



Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias Físico Matemáticas

Caracterización de tumores en mamografías utilizando la
métrica Euclidiana y de Manhattan

Tesis presentada al

Colegio de Física

como requisito parcial para la obtención del grado de

LICENCIADA EN FÍSICA

por

Lizbeth Giovana Rodríguez Sánchez

Asesorada por

Dr. Cristian Heber Zepeda Fernández

Dr. José Gerardo Suárez García

Puebla Pue.
9 de noviembre de 2023



Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias Físico Matemáticas

Caracterización de tumores en mamografías utilizando la
métrica Euclidiana y de Manhattan

Tesis presentada al

Colegio de Física

como requisito parcial para la obtención del grado de

LICENCIADA EN FÍSICA

por

Lizbeth Giovana Rodríguez Sánchez

Asesorada por

Dr. Cristian Heber Zepeda Fernández

Dr. José Gerardo Suárez García

Puebla Pue.
9 de noviembre de 2023

Título: Caracterización de tumores en mamografías utilizando la métrica Euclidiana y de Manhattan

Estudiante: LIZBETH GIOVANA RODRÍGUEZ SÁNCHEZ

COMITÉ

Dr. Eduardo Moreno Barbosa
Presidente

Dr. Benito De Celis Alonso
Secretario

Dr. Lucio Fidel Rebolledo Herrera
Vocal

M. C. Margarita Amaro Aranda
Vocal

Dr. Cristian Heber Zepeda Fernández
Dr. José Gerardo Suárez García
Asesores

Dedicatoria

*"Si he visto más lejos, es poniéndome sobre los hombros de Gigantes".
Sir Isaac Newton, 1675*

Mis padres, José Antonio y Micaela han sido esos Gigantes que me han permitido ver mas allá del horizonte; por ello les dedico esta tesis, fruto del esfuerzo, trabajo y dedicación que ellos me enseñaron a tener.

Tambien va para mi hermano, Brayan, quien me ha acompañado en cada paso que he dado.

Agradecimientos

La gratitud es uno de los valores más nobles y necesarios en el día a día. Agradecer es reconocer los aportes que los demás hacen por uno mismo, por ello, deseo expresar mi mayor agradecimiento a Dios y a mi propia vida; pues son los pilares que me permiten estar en este mundo y disfrutar de él, así como desempeñarme en esta apasionante carrera y de esta manera, cumplir mi sueño anhelado: ser científica.

No encuentro las palabras suficientes para expresar todo lo que agradezco a mi familia. Mis padres, Micaela y José Antonio, quienes siempre han confiado en mí y en cada una de las decisiones que he tomado a lo largo de mi vida, me han apoyado con la seguridad de que los pasos que doy son los correctos, y han sido testigos de cada uno de mis logros y fracasos. Y mi hermano Brayan, quien ha brindado todo el sentido a mi vida, cuya compañía llena de cariño, compañerismo y amor han sido un tesoro invaluable en mi existencia.

A la Facultad de Ciencias Físico Matemáticas junto a cada uno de los profesores quienes fueron parte de mi formación profesional, porque cada clase que brindaron se convirtió en una herramienta invaluable y una oportunidad para crecer. Agradezco su paciencia y compromiso por la educación brindada. Específicamente, a mi asesor, Dr. Heber, por guiar mis esfuerzos y su apoyo durante la elaboración de esta tesis; su habilidad de enseñanza, su paciencia y confianza así como la amistad que ha surgido en el proceso, han sido fundamentales para mí. Así mismo, gracias al cuerpo académico de Física Médica, profesores y compañeros, que me acompañaron en el camino y de quienes pude aprender mucho.

Gracias también a Sofía, quien es como una hermana para mí. Cuya presencia en mi vida es fundamental así como lo ha sido a lo largo de la licenciatura, la cual cursamos juntas; con quien he compartido muchas vivencias académicas y personales, y cuyos aportes para esta tesis son inestimables. Mis amigos Salvador, Iván, Víctor, Iñaki y Edmundo me ayudaron a vivir esta etapa universitaria en su mejor versión; el equipo académico que creamos, los momentos vividos y la ayuda que brindaron hacia mi persona fue un aporte muy grande dentro de mi formación.

Por último, pero no menos importante, quiero expresar mi más sincero agradecimiento a Manuel, por estar a mi lado mientras elaboraba este trabajo. Su apoyo académico y compañía emocional sostuvieron gran parte de mi persona. Aportó más de lo que podría imaginar.

Resumen

El cáncer de mama es una enfermedad ligada al desarrollo de tumores en los senos y una de las principales causas de muerte alrededor del mundo. La detección temprana juega un papel importante al momento de enfrentar este desafío de salud. Por ello, en años recientes se han creado modelos y recursos altamente eficientes para el diagnóstico del cáncer de mama. Uno de ellos es el procesamiento de imágenes médicas tales como mamografías, mediante enfoques como el aprendizaje profundo y la segmentación de imágenes.

En este trabajo se utilizaron herramientas de programación tales como ROOT (del CERN) y el lenguaje C++ para procesar y parametrizar imágenes mamográficas, de acuerdo con su intensidad de brillo. Con ello, se seleccionaron muestras de tejido tumoral y sano en cada imagen y se realizaron cálculos de las métricas Euclidiana y de Manhattan. Posteriormente, con los datos obtenidos se realizó un análisis comparativo con la finalidad de diferenciar los tejidos. Los resultados sugieren la posibilidad de diferenciar los tumores malignos del resto de la imagen, debido a que los valores obtenidos para este tipo de tejido son mayores, comparados con los del tejido sano. Esto podría permitir en el futuro ser utilizado como auxiliar para el diagnóstico de cáncer de mama.

Índice general

1. Introducción	1
1.1. Planteamiento del problema	1
1.2. Estado del Arte	2
1.3. Hipótesis	3
1.4. Objetivo General	3
1.5. Objetivos específicos	3
1.6. Estructura de la Tesis	3
2. Marco Teórico	5
2.1. Tumores y cáncer	5
2.1.1. Tumor mamario	6
2.1.2. Métodos de detección de tumores	7
2.2. Mamografías	7
2.2.1. Tipos de mamografías	9
2.2.2. Interpretación radiológica de una mamografía	10
2.3. Imagen digital	12
2.4. Variedades Diferenciables	13
2.5. Métricas	15
3. Metodología	19
3.1. Mamografías	19
3.2. Selección de Imágenes y normalización	20
3.3. Obtención de la métrica Euclidiana y de Manhattan	23
3.4. Cálculo de cocientes	25
4. Resultados y discusión	27
4.1. Resultados de los valores de las métricas	27
4.2. Resultados del cálculo de cocientes	29
4.3. Discusión	34
5. Conclusiones	37
Bibliografía	39

Capítulo 1

Introducción

En este capítulo se enuncia y justifica el problema a desarrollar; se menciona el estado del arte, se formula una hipótesis que fungirá como base para la elaboración de esta tesis, además de establecer sus objetivos principales y específicos. Finalmente, se incluye un resumen de la estructura de este trabajo.

1.1. Planteamiento del problema

El cáncer de mama es una enfermedad relacionada a la formación de tumores, masas celulares o carcinomas en el tejido mamario [1].

De acuerdo a la Organización Mundial de la Salud (OMS), esta enfermedad es una de las principales causas de muerte alrededor del mundo; se estima que en 2020 ocurrieron 685 000 muertes relacionadas con el cáncer de mama. En México, en 2020, la tasa de incidencia fue de 40.5, con tasa de mortalidad de 10.6 por cada 100 mil mujeres, siendo la enfermedad más frecuente de ese año, con el 15.3 % del total de enfermedades, un estimado de 29 mil 929 nuevos casos y siete mil 931 muertes [2]. Para 2021 la tasa nacional de mortalidad aumentó a 18 por cada 100 mil para mujeres de 20 años y más, esto es, la muerte de 7 mil 973 personas de las cuales el 99.4 % eran mujeres y el 0.6 % hombres, según el instituto nacional de estadística y geografía (INEGI) [3]. Sin embargo, pese a que la tasa de prevalencia es sumamente significativa, la probabilidad de supervivencia es del 90 % siempre que ocurra una detección temprana y el seguimiento de un tratamiento eficaz [1].

El interés por desarrollar nuevas técnicas de diagnóstico médico han llevado a la ciencia a desarrollar modelos y herramientas útiles para la detección del cáncer de mama, proponiendo métodos de obtención y procesamiento de imágenes, tales como la mamografía. Una mamografía consiste en una radiografía de las mamas. Es un método de diagnóstico por imagen que permite detectar la presencia de tumores que no son sensibles al tacto o no causan síntomas. Se trata de una de las mejores alternativas para detección de cáncer de mama, además de favorecer la interacción con la paciente, al ser un método no invasivo. El procesamiento de imágenes mamográficas (que en adelante también llamaremos MI, abreviatura de mammographic image) con medios computarizados, brinda y facilita la obtención de información importante respecto al estado de los tumores presentes, con detalles como tamaño, ubicación y caracterización. Previamente se ha trabajado con MI, analizando la diferencia de intensidades en la escala de grises entre tejido tumoral y de fondo, con el fin de obtener las características antes mencionadas con la ayuda del procesamiento de imágenes.

Por todo lo anterior, se propone el uso de análisis de datos aplicado a mamografías digitales, utilizando recursos de programación y geometría diferencial aplicada, con la finalidad de que, mediante el cálculo de métricas y su análisis comparativo, sea posible la caracterización de tumores malignos respecto a tejido sano.

1.2. Estado del Arte

La búsqueda de nuevos métodos de diagnóstico para la prevención y detección temprana del cáncer de mamá ha llevado a ciencias como la física y la ingeniería aplicadas a la medicina, a desarrollar procedimientos de análisis de mamografías mediante técnicas computacionales. La revisión de estos, nos ha permitido seleccionar materiales de interés, mismos que se convirtieron en referentes teóricos de esta tesis.

A partir de la base de datos mini-MIAS (Mini- Mammographic Image Analysis Society) del Archivo Piloto Europeo de Procesamiento de Imágenes(PEIPA) [4], que contempla un conjunto de mamografías digitales con datos asociados disponibles para pruebas e investigaciones, se han realizado análisis computacionales bajo diferentes enfoques y objetivos. Por ejemplo, en 2020 Zepeda et al. en su investigación “Tumor and microcalcification characterization using Entropy, Fractal Dimension and intensity values statistical analysis in a mammography”[5] lograron diferenciar tejido tumoral y calcificaciones de las mamas mediante el software ROOT y la transformación de mamografías en mapas de intensidades, determinando un valor de intensidades para dichos tejidos. De esta investigación destaca la conclusión de que el valor de intensidad de brillo de tumores es mayor a 0.7 en una escala de intensidades de 0 a 1 y que la distribución de esos valores es consistente para cualquier tejido sano. Así mismo, en 2022 Lopez, publicó su tesis de licenciatura bajo el título “Separación de regiones de interés para el diagnóstico y tejido de fondo en mamografías mediante análisis de datos”[6] misma que comprende el uso de análisis de datos y la digitalización de mamografías para caracterizar cuantitativamente el tejido de fondo de las MI. También se publicó un artículo en 2020 titulado: “Characterization of mammographic masses based on local photometric attributes”[7], en el que los autores Rabidas y Arif, desarrollaron una técnica para identificar lesiones en los senos y determinar si son malignas o no.

De igual manera, se han implementado modelos matemáticos y algoritmos, tales como en el artículo de 2020 “Novel mathematical model of breast cancer diagnostics using an Associative Pattern Classification” [8], en el que sus autores, Santiago-Montero et al. utilizan redes neuronales y la clasificación de patrones asociativos aplicados a una base de datos de mamografías, para detectar en ella posibles casos de cáncer de mama. Chiao et al. en su artículo de 2019 “Detection and classification the breast tumors using mask R-CNN on sonograms”[9] utilizaron redes neuronales convolucionales y la segmentación de instancia, aplicadas a sonogramas para detectar lesiones y tumores mamarios.

Además en 2022, mediante el uso de la base de datos Mini-MIAS y redes neuronales, Kavitha et al. desarrollaron en su artículo “Deep Learning Based Capsule Neural Network Model for Breast Cancer Diagnosis Using Mammogram Images”[10] el uso de aprendizaje profundo (Deep learning) para detectar cáncer de mama.

En cuanto al uso de métricas o distancias, en el artículo de 2021, “Shared Information-Based Late Fusion for Four Mammogram Views Retrieval using Data-driven Distance Selection” de Jouirou et al [11], mediante la técnica de recuperación de mamografías basada en contenido (CBMR) se utilizaron algunas distancias como la Euclidiana, de Manhattan, de Canberra, etc. como una métrica de similitud y con ello compararon semejanzas entre mamografías. También se ha utilizado la métrica Euclidiana como criterio de similitud en técnicas de diagnóstico de cáncer de mama mediante redes convolucionales, como lo hicieron Chunhapran et al. en su artículo “ Multi-View Combination using Mutual Information and 3-D Euclidean Distance for Breast Cancer Classification”[12].

De acuerdo con las referencias recopiladas, en los últimos años no se ha implementado algún modelo matemático relacionado a las distancias Euclidiana y de Manhattan desde el enfoque de análisis de datos o bajo la aplicación de variedades diferenciales, para la caracterización de tumores malignos y tejido sano. Por lo que la idea propuesta en esta tesis es relevante para continuar el desarrollo de herramientas de diagnóstico no invasivas y eficaces.

1.3. Hipótesis

A partir de la aplicación de variedades diferenciales a mamografías digitales y calcular métricas basadas en los valores de intensidad de brillo de áreas tumorales, es posible su caracterización, anticipando que los valores de las métricas sean más altos en los tumores debido a su mayor intensidad.

1.4. Objetivo General

Se busca caracterizar las regiones de tejido maligno y sano, a partir de la comparación de los valores de métrica Euclidiana y Manhattan.

1.5. Objetivos específicos

- Disminuir artefactos.
- Parametrizar y normalizar las MI seleccionadas.
- Delimitar la sección del tejido maligno y sano de cada mamografía.
- Obtener la métrica Euclidiana y de Manhattan de cada región y para cada mamografía.
- Obtener los cocientes de relación de cada métrica.
- Comparar los valores de las métricas y determinar la caracterización.

1.6. Estructura de la Tesis

El capítulo dos de esta tesis presenta las bases teóricas necesarias para la comprensión de la investigación. La primera sección muestra información acerca de la glándula mamaria así como de los tumores mamarios, su clasificación y los métodos de diagnóstico utilizados, haciendo énfasis a los métodos de detección de tumores por imágenes médicas. Posteriormente se profundiza acerca de la mamografía, explicando los fundamentos físicos necesarios, como es la absorción diferencial, además de los tipos de mamografías y la clasificación BI-RADS a considerar al momento de realizar un diagnóstico. Este capítulo también comprende los fundamentos matemáticos importantes y aplicados a nuestro trabajo. Se explica una introducción acerca de las variedades diferenciales y su aplicación a matrices en espacios \mathbb{R}^n . Finalmente, se incluyen los conceptos de métricas, el tipo euclidiano y de Manhattan.

El capítulo tres presenta el procedimiento seguido para realizar esta investigación; explica la base de datos utilizada, características de las imágenes de muestra, así como el proceso de selección de MI, parametrización y obtención de métricas detallando el tipo de programa y recursos utilizados.

En el capítulo cuatro se enlistan los resultados obtenidos mediante dos formas: utilizando gráficas de puntos para comparación entre muestras, y utilizando histogramas para un resultado más generalizado. La discusión de dichos resultados también se incluye en él. Por último, el capítulo cinco consiste en las conclusiones generadas a partir de todos los capítulos previos.

Capítulo 2

Marco Teórico

Este capítulo incluye toda la información previa necesaria que sirve como base de la tesis. Consiste de una descripción clínica acerca de las mamas, los tumores y el cáncer. Además, explica los fundamentos de las mamografías, así como los aspectos matemáticos utilizados para esta investigación, tales como las variedades diferenciales y los espacios métricos.

2.1. Tumores y cáncer

El seno mamario, o la mama, es una glándula de secreción externa presente en el cuerpo humano tanto femenino como masculino. Las mamas presentan una forma semi-cónica o de “lágrima” con un diámetro de entre 10 y 12 cm, además de un espesor de 7 cm en promedio. Generalmente son asimétricas y de acuerdo con factores genéticos, geográficos, nutricionales, de raza y hormonales, pueden existir variedades anatómicas tales como tamaño y forma. Esto quiere decir que más de una mujer no tendrá las mismas características de sus senos, así como una misma mujer presentará en diferentes etapas de su vida variaciones en sus mamas, por ejemplo, durante el embarazo y la menopausia. La glándula mamaria se estructura por la glándula en sí, además de tejido conjuntivo que forman el estroma de la mama [13]:

- **Tejido glandular.** Comprende los lóbulos, lobulillos y conductos de la mama.

- **Tejido conjuntivo.** Se conforma de tejido graso o adiposo y tejido denso fibroso. Rodea la glándula mamaria y proporciona la mayor parte del volumen total de las mamas. El tejido denso fibroso brinda soporte al seno; el tejido graso comprende células adiposas y es el que da tamaño y forma a las mamas. Se dice que la mama es densa, cuando posee gran cantidad de tejido conjuntivo en comparación con el glandular.

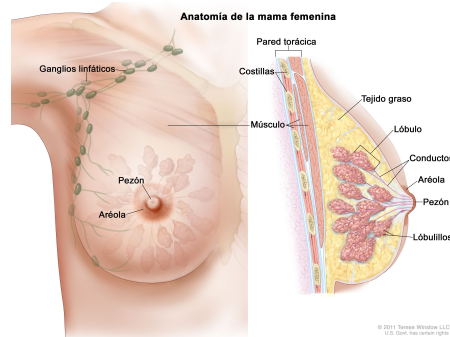


Figura 2.1: **Anatomía de la mama femenina.** En la zona externa se observa el pezón y la areola, mientras que internamente se muestran los conductos, ganglios, lóbulos, lobulillos y estroma. Imagen de *Instituto Nacional del Cáncer*. [14]

La Fig. 2.1 ilustra la composición mamaria con los respectivos tipos de tejidos que conforman a los senos.

Tanto en hombres como en mujeres, la estructura mamaria desde el nacimiento hasta la pubertad es la misma; en el caso de la anatomía masculina, el aumento en la producción de testosterona y la ausencia de progesterona y estrógenos inhibe el desarrollo glandular de la mama. En esta investigación se ha enfocado el estudio teórico a los casos de cáncer de mama en mujeres. Según la organización mundial de la salud, solo entre el 0.5 y el 1% de los casos totales corresponden a varones [15].

2.1.1. Tumor mamario

Un tumor mamario es una masa, inflamación o protuberancia de tejido anormal en la mama, que se origina cuando las células se multiplican descontroladamente, comúnmente derivado de un proceso citológico alterado y un desajuste hormonal. También puede ser un conglomerado de depósitos de calcio, o microcalcificaciones en región mamaria [16]. Los tumores pueden ser benignos o malignos.

Tumores benignos

Estos tumores no poseen naturaleza cancerígena ni la capacidad para diseminarse. Están compuestos propio tejido mamario y comúnmente aparecen por causas de desbalance hormonal. A excepción del carcinoma lobulillar *in situ*, son poco relacionados con aumento en el riesgo de cáncer.

Tumores malignos: Cáncer de mama

Los tumores malignos de la mama son cancerosos y se diseminan al resto del organismo, tienden a ser agresivos e invadir tejidos circundantes. Estos tumores difícilmente son palpables y, contrario a la mayoría de tumoraciones benignas, no presenta signos o síntomas hasta estar en etapas avanzadas. Anatómicamente se pueden describir como nódulos localizados en el revestimiento de los conductos el 85% de los casos, o en los lóbulos el 15% de las veces. Estos no poseen una forma bien definida, y sus bordes son irregulares, además de presentar una consistencia firme y dura, detectable solo con métodos de imagen o quirúrgicos. Cuando logran ser palpables, su ubicación es fija, y no se mueven entre tejidos.

Durante la primera etapa del cáncer de mama se presenta solo en alguna parte del tejido glandular dependiendo el caso, sin haberse diseminado a otras partes de la mama. Entonces se dice que el cáncer se encuentra *in situ* o en estadio 0. Posteriormente cuando comienza a invadir tejido

circundante se denomina cáncer de mama invasivo. La etapa posterior ocurre cuando se propaga mediante los ganglios linfáticos y se llama cáncer metastásico regional. Finalmente, si el cáncer llega a avanzar a otros órganos del cuerpo mediante el sistema linfático y el torrente sanguíneo, se denomina como metástasis distante, y esta es la última etapa de desarrollo del cáncer de mama [17]. El carcinoma ductal invasivo es el tipo de de cáncer de mama más común.

Los factores de riesgo de desarrollar esta afección, abarcan varios aspectos. Algunos factores son la edad, predisposición genética, antecedentes familiares, uso de medicamentos hormonales, alcoholismo, vida sedentaria y exposición a la radiación ionizante o tratamientos previos de radioterapia. [18].

2.1.2. Métodos de detección de tumores

Para detectar tumoraciones mamarias, la valoración clínica es fundamental. Genralmente la realiza el médico ginecólogo, y consiste de los siguientes procedimientos:

- **Interrogatorio médico.** Debe incluir antecedentes familiares de casos de cáncer de mama así como numero de embarazos, lactancia, menopausia, tratamientos hormonales, síntomas y su evolución, etc.
- **Exploración física.** Consiste en una inspección y palpación de la mama, con el fin de encontrar bultos o alteraciones en la forma y tamaño del seno.
- **Apoyos diagnósticos de imagen.** Los principales incluyen mastografía o mamografía, ultrasonido y resonancia magnética. El método de imagen lo decidirá el médico, de acuerdo a síntomas, antecedentes y el resultado de la exploración física.

Además de detectar tumores mamarios, la valoración clínica permite establecer la sospecha de cáncer, sin embargo el diagnóstico y confirmación de cáncer de mama se establece mediante un procedimiento histopatológico, como la biopsia [13].

El diagnóstico por imagen cumple un rol importante para la detección temprana del cáncer. Junto a ello, dentro de la investigación médica oncológica se desarrolla y aplica el procesamiento de imágenes médicas. Siguiendo la idea de aplicar conocimientos matemáticos y de programación, como el aprendizaje profundo y la segmentación de imágenes, permite mejorar la precisión y comprensión diagnóstica hacia los pacientes. Esta tesis se ha enfocado en el uso de mamografías para dicho fin.

2.2. Mamografías

Una mamografía o mastografía es una imagen médica obtenida mediante la aplicación de rayos X sobre las mamas. Su función principal radica en la visualización interna de los senos para que de esta manera se pueda detectar o diagnosticar tumoraciones mamarias.

Dentro del campo de la imagenología, la mamografía representa un gran desafío, pues la mama es un órgano radiológicamente complejo. Se trata de una glándula conformada por diferentes tejidos blandos que incluyen el tejido graso, el tejido glandular y el tejido fibroso los cuales se entrelazan de diferentes maneras, predominando distintos tipos de tejido en función de la paciente. Debido a estas diferencias de composición, es complicado definir a una mama "normal". Además de lo anterior, al ser un tejido blando tiene menor densidad en comparación con otros tejidos, como los óseos; esto significa que la mama absorbe más radiación lo que dificulta el contraste entre sus elementos, que podrían ocultar masas o calcificaciones.

Considerando todas las condiciones anatómicas que presentan los senos, y al ser altamente susceptibles a los efectos de la radiación sobre los tejidos, se busca seguir el principio ALARA (*As Low As Reasonably Achievable*). Esto es recibir la dosis más baja como sea razonablemente posible durante el examen, pero permitiendo que la imagen tenga el mayor contraste posible. Para lograrlo

es necesario trabajar con valores de kilovoltaje máximo (Kvp) menores en comparación con los de una radiografía convencional de cualquier otra zona anatómica. Además de emplear haces de rayos X suaves, que poseen menor energía y se absorben mejor en los tejidos, permitiendo así un mayor contraste y mejor diferenciación entre ellos [19].

Para realizar una mamografía, se emplea una máquina de rayos X llamada mastógrafo (Fig. 2.2). El sistema se centra en el tubo de rayos X, que se encuentra montado sobre un brazo giratorio (brazo C), junto a este, radica la placa de compresión mamaria, una rejilla móvil y el receptor de imagen que puede ser una película de rayos X, pantallas o detectores capaces de captar la radiación y convertirla en imágenes; para regular la cantidad de radiación utilizada se usa un dispositivo de control de exposición automático [19].

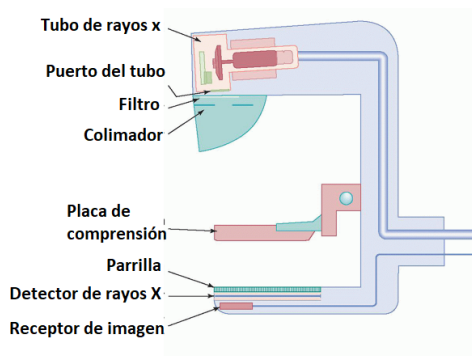


Figura 2.2: **Mamógrafo**. Se puede observar el tubo de rayos X, conectado a la placa de compresión, al detector de rayos X y al receptor de imagen mediante el brazo giratorio. Imagen de *Radiology Key* [20].

Durante la toma de mamografía, la mama se comprime entre la placa de compresión y el receptor, donde es irradiada. Es importante realizar la compresión mamaria pues ayuda a extender los tejidos y de esa manera poder detallar mejor la imagen, además de reducir el tiempo de exposición y la dosis de radiación necesaria, minimizando riesgos y problemas en la calidad de la imagen obtenida. Una unidad para medir la dosis absorbida es el Grey (Gy); un Grey equivale a la absorción de un joule de energía por un kilogramo de materia; durante una mamografía estandar, la dosis absorbida promedio de radiación ionizante en los tejidos fibroglandulares radiosensibles, conocida como dosis glandular media es de aproximadamente 2 mGy por exposición [19].

Absorción diferencial

Una imagen radiológica es creada como producto de la absorción diferencial. Esta consiste en la diferencia entre los rayos X que penetran los tejidos sin tener interacción alguna con sus átomos, y los rayos X absorbidos por el efecto fotoeléctrico.

Dentro de la radiología, los efectos Compton y fotoeléctrico y los rayos X que no interaccionan juegan un papel importante:

- **Efecto Fotoeléctrico.** Los rayos X que al interaccionar producen este efecto, proporcionan información diagnóstica. Producen las zonas claras o de mayor brillo en una radiografía. Los tejidos representados con ello se consideran radioopacos.
- **Rayos X sin interacción.** Estos rayos penetran en el cuerpo pero no interaccionan con los átomos del mismo. Producen las zonas oscuras o con menor brillo de la imagen. Los tejidos que son penetrados y representados con poco brillo se llaman radiolucientes.

- **Efecto Compton.** Un rayo X disperso por efecto Compton es detectado por el receptor de imagen de un mamógrafo como un rayo proveniente de la fuente. Produce un velo en la imagen, es decir, genera áreas atenuadas ajenas a la región anatómica estudiada.

La absorción diferencial dependerá del valor de KVp. Para tejidos blandos como las mamas, un KVp reducido maximizará la absorción [21].

2.2.1. Tipos de mamografías

Se puede diferenciar clínicamente entre dos tipos de mamografías:

- **De detección:** Se aplica a una paciente sin síntomas y se efectúa sobre dos vistas o proyecciones: una cráneo-caudal y otra oblicua mediolateral.
- **De diagnóstico:** Es utilizada cuando existe algún síntoma, lo que conlleva a aplicarla sobre más vistas. [22, 23].

También se puede clasificar de acuerdo a las técnicas de obtención, ya sea digital o convencional.

Mamografía convencional

También llamada técnica de Screen-Film Mammography (SFM) o mamografía analógica, es la técnica más utilizada a lo largo de los años, y la primera técnica de mamografía desarrollada. Este tipo de mamografía utiliza un mamógrafo que posee como detector un cassette de películas de pantalla sensible a los fotones de radiación que atraviesan la mama y que almacena la radiación detectada en forma de imagen.

Al ser un procedimiento analógico, es decir, usando señales continuas y variables, la obtención de una mamografía convencional requiere de un proceso adicional después de la exposición radiológica. La película del detector debe someterse a un proceso de revelado que consiste en sumergirla en productos químicos especiales, tales como hidroquinona y metol para revelar y tiosulfato de sodio para fijar la imagen latente u oculta, registrada sobre la película. Una SFM implica un costo bajo comparado con la técnica digital, haciendo de ella una opción accesible para muchos centros de atención médica; sin embargo, la mamografía analógica puede presentar limitaciones pues no permite manipular la imagen de forma digital, la calidad de la imagen dependerá de muchos factores físicos al momento de generar la mamografía; además de requerir una dosis más alta de radiación [19, 24].

Mamografía digital

Una mamografía digital utiliza detectores electrónicos que absorben la radiación y la procesan, convirtiéndola en imágenes constituidas por píxeles, a diferencia de una convencional que crea una imagen continua.

Para obtener imágenes mamográficas digitales se usa un mamógrafo similar al utilizado en mamografías análogas; los rayos X atraviesan la mama y llegan a un detector que produce luminosidad, que es transmitida mediante fibra óptica hasta un dispositivo de carga acoplada. Este es un sistema de detección digital que se conforma de una matriz de píxeles, capaces de capturar la luz y convertirla en señales eléctricas. Dependiendo de su intensidad, se generará una imagen digital que se transfiere a un equipo de cómputo para su procesamiento y visualización. Este tipo de mamografías presenta gran ventaja frente a una SFM, ya que permite una visualización directa y la manipulación de la misma en cuestión de brillo y contraste, porque sus valores se pueden ajustar dependiendo las necesidades de diagnóstico, además de poder almacenarlas electrónicamente, permitiendo la creación de una base de datos, facilitando su acceso y gestión. También ofrece mayor precisión al momento de realizar un diagnóstico debido a mayor claridad y resolución en la imagen, en comparación a una SFM. Otra ventaja importante es la disminución de la exposición

a la radiación. Un examen de mamología digital tiende a tener una dosis glandular media 20 % menor que una convencional. Existen dos categorías de mamografía digital, que son [24, 25]:

- **Mamografía digital directa.** También se llama de campo completo (DR-M). El receptor de imagen del mamógrafo es un detector que captura los rayos X y procesa en forma de señales eléctricas para formar una imagen digital inmedita.
- **Mamografía digital indirecta o computada.** Se llama de método indirecto (CR-M). El proceso de obtención es similar al de una mamografía convencional. Los rayos X atraviesan el seno y llegan a una placa de fósforo fotoestimulable que genera una imagen latente. Al introducir la placa en un lector, este la procesa mediante fotoestimulación láser. Finalmente se genera una imagen digital. Dentro de esta clasificación también se incluyen las mamografías analógicas que han sido digitalizadas.

La base de datos mini-MIAS [4] provee un conjunto de mamografías digitales computadas, debido a que originalmente fueron obtenidas mediante el método analógico y posteriormente digitalizadas. Por lo tanto, el tipo CR-M de mamografías fueron las utilizadas en esta tesis.

2.2.2. Interpretación radiológica de una mamografía

La mamografía se acompaña de un informe con una interpretación del estado de la mama de acuerdo a los elementos encontrados. Para ello se siguen los criterios del Breast Imaging Reporting and Data System (BI-RADS), un sistema de clasificación creado en 1993 por el Colegio Americano de Radiología que tiene como objetivo estandarizar la descripción de resultados hallados en una mamografía. Dentro del sistema BI-RADS, se pueden clasificar diversos elementos, como la composición mamaria, nódulos y microcalcificaciones, además de establecer una categoría de evaluación que determina el estado de la mama [26].

Composición mamaria

Las mamas pueden componerse de tejido fibroglandular, fibroadiposo y adiposo. De acuerdo a la cantidad de tejido fibroso presente en el seno, se determina su densidad mamaria. El sistema BI-RADS reconoce cuatro categorías de densidad:

- **Tipo A: Senos grasos.** Se conforman en su totalidad de tejido adiposo.
- **Tipo B: Senos de densidad media.** Se conforman de tejido adiposo y algunas regiones de tejido fibroglandular.
- **Tipo C: Senos de densidad heterogénea.** Poseen en su mayoría tejido fibroglandular con algunas regiones de tejido graso.
- **Tipo D: Senos densos .** Se conforman en su totalidad de tejido fibroglandular.

La densidad de los senos no depende ni se puede deducir mediante la forma y el tamaño de los senos. La densidad mamaria de tipo A tiende a ser más común en mujeres mayores, mientras que las de tipo B, C y D presominan en mujeres jóvenes. La densidad mamaria puede disminuir con la edad [27].

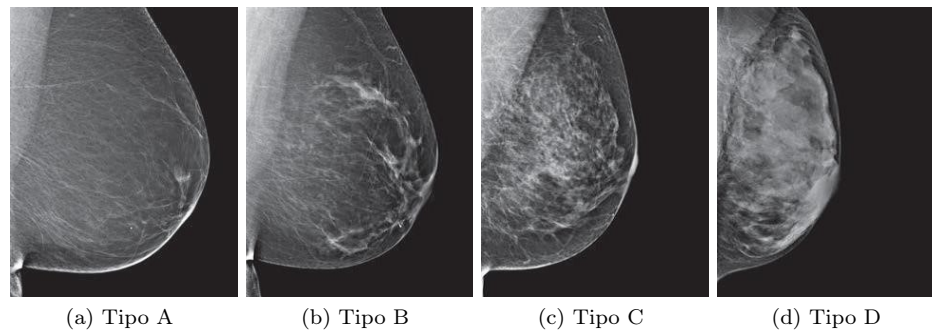


Figura 2.3: **Clasificación de la densidad mamaria.** La mama tipo A se conforma en su totalidad de tejido graso. La mama tipo B posee una densidad media, regulada entre tejido adiposo y glandular. En el tipo C predomina el tejido fibroglandular. La mama tipo D se conforma en su totalidad por tejido fibroso glandular. Imágenes de Sociedad Americana del Cáncer. *Densidad de los senos e informe de su mamograma* [27].

Las mamas de tipo C y D dificultan la detección de anomalías, pues la alta densidad refleja alta intensidad o brillo en la imagen [27].

Tipos de anormalidades

Se pueden clasificar dos tipos de anormalidades mamarias: los nódulos o masas, y las microcalcificaciones [26].

- **Nódulos o masas:** Se trata de lesiones que pueden ser sólidas o quísticas (líquidas). Son visibles en una o más proyecciones de la MI. Cuando solo son visibles en una proyección, se llama asimetría y se refiere a un cambio en la densidad mamaria. Un nódulo o masa es visible en dos o más proyecciones. Respecto al tejido circundante, se define su forma, contorno y densidad de la siguiente manera:

Forma: Puede ser redonda, oval, lobulada o irregular.

Contorno: Que puede ser bien definido, oscurecido, mal definido, microlobulado o espiculado.

Densidad: se describen como superior, similar o inferior.

- **Microcalcificaciones:** Son lesiones compuestas de calcio que pueden ser de tamaño grande (macrocalcificaciones) o en forma de partículas. Aunque comúnmente benignas, si se encuentran agrupadas generan sospecha de malignidad. El sistema BI-RADS establece la distribución de las microcalcificaciones como difusa, regional, agrupadas, lineales, segmentarias. Luego, esta se puede clasificar de dos formas:

Clasificación típicamente benignas: Las microcalcificaciones en su mayoría se encuentran en esta clasificación. Son vasculares, cutáneas, lechada de cal, lineales gruesas, pop-corn, distróficas, puntiformes, en anillo.

Clasificación sospechosas: Presentan sospecha de malignidad y necesitan un estudio adicional. Son las heterogéneas, amorfas, líneas finas y ramificadas [28].

Categorías BI-RADS

De acuerdo a los hallazgos de la mamografía y para describir los resultados y seguimientos, se establecen siete categorías numeradas del 0 al 6 [29]:

- **BI-RADS 0:** “Estado incompleto”, se necesitan estudios de imagen adicionales.

- **BI-RADS 1:** “Hallazgo negativo”, no se encontraron anomalías respecto a las mamas, simetría, ausencia de nódulos o calcificaciones.
- **BI-RADS 2:** “Hallazgo benigno”, se encontraron calcificaciones benignas o fibroadenomas.
- **BI-RADS 3:** “Hallazgo posiblemente benigno”, presencia de anomalías con 98 % de posibilidad de ser no cancerosos con seguimiento a estudios por imagen a corto plazo.
- **BI-RADS 4:** “Anomalía sospechosa”, hallazgo de 2 a 94 % de probabilidad de cáncer, recomendación de biopsia.
- **BI-RADS 5:** “Anomalía que sugiere firmemente que se trata de un hallazgo maligno”, con 95 % de certeza de malignidad. Complemento con biopsia y acciones de tratamiento adecuadas.
- **BI-RADS 6:** Ya se ha demostrado la malignidad mediante biopsia y esta clasificación da seguimiento al tratamiento.

Artefactos

Una mamografía también puede presentar artefactos, que son elementos, distorsiones o errores presentes en la imagen, pero que no poseen relación con la región anatómica analizada. Estos pueden ser causados por diversos factores y provocan alteraciones en la fidelidad y calidad del estudio realizado. Algunos motivos de artefactos son [19]:

- **Relacionados al paciente:** Movimientos del paciente, productos de higiene utilizados, mamas pequeñas.
- **De procesamiento:** Marcas de rodillos, polvo, pelusa, pliegues de la piel, huellas digitales, movimiento del equipo, diferencia de presión.
- **Relacionados al equipo mamográfico:** Irregularidad por mala calibración del detector, líneas por desalineación, vibración, polvo.
- **Relacionados al software (mamografía digital):** Pérdida de definición, mala lectura del detector, algoritmos de lectura incompletos .

2.3. Imagen digital

Una imagen digital es una representación plana de un conjunto de información visual, traducida a datos digitales utilizando elementos llamados bits. Estas unidades contienen información binaria y al combinarlos crean píxeles. Los píxeles son la unidad de una imagen digital. Estos datos se pueden procesar mediante dispositivos electrónicos [30].

Una imagen digital se puede representar a través de una función parametrizada $f(m, n)$, donde m y n son coordenadas espaciales de los píxeles y f es la intensidad de brillo. Para una imagen monocromática, como puede ser en escala de grises de 8 bits, la intensidad del brillo se mide en tonos de gris en una escala de valores del 0 al 255. Esto porque al tener 8 bits, se pueden representar $2^8 = 256$ combinaciones únicas en el sistema binario, que van desde 00000000 (negro) hasta 11111111 (blanco). La convención de coordenadas para una imagen se presenta en la Fig. 2.4:



Figura 2.4: **Convención de coordenadas de una imagen** $f(m, n)$. El centro u origen se encuentra en la esquina superior izquierda con sentido positivo de izquierda a derecha en n y de arriba a abajo en m . Imagen de *Luyoun* [31].

También se puede entender a la imagen digital como una matriz de $m \times n$ dimensiones, donde cada elemento de la matriz representa un valor del brillo de ese pixel. Esta convención es relevante para la subsección “Variedades diferenciales aplicadas a matrices” [32].

Una imagen en formato de archivo .pgm (Portable Gray Map), es un archivo digital que contiene información de la imagen en escala de grises en 2 dimensiones, comúnmente de 8 o 16 bits. Lo que implica que puede tener 2^8 y 2^{16} niveles de grises. En este formato de imagen se presentan las muestras mamográficas de esta tesis.

Procesamiento digital de imagen

Consiste en el conjunto de técnicas aplicadas a imágenes digitales para manipularlas, analizarlas y obtener información mediante una computadora. Estas técnicas pueden abarcar, desde mejorar la calidad de la imagen, por ejemplo, al disminuir ruido y artefactos, o modificar nitidez, color y brillo (procesamiento de bajo nivel), localizar y describir objetos en la imagen, por ejemplo la segmentación (procesamiento de nivel medio) hasta el análisis de los elementos de la imagen (procesamiento de alto nivel) [32].

2.4. Variedades Diferenciables

Para comprender la noción de variedad diferenciable, es necesario introducir los siguientes conceptos:

Carta

Considerando un conjunto M , una carta en M es un par (U, ϕ) donde U es un subconjunto de M y ϕ es una función biyectiva que mapea los elementos de U , uno a uno a un subconjunto abierto de \mathbb{R}^n . La función ϕ se conoce como función de coordenadas o parametrización [33].

Atlas

Se trata de una colección de cartas compatibles entre sí: $\mathcal{A} = \{(U_\alpha, \phi_\alpha)\}_{\alpha \in A}$

La compatibilidad de cartas de un conjunto M se establece cuando dos cartas, (U, ϕ) y (V, ψ) con $U, V \in M$ cumplen con:

- $U \cap V \neq \emptyset$
- $\phi(U \cap V)$ y $\psi(U \cap V)$ son abiertos en \mathbb{R}^n .
- Las funciones de cambio de coordenadas $\psi \circ \phi^{-1}$ y $\phi \circ \psi^{-1}$ son diferenciables.

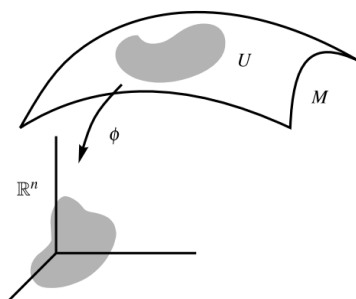


Figura 2.5: **Parametrización de M .** En una carta sobre M , ϕ mapea los elementos de U a un subconjunto de \mathbb{R}^n . Imagen de Gerardo Torres del Castillo, *Differentiable Manifolds: A Theoretical Physics Approach* [33].

Un atlas diferenciable cumple las siguientes condiciones:

- $\cup_{\alpha \in A} U_{\alpha} = M$ (Cobertura total)
- Las cartas son compatibles para todo par de ellas.

Una estructura diferenciable sobre M es una asignación de un atlas \mathcal{A} que es maximal para las dos condiciones enlistadas anteriormente.

Definición de variedad diferenciable

Una variedad diferenciable de dimensión n se define como un par (M, \mathcal{A}) , donde M es un conjunto y \mathcal{A} es una estructura diferenciable definida en M .

A continuación, se muestra un ejemplo de aplicación de variedades que será útil para el análisis de datos de las MI.

Aplicación de variedades diferenciables a matrices

El conjunto $\mathcal{M}(m \times n, \mathbb{R})$, que consiste en todas las matrices reales de tamaño $m \times n$, puede ser equipado con una estructura de variedad diferenciable. Sea A una matriz, tal que pertenezca al conjunto $\mathcal{M}(m \times n, \mathbb{R})$ y se escriba como:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

La función $\phi : (m \times n, \mathbb{R}) \rightarrow \mathbb{R}^{mn}$ se puede definir por:

$$\phi(A) = (a_{11}, a_{12}, \dots, a_{1n}, a_{21}, a_{22}, \dots, a_{2n}, \dots, a_{m1}, a_{m2}, \dots, a_{mn})$$

que es la estructura diferenciable estándar para el conjunto de matrices $\mathcal{M}(m \times n, \mathbb{R})$ [34].

Esta estructura es útil para aplicar operaciones de distancia a las MI, si se consideran como matrices de 1024×1024 , tal como se define en la sección sobre imagen digital.

Por ejemplo, sea A una matriz de 1024×1024 que representa una MI. Los valores de m y n son entonces de 1024 cada uno, que son las dimensiones espaciales de la imagen en pixeles, por lo que esas mismas dimensiones son asignadas a la matriz. Cada elemento de A alberga el valor

de intensidad correspondiente a un píxel de la imagen original. Entonces una MI está representada por:

$$A = \begin{pmatrix} z_{11} & z_{12} & \cdots & z_{1,1024} \\ z_{21} & z_{22} & \cdots & z_{2,1024} \\ \vdots & \vdots & \ddots & \vdots \\ z_{1024,1} & z_{1024,2} & \cdots & z_{1024,1024} \end{pmatrix}$$

Ahora, sea $\phi(A)$ una estructura diferenciable aplicada a A , tal que se obtiene:

$$\phi(A) = (z_{11}, z_{12}, \dots, z_{1,1024}, z_{21}, z_{22}, \dots, z_{2,1024}, \dots, z_{1024,1}, z_{1024,2}, \dots, z_{1024,1024})$$

bajo esa estructura, es posible operar sobre el conjunto de intensidades de cada MI, ya que al realizar el mapeo a $\mathbb{R}^{1024 \times 1024}$ estas intensidades representan las 1024×1024 coordenadas del punto asociado a la matriz A . Entonces, se puede calcular las distancias Euclidiana y de Manhattan del origen de ese hiperespacio, al punto A , que es la MI.

2.5. Métricas

Una métrica es una función que describe la distancia entre dos elementos de un conjunto. Como parte de la geometría de un espacio euclidiano, algunas métricas como la Euclidiana y la de Manhattan pueden resultar útiles para aplicarlas al procesamiento de imágenes, como se ha aplicado en esta tesis.

Definición de distancia y espacio métrico

Sea d , una función definida y aplicada en el conjunto no vacío M , tal que $d : M \times M \rightarrow \mathbb{R}$ y que a cada par de puntos $x, y \in M$ asocia un número real $d(x, y)$ con las siguientes propiedades:

1. $d(x, y) \geq 0$
2. $d(x, y) = 0 \iff x = y$
3. $d(x, y) = d(y, x) \forall x, y \in M$ (Propiedad de simetría)
4. $d(x, y) \leq d(x, z) + d(z, y) \forall x, y, z \in M$ (Propiedad de desigualdad triangular)

A d se le nombra función distancia o métrica en M . Cuando se asocia al conjunto M con la función d , se forma un par (M, d) que define a un espacio métrico.[35]

Métrica Euclidiana

Es la distancia que hay entre dos puntos existentes dentro de un espacio euclidiano de n dimensiones; también se puede definir como la longitud de la línea recta más corta que une a ambos puntos.

La base de esta métrica surge del Teorema de Pitágoras, que establece que el valor del cuadrado de la longitud de la hipotenusa, es la suma del cuadrado de sus catetos, es decir:

$$c^2 = a^2 + b^2 \tag{2.1}$$

Considerando la hipotenusa de un triángulo rectángulo como la métrica Euclidiana de los dos vectores proyectados sobre los ejes directores (los catetos).

Para un plano euclidiano, esto es, en dos dimensiones, la ecuación 2.1 nos permite calcular la distancia Euclidiana, de la siguiente manera:

Sea d la hipotenusa de un triángulo rectángulo, y $(q_1 - p_1), (q_2 - p_2)$ los catetos respectivos.

Entonces obtenemos:

$$d^2 = (q_1 - p_1)^2 + (q_2 - p_2)^2$$

Y finalmente la ecuación de la métrica Euclidiana para 2 dimensiones:

$$d = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2}$$

Para un espacio de n dimensiones, la distancia Euclidiana d_e de los puntos $P(p_1, p_2, \dots, p_n)$, $Q(q_1, q_2, \dots, q_n)$ se define como [36, 37] :

$$d_e(P, Q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2} \quad (2.2)$$

Esta tesis trabaja con imágenes digitalizadas cuya dimensión de interés resulta ser la intensidad.

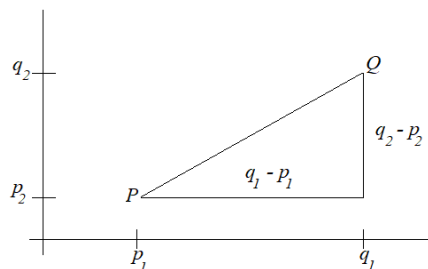


Figura 2.6: **Triángulo Rectángulo.** Sus catetos son $(q_1 - p_1)$, $(q_2 - p_2)$ respectivamente, y la hipotenusa es la distancia del punto P al punto Q . A partir de este se puede calcular las ecuaciones de distintas métricas, como la Euclidiana y la de Manhattan.

Para aplicar esta métrica, se considera que z representa en valor $q_n - p_n$ donde p_n representa un punto de referencia al origen, el cual siempre vale 0 y q_n el valor de intensidad del elemento n . Por lo que a este parametro le asignamos z . Entonces la ecuación 2.2 se reduce a

$$d_e(z) = \sqrt{z_1^2 + z_2^2 + \dots + z_n^2} \quad (2.3)$$

Métrica de Manhattan

Dentro de la geometría de Manhattan, del taxista, o de cityblock las distancias entre un punto y otro se miden siguiendo un camino que sólo puede direccionarse en vertical u horizontal. Por ejemplo, un taxista dentro de una ciudad meticulosamente construida como una cuadrícula, se trasladaría rodeando las cuadradas, debido a que no puede pasar sobre los edificios. De aquí que a esta distancia se le llame Manhattan o del taxista [36, 37]. La Fig. 2.7 ilustra mejor esta idea:

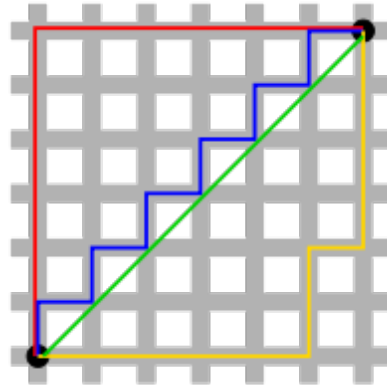


Figura 2.7: **Distancia de Manhattan.** Las líneas roja, azul y amarilla siguen la geometría de Manhattan, mientras que la verde continúa la idea de la distancia Euclidiana, vista en la sección anterior. Imagen de *Wikipedia* [38].

Para definir la ecuación que dicta la métrica de Manhattan en dos dimensiones, retomaremos el triángulo rectángulo de la sección anterior. Esta vez la distancia será la suma del valor absoluto de los catetos (recordemos que una propiedad de la función distancia es la no negatividad) es decir:

$$d_M = |q_1 - p_1| + |q_2 - p_2|$$

La distancia del taxista P_M del punto $P(p_1, p_2, \dots, p_n)$ al punto $Q(q_1, q_2, \dots, q_n)$ de un espacio de n dimensiones, se define como:

$$d_M = |q_1 - p_1| + |q_2 - p_2| + \dots + |q_n - p_n| \quad (2.4)$$

Al igual que con la distancia Euclidiana, si se considera que z representa en valor $q_n - p_n$ donde p_n representa un punto de referencia al origen, el cual siempre vale 0 y q_n el valor de intensidad del elemento n los cálculos para obtener esta métrica en las intensidades (z) de muestras MI se obtendrán siguiendo la ecuación 2.4 modificada para z :

$$d_M(z) = |z_1| + |z_1| + \dots + |z_n| \quad (2.5)$$

Capítulo 3

Metodología

En este capítulo se describen los pasos llevados a cabo para obtener el valor de la métrica Euclidiana y de Manhattan correspondientes a tejidos malignos y sano (de fondo) de la mama; este método incluye la obtención de las MI, su cambio de formato y normalización, así como su tratamiento utilizando datos de diagnóstico y su valor de intensidades para finalmente realizar los cálculos de interés.

3.1. Mamografías

Las imágenes utilizadas se recopilieron de la base de datos de mamografías digitales mini-MIAS (Mini-Mammographic Image Analysis Society) disponibles en el llamado Archivo Piloto Europeo de Procesamiento de Imágenes/Pilot European Image Processing Archive (PEIPA, por sus siglas en inglés) [4]. Esta base de datos es de libre acceso para investigaciones científicas dentro del campo de la detección del cáncer de mama. En su totalidad, la base de datos proporciona información acerca de cada imagen en un listado de siete columnas, que comprenden:

1. Imágenes mamográficas

Consiste en 322 MI pertenecientes a 161 pacientes. Estas se encuentran numeradas, enlistadas y organizadas en pares, donde la numeración par corresponde a las mamografías de la mama izquierda, y la impar a las mamografías de la mama derecha de una sola paciente. Las MI muestran diferentes anomalías mamarias, incluyendo tumores benignos, tumores malignos y masas no tumorales. Todas las imágenes tienen un tamaño normalizado de 1024×1024 píxeles con origen en la esquina inferior izquierda y se encuentran en formato PGM en 8 bits.

2. Carácter de tejido

Indica el tipo de tejido de fondo que predomina en la mama de acuerdo a su densidad, y pueden ser graso (F), graso glandular (G) o denso glandular (D). Estas categorías se relacionan con la densidad mamaria que se menciona en el sistema BI-RADS. El tejido graso (F) se relaciona con la mama tipo A. El carácter graso glandular (G) con la composición tipo B. Finalmente, la densidad mamaria de tipo C y D, al ser de “alta densidad” se relaciona con el tejido denso glandular (D).

3. Tipo de anormalidad

Es la clase de anormalidad que está presente en la mama. Esta puede ser una calcificación (CALC), masas bien definidas (CIRC), masas espiculadas (SPIC), otras masas mal definidas (MISC), distorsión arquitectónica (ARCO), asimetrías (ASYM) o tejido normal (NORMA).

4. Gravedad de la anormalidad

El diagnóstico médico determinado por médicos especialistas en cada anormalidad; este puede

ser benigno (B) o maligno (M).

5. Valor x del centro de la anormalidad

En una escala no decimal del 0 al 1024, el valor de la coordenada x del centro de la anormalidad.

6. Valor y del centro de la anormalidad

En la misma escala no decimal de 0 a 1024, en este caso es el valor para y del centro de la anormalidad.

7. Radio de la anormalidad

Radio aproximado de un círculo que encierra a la anormalidad y cuyo centro se ubica en las coordenadas x y y . Sigue la misma escala que las columnas 6 y 7, por lo que se mide en pixeles.

3.2. Selección de Imágenes y normalización

Se seleccionaron todas las MI cuya gravedad de anormalidad es maligna (M) de la base de datos. Esto es, 48 MI, de las cuales 18 corresponden a tejido de fondo graso, 15 a denso glandular y 15 a graso glandular. No se distinguió entre los tipos de anormalidades.

La organización de las MI consideró un listado donde las primeras MI fueron las correspondientes al tejido de fondo denso, seguidas de las de tejido de fondo graso y finalmente las de tejido de fondo graso glandular. Se etiquetaron de forma numerica ascendente. Pese a que en este paso se agruparon las MI de acuerdo al tejido de fondo, en adelante se consideraron sin distinción de tejido.

Muestras mamográficas		
Número de etiqueta	Clave según mini-MIAS	Carácter de tejido
1	mdb058	D
2	mdb102	D
3	mdb105	D
4	mdb110	D
5	mdb111	D
6	mdb125	D
7	mdb130	D
8	mdb170	D
9	mdb171	D
10	mdb179	D
11	mdb202	D
12	mdb239	D
13	mbd241	D
14	mdb249	D
15	mdb253	D

Tabla 3.1: Muestras mamográficas seleccionadas en función de la gravedad de anormalidad (Maligna) y tejido de fondo denso glandular.

Muestras mamográficas		
Número de etiqueta	Clave según mini-MIAS	Carácter de tejido
16	mdb028	F
17	mdb075	F
18	mdb092	F
19	mdb095	F
20	mdb134	F
21	mdb141	F
22	mdb144	F
23	mdb148	F
24	mdb155	F
25	mdb158	F
26	mdb184	F
27	mdb206	F
28	mdb231	F
29	mdb238	F
30	mdb256	F
31	mdb267	F
32	mdb271	F
33	mdb274	F
34	mdb023	G
35	mbd072	G
36	mdb090	G
37	mdb115	G
38	mdb117	G
39	mdb120	G
40	mdb124	G
41	mdb178	G
42	mdb181	G
43	mdb209	G
44	mdb211	G
45	mdb213	G
46	mbd264	G
47	mdb265	G
48	mbd270	G

Tabla 3.2: Muestras mamográficas seleccionadas en función de la gravedad de anormalidad (Maligna) y tejido de fondo graso y graso glandular.

La tabla 3.1 muestra las MI de tejido de fondo denso etiquetadas de la 1 a la 15. La tabla 3.2 muestra las MI 16 a la 48, siendo las que poseen composición grasa y grasa glandular.

Posterior a la descarga de estas imágenes se utilizó un programa convertidor de formatos disponible en la web, llamado "Convertio"[40] para cambiar de formato .pgm a .png; esta conversión se debió a que el formato original no es admitido por los programas necesarios para edición y procesamiento.

La Fig. 3.1 ilustra una de las MI descargadas de la mini-MIAS. Como se puede observar, la MI original presenta artefactos que limitan la escala de intensidades que nos interesa analizar, es decir, los artefactos representan el color "más blanco" de la imagen, por lo que al momento de parametrizar y medir la intensidad de cada bin, la información que se obtendría sería uno de estos como uno de los puntos más brillantes de la MI, lo cual es incorrecto. Por lo que después del cambio de formato, se utilizó la aplicación "PicsArt"[41] para cortar y retirar manualmente estos

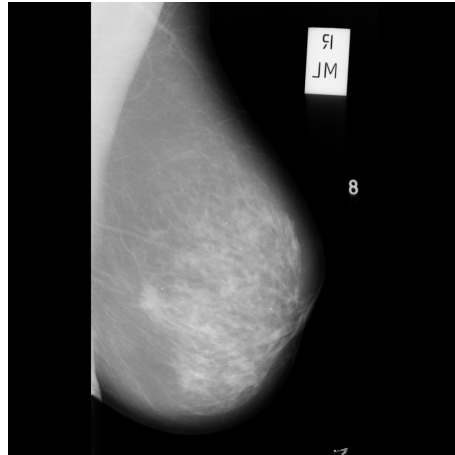


Figura 3.1: **Mamografía de seno mdb058**. Esta mamografía posee tejido denso glandular de fondo. Presenta una anomalía clasificada como “otras masas mal definidas” y gravedad como maligna. Obtenida de la base de datos mini-MIAS [4].

artefactos, que en su mayoría consistieron en etiquetas con los datos de la MI, letras, números y partes del pectoral, brazo o axila de la paciente.

La figura 3.2 muestra la MI que se presentó en la figura 3.1 pero sin artefactos; se realizó este procedimiento para las 48 mamografías.

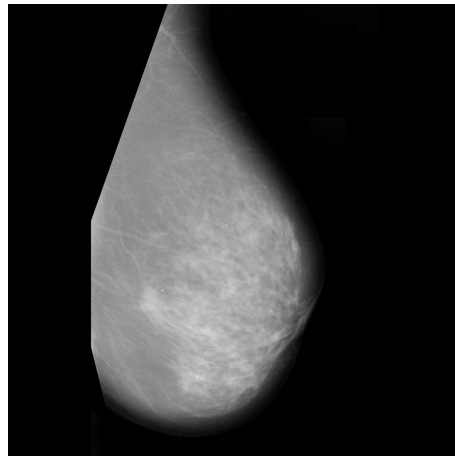


Figura 3.2: **MI sin artefactos**. Edición de la mamografía original obtenida de la base de datos mini-MIAS [4]

Una vez que las imágenes han sido editadas, se procedió a su normalización utilizando un código de programación en *C++*, y con el framework utilizado para análisis de datos del Centro Europeo para la Investigación Nuclear (CERN), ROOT versión 6.26/06 [42]. Este procedimiento consistió en estandarizar las coordenadas espaciales de las imágenes a una unidad, con la finalidad de facilitar su procesamiento. Las coordenadas inicialmente se guiaban bajo los píxeles que conforman la imagen, por lo que iban de 0 a 1024 en ambos ejes, pero mediante una transformación lineal de las coordenadas se ajustaron para que ahora vayan de 0 a 1 en unidades arbitrarias (u.a.). La clase utilizada para realizar esta normalización fue *TASImage*[43], interfaz que permite la manipulación de imágenes en el campo del análisis y procesamiento de imagen. Posterior a esta normalización,

cada pixel posee un tamaño de $\frac{1}{1024}$, esto quiere decir que cada pixel es una porción de tamaño $0,0009765 \times 0,0009765$ de la imagen total.

Además de una normalización espacial, se creó una nueva “Imagen parametrizada” mediante un mapa de intensidades, con la finalidad de obtener la información de la intensidad de brillo de cada pixel. Esto consistió en generar una matriz de tamaño 1024×1024 donde cada elemento representa un pixel de la imagen, mismo que almacena el valor de intensidad del brillo correspondiente. Estos valores siguen un rango que va desde 0 que representa al negro o a la ausencia de brillo (que serían zonas donde los rayos x no fueron absorbidos por la mama) hasta 1, que es el blanco, o el nivel de máximo brillo. Entre estos valores, existirán aquellos para tonos de grises. Sobre este nuevo espacio fase, se llevó a cabo el procedimiento posterior.

Como resultado se obtuvo una nueva imagen, conformada por un mapa de intensidades que mide el nivel de brillo en cada punto de la imagen y siguiendo una escala de coordenadas que van del 0 al 1 con unidades arbitrarias (u.a.). La figura 3.3 representa la nueva imagen normalizada y parametrizada de la Fig. 3.2.

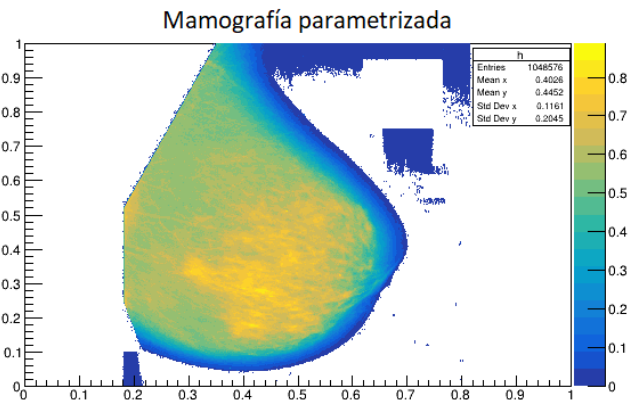


Figura 3.3: **Mapa de intensidad de mamografía 1.** Se observa una distribución de acuerdo a la intensidad de brillo en cada región de la imagen. (Imagen con unidades arbitrarias u.a.)

Además, se obtiene un archivo de datos con extensión .dat, que enlista cada uno de los 1, 048, 576 pixeles y en tres columnas: sus nuevas coordenadas x y y y el valor de intensidad de brillo correspondiente a cada pixel. Asimismo, la consola del programa muestra el punto más brillante de la imagen, sus coordenadas en la escala anterior y las normalizadas, y el valor de intensidad. Entonces a cada una de las 48 MI, le corresponde la MI original, la imagen redigitalizada y un archivo de datos de intensidades.

Este proceso permitió la manipulación de los datos, con la finalidad de obtener una muestra de cada imagen y poder aplicar en ella los aspectos matemáticos de interés. Esta transformación conservó toda la información original de la MI.

3.3. Obtención de la métrica Euclidiana y de Manhattan

El siguiente paso realizado, consistió en la aplicación de los fundamentos matemáticos incluidos en el capítulo dos. Al considerar la MI redigitalizada como una matriz, llamémosle matriz A , tenemos la función $\phi(A)$ que permite utilizar los valores de intensidad almacenados en esa matriz, como coordenadas que nos lleven a obtener las métricas Euclidiana y de Manhattan.

Para ello, se comenzó por delimitar el par de muestras de cada MI. Se buscaba que este par fuera del mismo tamaño, con la diferencia de que una muestra corresponda a la sección del tumor y la otra a una sección del tejido de fondo. Se emplearon los datos de radio tumoral ya establecidos

de la base de datos, pero transformados a un valor normalizado, es decir, el radio normalizado R fue igual a $\frac{r}{1024}$, donde r fue igual al radio en coordenadas originales. Cada mamografía contó con una sección delimitada diferente, pues dependía de su valor de radio.

El centro del tumor era un dato ya conocido, que el mismo programa proporcionó como el punto más brillante, a excepción de los sujetos 17, 27, 35, 37, 42, 46 y 48 que presentaban artefactos dentro de la región anatómica en la MI, y que eran detectados como el punto más brillante. Para estos sujetos se utilizaron las coordenadas normalizadas del centro de la base de datos.

Para obtener un centro para la muestra del tejido de fondo, se utilizó el archivo de datos correspondiente a cada MI, en él se buscó algún pixel que tuviera una intensidad menor a 0.7, pues como se concluye en el trabajo “Tumor and microcalcification characterization using Entropy, Fractal Dimension and intensity values statistical analysis in a mammography”[5], la intensidad de los tumores poseen un valor mayor o igual a 0.7 en una escala del 0 al 1. Además, concluye que la distribución de valores de intensidad en tejido mamario es consistente para cualquier tejido sano.

La Fig. 3.4 ilustra las dos muestras seleccionadas para la MI número 1, encerradas en un cuadrado rojo. Ambas regiones son del mismo tamaño.

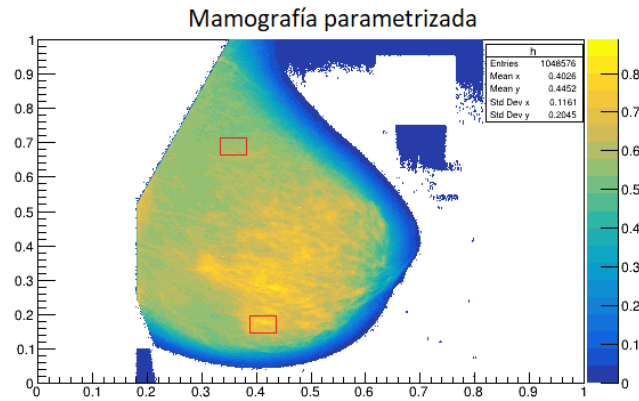


Figura 3.4: **Muestras seleccionadas en la mamografía 1.** La selección de las regiones dependió del tamaño del tumor proporcionado por la base de datos y los valores de intensidad de cada tipo de tejido.(Imagen con unidades arbitrarias u.a.)

Una vez conocidas ambas muestras, se empleó un código de programación, nuevamente en lenguaje C++ y mediante ROOT, en el que al ingresar valores de las coordenadas x y y de ambos puntos de muestra, calcularía el valor de la métrica Euclidiana y de Manhattan, siguiendo las ecuaciones 2.3 y 2.5 respectivamente:

$$d_e(z) = \sqrt{z_1^2 + z_2^2 + \dots + z_n^2}$$

$$d_M(z) = |z_1| + |z_2| + \dots + |z_n|$$

Se incluyeron los valores de intensidades de cada uno de los pixeles ubicados entre las coordenadas $[x_1, x_2]$ y $[y_1, y_2]$, donde x_1 es $x - R$, x_2 es $x + R$, y_1 es $y - R$ y y_2 es $y + R$.

Los valores de las métricas se almacenaron en un archivo de datos final. Ese archivo se utilizó para comparar los valores de los dos tipos de métricas y generar recursos gráficos que permitan ilustrar los resultados del siguiente capítulo.

3.4. Cálculo de cocientes

Para determinar la relación de los tejidos, se calculó el cociente entre los valores de cada métrica en los diferentes tejidos:

$$R_m = \frac{\text{Métrica del tejido sano}}{\text{Métrica del Tumor}}$$

Así como la relación entre las relaciones R_m de la métrica Euclidiana y la de Manhattan:

$$R = \frac{R_m \text{ métrica Euclidiana}}{R_m \text{ métrica de Manhattan}}$$

Para obtener una medida representativa de los resultados así como su variabilidad, se les calculó la media \bar{X} y la desviación estándar σ .

La ecuación de la media:

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$$

y el error estándar de la media SEM :

$$SEM = \frac{\sigma}{\sqrt{N}}$$

donde N es el número total de MI (48), X_i es cada uno de los cocientes y σ es la desviación estándar:

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N}}$$

Capítulo 4

Resultados y discusión

En este capítulo se muestran los resultados obtenidos a partir de concluir cada paso de la metodología; se presenta un listado y gráficas de datos que permiten realizar un análisis comparativo entre tejidos sanos y tumorales.

4.1. Resultados de los valores de las métricas

Como el objetivo es hacer una caracterización de tumores, se presenta un listado con los valores de métricas obtenidos, además de que dichos datos fueron empleados para hacer una comparativa mediante dos formas: una gráfica de puntos y el cálculo de su relación mediante cocientes.

La Tabla 4.1 muestra un listado de los valores resultantes del cálculo de las métricas Euclidiana y de Manhattan tanto de la sección tumoral como de la sección de tejido sano (o de fondo); cada MI está etiquetada y tiene sus cuatro valores agrupados por filas.

Valores de las métricas Euclidiana (ME) y de Manhattan (MM) de tumores y tejido sano				
MI	ME tumor	MM Tumor	ME Sano	MM Sano
1	44.3015	2477.26	30.9531	1685.76
2	61.4038	4659.59	38.6526	2938.73
3	180.694	35589.6	124.313	24157.4
4	82.2815	8341.43	45.9735	4541.05
5	155.973	33053	101.644	19575.3
6	93.5688	11114.6	46.8035	5517.77
7	45.447	2543.21	14.1484	703.336
8	100.98	16381.4	72.7345	11534
9	113.7229	14100.2	43.7208	5310.34
10	121.7961	16316.5	53.0073	5459.39
11	37.5995	2626.15	26.4654	1868.39
12	70.9347	5674.13	33.958	2661.8
13	60.5828	4633.07	21.2951	1428.55
14	68.13478	6393.82	52.3451	4918.14
15	47.037	2609.96	14.8358	773.766

Resultados y discusión

4.1 Resultados de los valores de las métricas

Valores de las métricas Euclidiana (ME) y de Manhattan (MM) de tumores y tejido sano				
MI	ME tumor	MM Tumor	ME Sano	MM Sano
16	81.0489	8993.93	46.8089	5230.93
17	26.7057	1193.2	25.2053	1146.26
18	50.2439	4361.44	32.4458	2743.91
19	42.1797	2452.31	30.6754	1792.83
20	66.4794	6444.56	47.6572	4664.43
21	32.1198	1804.02	26.4144	1520.18
22	35.8957	1954.42	29.035	1581.76
23	188.657	56656.9	131.426	44369.6
24	124.693	23634.2	89.5445	15847
25	105.641	18458.4	76.7701	12252.6
26	170.234	38618.5	127.257	28371.3
27	21.0735	703.891	18.5946	612.844
28	48.6032	4246.45	36.1747	3157.18
29	22.8491	775.18	12.8098	432.148
30	49.7518	3653.66	39.5614	2895.41
31	78.576	8827.75	33.1914	3260.3
32	87.0627	10050	68.5946	9325.42
33	138.462	33790.4	86.7206	19792.6
34	46.5352	2718.09	34.0889	1988.99
35	36.1032	1980.88	14.1401	757.141
36	81.3998	7972.39	44.2302	4304.04
37	163.018	36201.7	125.037	28975.2
38	99.0236	16179.1	64.3456	9840.7
39	114.302	18048.8	77.3438	12047.9
40	54.966	3608.69	37.1481	2394.48
41	102.22	14299.8	77.2927	9990.69
42	67.2055	7192.71	62.3717	6716.94
43	124.0383	21279.4	67.5995	11119.6
44	22.9515	619.52	15.4127	407.645
45	54.0879	4846.96	36.0836	3134.52
46	39.3747	2787.77	33.0418	2391.19
47	92.5192	11072.5	64.6135	7147.12
48	90.3173	12862.6	87.4297	12613.9

Tabla 4.1: **Valores numéricos de las métricas Euclidiana y de Manhattan.** La segunda y tercera columna corresponden a los valores de las métricas Euclidiana y de Manhattan de las muestras tumorales. La cuarta y quinta columna corresponden a las de tejido sano.

Las Fig. 4.1 es una gráfica de puntos que presenta los valores de la métrica Euclidiana para ambos tipos de tejido. Estas gráficas están seccionadas de acuerdo al tipo de tejido de fondo. Así mismo, la Fig 4.2 muestra los valores de la métrica de Manhattan para ambas muestras de tejido, y la comparación entre sí para cada paciente. En estos gráficos existen cuatro variables que permiten visualizar la relación entre mismo tipo de métricas y los tejidos con escalas adaptadas a cada una.

En ambas gráficas, las MI se ordenaron a lo largo del eje x siguiendo la numeración ascendente que se les asignó al momento de seleccionarlas, asociando a cada valor de x un par de valores de métricas para cada tipo de tejido. Además de seccionarlas para cada tipo de composición mamaria. Así se generó una estructura visual que permite un mejor análisis y una comparación entre datos más sencilla.

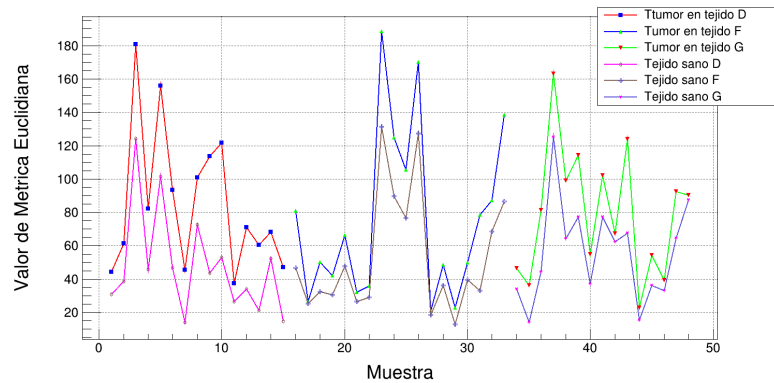


Figura 4.1: Comparación de métrica Euclidiana entre tumores y tejido de fondo

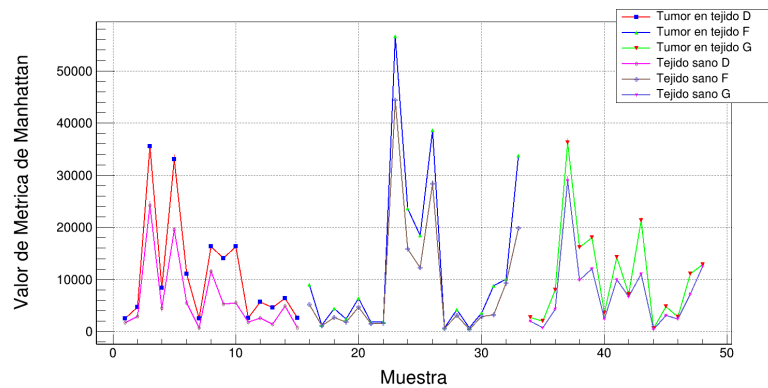


Figura 4.2: Comparación de métrica de Manhattan entre tumores y tejido de fondo

4.2. Resultados del cálculo de cocientes

Los valores de la relación entre ambos tejidos son presentados por medio de una tabla y un gráfico. La tabla 4.2 cuenta con tres columnas: la primera es del número de muestra, la siguiente es del cociente del valor de métrica Euclidiana del tejido sano entre el valor de métrica de Euclidiana del tumor y la última columna, que incluye cociente del valor de métrica de Manhattan del tejido sano entre el valor de métrica de Manhattan del tumor.

Cocientes de métricas de tejido sano entre tejido del tumor		
MI	Métrica Euclidiana	Métrica de Manhattan
1	0.6986	0.6804
2	0.6294	0.6306
3	0.6879	0.6787
4	0.5587	0.5443
5	0.6516	0.5922

Resultados y discusión
4.2 Resultados del cálculo de cocientes

Cocientes porcentuales de métricas de tejido sano entre tejido del tumor		
MI	Métrica Euclidiana	Métrica de Manhattan
6	0.5002	0.4964
7	0.3113	0.2765
8	0.7202	0.7040
9	0.3844	0.3766
10	0.4352	0.3345
11	0.7038	0.7114
12	0.4787	0.4691
13	0.3515	0.3083
14	0.7682	0.7692
15	0.3154	0.2964
16	0.5775	0.5816
17	0.9438	0.9606
18	0.6457	0.6291
19	0.7272	0.7310
20	0.7168	0.7237
21	0.8223	0.8426
22	0.8088	0.8093
23	0.6966	0.7831
24	0.7181	0.6705
25	0.7267	0.6637
26	0.7475	0.7346
27	0.8823	0.8706
28	0.7442	0.7434
29	0.5606	0.5574
30	0.7951	0.7924
31	0.4224	0.3693
32	0.7878	0.9279
33	0.6263	0.5857
34	0.7325	0.7317
35	0.3916	0.3822
36	0.5433	0.5398
37	0.7670	0.8003
38	0.6498	0.6082
39	0.6766	0.6675
40	0.6758	0.6635
41	0.7561	0.6986
42	0.9280	0.9338
43	0.5449	0.5225
44	0.6715	0.6580
45	0.6671	0.6466
46	0.8391	0.8577
47	0.6983	0.6454
48	0.9680	0.9806

Tabla 4.2: **Relación porcentual de las mediciones de las métricas Euclidiana y de Manhattan.**

Los valores mostrados en la Tabla 4.2 se presentan organizados gráficamente en la Fig. 4.3. Se encuentra un par de valores para cada muestra mamográfica, uno corresponde a la métrica

Euclidiana y el otro a la métrica Manhattan, y los datos corresponden a los errores relativos de las métricas de tumores respecto a las de tejido sano.

La media de los valores del cociente del tejido sano entre el tejido del tumor para la métrica Euclidiana fue:

$$\bar{X} = 0,6595 \pm 0,0227(SEM)$$

y su desviación estandar σ :

$$\sigma = 0,1578$$

para la métrica de Manhattan:

$$\bar{X} = 0,6496 \pm 0,0252(SEM)$$

y su desviación estandar σ :

$$\sigma = 0,1746$$

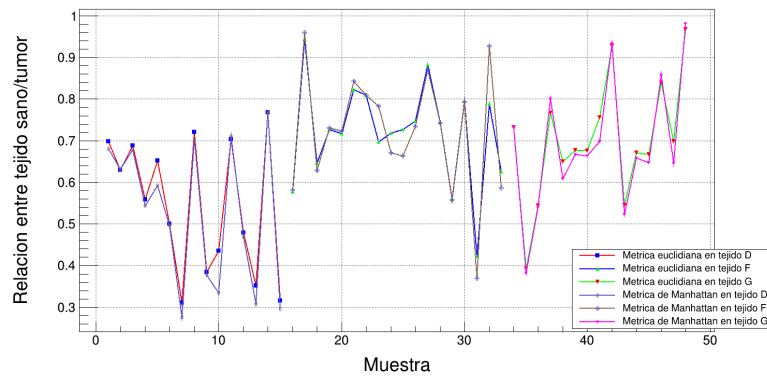


Figura 4.3: Comparación de los cocientes de tejidos de acuerdo a diferentes tipos de métricas.

La distribución de los valores de estos cocientes se pueden observar en los histogramas de las Fig. 4.4 y 4.5.

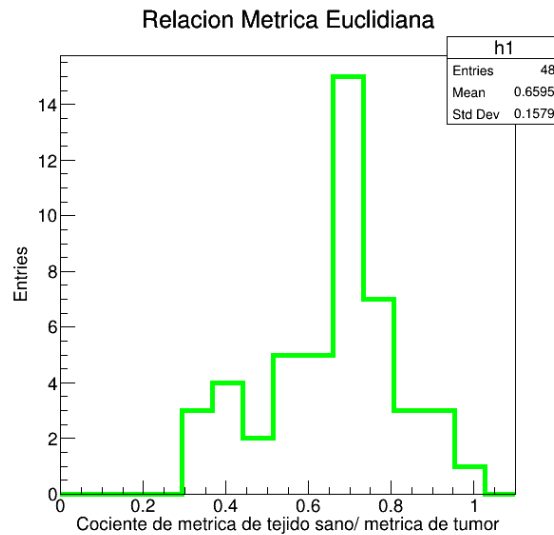


Figura 4.4: Histograma de los cocientes de los valores de métrica Euclidiana de tejido sano entre la metrica Euclidiana de tumor.

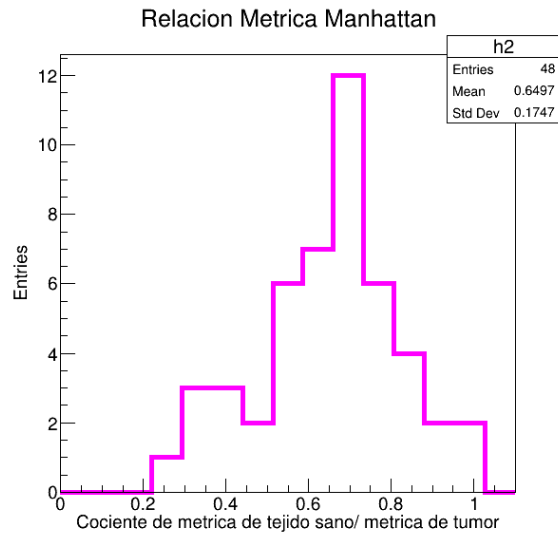


Figura 4.5: Histograma de los cocientes de los valores de métrica de Manhattan de tejido sano entre la métrica de Manhattan de tumor.

También se hizo el cálculo cociente de las dos relaciones calculadas para las métricas Euclidiana y de Manhattan. Este cálculo implica la comparación de los dos tipos de métricas mediante los cocientes calculados anteriormente. La tabla 4.3 enlista el resultado de realizar dicha operación.

Cocientes de la relación de la métrica Euclidiana entre la de Manhattan	
MI	Cociente
1	0.9739
2	1.0019
3	0.9866
4	0.9743
5	0.9087
6	0.9924
7	0.8883
8	0.9775
9	0.9796
10	0.7688
11	1.0107
12	0.9799
13	0.8771
14	1.0012
15	0.9399
16	1.0070
17	1.0178
18	0.9742
19	1.0052
20	1.0096

Cocientes de la relación de la métrica Euclidiana entre la de Manhattan	
MI	Cociente
21	1.0246
22	1.0005
23	1.1241
24	0.9337
25	0.9134
26	0.9827
27	0.9867
28	0.9989
29	0.9943
30	0.9965
31	0.8743
32	1.1777
33	0.9352
34	0.9989
35	0.9759
36	0.9935
37	1.0435
38	0.9360
39	0.9864
40	0.9817
41	0.9239
42	1.0062
43	0.9588
44	0.9798
45	0.9693
46	1.0221
47	0.9242
48	1.0130

Tabla 4.3: **Relación de los cocientes de las métricas Euclidiana y de Manhattan.**

La Fig. 4.6 es una gráfica de puntos que muestra el valor del cociente de cada sujeto de estudio en una escala del 0 al 1. Esta seccionada de acuerdo a los tres tipos de composición de tejido. La Fig. 4.7 es un histograma que ilustra las frecuencias de los cocientes de métricas calculados.

La media para estos valores fue de:

$$\bar{X} = 0,9777 \pm 0,0087(SEM)$$

y su desviación estandar σ :

$$\sigma = 0,0605$$

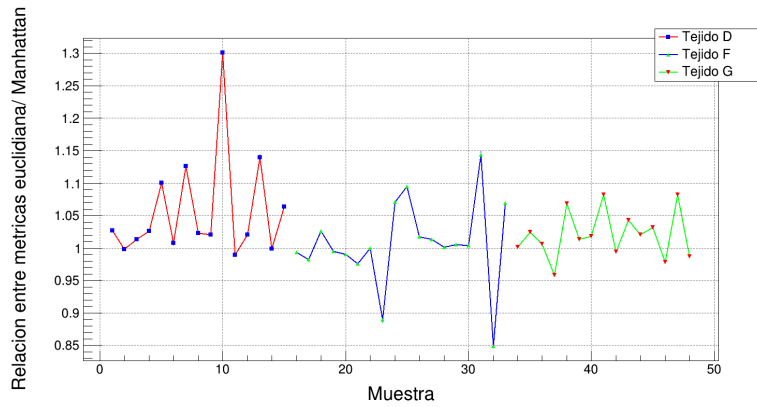


Figura 4.6: Comparación de los cocientes de tipos de métricas.

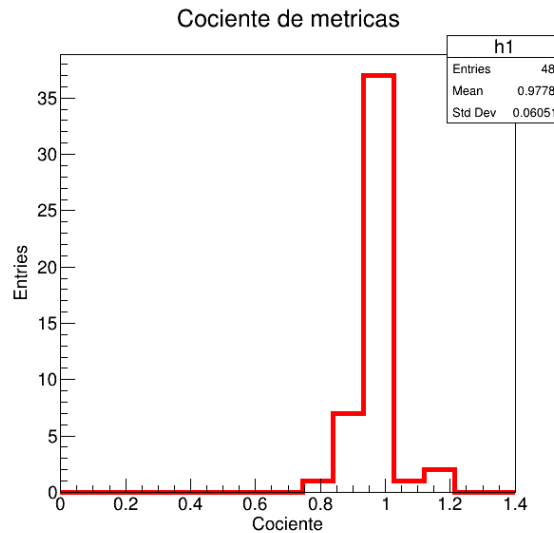


Figura 4.7: Histograma de los cocientes de los valores de métrica Euclidiana entre la métrica de Manhattan.

4.3. Discusión

Respecto a la lista de valores de la tabla 4.1 y a las gráficas de puntos de las Fig. 4.1 y 4.2, las medidas de distancias de cada MI son totalmente independientes una de la otra, y no siguen un rango de valores; esto se debe a que cada muestra de las MI poseen un tamaño específico según los radios, entonces la variación viene a partir de que cada uno de los 48 tumores considerados tienen radios diferentes. La caracterización que se busca, compara cada métrica pero solo con su homóloga perteneciente a la misma mamografía. La comparación de la escala de ambos graficos también nos permite ver que las medidas de la métrica Euclidiana, tanto de tumores como de tejido de fondo, son muy pequeñas respecto a las de la métrica de Manhattan. Estos resultados coinciden con las definiciones de ambas métricas, pues al ser la métrica Euclidiana la distancia más corta que hay entre dos puntos, se podía esperar valores más bajos en las mediciones, contrario a la distancia Manhattan, pues al seguir solo “movimientos” ortogonales y no diagonales, el “recorrido” a seguir tiende a ser más largo. Sin embargo, también se puede observar que la forma de las gráficas

evidencia una distribución similar en cada métrica, es decir, conforme el valor de una distancia cambia, el valor de su homóloga en la otra distancia también lo hace, de forma casi proporcional, por lo que la relación entre los datos de cada una de las métricas es similar para todas las imágenes mamográficas.

Además, de la Fig. 4.1 podemos observar que el comportamiento de los valores de la métrica Euclidiana para el tejido sano es muy parecido al del tumor pero en menor magnitud, es decir, los trazos de la gráfica tienen la misma forma y comportamiento. Lo mismo sucede con la Fig. 4.2 que indica que los valores de la métrica de Manhattan se comporta muy parecido en ambos tejidos, pero con una diferencia de magnitud en dichos valores. El motivo principal radica en que al tener dos regiones de tejido con características similares pero con diferentes niveles de intensidad, los valores de medición pueden ser múltiplos entre sí; siendo las muestras de tejido tumoral, que al tener mayor densidad mamaria y reflejar una imagen más brillante, poseen los valores de métricas más grandes.

De los valores de la tabla 4.2 y la gráfica de la Fig. 4.3 podemos hablar de cómo se relacionan los dos tipos de tejido entre sí, en cada una de las métricas. Primero, observemos que los valores de razón o cociente de las métricas tejido sano entre las del tumor tanto para la métrica Euclidiana como la de Manhattan son menores a 1. Es decir, los valores del tejido sano, indican ser menores que los del tumor, tal como se observaba en las gráficas de las Fig. 4.1 y 4.2. El comportamiento de ambos cocientes para cada sujeto fue muy similar, y en general el comportamiento de la gráfica se muestra igual para cada tipo de métrica. El promedio o media de esta relación para la métrica Euclidiana, indica que los valores del tejido sano son $0,6595 \pm 0,0227(SEM)$ veces los del tejido del tumor. En otras palabras, la métrica Euclidiana del tejido del tumor es mayor que la del tejido sano. Para la métrica de Manhattan, el cociente promedio entre los tejidos indica que los valores del tejido sano son $0,6496 \pm 0,0252(SEM)$ veces los del tejido del tumor, siendo, al igual que en la otra métrica, mayores en las muestras de tumores. En el histograma de la Fig. 4.4 se puede observar que la dispersión de los datos de relación entre tejidos respecto a la métrica Euclidiana tiene mayor frecuencia entre 0.6 y 0.8, que es el mismo para la métrica Manhattan de la Fig. 4.5, esto es, que en la mayoría de MI, las métricas de tejidos sanos son entre 0.6 y 0.8 veces las de tejido del tumor. En general, se puede observar que estos valores muestran una diferenciación entre los dos tipos de tejido.

Finalmente, la Fig. 4.6 muestra que la relación entre los cocientes de cada una de las métricas se aproximan a 1. Así mismo, la media de esta relación, que es de $0,9777 \pm 0,0087(SEM)$ nos indica que, la métrica de Manhattan y la Euclidiana se comportan de forma consistente en ambas muestras de tejido. O sea, que la manera en la que estas dos métricas se relacionan no indica diferencia alguna. El histograma de la Fig. 4.7 también puede indicar que la frecuencia predominante de estos valores es en el intervalo de 0.9 a 1. Entonces, podemos decir que la información que proporcionan ambos tipos de métricas es comparable, y que la caracterización de tumores es posible sin importar el tipo de métrica (Euclidiana o de Manhattan) utilizada.

Capítulo 5

Conclusiones

Esta tesis buscó caracterizar partes de tejido maligno mediante el cálculo de métricas geométricas utilizando el valor de intensidad de estos tejidos y los de fondo para una muestra de mamografías obtenidas de la base de datos mini-MIAS. A partir de la metodología seguida y los resultados obtenidos, se puede concluir que se cumplen con los objetivos y la hipótesis planteada al comienzo de la investigación.

Mediante los conceptos incluidos en el marco teórico, se tiene que al definir a una imagen digital como una matriz, se le puede aplicar una función de diferenciabilidad, de manera que se obtiene un conjunto de puntos, cuyos valores son las intensidades de brillo de los píxeles, que posteriormente se pudieron utilizar para cálculos de distancias geométricas. Fue importante darle un tratamiento previo a las MI, especialmente al momento de redigitalización ya que los artefactos interfieren al momento de buscar las regiones de mayor brillo, por lo que eliminarlos resulta útil.

Del análisis realizado a los resultados de los cálculos de distancias de las regiones mamográficas se puede concluir que la caracterización entre tejidos es válida, ya que el comportamiento de los datos refleja características propias de cada tipo de densidad de tejido, y al mostrar una relación de 0.65 y 0.64 para las métricas Euclidiana y de Manhattan respectivamente, también se logró deducir que la caracterización entre tumores es posible, sin importar el tipo de métrica calculada, pues el comportamiento con ambas es muy similar con una relación de 0.97. Sin embargo, cabe recalcar que este trabajo se centra únicamente en la distinción entre un tejido maligno contra el tejido “sano” de su misma imagen, y no distingue de tipos de tejido de fondo o nivel de malignidad del tumor. Para mayor precisión en los resultados, se debería incluir una muestra de MI más grande.

Se propone para trabajos posteriores hacer este análisis geométrico aplicado a tumores benignos, con la finalidad de comparar el comportamiento de los datos de métricas de ese tipo de tumores con los expuestos en este trabajo. También se puede complementar con un análisis cuantitativo de los radios de los tumores, de manera que se tenga el análisis completo en un algoritmo, y no sea necesario hacer uso de datos preestablecidos en la base de datos. Se puede complementar el trabajo si se realiza una prueba más detallada de este trabajo en la cual se pueda distinguir entre tipos de tejido de fondo, y así determinar la caracterización individualmente. De esta manera se podría tener mayor complementación y comparación entre los métodos de diagnóstico convencionales y las nuevas propuestas.

Si bien el cáncer de mama es un problema de salud grave al ser una de las principales causas de muerte a nivel mundial y la detección temprana es crucial para marcar la diferencia entre sobrevivir o no ante ello, es claro que la mamografía es una excelente herramienta para diagnóstico médico, especialmente al ser no invasiva, y en su versión digital, permite implementar métodos y técnicas que facilitan la labor.

El trabajo desarrollado en esta tesis permitiría ser un auxiliar más en cuanto al diagnóstico por imagen, sin dejar de lado que para garantizar asertividad en ello, se debe seguir el lenguaje universal para mamografías BI-RADS, y con ello el previo análisis cualitativo de la MI.

Bibliografía

- [1] ORGANIZACIÓN MUNDIAL DE LA SALUD (2022, 3 febrero) *Cáncer* Organización mundial de la salud. Recuperado 1 de marzo de 2023, de <http://www.who.int/es/news-room/fact-sheets/detail/cancer>
- [2] MARTÍNEZ SÁNCHEZ, Y. L. & MEDRANO GUZMÁN, R. (2022, 24 octubre). *Epidemiología del cáncer de mama*. gob.mx. Recuperado 1 de marzo de 2023, de <https://www.gob.mx/imss/articulos/epidemiologia-del-cancer-de-mama-318014>
- [3] INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA (2022, 17 octubre) *Estadísticas a propósito del día internacional de lucha contra el cáncer de mama (19 de octubre)* [Comunicado de prensa] www.inegi.org.mx/contenidos/saladeprensa/aproposito/2022/EAP_CANMAMA22.pdf
- [4] PILOT EUROPEAN IMAGE PROCESSING ARCHIVE. (s.f.) *The mini-MIAS data base of mammograms*. Base de datos. <http://peipa.essex.ac.uk/info/mias.html>
- [5] ZEPEDA, C.H, VÁZQUEZ, M.G, MORENO, E., DE CELIS, B. AND RODRÍGUEZ, M. (2021) *Tumor and microcalcification characterization using Entropy, Fractal Dimension and intensity values statistical analysis in a mammography*. DOI: 10.48550/arXiv.2101.11090
- [6] LÓPEZ, A. (2022) *Separación de regiones de interés para el diagnóstico y tejido de fondo en mamografías mediante análisis de datos*. [Tesis para obtener el grado de licenciado, Benemérita Universidad Autónoma de Puebla] Repositorio Institucional de Acceso Abierto BUAP. <https://repositorioinstitucional.buap.mx/>
- [7] RABIDAS, R., & ARIF, W., (2020). *Characterization of mammographic masses based on local photometric attributes*. Multimedia Tools and Applications, 79(29-30), 21967-21985. <https://doi.org/10.1007/s11042-020-08959-7>
- [8] SANTIAGO-MONTERO R, SOSSA H, GUTIÉRREZ-HERNÁNDEZ DA, ZAMUDIO V, HERNÁNDEZ-BAUTISTA I & VALADEZ-GODÍNEZ S. *Novel Mathematical Model of Breast Cancer Diagnostics Using an Associative Pattern Classification*. *Diagnostics (Basel)*. 2020 Mar 1;10(3):136. doi: 10.3390/diagnostics10030136. PMID: 32121569; PMCID: PMC7151177.
- [9] CHIAO JY, CHEN KY, LIAO KY, HSIEH PH, ZHANG G, & HUANG TC. *Detection and classification the breast tumors using mask R-CNN on sonograms*. *Medicine (Baltimore)*. 2019 May;98(19):e15200. doi: 10.1097/MD.00000000000015200. PMID: 31083152; PMCID: PMC6531264.
- [10] KAVITHA, T., MATHAI, P.P., KARTHIKEYAN, C. ET AL. *Deep Learning Based Capsule Neural Network Model for Breast Cancer Diagnosis Using Mammogram Images*. *Interdiscip Sci Comput Life Sci* 14, 113–129 (2022). <https://doi.org/10.1007/s12539-021-00467-y>

- [11] JOUIROU, A., BAÁZAOU, A., AND BARHOUMI, W. (2021). *Shared Information-Based Late Fusion for Four Mammogram Views Retrieval using Data-driven Distance Selection*. In Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2021) - Volume 4: VISAPP; ISBN 978-989-758-488-6; ISSN 2184-4321, DOI: 10.5220/0010261701440155
- [12] CHUNHAPRAN, O., NOOLEK, D., LABCHAROENWONGS, P., & YAMPAKA, T. (2022). *Multi-View Combination using Mutual Information and 3-D Euclidean Distance for Breast Cancer Classification*. Information Technology Journal, 18(2), 44-54. ISSN: 1685-8573
- [13] OCAMPO LE ROYAL R, & ARCHUNDIA GARCÍA A. (2013) *Tumor mamario*. Archundia García A(Ed.), Cirugía 2. McGraw Hill. <https://accessmedicina.mhmedical.com/content.aspx?bookid=1434§ionid=94946138>
- [14] DICCIONARIO DE CÁNCER DEL NCI. (s.f.-b). Instituto Nacional del Cáncer. <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/glandula-mamaria>
- [15] AMERICAN CANCER SOCIETY (2019, 18 septiembre) *¿Qué es el cáncer de seno?* American Cancer Society. Recuperado 11 de marzo de 2023, de <https://www.cancer.org/es/cancer/tipos/cancer-de-seno/acerca/que-es-el-cancer-de-seno.html>
- [16] DICCIONARIO DE CÁNCER DEL NCI. (s.f.). Instituto Nacional del Cáncer. <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/tumor>
- [17] WORLD HEALTH ORGANIZATION: WHO. (2023). *Cáncer de mama*. www.who.int. <https://www.who.int/es/news-room/factsheets/detail/breast-cancer>
- [18] GIULIANO A. E. & HURVITZ S. A. (2022) *Carcinoma de mama en mujeres*. Papadakis M.A., McPhee S.J., Rabow M.W., & McQuaid K.R. (Eds.), Diagnóstico clínico y tratamiento 2022. McGraw Hill <https://accessmedicina.bibliotecaabuap.elogim.com/content.aspxbookid=3153§ionid=268039478>
- [19] HENDRICK, R. E., IKEDA, D. M., & MAYAKE, K. K. (2017). *Breast Imaging: the Requisites* (3.a ed.). Elsevier (Obra original publicada 2004) <https://clinicalkey.bibliotecaabuap.elogim.com/#!/content/book/3-s2.0-B978032332904000001X>
- [20] THEMES, U. (2021, 16 mayo). *Breast imaging: Mammography*. Radiology Key. <https://radiologykey.com/breast-imaging-mammography/>
- [21] CARLYLE BUSHONG S. (2010). *Manual de radiología para técnicos*. Barcelona, España: ELSEVIER.
- [22] AMERICAN CANCER SOCIETY (2020, 5 marzo) *Conceptos básicos del mamograma*. American Cancer Society. <https://www.cancer.org/es/cancer/tipos/cancer-de-seno/pruebas-de-deteccion-y-deteccion-temprana-del-cancer-de-seno/mamogramas/conceptos-basicos-del-mamograma.html>
- [23] *Mamografía*. (s. f.). National Institute of Biomedical Imaging and Bioengineering. <https://www.nibib.nih.gov/espanol/temas-cientificos/mamograf%C3%ADa>

- [24] Mamografía. Analógica y digital. Historia, evolución. (2020). *Revista Argentina de Mastología*. Recuperado 8 de junio de 2023, de https://www.revistasamas.org.ar/revistas/2020_v39_n141/06.pdf
- [25] CENETEC-SALUD (2021, 15 abril). *Eficacia, seguridad y evaluación económica de la mamografía digital directa contra la mamografía digital indirecta o computada*. Gobierno de México <https://www.gob.mx/salud/cenetec/articulos/eficacia-seguridad-y-evaluacion-economica-de-la-mamografia-digital-directa-drm-contra-la-mamografia-digital-indirecta-o-computada-crm>
- [26] VILLAN, L. A., SANTALLA, A., CRIADO, M. S. L., GONZÁLEZ-PÉREZ, I., CALDERÓN, M., GALLO, J., & PARRA, J. Z. (2011). *Clasificación radiológica y manejo de las lesiones mamarias*. *Clínica e Investigación en Ginecología y Obstetricia*, 38(4), 141-149. <https://doi.org/10.1016/j.gine.2010.10.016>
- [27] *Densidad de los senos e informe de su mamograma*. (s. f.-c). American Cancer Society. <https://www.cancer.org/es/cancer/tipos/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>
- [28] HERNÁNDEZ, P. L. A., ESTRADA, T. T., PIZARRO, A., CISTERNAS, M. L. D., & TAPIA, C. S. (2016). *Calcificaciones mamarias: descripción y clasificación según la 5.a edición BI-RADS*. *Revista Chilena de Radiología*, 22(2), 80-91. <https://doi.org/10.1016/j.rchira.2016.06.004>
- [29] *Como entender su informe de mamograma*. (2019, 3 octubre). American Cancer Society. Recuperado 26 de junio de 2023, de <https://www.cancer.org/es/cancer/tipos/cancer-de-seno/pruebas-de-deteccion-y-deteccion-temprana-del-cancer-de-seno/mamogramas/como-entender-su-informe-de-mamograma.html>
- [30] GÓMEZ, M. (2013). *Historia(s) de la Imagen Digital*. InterArtive.org (eJournal/eMagazine edición) (Interartive2008). ISSN 2013-679X. Consultado el 9 de agosto de 2023 en <https://interartive.org/2017/04/historias-de-la-imagen-digital-marisa-gomez>
- [31] LUYOUN. (2015, 30 junio). *Gato abisinio escala de Grises*. Pixabay. <https://pixabay.com/es/photos/gato-abisinio-escala-de-grises-825302/>
- [32] GONZALEZ, R. C., & WOODS, R.E. (2018) *Digital Image Processing, Global Edition*. (4.a ed.). Pearson UK
- [33] DEL CASTILLO, G. F. T. (2012). *Differentiable manifolds*. En Birkhäuser Boston eBooks. <https://doi.org/10.1007/978-0-8176-8271-2>
- [34] LUCAS, P (s. f.) *Geometría y Topología* Open Course Ware Universidad de Murcia <https://www.um.es/documents/4874468/11782809/01-variedadesdiferenciabllessubvariedades-v100901.pdf/26b08c6c-2090-4f9c-9035-c7dc7fbda8ad>
- [35] HERRERO, P. J. (s. f.) *Topología de Espacios métricos* Open Course Ware Universidad de Murcia <https://www.um.es/documents/4874468/11035667/cap-1-esp-met.pdf/6c66d56e-c01d-4270-a56c-50ec184a830e>
- [36] CORTADA A. *La geometría taxicab: un mundo donde los círculos son cuadrados*. En: TEMat, 6 (2022), págs. 1-15. issn: 2530-9633. <https://temat.es/articulo/2022-p1>.

- [37] PARRA, W. & CÁRDENAS, R. A. (2013). *Estudio de la métrica de Manhattan. Segmentos, rectas, rayos, circunferencias y algunos lugares geométricos en la geometría del taxista*. <http://hdl.handle.net/20.500.12209/2188>.
- [38] COLABORADORES DE WIKIPEDIA. (2006). *Geometría del taxista*. Wikipedia, la enciclopedia libre. https://es.wikipedia.org/wiki/Geometr%C3%ADa_del_taxista#/media/Archivo:Manhattan_distance.svg
- [39] WOLFRAM RESEARCH, INC. (s.f.). *Relative error*. Wolfram MathWorld. <https://mathworld.wolfram.com/RelativeError.html>
- [40] Convertidor de formato <https://convertio.co/es/pgm-jpg/>
- [41] Editor de imágenes <https://picsart.com/es>
- [42] <https://root.cern> (c) 1995-2021, The ROOT Team; conception: R. Brun, F. Rademakers
Built for linuxx8664gcc on Jul 28 2022, 18:08:51 From tags/v6-26-06@v6-26-06 With c++
(Ubuntu 9.4.0-1ubuntu120.04.1) 9.4.0
- [43] *ROOT: TASIImage Class Reference*. (s. f.). ROOT: Reference guide. <https://root.cern.ch/doc/master/classTASIImage.html>