

- [1] Sohrab Behnia y col. "ULTRASOUND THERMOTHERAPY OF BREAST: THEORETICAL DESIGN OF THE TRANSDUCER AND NUMERICAL SIMULATION OF THE PROCEDURE". En: *Ultraschall in der Medizin - European Journal of Ultrasound* 26 (sep. de 2005). DOI: 10.1055/s-2005-917369 (vid. pág. 22).
- [2] Martin L.J. Yaffe M. et al. Boyd N.F. "Mammographic density". En: *Breast Cancer Research* 11, S4 <https://doi.org/10.1186/bcr2423> (2009) (vid. pág. 17).
- [3] Martin LJ Sun L Stone J Fishell E Jong RA Hislop G Chiarelli A Minkin S Yaffe MJ. Boyd NF Guo H. "Mammographic density and the risk and detection of breast cancer". En: *The New England journal of medicine*, 356(3) // <https://doi.org/10.1056/NEJMoa062790> (2007), 227–236 (vid. pág. 17).
- [4] Melcior Sentís i Crivellé. "An approach to breast density". En: *Journal of Breast Science* <https://www.elsevier.es/es-revista-revista-senologia-patologia-mamaria-131-articulo-la-densidad-mamaria-una-aproximacion-S0214158214000140> (2014) (vid. pág. 17).
- [5] "Cáncer". En: <https://www.who.int/es/news-room/fact-sheets/detail/cancer> (2021) (vid. pág. 12).
- [6] "Densidad de los senos e informe de su mamograma". En: *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> (2022) (vid. pág. 16).
- [7] YENI RODRÍGUEZ; CLARENA ZULUAGA; SANDRA XIMENA FRANCO. "Valoración del riesgo de cáncer de seno y probabilidad de mutaciones en genes

- BRCA 1-2". En: *REVISTA MÉDICA / CLÍNICA DEL COUNTRY* (2013) (vid. pág. 28).
- [8] Fernando Gómez D. Cirujano Oncólogo de Mama. "Evaluación cuantitativa del riesgo de cáncer de mama". En: *Rev. Med. Clin. Condes* (2006) (vid. pág. 28).
- [9] Mathworks: bware. "<https://la.mathworks.com/help/images/ref/bwarea.html>". En: () (vid. pág. 23).
- [10] Michael Bronskill Martin J. Yaffe Neb Duric Salomon Minkin Norman F. Boyd Lisa J. Martin. "Breast Tissue Composition and Susceptibility to Breast Cancer". En: *Journal of the National Cancer Institute, Volume 102, Issue 16* (2010), 1224–1237 (vid. pág. 14).
- [11] Observatorio Nacional de Cáncer, Guía Metodológica. "<https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/ED/GCFI/guia-ross-cancer.pdf>". En: () (vid. pág. 12).
- [12] Organización panamericana de la salud. "<https://www.paho.org/es/temas/cancer-mama>". En: () (vid. pág. 12).
- [13] V. Paulina Neira. "Densidad mamaria y riesgo de cáncer mamario". En: *Revista Médica Clínica Las Condes* 24 (2013), págs. 122-130. URL: <https://www.sciencedirect.com/science/article/pii/S0716864013701378> (vid. pág. 18).
- [14] "Prevención y seguimiento del cáncer de mama, según categorización de factores de riesgo y nivel de atención". En: *Revista Médica Clínica Las Condes* (2013), págs. 578,587 (vid. pág. 10).
- [15] Instituto Nacional de Salud. "INFORME DE EVENTO CÁNCER DE MAMA Y CUELLO UTERINO EN COLOMBIA, 2018". En: (2018), pág. 3 (vid. pág. 10).