



**ESPE**  
UNIVERSIDAD DE LAS FUERZAS ARMADAS  
INNOVACIÓN PARA LA EXCELENCIA

**DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN**

**CARRERA DE INGENIERÍA EN SISTEMAS E INFORMÁTICA**

**TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL  
TÍTULO DE INGENIERO EN SISTEMAS E INFORMÁTICA**

**TEMA: “IMPLEMENTACIÓN DE UNA ONTOLOGÍA DE  
PATRONES DE IMÁGENES DE MAMOGRAFÍAS EN EL DOMINIO  
DE LA TEXTURA SEGÚN LOS COEFICIENTES DE HARALICK.”**

**AUTORES: CARRILLO GALLARDO, JAIME VINICIO  
ESCORZA GUAÑA, EDWIN ALCIDES**

**DIRECTOR: ING. DÍAZ RODRIGUEZ, OSWALDO EFRAÍN**

**SANGOLQUÍ**

**2020**



**ESPE**  
UNIVERSIDAD DE LAS FUERZAS ARMADAS  
INNOVACIÓN PARA LA EXCELENCIA

**DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN**  
**CARRERA DE INGENIERÍA EN SISTEMAS E INFORMÁTICA**

**CERTIFICACIÓN**

Certifico que el trabajo de titulación, “IMPLEMENTACIÓN DE UNA ONTOLOGÍA DE PATRONES DE IMÁGENES DE MAMOGRAFÍAS EN EL DOMINIO DE LA TEXTURA SEGÚN LOS COEFICIENTES DE HARALICK.” fue realizado por los señores Carrillo Gallardo, Jaime Vinicio y Escorza Guaña, Edwin Alcides, el mismo que ha sido revisado en su totalidad, analizado por la herramienta de verificación de similitud de contenido; por lo tanto cumple con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de Fuerzas Armadas ESPE, razón por la cual me permito acreditar y autorizar para que lo sustente públicamente.

Sangolquí, 19 de febrero del 2020

Firma:

Ing. Oswaldo Efraín Díaz Rodríguez  
C. C. 0400652020



DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN  
CARRERA DE INGENIERÍA EN SISTEMAS E INFORMÁTICA

**AUTORÍA DE RESPONSABILIDAD**

Nosotros, **Carrillo Gallardo, Jaime Vinicio** y **Escorza Guaña, Edwin Alcides**, declaramos que el contenido, ideas y criterios del trabajo de titulación: **“IMPLEMENTACIÓN DE UNA ONTOLOGÍA DE PATRONES DE IMÁGENES DE MAMOGRAFÍAS EN EL DOMINIO DE LA TEXTURA SEGÚN LOS COEFICIENTES DE HARALICK.”** es de nuestra autoría y responsabilidad, cumpliendo con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Consecuentemente el contenido de la investigación mencionada es veraz.

Sangolquí, 19 de febrero del 2020



---

Jaime Vinicio Carrillo Gallardo  
C.C.: 1721670444



---

Edwin Alcides Escorza Guaña  
C.C.: 1725023566



# ESPE

UNIVERSIDAD DE LAS FUERZAS ARMADAS  
INNOVACIÓN PARA LA EXCELENCIA

DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

CARRERA DE INGENIERÍA EN SISTEMAS E INFORMÁTICA

## AUTORIZACIÓN

Nosotros, **Carrillo Gallardo, Jaime Vinicio y Escorza Guaña, Edwin Alcides**, autorizamos a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: **“IMPLEMENTACIÓN DE UNA ONTOLOGÍA DE PATRONES DE IMÁGENES DE MAMOGRAFÍAS EN EL DOMINIO DE LA TEXTURA SEGÚN LOS COEFICIENTES DE HARALICK.”** en el Repositorio Institucional, cuyo contenido, ideas y criterios son de nuestra responsabilidad.

Sangolquí, 19 de febrero del 2020

---

Jaime Vinicio Carrillo Gallardo  
C.C.: 1721670444

---

Edwin Alcides Escorza Guaña  
C.C.: 1725023566

## **DEDICATORIA**

Este trabajo está dedicado en primer lugar a Dios, por darnos la fortaleza de concluir la carrera, a nuestros padres por el apoyo incondicional que nos brindaron diariamente, y hacer de nosotros unas mejores personas, a toda la familia que aportaron y estuvieron en los tiempos buenos y malos siendo así lo más valioso que Dios nos ha dado.

## **AGRADECIMIENTOS**

Agradecemos a Dios por ayudarnos a alcanzar una de las grandes metas.

A todos los familiares por sus consejos y ánimos que nos brindaron.

A los docentes que aportaron con nuestra formación académica.

Al tutor de tesis el Ing. Oswaldo Díaz que nos orientó y colaboro durante todo el proyecto de grado.

A nuestros compañeros por brindarnos su amistad en la universidad.

A la universidad de las fuerzas Armadas ESPE por abrirnos las puertas y cumplir nuestras metas, otorgándonos conocimientos y formándonos como profesionales

## ÍNDICE

CARÁTULA	
CERTIFICACIÓN .....	i
AUTORÍA DE RESPONSABILIDAD.....	ii
AUTORIZACIÓN .....	iii
DEDICATORIA.....	iii
AGRADECIMIENTOS .....	v
ÍNDICE .....	vi
ÍNDICE DE TABLAS .....	xii
ÍNDICE DE FIGURAS .....	xiii
RESUMEN .....	xvi
ABSTRACT .....	xviii
CAPITULO I	
PLANTEAMIENTO DEL PROBLEMA	
1.1. Antecedentes .....	1
1.2. Problemática .....	4
1.2.1 Contextualización del Problema .....	4
1.3. Formulación del problema .....	4
1.4. Justificación.....	5

1.5. Objetivos...	6
1.5.1 Objetivo General	6
1.5.2 Objetivos Específicos	6
1.6. Alcance	7

## CAPITULO II

### MARCO TEÓRICO

2.1. Introducción	9
2.2. Ontología	9
2.2.1. Definición de Ontología según la inteligencia artificial	10
2.2.2. Historia de la ontología	11
2.2.3. Componentes de una ontología	12
2.2.4. Taxonomía	13
2.2.5. Construcción de una ontología	14
2.2.6. Clasificación de las ontologías según el conocimiento	15
2.3. Metodología para desarrollar la ontología	16
2.3.1. Metodología Cyc	16
2.3.2. Metodología de Construcción de Ontologías de Uschold y King	17
2.3.3. Metodología de Construcción de Ontologías de Grüninger y Fox	18
2.3.4. Metodología KACTUS	19
2.3.5. Metodología SENSUS	19

2.3.6. Metodología On-To-Knowledge.....	20
2.3.7. Terminae .....	20
2.3.8. Methontology.....	21
2.4. Cáncer de mama.....	22
2.5. Categorías de Riesgo.....	24
2.5.1. Riesgo de cáncer de mama de la población en general.....	24
2.5.2. Riesgo moderado de cáncer de mama.....	24
2.5.3. Alto riesgo de cáncer de mama.....	25
2.5.4. Altísimo Riesgo de cáncer de mama.....	26
2.6. Métodos de diagnóstico del cáncer de mama .....	26
2.6.1. Exploración clínica de las mamas .....	27
2.6.2. Emisión de positrones .....	27
2.6.3. Mamografía .....	27
2.6.4. Ecografía ... ..	29
2.6.5. Punción con aguja fina (PAAF) y Biopsia con aguja gruesa (BAG).....	29
2.6.6. Resonancia Magnética.....	30
2.6.7. Biopsia Quirúrgica .....	30
2.6.8. Tratamientos .....	30
2.7. Composición de la mama.....	32
2.8. La textura .....	33

2.8.1. Texturas Naturales .....	33
2.8.2. Texturas artificiales .....	34
2.8.3. Texturas según coeficientes Haralick.....	34
2.9. Matriz de coocurrencia .....	34
2.9.1. Propiedades de la matriz de coocurrencia .....	35
2.9.2. Cálculo de la matriz de coocurrencia .....	36
2.10. Clasificación de imágenes.....	40
2.11. Coeficientes Haralick.....	41
2.11.1. Auto correlación.....	41
2.11.2. Densidad .....	41
2.11.3. Contraste .....	41
2.11.4. Correlación .....	42
2.11.5. Disimilitud .....	42
2.11.6. Energía .....	43
2.11.7. Entropía.....	43
2.11.8. Homogeneidad .....	44
2.11.9. Media.....	44
2.11.10 Desviación estándar .....	45
2.12. Patrones de textura presentes en las mamografías .....	45
2.12.1 Nódulos .....	45

2.12.2 Forma .....	45
2.12.3 Margen .....	46
2.12.4 Densidad .....	47
2.12.5 Calcificaciones .....	49
2.12.6 Ganglio linfático intramamario .....	55
2.12.7 Lesión cutánea .....	56
2.12.8 Conducto dilatado solitario .....	56
2.12.9 Ubicación de la lesión.....	57

### CAPITULO III

#### MODELO CONCEPTUAL DE LA ONTOLOGÍA

3.1 Introducción .....	59
3.2 Directrices para el diseño de la ontología.....	60
3.2.1 Niveles de grises .....	60
3.2.2 Pixeles en una imagen .....	61
3.2.3 Lista de coeficientes Haralick seleccionados .....	63
3.2.4 Rangos de los coeficientes Haralick.....	63
3.3 Diseño de la ontología con la metodología Methontology .....	64
3.3.1 Especificación de requisitos .....	65
3.3.2 Conceptualización .....	66
3.3.3 Formalización e implementación .....	68

3.3.4 Mantenimiento.....	68
--------------------------	----

3.4 Diseño de la Ontología previo a la entrada al software .....	69
--	----

## CAPITULO IV

### IMPLEMENTACIÓN DE LA ONTOLOGÍA PARA EL DOMINIO DE LOS COEFICIENTES HARALICK.

4.1. Explicación de librerías utilizadas.....	71
---	----

4.2. Descripción inicial del programa.....	73
--	----

4.3. Explicación del código fuente .....	76
--	----

4.4. Ejecución del programa.....	82
----------------------------------	----

4.5. Verificación de los coeficientes Haralick.....	85
---	----

## CAPITULO V

### CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones.....	88
------------------------	----

5.2. Recomendaciones.....	89
---------------------------	----

5.3. Trabajos Futuros .....	90
-----------------------------	----

BIBLIOGRAFÍA .....	91
--------------------	----

## ANEXOS

**ÍNDICE DE TABLAS**

Tabla 1 Preguntas de Investigación .....	8
Tabla 2 Porcentaje de riesgo de contraer cáncer de mama .....	24
Tabla 3 Matriz de distancia entre píxeles .....	37
Tabla 4 Matriz simétrica .....	38
Tabla 5 Matriz de probabilidades .....	39
Tabla 6 Construcción de la matriz de probabilidades .....	40
Tabla 7 Rangos de los coeficientes de Haralicks Patrón 1 .....	63
Tabla 8 Rangos de los coeficientes de Haralicks Patrón 2 .....	64
Tabla 9 Especificación de requisitos .....	65

## ÍNDICE DE FIGURAS

Figura 1	Modelo conceptual de la ontología.....	3
Figura 2	Composición de la mama.....	32
Figura 3	Texturas naturales.....	33
Figura 4	Texturas artificiales .....	34
Figura 5	Representación de los niveles de grises en una ventana .....	36
Figura 6	Formas de los nódulos (Ovalada, redonda, irregular) .....	46
Figura 7	Margen de los nódulos (Circunscrito, Oscurecido).....	47
Figura 8	Margen de los nódulos (Microlobulado, Indefinido, Espiculado) .....	47
Figura 9	Densidad de los nódulos (Hiperdenso, Isodenso).....	48
Figura 10	Densidad de los nódulos (Hipodenso, Contenido adiposo).....	49
Figura 11	Calcificación cutánea .....	50
Figura 12	Calcificaciones Vasculares.....	50
Figura 13	Macrocalcificaciones .....	51
Figura 14	Calcificaciones de líneas gruesas y redondas .....	52
Figura 15	Calcificaciones anulares y distróficas.....	52
Figura 16	Calcificaciones de leche cálcica e hilos de sutura.....	53
Figura 17	Calcificaciones con morfologías amorfas y heterogéneas .....	54
Figura 18	Calcificaciones con morfología finas pleomorfas y lineales finas .....	55
Figura 19	Ganglio linfático intramamario.....	55
Figura 20	Lesión cutánea.....	56
Figura 21	Conducto dilatado solitario .....	57

Figura 22	Ubicación de la lesión de la mama.....	58
Figura 23	Modelo conceptual de la ontología - Detallado .....	60
Figura 24	Distintos niveles de grises.....	61
Figura 25	Imagen dividida en píxeles (8 x 8).....	62
Figura 26	Imagen dividida en 16 celdas con valores de tonos de grises .....	62
Figura 27	Tareas de la actividad de conceptualización de Methontology .....	67
Figura 28	Proceso de almacenamiento de información en la base de datos .....	68
Figura 29	Diagrama de flujo del proceso realizado por la ontología.....	70
Figura 30	Interfaz gráfica de la ontología .....	73
Figura 31	Opción Cargar Imagen .....	74
Figura 32	Opción Dibujar Imagen.....	74
Figura 33	Opción coocurrencia .....	75
Figura 34	Opción sacar Haralicks .....	75
Figura 35	Opción Cerrar.....	76
Figura 36	Clase inicial (Interfaz gráfica) .....	77
Figura 37	Coeficientes Haralick.....	78
Figura 38	Clase para obtener la matriz de coocurrencia.....	79
Figura 39	Función para dibujar una imagen.....	80
Figura 40	Función Haralicks.....	81
Figura 41	Proceso de almacenamiento .....	82
Figura 42	Seleccionar la imagen para procesarla .....	82
Figura 43	Obtener la matriz de coocurrencia .....	83
Figura 44	Obtener Haralicks.....	83

Figura 45 Selección de archivo JSON con metadata de mamografía.....	84
Figura 46 Almacenamiento de los datos .....	84
Figura 47 Menú verificar.....	85
Figura 48 Ingreso de coeficientes Haralick.....	86
Figura 49 Resultado de coincidencia con los coeficientes Haralick ingresados. .	86
Figura 50 Imagen dibujada basada en la matriz de grises. ....	87

## RESUMEN

El presente trabajo tiene como finalidad proporcionar un prototipo de una ontología desarrollada en Python, que establezca mediante los patrones reconocidos y procesados de una mamografía, el BI-RADS (Clasificación de 0 a 6) para ayudar a dar una alerta temprana de un posible cáncer de mama.

Es decir, esta investigación va a comprender el desarrollo e implementación de un intérprete (ontología) de patrones de textura de una mamografía que sirva como evaluador para un diagnóstico de cáncer de mama, dicho intérprete empleará los coeficientes de textura Haralick. Los resultados serán guardados en una estructura JSON.

Para emitir un resultado aproximado del problema mamario detectado, se utilizará la escala BI-RADS y así dar un diagnóstico previo, utilizando un rango de parámetros de cada caso ingresado para obtener un porcentaje de aproximación al caso detectado según el experto.

Palabras clave:

- **ONTOLOGÍA**
- **BI-RADS**
- **CANCER DE MAMA**
- **MATRIZ DE COOCURRENCIA**
- **COEFICIENTES HARALICK**

## **ABSTRACT**

The purpose of this work is to provide a prototype of an ontology developed in Python, which establishes through the recognized and processed patterns of a mammogram, the BI-RADS (Classification 0 to 6) to help give an early warning of a possible cancer of breast

That is, this research will include the development and implementation of an interpreter (ontology) of mammography texture patterns that serve as an evaluator for a diagnosis of breast cancer, the interpreter will use Haralick texture coefficients. The results will be saved in a JSON structure.

To emit an approximate result of the detected breast problem, the BI-RADS scale will be used and thus give a previous diagnosis, using a range of parameters of each case entered to obtain a percentage of approximation to the case detected according to the expert.

Keywords:

- **ONTOLOGY**
- **BI-RADS**
- **BREAST CANCER**
- **COOCURRENCE MATRIX**
- **HARALICK COEFFICIENTS**

## CAPITULO I

### PLANTEAMIENTO DEL PROBLEMA

El presente trabajo tiene como finalidad proporcionar un prototipo de una ontología desarrollada en Python, que establezca mediante los patrones reconocidos y procesados de una mamografía, el BI-RADS (Clasificación de 0 a 6) para ayudar a dar una alerta temprana de un posible cáncer de mama.

#### 1.1. Antecedentes

El cáncer de mama es una enfermedad que frecuentemente se presenta en las mujeres a nivel mundial porque constituye el 16% de todos los cánceres femeninos (UNIVERSIDAD ESTATAL PENÍNSULA DE SANTA ELENA, 2017).

Según la Organización Mundial de la Salud (OMS), en el mundo ha registrado cerca de 500.000 nuevos casos de cáncer por año llegando a producir alrededor de 250.000 muertes por este padecimiento (UNIVERSIDAD ESTATAL PENÍNSULA DE SANTA ELENA, 2017). Ante la presencia de altas cifras de casos, muertes asociadas al cáncer de mama, se ha identificado como un padecimiento que forma parte del mundo desarrollado, pero las incidencias varían notablemente a nivel mundial.

Esta enfermedad se encuentra en el lugar 12 en el cuadro de factores de muerte, impacta a las mujeres de todas las edades sin excepción, pero la mayoría de los casos se presentan a partir de los 40 años en adelante (INEC, 2017).

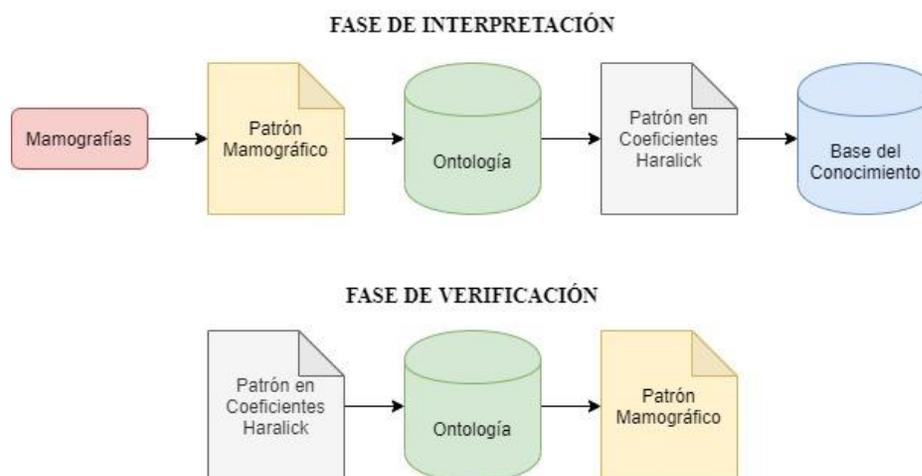
Según los datos proporcionados por el Instituto Nacional de Estadística y Censo a partir del 2012 en el Ecuador se registró 2760 muertes por causa de esta enfermedad (INEC, 2017).

Se han identificado que existen diversos métodos que se aplican para la detección, control temprano del cáncer de mama, donde se destaca el uso de mamografía, que se caracteriza por ser factible de aplicarlo, bajo costo y resultados más precisos (MIERCOLA, 2012).

En el Proyecto REDU 013 que se titula “Desarrollo de un sistema inteligente para el análisis automático de lesiones mamarias utilizando técnicas de inteligencia artificial y otras técnicas computacionales de análisis de imágenes digitales” con el apoyo de varias instituciones públicas, de igual forma las privadas donde se encuentra inmersa la Universidad de las Fuerzas Armadas-ESPE.

El objetivo del proyecto REDU es estimular el desarrollo de Proyectos de Investigación e Innovación para el perfeccionamiento del conocimiento de las diversas áreas en el fomento a la Ciencia y Tecnología en el Ecuador.

Seguidamente se va a presentar un esquema enfocado a un modelo conceptual que se busca alcanzar a través del proyecto denominado “Desarrollo de un sistema inteligente para análisis automático de lesiones mamarias utilizando técnicas de Inteligencia Artificial y otras técnicas computacionales de imágenes”.



**Figura 1.** Modelo conceptual de la ontología

En la Figura 1 se expone la obtención de patrones mamográficos que existen en las mamografías, por ende, estos son un conjunto de datos que no están controlados, de cada mamografía se recibe varios patrones con sus respectivos metadatos. La ontología se encarga de transformar los datos y los metadatos de los patrones mamográficos en valores tabulares de acuerdo con los coeficientes de Haralick para alimentar la base del conocimiento.

Para la implementación de la ontología se aplica a todas las imágenes de patrones que se extraen de las mamografías; se cuenta con el marco teórico sobre la indagación de textura según los coeficientes de Haralick, además de la formación académica que se obtuvo de la carrera de Ingeniería en Sistemas e Informática de la Universidad de las Fuerzas Armadas ESPE, este proyecto se lo presenta como proyecto previo a la graduación, en él se propone realizar dicha ontología para generar registros, en su estructura sintáctica y semántica que formarán parte del sistema inteligente para la detección del cáncer de mama.

## **1.2. Problemática**

### **1.2.1 Contextualización del Problema**

El análisis y dictamen del cáncer de mama se lo hace a través de imágenes (mamografías, ecografías, ...) que son interpretadas por un especialista en este tipo de cáncer, los diagnósticos obtenidos en ciertos casos no son precisos por el error humano del médico, o porque existen escasos detalles que se pueden obtener de una imagen a simple vista.

Por lo tanto, se proyecta desarrollar un sistema inteligente que facilite el análisis automático a las lesiones mamarias empleando un conjunto de técnicas de inteligencia artificial, de igual forma técnicas computacionales para la evaluación de imágenes digitales; el requerimiento del sistema es que cuente con elementos de sintaxis, semántica, conocimiento sobre los aspectos de textura que arrojan las imágenes de los patrones mamográficos para que se pueda detectar el nivel de cáncer en el paciente.

### **1.3. Formulación del problema**

Se ha identificado que los resultados que las mamografías entregan son poco efectivos, porque los resultados son entregados en base al conocimiento, experiencia que poseen los médicos, este tipo de valoración puede ser validado a través de la ejecución de los exámenes mamarios, en la actualidad no existe un método o mecanismo de valoración de las mamografías a través del análisis de textura (Coeficientes de Haralick) que permitan determinar el nivel de BI-RADS del cáncer de mama.

## 1.4. Justificación

El cáncer de mama se define como una enfermedad en que se establecen células malignas al interior de los tejidos de la mama, por lo cual la forma para que se pueda detectar estas anomalías es la realización de un autodiagnóstico mamario, que es la identificación de las masas extrañas que se han formado en los tejidos mamarios empleando el tacto, se considera como uno de los principales indicadores de la existencia de cáncer de mama (maligno o benigno) (MAYO CLINIC, 2018).

Otra forma de detección es mediante una exploración clínica (empleada esencialmente en mujeres jóvenes) y mediante el uso de medios de diagnóstico como son las mamografías (utilizada esencialmente en mujeres mayores de 40 años) (RadiologyInfo.org, 2018).

Al ser la mamografía uno de los exámenes más precisos y utilizados por mujeres, hay que considerar los resultados falsos negativos y los falsos positivos, es así que surge la idea de obtener resultados más específicos realizando un filtrado y clasificación de los patrones que se consiguen de una mamografía convirtiendo la imagen en valores numéricos que permitan el manejo de los mismos de acuerdo a la necesidad (Febles, 2012).

El proyecto se lo realizará mediante un modelo que servirá como intérprete entre los patrones obtenidos de la mamografía frente a los coeficientes Haralick de textura analizados.

El problema radica en que se requiere un gran número de casos documentados (prototipos) para desarrollar y entrenar un clasificador o un intérprete y hacer frente al problema con ciertas garantías.

En la presente investigación se busca el desarrollo de una ontología que sirva de intérprete entre los patrones de las imágenes de las mamografías utilizando coeficientes de textura Haralick que ayude a convertir las imágenes en datos numéricos que posteriormente permita que se identifique el aspecto de lesiones cancerígenas u otros diagnósticos.

## **1.5. Objetivos**

### **1.5.1 Objetivo General**

Implementar una ontología para interpretar los patrones de imágenes extraídas de las mamografías generando un registro numérico de los mismos basados en los coeficientes de textura Haralick que alimente una base del conocimiento que permita el diagnóstico de cáncer de mama.

### **1.5.2 Objetivos Específicos**

- Revisar el marco teórico de los coeficientes de textura Haralick.
- Definir los límites, alcances y el dominio de la ontología.
- Conocer los patrones mamográficos en el ámbito médico y lo que nos indica cada uno.
- Implementar la ontología utilizando el lenguaje de programación Python, utilizando librerías asociadas al procesamiento de imágenes y matrices.

- Obtener un resultado aproximado del problema mamario según BI-RADS, en base a los coeficientes de textura Haralick registrados.

## 1.6. Alcance

El desarrollo tiene como propósito generar los registros semánticos, sintácticamente estructurados para alimentar la base del conocimiento, a través de una ontología que tendrá como insumos los patrones extraídos de las mamografías.

Esta investigación comprende el desarrollo e implementación de un intérprete (ontología) de patrones de textura de una mamografía que sirva como evaluador para un diagnóstico de cáncer de mama, dicho intérprete empleara los coeficientes de textura Haralick. Los resultados serán guardados en una estructura JSON, de acuerdo con el diseño de la base de conocimiento.

Para emitir un resultado aproximado del problema mamario detectado, se utilizará la escala BI-RADS y así dar un diagnóstico previo, utilizando un rango de parámetros de cada caso ingresado para finalmente obtener un porcentaje de aproximación al caso detectado según el experto.

Para hacer la validación del funcionamiento de la ontología, primero se debe tener alimentada la base interna de la ontología con casos reales ya procesados. Después se ingresa los coeficientes de un caso que se desea analizar; es ahí cuando se realiza una búsqueda en la base de datos para determinar el caso al cual más se asemeje, dependiendo de la distancia que exista entre los coeficientes Haralick ingresados frente a los coeficientes Haralick que se encuentran en la base del

conocimiento. Al finalizar este proceso, se obtiene una relación en porcentaje del caso al que pertenece y seguidamente se muestra la imagen del patrón al cual está asociado, esto mediante la matriz de grises que se encuentra almacenada en la base de datos.

Para delinear de forma adecuada el alcance de la investigación planteada, se proponen varias preguntas de investigación asociadas a los objetivos específicos, tal como se muestra en la tabla 1.

**Tabla 1**  
*Preguntas de Investigación*

<b>Objetivo específico</b>	<b>Pregunta de investigación</b>
Revisar el marco teórico de los coeficientes de textura Haralick.	¿Se ha utilizado el análisis de textura Haralick en una mamografía para la detección de cáncer?
Definir los límites, alcances y el dominio de la ontología	¿Se ha realizado la implementación de una ontología para el análisis de textura utilizando imágenes extraídos de una mamografía?
Conocer los patrones mamográficos en el ámbito médico y lo que nos indica cada uno.	¿Qué tipos de problemas mamarios existen y cuáles son sus características?
Implementar la ontología utilizando el lenguaje de programación Python, utilizando librerías asociadas al procesamiento de imágenes y matrices.	¿En qué lenguajes de programación se han implementado las ontologías?
Obtener un resultado aproximado del problema mamario según BI-RADS, en base a los coeficientes de textura Haralick registrados.	¿Existe algún proceso para realizar el análisis de la textura mamográfica a través de coeficientes Haralick?

## **CAPITULO II**

### **MARCO TEÓRICO**

#### **2.1. Introducción**

Dentro del marco teórico se describe a detalle los aspectos teóricos y definiciones más relevantes que deben ser analizados, comprendidos para que se pueda diseñar una ontología; conjuntamente se describe lo más relevante en lo que tiene que ver con el cáncer de mama, adicionalmente se ha descrito las herramientas que se emplean en el proceso de diseño y realización de la ontología.

#### **2.2. Ontología**

La palabra ontología se puntualiza como el esclarecimiento sistemático al interior del área de la inteligencia artificial, es necesario que se definan a los respectivos conceptos básicos, de igual forma las que son parte de un vocabulario de un espacio temático para que se puedan definir las respectivas reglas que permitan la combinación de términos, relaciones para que se extienda el vocabulario.

Es importante señalar que la ontología no solo contiene términos determinados claramente, sino que se pueden presentar otros términos que pueden definirse a través del uso de reglas planteadas para que existan extensiones en el vocabulario; de tal forma se considera como una explicación explícita para una conceptualización simultánea (Pérez; Puentes; Guzmán, 2011).

Un sinónimo que frecuentemente se emplea para describir a una ontología es una generalización porque una ontología se conforma de una descripción formal y a la vez explícita sobre una conceptualización que se enfoca en narrar un modelo sobre un fenómeno que ocurre en el mundo en el que se descubre los conocimientos de mayor importancia, pero permite que se explique los conceptos ya empleados, junto con las restricciones existen sobre los fenómenos que ya están definidos, adicionalmente la ontología posee una característica formal relacionada a la idea enfocada a que una ontología posee un conocimiento consensuado, de tal forma una ontología no solo es para un sujeto, más bien requiere que se acepte en un grupo. (Chantal, 2002).

### **2.2.1. Definición de Ontología según la inteligencia artificial**

Se ha identificado una primera definición de ontología que se presentó en 1991 por Neches que posee una relación a la inteligencia artificial donde se muestra que se encarga de la definición de términos básicos que poseen relaciones para que se conforme un vocabulario para un área específica, de igual forma un conjunto de reglas que dan paso a la combinación de los términos, y las respectivas relaciones que permitan la definición de las respectivas extensiones en los vocabularios. (Neches, Fikes, & Finin, 1991)

Pero en 1993 se presenta una ampliación a la definición proporcionada por (Gruber, 1993) donde expresa que una ontología se puntualiza como una descripción rotunda para un concepto, por lo cual este término procede de la filosofía, en la cual una ontología se considera como un recuento sistemático sobre la existencia. Por lo cual,

en los sistemas de Inteligencia Artificial, lo que coexiste puede ser representado, de tal forma el conocimiento que posee un dominio posee una representación a través de un formalismo de carácter declarativo, por ende, un conjunto de objetos tiene la capacidad para que sea representado que toma el nombre de universo del discurso. Por lo cual este tipo de conjunto de objetos, de igual forma las relaciones establecidas entre ellos se reflejan al interior de un vocabulario el cual se representa un conocimiento en base a un sistema de conocimiento.

Asimismo, al interior del contexto de IA, se puede describir a una ontología que posee un programa que es parte de un conjunto de términos, de esta manera las definiciones poseen varias asociaciones con los respectivos nombres de formas que posee el universo en la disertación a través del uso de textos perceptibles por parte de los humanos, para que se detallen los significados de los nombres, de igual la presencia de axiomas serias que restringen la interpretación, un adecuado uso de los términos, a nivel formal una ontología se considera como una teoría lógica.

### **2.2.2. Historia de la ontología**

Etimológicamente la palabra ontología proviene del griego que representa “estudio del ser”, lo cual significa que se enfoca en la búsqueda de la interpretación de la forma que funciona el mundo a través de varios objetos que necesitan de una asignación de un concepto para que se pueda comprender un comportamiento de un entorno, por lo cual cuando se relaciona al área de la inteligencia artificial se puede comentar que una ontología se considera como una ciencia que evalúa a los

componentes, conductas que posee un objeto y los respectivos principios a los que se rige (Fernández , 2003).

### 2.2.3. Componentes de una ontología

En una ontología se ha identificado que los elementos básicos que entregan una representación explícita a la información para la construcción que normalmente se emplean varios componentes que son: principios, funciones, relaciones, demandas y clases (Fernández , 2003) (Sánchez, 2007):

- **Clases:** Es un tipo de ontología en la cual se ejecuta una descripción a la respectiva información en un dominio específico definido en las taxonomías, porque se considera como parte en la descripción de un proceso, labor, función entre otros.
- **Relaciones:** Es la representación de la interacción junto con el respectivo enlace con los conceptos que posee el dominio, para dar paso a formar parte de la taxonomía del dominio, en el cual se pueden enumerar las respectivas relaciones que determinan la respectiva equivalencia, de igual forma la taxonomía, intención, funcionalidad.
- **Funciones:** Se considera como un tipo de relación en el que se reconoce a un elemento cualquiera para realizar un adecuado cálculo de una función en la que se requiere realizar un cálculo a una función en la que se involucra varios elementos que son parte de una ontología.
- **Axiomas:** Son definiciones que poseen un valor de verdad que permiten el establecimiento de las respectivas restricciones, argumentos a las relaciones, donde se registra la información y se concluye la nueva.

- **Instancias:** Permiten la representación de los respectivos objetos específicos al interior de un dominio.

#### 2.2.4. Taxonomía

Es la ciencia que se encarga de estudiar la clasificación de sistemas jerarquizados, los niveles de dicho sistema se los nombra como categoría taxonómica, cuando los niveles de clasificación se tornan más complejos aparecen categorías intermedias, desde el ámbito de las ontologías la taxonomía se cataloga como una organización ontológica donde se agrupan las entidades, es decir que nos sirven para modelar diseños de bases de datos basados en conocimientos semánticos (Fernández , 2003) (Arija, 2012).

La ontología en esta tesis se lo utiliza como un elemento que ayude a representar imágenes en coeficientes numéricos, se relaciona los números con la textura de una imagen, la entrada es una imagen de una mamografía la cual será descrita en varios coeficientes numéricos de textura a través de la matriz de coocurrencia que se genera de una matriz de grises, los valores de la textura tendrán un valor de aproximación muy próximo a la realidad ya que los niveles de grises están en la escala de 0-255 y las imágenes con las cual se trabajará son de 256x256 píxeles.

La ontología iniciará con un conocimiento básico ya que no se han realizado cálculos previos de coeficientes numéricos de textura para mamografías, por lo que conforme avance su alimentación de conocimiento la ontología y sus representaciones irán perfeccionando.

### 2.2.5. Construcción de una ontología

EL proceso para construir una ontología es algo más que una actividad de ingeniería cada grupo de desarrollo es diferente por lo que cada uno elige la manera de construir con sus propios principios y criterios de diseño, pero cuando existe un falto de métodos generalizados para la definición de una ontología la consecuencia es que se puede repetir ontologías existentes además pueden tener los siguientes problemas (Pérez; Puentes; Guzmán, 2011):

- Los modelos conceptuales que posee una ontología se encuentran sobrentendidos al interior del código de la implementación.
- Los diversos compromisos ontológicos al igual que los respectivos criterios para el diseño se encuentran incluidos y claros en los respectivos códigos de la ontología.
- Los especialistas en el dominio, al igual que los usuarios finales que no logran una adecuada comprensión de las ontologías que se encuentran codificadas como parte de un lenguaje ontológico en el cual se consideran como las bases en un conocimiento tradicional.
- La codificación directa es el resultado de un proceso de adquisición del conocimiento que se considera como un paso muy repentino, en especial para el caso de las ontologías complejas.
- Las respectivas preferencias que posee el responsable del desarrollo de la ontología en un lenguaje particular generan condicionamientos durante la implementación del conocimiento que se ha adquirido.

- Los responsables del desarrollo de ontologías que no poseen experiencia en los lenguajes en los que se codifican las ontologías presentan el riesgo de tener dificultades para comprender las ontologías que se han creado o implementado porque el uso de herramientas tradicionales ontológicas permite que se enfoquen en varios aspectos de implementación que en los temas de diseño.

Hasta la fecha no existe una metodología estándar que permita el perfeccionamiento de una ontología, pero en la medida que crece el uso de las ontologías, se presentan varias metodologías que permiten el desarrollo donde han identificado a varios pasos que deben seguir para que se construyan las respectivas ontologías, varias de estas metodologías se detallan a continuación:

#### **2.2.6. Clasificación de las ontologías según el conocimiento**

Se ha identificado que existen 4 tipos de ontologías esenciales en base al alcance (Sánchez, 2007):

- **Ontologías de dominio:** Pose un conocimiento técnico al interior de un campo específico como lo es la medicina.
- **Ontologías genéricas:** Es la representación de conceptos generales tales como las estructuras y los respectivos procesos.
- **Ontología de la aplicación:** Su uso permite que se lo realice en cualquier tipo de aplicación como ejemplo se puede mencionar al campo geográfico que permite que se identifique las fallas tectónicas.
- **Ontologías técnicas básicas:** Permite que se especifique las características sobre las funciones, procesos.

## 2.3. Metodología para desarrollar la ontología

Las ontologías de dominio poseen conceptos más específicos en relación con las ontologías de referencia centrales, que contienen el concepto fundamental de dominio. Las ontologías fundamentales se pueden ver como metodologías que detallan las concepciones de nivel superior o primitivos que se emplean para la definición de otras ontologías. En general las ontologías no están dedicadas a un dominio específico, por lo que sus conceptos pueden ser tan generales. como las de ontologías de referencia centrales (Roussey,2007).

Existen varios tipos de metodologías que se van a explicar a continuación:

- CYC
- Uschold y King
- Grüninger y Fox
- Kactus
- Methontology
- Sensus
- On-To-Knowledge
- Terminae

### 2.3.1. Metodología Cyc

Como primer punto se debe extraer manualmente el conocimiento común que está implícito en diferentes fuentes. Seguidamente tras haber obtenido suficiente información sobre la ontología, se puede obtener un nuevo conocimiento común a través del uso de herramientas que permitan el procesamiento del lenguaje natural o a nivel del aprendizaje computacional. Por lo cual permite la construcción de la ontología Cyc, este tipo de metodología sugiere un conjunto de pasos que se explican a continuación (Fernández , 2003) (Valencia, 2005):

- Simbolización manual al conocimiento tácito y evidente que se obtuvo a través del uso de varias fuentes.
- Simbolización de los respectivos conocimientos empleados varias herramientas de software.
- Representación del mayor porcentaje en la codificación de las respectivas herramientas.

### **2.3.2. Metodología de Construcción de Ontologías de Uschold y King**

Esta metodología propone un conjunto de pasos para el desarrollo de las ontologías, los cuales son: establecer el fin, conseguir los conceptos , las respectivas relaciones entre los conceptos, los términos que se han empleado para tratar a los conceptos, las relaciones, recopila la ontología, por lo cual es necesario que se documente, evalúe y se lo pueda aplicar a otras ontologías para que se creen nuevas, Asimismo se creó la Enterprise Ontology, la cual sugiere los subsiguientes pasos (Valencia, 2005) (Fernández , 2003):

- Establecer el fin
- Lograr la ontología
- Sistematización
- Constitución de las ontologías existentes
- Valoración
- Expedientes.

### **2.3.3. Metodología de Construcción de Ontologías de Gruninger y Fox**

Esta metodología muestra que el primer paso es la identificación de las posibles aplicaciones en las que se empleara la ontología. A continuación, se emplea un grupo de preguntas en el lenguaje común, que se denominan aspectos de competencia para que se determine el área de la ontología.

Se emplean este tipo de preguntas para la extracción de los conceptos esenciales, de igual forma las propiedades, relaciones, axiomas que se perfeccionan formalmente en Prolog. Consiguientemente, este tipo de metodología es formal porque aprovecha la robustez que posee la lógica clásica, a la vez es se usa como una guía que permite la transformación de los escenarios informales, al interior de los modelos computables.

Esta metodología se la empleo para la elaboración de la ontología TOVE, la cual sugiere los siguientes pasos (Gruninger, 1995) (Sanchez, Comas, & García, 2017):

- Espacios motivantes.
- Asuntos informales de competencia.
- Procesos formales
- Temas formales de competencia
- Propositiones formales
- Propositiones de completitud

#### **2.3.4. Metodología KACTUS**

Esta metodología permite la construcción de una ontología en base al conocimiento a través del uso de un proceso de abstracción, por lo cual cuantas más aplicaciones se creen las ontologías se transforman en más generales, pero se alejan más de una base de conocimiento. Pero, en otras palabras, cuando se propone una construcción de una base de conocimiento para una aplicación determinada. Posteriormente cuando se requiere de una base nueva de conocimientos al interior de un dominio parecido, se procede a generalizar la inicial base de conocimiento al interior de una ontología para que se adapte a las dos aplicaciones, y asimismo continuamente. Por lo cual la ontología representa al conocimiento consensuado requerido para cada una de las aplicaciones. Esta metodología ha permitido la construcción de una ontología que permita el diagnóstico de fallo, y por lo cual recomienda los siguientes pasos (Bernaras, Laresgoiti, & Correa, 1996) (Hernández & Saiz, 2007):

- Descripción de la aplicación.
- Croquis previo que se basa en condiciones ontológicas top-level notables
- Arreglo y organización de la ontología

#### **2.3.5. Metodología SENSUS**

La metodología basada en Sensus es un enfoque top-down para diseñar ontologías específicas del dominio a partir de ontologías de gran tamaño. Los autores proponen identificar un conjunto de términos semilla que son importantes en un dominio particular. Estos términos se incorporan de forma manual a una ontología que tiene un dominio amplio, pero los usuarios seleccionan de forma automática a los términos más

importancia que permitan la descripción del dominio, y acoten a la ontología Sensus, por lo cual un algoritmo entrega un grupo de términos ordenados de una forma ordenada para que se describa un dominio, que pueda ser empleado como parte de una estructura que sea base del conocimiento (Swartout, Patil, Knight, & Rus , 1997).

### **2.3.6. Metodología On-To-Knowledge**

Este tipo de metodología entrega un conjunto de guía para que se introduzcan conceptos, herramientas que permitan la gestión para el conocimiento en el aspecto laboral, aportando ayuda a los respectivos buscadores para que lo puedan presentar de una forma efectiva, eficiente. Esta metodología permite que se identifique los propósitos que deben ser extraídos a través del uso de las respectivas herramientas para la gestión del conocimiento, el cual se basa en el análisis a los diversos escenarios en los que se van a aplicar y los varios roles que los entes van a realizar para la generación del conocimiento en la ontología, por lo cual esta metodología propone los siguientes pasos (Staab, Studer, Schnurr, & Sure, 2001)

- Estudio de viabilidad
- Comienzo
- Refinamiento
- Evaluación

### **2.3.7. Terminae**

Terminae es una metodología que se emplea como una herramienta que permite la creación de ontologías donde se toman a los diversos textos, se realiza una primera

instancia a través de un estudio lingüístico a los textos, se lo realiza a través del uso de diversas herramientas en el desarrollo del lenguaje natural.

En una ontología se emplean dos herramientas las cuales son Syntex y Camaleón por lo cual Syntex permite el reconocimiento de las relaciones, términos, pero el Camaleón permite que se identifique a los roles. Su uso permite que se mantenga el respectivo significado que las palabras poseen, de igual forma el contexto de las frases para un dominio específico, por lo cual una ontología posee dos actividades principales las cuales son: (García, 2006) (Valencia, 2005):

- Los resultados que Syntex entrega se emplean para que se determine un conocimiento y se pueda identificar a las inferencias de este.
- Obtener de forma sistemática el texto para que proporcione el conocimiento que sea posible.

### **2.3.8. Methontology**

Methontology se define como una metodología que es la base para la elaboración de ontologías que parten desde cero, o a su vez se lo pueden emplear otras ontologías previamente a la realización de una reingeniería; de igual forma una metodología permite que se desarrolle a un nivel de conocimiento donde se ha identificado un respectivo procedimiento para la creación de una ontología en la cual es necesario que se incluya una conceptualización, aplicación, gestión para la respectiva configuración, unificación y valoración.

El proceso de la ontología durante en el respectivo ciclo de vida en base a los prototipos que van cambiando, la metodología Methontology detalla los pasos que deben aplicarse para cada actividad, de igual forma los productos que se logran, la respectiva evaluación para que valide el funcionamiento. Este tipo de metodología se ha empleado para la construcción de varias ontologías tales como la ontología química, de hardware y software, etc.

Los pasos que Methontology sugiere se explican a continuación (Valencia, 2005):

- Descripción.
- Conceptualización.
- Formalización.
- Ejecución.
- Sostenimiento.

## **2.4. Cáncer de mama**

El cáncer de mamá es la primer causa de muerte en las mujeres, los casos con mayor incidencias se ha dado en los países desarrollados, en menor medida se presenta en los países en desarrollo, esto se debe a los factores demográficas, los cambios hormonales y los estilos de vida de cada país (Icaza, Nuñez, & Bugueño , 2017), el cáncer es ya considerado como un complicación de salud pública que se produce en un alto porcentaje por la presencia de los factores genéticos y ambientales, y en un bajo porcentaje se presenta por la mutación germinal al interior del gen de predisposición a que se produzca el cáncer, se ha identificado que otros tipos de

factores inciden en el apareamiento del cáncer entre los cuales son la actividad física, los hábitos de alimentación, la duración de lactancia que le den a los hijos, el consumo de sustancias estupefacientes, hábitos de fumar, tomar y la exposición a radiación ionizante alcohol se calcula que hay una probabilidad de que 1 de cada 9 mujeres pueda desarrollar cáncer (Domelles , Santos da Silva, Oliveira, Goldim, & Asthon-Prolla, 2015).

El cáncer de mama se compone de un grupo de tumores que muestra un comportamiento inusual, el procedimiento de los tumores es difícil de identificar para realizar un diagnóstico temprano, los estudios que se han realizado han permitido especializar los tratamientos para cada paciente pero no es una regla general, especializar permita dar un diagnóstico más preciso a los pacientes con alto riesgo de poseer cáncer (Martinez & Castro, 2018).

Se ha realizado esfuerzos para reducir la muerte a causa de este terrible padecimiento, como no existe aún la cura para el cáncer, pero se busca que se diagnostique tempranamente la enfermedad para que se puedan realizar los tratamientos oportunamente, en los últimos tiempos se ha logrado a través de las mamografías reconocer cánceres mamarios más pequeños, reduciendo el nivel de mortalidad por esta causa. En estos años se han registrado factores de riesgo al que están expuestos, se debe conocer estos factores para poder determinar grupos de mujeres de alto riesgo cuantificando su riesgo que permita desarrollar estrategias de prevención para el cáncer de mama; para esto se ha definido las siguientes categorías de riesgo de adquirir cáncer (Peralta, 2011)

## 2.5. Categorías de Riesgo

### 2.5.1. Riesgo de cáncer de mama de la población en general

El riesgo que tienen estas personas tiene una variabilidad según la edad como se muestra en la siguiente tabla:

**Tabla 2**  
*Porcentaje de riesgo de contraer cáncer de mama*

Edad	Porcentaje de riesgo
30	0.5%
50	2%
70	7%

Fuente: (Peralta, 2011)

### 2.5.2. Riesgo moderado de cáncer de mama

Este grupo es conformado por mujeres que tienen el doble de riesgo de poseer cáncer entre los factores se presenta los que se indican a continuación (Peralta, 2011):

- **Edad:** Es una relación directamente proporcional a mayor edad existe mayor riesgo de cáncer con un riesgo relativo a 1.2 a 2.0.
- **Factores reproductivos:** Se aumenta el riesgo en un valor mínimo por la menopausia tardía, tener un parto tardío después de los 30 años entre otros.
- **Enfermedades mamarias benignas proliferativas sin atipias:** Si en los resultados de una biopsia se presentan tumores papilares, pero sin atipias esto significa que el riesgo de contraer cáncer es menor.

- **Terapia de sustitución hormonal en la menopausia:** Cuando una mujer necesita un tratamiento por la menopausia lo debe hacer de una manera correcta para evitar el aumento de riesgo de contraer cáncer.
- **Involución mamaria incompleta:** Se denomina involución mamaria a la reducción de ducto-lobulillares a lo largo de la vida de las mujeres se presenta el riesgo de cáncer cuando no existe involución.
- **Sobrepeso después de la menopausia:** La enzima aromatasa produce niveles altos de estrógenos en el tejido adiposo elevando el riesgo a un 20%.
- **Alcohol:** La ingesta de alcohol acrecienta el riesgo de contraer cáncer de mama, pero esta se da cuando se consume la siguiente dosis de alcohol sobre 8 grados hasta 12 grados, si el nivel se excede de los 12 grados entonces el nivel de riesgo se duplica.
- **Mayor densidad mamográfica:** Si se detecta una mayor densidad en las mamografías, pero depende del nivel de extensión y grado de la densidad.

### 2.5.3. Alto riesgo de cáncer de mama

En la siguiente categoría el riesgo que tiene una mujer de contraer cáncer es de tres a diez veces más, existen varios grupos que pertenecen a esta categoría que se muestra a continuación:

- **Herencia:** En la mayoría de los casos de personas con cáncer no se presenta por la transmisión hereditaria, se estima que entre el 5% a 10% tienen cáncer hereditario por tener los mismos genes. El riesgo se incrementa cuando la consanguineidad es entre hermanas, madre o abuela, si el número de afectados

en la familia es concurrente entonces hay más posibilidades de que se presente el cáncer, aunque varios estudios indican que el estar en contacto con el mismo medio ambiente que una persona que presenta cáncer podría tener alguna posibilidad de contraer por estar expuestas a los mismos factores.

- **Antecedentes de lesiones precursoras:** Las lesiones predecesoras son benignas como la Neoplasia lobulillar, hiperplasia ductal, atipia plana.
- **Antecedente personal de cáncer de mama:** Esto ocurre cuando ya la persona tuvo cáncer presentando el riesgo de presentar un segundo en la mama contralateral.
- **Densidad mamográfica aumentada:** Si la mamografía registra una densidad mayor a 50% en el tejido mamario.

#### 2.5.4. Altísimo Riesgo de cáncer de mama

- **Portadoras de mutaciones de alta penetrancia en genes BRCA 1 BRCA 2:** Este se debe cuando el cáncer es hereditario.
- **Antecedentes de radioterapia al tórax (cáncer linfático):** Se da por la pérdida del funcionamiento de los cromosomas, son alteraciones a causa de las mutaciones se estima un riesgo de 50% a 80% de desarrollar el cáncer.

#### 2.6. Métodos de diagnóstico del cáncer de mama

Los médicos buscan diagnosticar de una manera temprana para que los resultados ayuden a establecer tratamientos óptimos, para este objetivo se utiliza los siguientes métodos (Asociación Española Contra el Cáncer, 2014):

### **2.6.1. Exploración clínica de las mamas**

Es el examen que se lo realiza anualmente mediante el tacto comprobando si existe alteraciones en la mama, como en la edad de 30 a 35 años no existe un alto riesgo de que se presente el cáncer entonces no se hace más exámenes, solo se lo hace si en la familia existió más casos.

### **2.6.2. Emisión de positrones**

El examen que se realiza por medio de positrones se basa en utilizar radiosondas apoyados en una cámara y una computadora para constatar el estado de los tejidos y órganos, la emisión de positrones emite un resultado de la actividad metabólica del tejido tumoral, superior al de la mamografía ya que los positrones no se ven afectados por la densidad del tejido glandular, existe cambios en las mamas por la cirugía mamaria, implantes, radioterapia (RadiologyInfo.org, 2017) (Martinez Villaseñor & Gerson Cwilich, 2006).

La emisión de positrones es un tipo de diagnóstico por imágenes de medicina nuclear que constituye una subespecialidad del ámbito de las imágenes médicas que funcionan mediante material radioactivo, esto se debe a que estos procesos permiten que se detecte actividades moleculares ayudando a la identificación de enfermedades de forma temprana, este modelo de diagnóstico no es invasivos y utilizan radiofármacos (RadiologyInfo.org, 2017).

### **2.6.3. Mamografía**

Es una técnica radiológica que consta de la parte axilar y planos profundos pretorácicos, la mamografía identifica tejidos normales o tejidos con alguna anomalía

los tres tejidos que se identifican en la mama son los siguientes adiposo, glandular, fibro-conectivo y pretende diferenciar entre un tejido normal y un tejido con anomalías.

Se ha convertido en el examen más eficaz para el diagnóstico temprano del cáncer ya que supera a la detección por exploración clínica, existen varias técnicas de mamografías (Argentina.gob.ar, 2010) (Asociación Española Contra el Cáncer, 2014) (Buffa, Gamarra, & Viniegra, 2011):

- **Mamografía convencional:** El proceso para obtener la mamografía es disparar fotones a través de la mama, la desventaja es que la mamografía obtenida no se puede modificar.
- **Mamografía digital:** Lo que se obtiene es una imagen basada en píxeles que se procesa en una computadora, la ventaja de que se lo realice en una computador es que se puede emitir un análisis sin la necesidad de efectuar impresiones, pero en el dictamen final no existe una gran diferencia con la mamografía convencional, en la mamografía digital se puede mejorar el diagnóstico al tener la posibilidad de manipular el contraste de la imagen, como consecuencia de la aparición de la mamografía digital implica que los profesionales médicos deben capacitarse en el ámbito tecnológico, además el cambio de procesos analógicos a digitales conlleva una gran inversión por lo que no todos los centros médicos están en la capacidad de adquirirlo.
- **Localización asistida por computadora(CAD):** Una vez que se realiza la lectura a través de la mamografía digital, se realiza un segundo análisis para aumentar la efectividad del diagnóstico, el CAD ayuda señalando los sitios de sospecha de un

indicio de cáncer, se recomienda que se utilice el CAD una vez que el especialista haya hecho su interpretación, se debe manejar de una manera adecuada el CAD, caso contrario esta herramienta se convierte en una desventaja y reduce el rendimiento del radiólogo.

- **Tomosíntesis:** El proceso de la tomosíntesis se basa en realizar 15 proyecciones de diferentes ángulos formando una mamografía digital más completa, posteriormente se realiza cortes en la mamografía digital de alrededor de 1 mm de forma paralela, la lectura se lo realiza en cada corte; la desventaja de esta técnica que es muy costoso.

#### **2.6.4. Ecografía**

La ecografía es una técnica que utiliza ondas sonoras que permite la evaluación de tejidos sin emplear radiaciones ionizantes, la ecografía muestra claramente los órganos y tejidos internos ya que las ecografías muestran imágenes macroscópicas, esta técnica no emite radiación, es ideal para realizar diagnósticos a mujeres jóvenes ya que ellas tienen mamas considerablemente densas (Bru, 1992) (Asociación Española Contra el Cáncer, 2014) (Walker & Cartwright, 2012) (Villa Veásquez, Pabón Romero, Jaramillo Botero, & Mantilla Suárez, 2006)..

#### **2.6.5. Punción con aguja final (PAAF) y Biopsia con aguja gruesa (BAG)**

PAAF es una técnica el cual su proceso se trata de adquirir una muestra de células mediante una jeringa se ejecutan movimientos para obtener una muestra mayor de representación celular para realizar un análisis citológico. BAG es una técnica igual que el PAAF que obtiene una muestra celular pero esta lo hace a través de agujas de

corte, en este caso se utiliza anestesia, la desventaja es por el tamaño de la muestra cómo se trata de un corte no se puede sacar grandes muestras por lo que si la lesión de la mama es grande se lo puede detectar, si la lesión es muy pequeña es difícil de diagnosticar (Vega, 2011).

#### **2.6.6. Resonancia Magnética**

La resonancia magnética, por sus siglas RM, es una técnica de imágenes utilizada por los médicos para obtener imágenes de los tejidos y los órganos del cuerpo, a través de la emisión de campos magnéticos y ondas de radio.

Este proceso es utilizado de forma no invasiva y sirve para realizar el análisis de los tejidos u órganos que tengan algún problema interno.

#### **2.6.7. Biopsia Quirúrgica**

La biopsia es una extracción del tejido benigno para una evaluación anatomopatológico, en la actualidad la biopsia ha quedado obsoleta por el avance de nuevas técnicas.

#### **2.6.8. Tratamientos**

Una vez que se le ha detectado cáncer de mama a una persona se presenta varios tratamientos que se muestran posteriormente (Asociación Española Contra el Cáncer, 2014):

- **Cirugía:** La cirugía es un tratamiento cuya ejecución varía según el tamaño del tumor, si el tumor no tiene un tamaño representativo entonces solo se extirpa el tumor, pero si el tumor es considerable se puede extirpar toda la glándula mamaria.

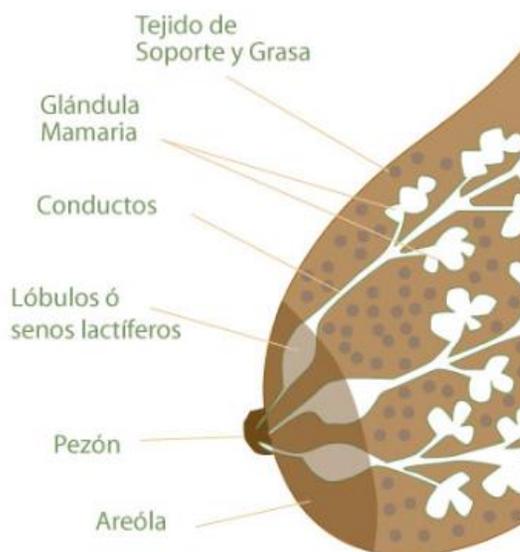
En el caso de extirpar toda la mama se puede optar por una reconstrucción mediante silicona o trasplantes miocutáneos.

- **Radioterapia:** Se aplica tras la realización de una cirugía, usa radiaciones ionizantes sobre la mama y en la pared torácica para eliminar las células cancerígenas que hubiesen quedado tras la operación.
- **Quimioterapia:** Es el tratamiento que se basa en la utilización de fármacos usualmente por vía intravenosa para eliminar las células cancerígenas que se hayan dispersado por todo el cuerpo, el tratamiento busca disminuir el tamaño del tumor antes de cualquier cirugía.
- **Tratamiento hormonal:** Las células cancerígenas poseen las mismas características de las células de la glándula mamaria por lo que al aplicarle hormonas complementarias como estrógenos, se ha demostrado que al aplicar hormonas se disminuye el riesgo de desarrollar cáncer nuevamente mejorando la salud del paciente.
- **Tratamientos Biológicos:** Las células de la mama emiten sustancias como proteínas, hormonas, cuando existe la presencia de algún tumor maligno estas sustancias se generan en mayor medida, por lo que biológicamente se combate con anticuerpos que ayudan a reducir las sustancias sobre las células malignas. Los anticuerpos tienen la capacidad de actuar solo sobre las células cancerígenas y no sobre las células normales, este tratamiento para su mayor eficacia se puede combinar con otros tratamientos.

## 2.7. Composición de la mama

La mama es una glándula mamaria con la función principal de alimentar a las crías en el caso de los mamíferos, la mama aparece de la profundidad de la piel el tejido que conforma la mama es en su mayor parte adiposo con una densidad fibroglandular. La densidad es bastante alta por lo que en ocasiones no se puede detectar nódulos, la densidad conlleva a que haya menos sensibilidad (D'Orsi, 2016) (Rivas, Urbina, Serpas , & García , 2006).

La piel que cubre la mama es más gruesa en la parte inferior en comparación con la parte superior de la mama. La glándula mamaria está compuesta generalmente de 20 a 25 lóbulos los cuales están ubicados junto al tórax en la misma dirección que el pezón, los lóbulos no son independientes, sino que están entrelazados por vasos sanguíneos (AEGO, 2010).



**Figura 2.** Composición de la mama

Fuente: (cirugiasdelamama.com, 2017)

## 2.8. La textura

La textura representa la parte externa de cualquier objeto, estas pueden ser artificiales o naturales, las texturas se asocian directamente con el color y la luz ya que estos dos aspectos le dan una diferente perspectiva a los objetos, también se debe considerar la relación que existe entre el ojo y el tacto, en ocasiones la textura visual no concuerda con la textura táctil, por lo que cuando se observa una textura visual de modo bidimensional no transmite lo que lo hace una textura táctil que puede ser áspera, suave, lisa (Caeiro, 2017). En otras palabras, la textura se la define como variado número de elementos visibles ordenados en un espacio de forma aleatoria donde los patrones son repetitivos. Las texturas se clasifican en dos grupos las texturas naturales y las texturas artificiales (Vergara, 2003):

### 2.8.1. Texturas Naturales

Las texturas naturales no poseen patrones que pueden ser identificados por ejemplo son todas las texturas creadas por la naturaleza que por lo general son patrones irregulares por ejemplo la madera, rocas, césped (Vergara, 2003).



**Figura 3.** Texturas naturales

Fuente: (shutterstock.com, 2018)

### 2.8.2. Texturas artificiales

Las texturas artificiales se rigen a ciertos patrones de manera uniforme y repetitiva.



**Figura 4.** Texturas artificiales

Fuente: (shutterstock.com, 2018)

### 2.8.3. Texturas según coeficientes Haralick

Haralick publicó en el año de 1973 un trabajo sobre el análisis textural mediante las matrices de co-ocurrencia (GLCM) y propuso 14 tipos de coeficientes basándose en la matriz anterior mencionada para realizar la caracterización de una textura (Haralick, 1973).

Este tipo de análisis se basa en la descripción de una textura a través métodos estadísticos obtenidos de la solución de la matriz de grises y la matriz de co-ocurrencia.

## 2.9. Matriz de coocurrencia

La matriz de coocurrencia indica la frecuencia con la que un nivel de gris aparece en una imagen relacionando con toda la imagen dentro de un píxel definitivo. La matriz de coocurrencia refleja el comportamiento que tienen los píxeles frente a sus píxeles cercanos en cada una de las coordenadas arriba, abajo, izquierda o derecha (Presutti, 2004).

### 2.9.1. Propiedades de la matriz de coocurrencia

- **Cuadrada:** Todos los píxeles deben tener el mismo volumen de vecinos por lo que en la matriz debe tener el mismo número de filas y columnas. La imagen puede llegar a tener valores de dos bits que simbolizan 4 celdas, una imagen de 8 bits que representa 256 celdas o píxeles, si la imagen es de 256 x 256 la imagen cuenta con 65536 celdas, si la imagen es de 16 bits.
  - **Simétrica:** La matriz simétrica quiere decir que las celdas que tienen valores opuestos en referencia a la diagonal de la matriz poseen el mismo valor es decir que la celda (5,4) debe tener el mismo valor que la celda (4,5).

Finalmente, la matriz de coocurrencia indica cual es el comportamiento de cada píxel de una ventana con respecto a los píxeles vecinos.

Para un mejor tratamiento de análisis de textura de una imagen se debe considerar los siguientes aspectos (Pinto, 2006):

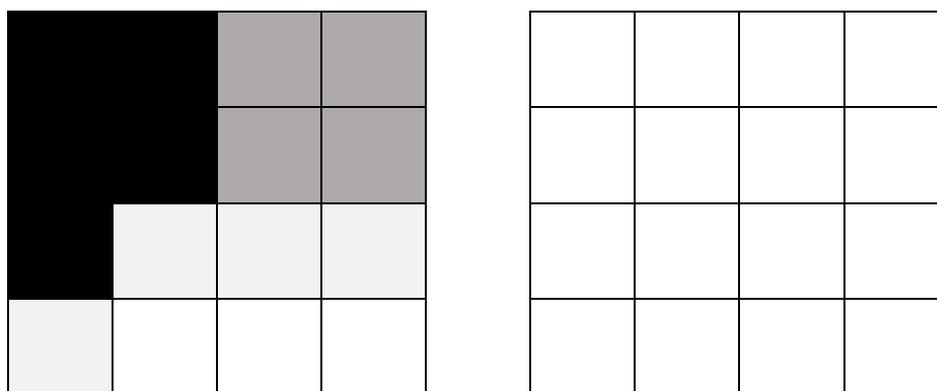
- Definir cuál es el tamaño de la imagen (ventana) que debe guardar con el número de niveles de grises establecidos.
- La distancia de píxel a píxel lo recomendable es hacerlo con los píxeles vecinos.
- Extraer la matriz simétrica
- Extraer la matriz de probabilidades (es el grado de incidencia que tiene un determinado píxel dentro de una ventana).
- Extraer la imagen de textura.

Para realizar un análisis a la textura en el momento de seleccionar las variables se debe elegir la que más se ajuste al caso que se vaya a tratar para evitar problemas durante el desarrollo, la textura representa una variable para las imágenes, para tener éxito en el análisis de textura no implica que se debe tomar todas las características de textura que existen ya que se debe realizar un consenso en cuales de las características son las más apropiadas (Presutti, 2004).

### 2.9.2. Cálculo de la matriz de coocurrencia

Los valores que se le da a la matriz de coocurrencia dependen del nivel de grises que se posea en la imagen, los valores son arbitrarios (Presutti, 2004).

Cada número puede representar el nivel de gris dependiendo del número de grises que se desee establecer, a esta matriz que se origina se lo denomina matriz de grises.



**Figura 5.** Representación de los niveles de grises en una ventana

Fuente: (researchgate.net, 2017)

Para sacar la matriz de las distancias entre píxeles se debe considerar los índices de la matriz que indican la posición de cada celda dentro de la matriz, por ejemplo si se posee el índice de la primera celda es (0,0), para determinar la distancia entre píxeles se toma el primer índice que es 0 y realiza la búsqueda de un píxel vecino a la derecha que contenga el valor del segundo índice que en este caso es 0, por lo tanto en la matriz se debe contar cuantas veces el valor cero contiene un vecino con un valor cero, si encaso se realiza la siguiente celda (0,1) entonces se debe verificar que valores de 0 tiene como vecino un valor uno, así se realiza sucesivamente hasta completar todos los valores de la matriz y obtener toda la matriz de distancias entre píxeles, dicha matriz es un paso a obtener la matriz de distancia entre píxeles (Presutti, 2004).

A continuación, se expone el proceso de obtención de la matriz de distancia entre píxeles a partir de la imagen del ejemplo anterior, que es una matriz de 4x4.

**Tabla 3**  
*Matriz de distancia entre píxeles*

(0,0)	(0,1)	(0,2)	(0,3)	2	2	1	0
(1,0)	(1,1)	(1,2)	(1,3)	0	2	0	0
(2,0)	(2,1)	(2,2)	(2,3)	0	0	3	1
(3,0)	(3,1)	(3,2)	(3,3)	0	0	0	1

El siguiente paso es obtener la matriz simétrica y se suma entre sí para la obtención de la relación correcta entre píxeles vecinos, ya que solo se obtenía los vecinos del lado derecho, con este proceso se obtiene ambas relaciones.

**Tabla 4**  
*Matriz simétrica*

2	2	1	0		2	0	0	0		4	2	1	0
0	2	0	0	+	2	2	0	0	=	2	4	0	0
0	0	3	1		1	0	3	0		1	0	6	1
0	0	0	1		0	0	1	1		0	0	1	2

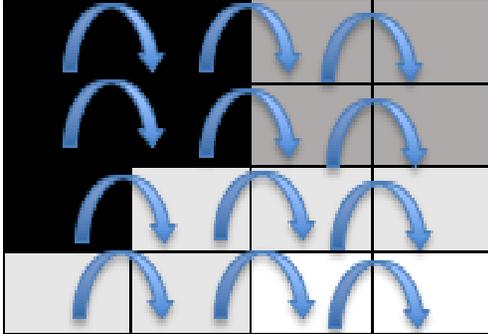
Ecuación para determinar la probabilidad de la matriz es decir el número de veces que ocurre un evento y el total se lo divide por el número de posibilidades de los eventos, seguidamente, se muestra la ecuación que calcula cada uno de los valores (Presutti, 2004):

$$P_{ij} = \frac{V_{i,j}}{\sum_{i,j=0}^{N-1} V_{i,j}}$$

- **i:** Número de filas
- **j:** Número de columnas
- **V:** Valor de la celda (i, j) en el pixel
- **P<sub>i, j</sub>:** probabilidad en la celda i, j
- **N:** número de filas o columnas

Para obtener la matriz de probabilidades se debe considerar cuantos píxeles vecinos existe dentro de la ventana, se considera píxeles vecinos a los que se encuentran en la parte derecha e izquierda de cada píxel como se muestra en seguida (Presutti, 2004):

**Tabla 5**  
*Matriz de probabilidades*



En la matriz se puede observar que existe 12 píxeles vecinos, pero como se debe considerar la parte de los píxeles vecinos hacia la izquierda se debe duplicar los valores, para la matriz de 4 x 4 en este caso todos los píxeles vecinos dan un total de 24, pero si la matriz es muy grande para determinar todas las combinaciones no se lo podrá realizar visualmente entonces se puede deducir la siguiente fórmula para evitar contar los píxeles.

El número de columnas ( $j$ ) menos uno multiplicado por el número de filas:

$$P = (j - 1) \times i$$

Una vez obtenido el número de píxeles vecinos se puede obtener la matriz de probabilidades que se basa en la cantidad total de los píxeles vecinos, para construir la matriz de probabilidades se debe utilizar la matriz de coocurrencia, si la primera celda de la matriz de coocurrencia tiene un valor de 2 el cual representa a la celda (0,0) dicho valor se debe dividir para el número total de píxeles vecinos que es 24, a continuación, se indica la construcción de la matriz de probabilidades:

**Tabla 6**  
*Construcción de la matriz de probabilidades*

4/24	2/24	1/24	0/24		0.16	0.083	0.042	0
2/24	4/24	0/24	0/24	=	0.083	0.16	0	0
1/24	0/24	6/24	1/24		0.042	0	0.25	0.042
0/24	0/24	1/24	2/24		0	0	0.042	0.083

Los datos que pueden dar la matriz de coocurrencia a simple vista es que: si la matriz diagonal no tiene diferencias en cuanto a los niveles de grises, por ende, se puede deducir que la matriz no tiene mucho contraste, si suma todos los valores de la diagonal principal, el resultado indica la probabilidad que un píxel tenga el mismo valor que el píxel vecino. Pero para tener mayor exactitud en conocer la textura de una imagen se puede utilizar las medidas de textura propuesta por Haralicks.

## 2.10. Clasificación de imágenes

La clasificación de las imágenes se la realiza utilizando una mamografía, se debe tomar una muestra, en la cual se visualice las texturas o patrones que se puedan presentar a través del uso de un proceso de extracción manual en el cual se realiza la clasificación de los patrones obtenidos, únicamente utilizando los nódulos de una mamografía.

La resolución de las imágenes obtenidas debe ser las mismas, por ende, deben ser cuadradas para que su procesamiento sea mejor, las medidas óptimas para cada patrón reconocido deben ser de 256px x 256px. Después de tener los patrones

caracterizados y previamente guardados, se debe llevar a cada imagen a una imagen de escala de grises, para posterior ser procesada como una matriz.

## **2.11. Coeficientes Haralick**

Los coeficientes Haralick son variables estadísticas de segundo orden, por cual narran a las características específicas que son: divergencia, energía, entropía, igualdad local, posibilidad máxima, matiz, jerarquía, reciprocidad.

### **2.11.1. Auto correlación**

Auto correlación en el aspecto técnico, la textura tiene relación con el tamaño espacial de la imagen respecto a todos los tonos de grises presentes en la imagen. Los tonos de grises de mayor profundidad dan un indicativo de que la imagen pertenece a una textura gruesa, si existe tonales de mayor claridad quiere decir que la textura es fina. La principal función de auto correlación es una característica que indica el tamaño de los primitivos tonales (Haralick, 1973).

### **2.11.2. Densidad**

La densidad es la relación que existe entre la masa y la superficie, esto es un caso tridimensional, como en la imagen se trata de una sola dimensión solo se puede definir dos particularidades que son el ancho y la altura (OIML, 2011).

### **2.11.3. Contraste**

El contraste mide la intensidad de los niveles de grises entre un píxel y su vecino además de las variaciones de las mismas, el valor del contraste se incrementa en la región donde tiene un alto contraste, igualmente el valor es grande si los mayores

valores están más alejados de la diagonal principal, y la probabilidad aumenta de una manera exponencial, a continuación, se muestra cómo se calcula el contraste utilizando la matriz de probabilidades (Pinto, 2006):

$$\text{Contraste} = \sum_{i,j=0}^{N-1} P_{ij} \times (i - j)^2$$

#### 2.11.4. Correlación

Una característica importante de la textura es la naturaleza repetitiva de la posición de los elementos de textura en la imagen. Se puede evaluar una función de correlación que mida una textura áspera. Sobre la base de la observación de la función de correlación, se calcula que algunas texturas son de naturaleza repetitiva, como las texturas de las telas. La característica de correlación de una imagen se utiliza para valorar la finura o la rugosidad de la textura presente en la imagen. Si la textura es áspera o suave, entonces la función de auto correlación bajará lentamente, si no, bajará muy rápidamente (Shodhganga, 2007)

#### 2.11.5. Disimilitud

Mide la media de la distribución de la diferencia de nivel de gris de la imagen. Un valor mayor implica una mayor disparidad en los valores de intensidad entre los valores vecinos (Brynnolfsson, 2017). Al igual que el contraste su relación es directamente proporcional al valor de la matriz de probabilidades (Pinto, 2006).

Se representa con la siguiente fórmula:

$$Disimilitud = \sum_{i,j=0}^{N-1} P_{i,j} \times |i - j|$$

#### 2.11.6. Energía

Es una medida de homogeneidad de una imagen. Una escena homogénea contendrá solo unos pocos niveles de gris, por lo que la matriz de grises tendrá unos pocos valores altos (Haralick, 1973). Como se mencionó anteriormente la matriz de coocurrencia simboliza las probabilidades que existen entre la similitud entre píxeles vecinos, y lo que hace la siguiente fórmula es aumentar los valores grandes y disminuir los pequeños.

$$Energía = \sum_{i,j=0}^{N-1} P_{i,j}^2$$

#### 2.11.7. Entropía

La entropía se trata de medir el nivel de aleatoriedad de los valores de un pixel que se encuentra en toda la imagen, la entropía toma un valor alto cuando las probabilidades en su mayoría son iguales, caso contrario la entropía llega a disminuir. Por lo tanto, una imagen homogénea dará como resultado un valor de entropía más bajo, mientras que una región heterogénea dará como resultado un valor de entropía más alto (Brynnolfsson,2017). Una imagen suave infiere una entropía baja y las

imágenes ásperas o toscas poseerán una entropía alta, a continuación, se presenta la ecuación correspondiente a la entropía (Pinto, 2006):

$$Entropía = \sum_{i,j}^{N-1} -P_{i,j} \ln(P_{i,j})$$

#### 2.11.8. Homogeneidad

La homogeneidad mide la similitud de los píxeles que existen entre sí, una imagen es homogénea cuando la diagonal principal de la matriz de coocurrencia son valores altos ya que si en la diagonal principal están los valores más grandes conforme los valores se alejen de la misma irá disminuyendo (Pinto, 2006).

$$Homogeneidad = \sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1 + (i - j)^2}$$

#### 2.11.9. Media

Este coeficiente se refiere a la frecuencia con la que se presenta un píxel vecino frente a un píxel de referencia (Pinto, 2006).

$$Media = \sum_{i,j=0}^{N-1} i \times P_{i,j}$$

### **2.11.10. Desviación estándar**

Es la desviación estándar de los niveles de grises en la ventana. Es alta cuando la desviación estándar en los niveles grises dentro de la ventana también es alta. Esta medida se basa en la media y la dispersión alrededor de la media de los valores de las celdas de la matriz de coocurrencia, como la varianza utiliza específicamente la combinación entre los píxeles de referencia y vecino, no es la misma que la varianza de los niveles de grises de la imagen original.

## **2.12. Patrones de textura presentes en las mamografías**

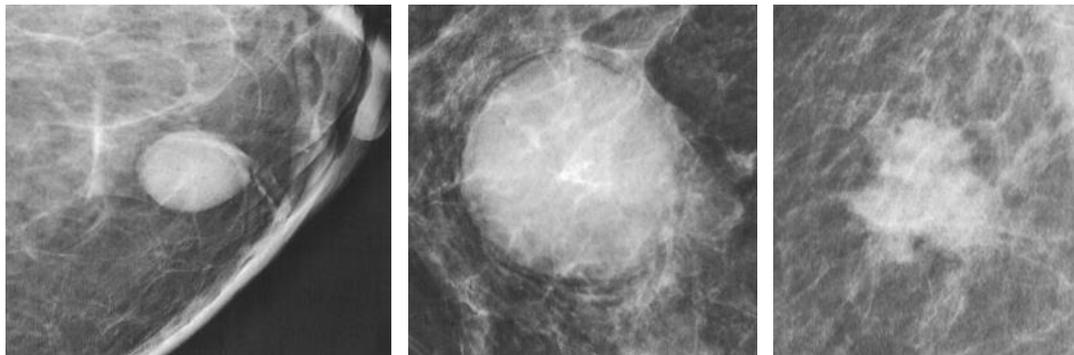
### **2.12.1 Nódulos**

Un nódulo mamario es un tumor que se presenta en el tejido, con frecuencia los nódulos tienden a ser benignos, por lo que puede llegar a presentar cáncer quien posea algún nódulo, cualquier tejido dentro de la mama tiene la eventualidad de presentar un nódulo, los nódulos en las mamografías se pueden detectar ya que presenta un borde completo o un borde convexo con una densidad en el centro mayor que en el perímetro, la presencia implica un control con un especialista de manera frecuente por la probabilidad de contraer cáncer, los nódulos presentan distintas formas que se indican a continuación (Gallego, 2005) (D'Orsi, 2016):

### **2.12.2 Forma**

- **Ovalada:** Tiene varias ondulaciones en forma de elipse u ovoide.
- **Redondeada:** Son formas circulares presentes en la mamografía.

- **Irregular:** Es un patrón amorfo, no tiene una forma definida, esto tiene carácter de sospechoso.

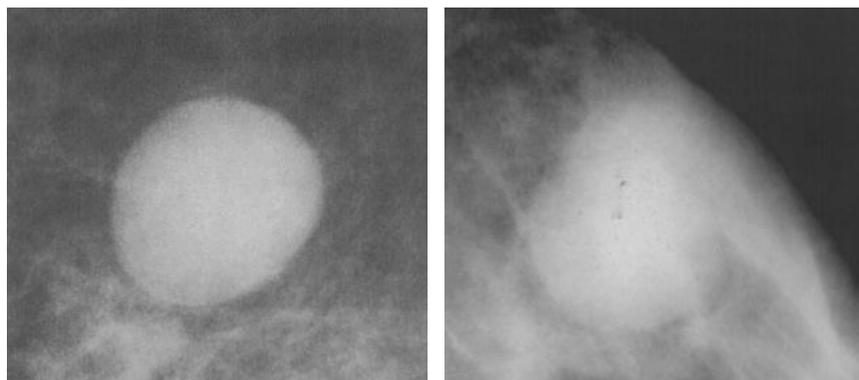


**Figura 6.** Formas de los nódulos (Ovalada, redonda, irregular)

Fuente: (elsevier.es.2017)

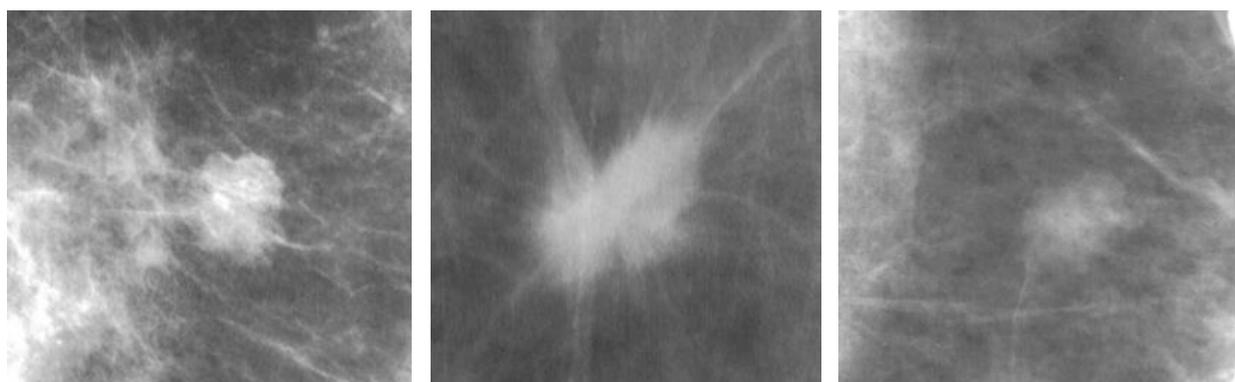
### 2.12.3 Margen

- **Circunscrito:** Para considerarse un patrón con un margen circunscrito, el perímetro debe estar adecuadamente definido y separado de los otros tejidos, al menos el margen debe estar determinado en  $\frac{3}{4}$  partes.
- **Oscurecido:** Al contrario del margen circunscrito, se supone oscurecido cuando más de  $\frac{1}{4}$  del margen no se percibe o está oculto por otros tejidos como el tejido fibroglandular.
- **Microlobulado:** Presenta como una especie de ondulaciones que no son constantes.
- **Indefinidos:** No se puede establecer un margen definido en su totalidad o parcialmente, el especialista considera benigno si el nódulo se origina del tejido de la mama.
- **Espiculado:** A partir centro del nódulo se despliegan líneas hacia la parte externa, dando la sensación de formar una estrella.



**Figura 7.** Margen de los nódulos (Circunscrito, Oscurecido)

Fuente: (elsevier.es.2017)



**Figura 8.** Margen de los nódulos (Microlobulado, Indefinido, Espiculado)

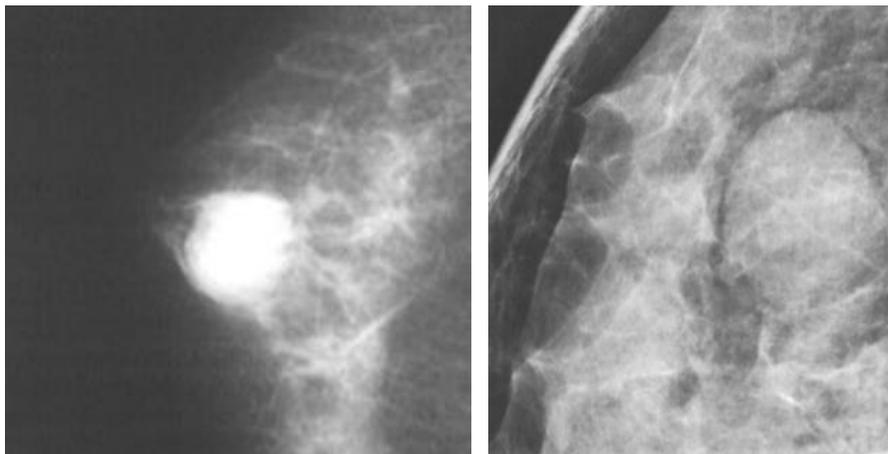
Fuente: (elsevier.es.2017)

#### 2.12.4 Densidad

En el aspecto de los nódulos presentes en una mamografía, la densidad se define con el contraste que existe entre el nódulo y el tejido mamario.

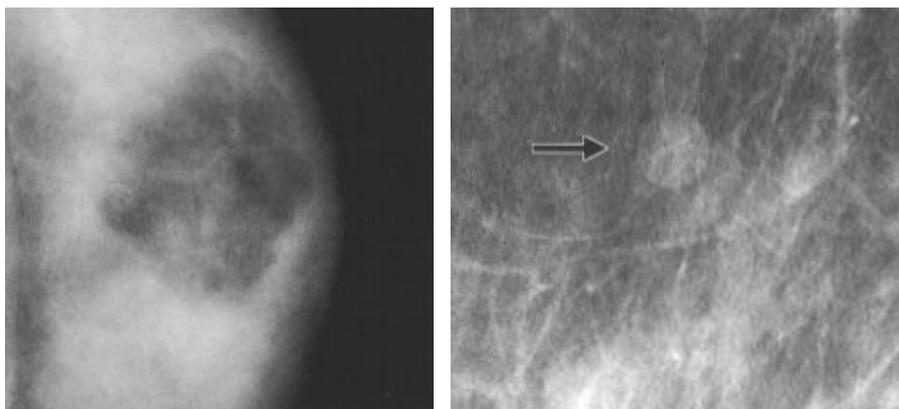
- **Hiperdenso:** El nódulo presenta un grado de contraste considerablemente mayor al del tejido mamario, por lo que a simple vista lo puede detectar un especialista.

- **Isodenso:** El nódulo isodenso tiene una consistencia alta como el hiperdenso, con la diferencia que el tejido de la mama tiene una similar densidad, que sobrelleva a perderse el nódulo.
- **Hipodenso:** Este nódulo es muy tenue que casi es imperceptible a la vista, ya que representa un volumen inferior al tejido de la mama, este nódulo es imposible que se desarrolle como cáncer.
- **Contenido Adiposo:** Este tipo de nódulo representa una lesión que tiene gran probabilidad de desarrollar cáncer, se razona que es adiposo cuando el nódulo está compuesto de grasa.



**Figura 9.** Densidad de los nódulos (Hiperdenso, Isodenso)

Fuente: (elsevier.es.2017)



**Figura 10.** Densidad de los nódulos (Hipodenso, Contenido adiposo)  
Fuente: (elsevier.es.2017)

### 2.12.5 Calcificaciones

Las calcificaciones mamarias son repositorios donde almacena ciertas cantidades calcio, estos casos son los más habituales que se presentan en una mamografía, la probabilidad de que aparezcan es directamente proporcional a la edad de la mujer; las calcificaciones la mayor parte proceden de un origen benigno debido a ciertos aspectos como: lesiones, traumas, secreciones. Para el especialista es importante diferenciar entre una calcificación benigna y maligna, las benignas pueden mostrar de mayor tamaño con margen redondo, en cambio las malignas son de un tamaño menor, difícil de detectar por lo que se debe ampliar la imagen para detectar estos casos, a continuación se muestra los distintos tipos de calcificaciones presentes en una mamografía (Noel, Aguilar, & Ramírez, 2004) (D'Orsi, 2016):

- **Cutáneas:** Poseen una densidad baja con el centro bastante notorio, comúnmente estas calcificaciones aparecen en la región sobre la axila, aparecen en grupos pequeños de calcificaciones asociadas, son poco probable que sean malignas, pero

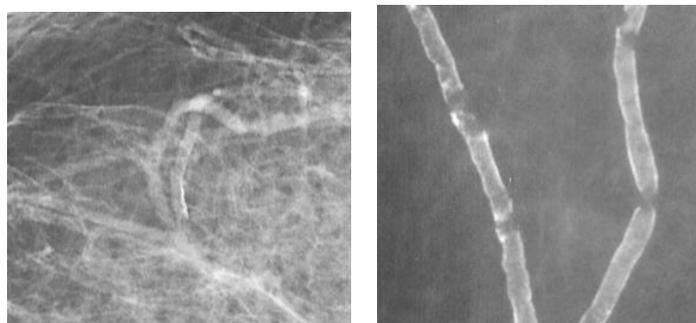
siempre se deben ejecutar proyecciones de diferente ángulo para descartar cualquier presencia de calcificación maligna.



**Figura 11.** Calcificación cutánea

Fuente: (elsevier.es.2017)

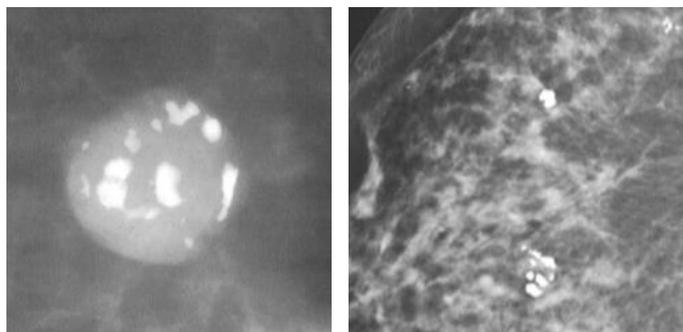
- **Vasculares:** Estas calcificaciones se denominan vasculares por la ubicación ya que se localizan en las arterias de la mama, su morfología es tubular y son continuas o discontinuas, dichas calcificaciones se exhiben con mayor frecuencia en personas con diabetes o con insuficiencia renal, no existe mayor complejidad en diferenciar una calcificación vascular, en caso de no diferenciar se procede a realizar una magnificación para detectar cualquier anomalía.



**Figura 12.** Calcificaciones Vasculares

Fuente: (elsevier.es.2017)

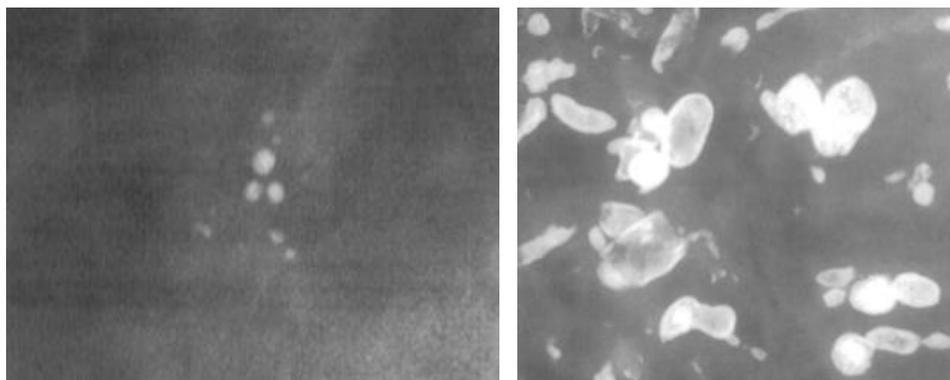
- **Macrocalcificaciones:** Tienen forma irregular, se compone de varias calcificaciones pequeñas, la ubicación usual de dichas calcificaciones es en el estroma mamario.



**Figura 13.** Macrocalcificaciones

Fuente: (diagnosticorojas.com, 2018)

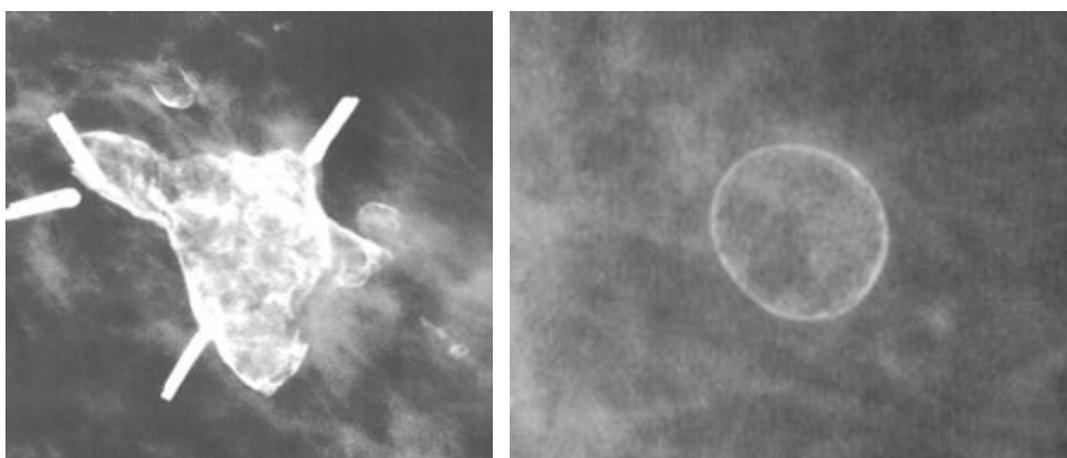
- **Lineales gruesas:** Poseen una forma lineal lisa y discontinua ubicada en los conductos mamarios, son intraductales las calcificaciones de líneas gruesas están en su mayoría en las mujeres mayores a 60 años.
- **Redondeadas:** Poseen una forma redonda de distintos tamaños, pero si el tamaño es menor a 0.5 mm se definen como depósitos de oxalato de calcio, si estas calcificaciones se encuentran separadas ubicadas en varias partes del tejido mamario se considera que es benigno, si se encuentran agrupadas se debe realizar un control sobre la paciente.



**Figura 14.** Calcificaciones de líneas gruesas y redondas

Fuente: (diagnosticorojas.com, 2018)

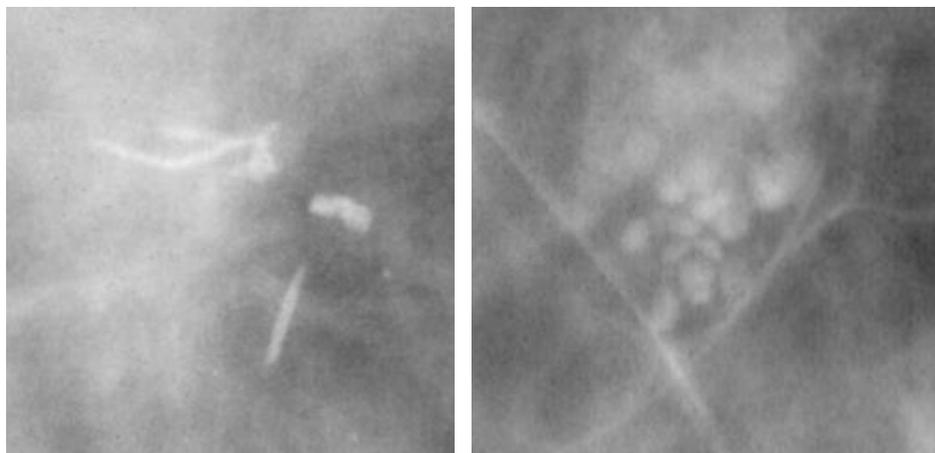
- **Anulares:** Tiene un margen definido y visible con un centro transparente, su forma es redonda sin encontrarse asociadas con otras del mismo tipo. Las calcificaciones anulares la mayor parte se encuentran en la parte externa de los quistes, solo en pocos casos el perímetro es de una línea gruesa.
- **Distróficas:** Es una calcificación grasa maligna, sus primordiales particularidades son la morfología irregular con un centro semi transparente, son más propensas a esta calcificación las mujeres que se realizan cirugías en un periodo de 3 a 5 años.



**Figura 15.** Calcificaciones anulares y distróficas

Fuente: (diagnosticorojas.com, 2018)

- **Leche cálcica:** Son pequeñas agrupaciones de quistes de forma redonda, poco visibles, en diferentes ángulos se pueden presentar como semilunares con un entorno curvilíneo definido, aunque esta calcificación es benigna puede presentarse junto a una maligna por lo que se debe realizar tomas de distintos ángulos y proyecciones.
- **Hilos de sutura:** Aparecen cuando la paciente ha recibido una intervención quirúrgica creando un repositorio de cúmulos cálcicos, la forma general es tubular y lineal.

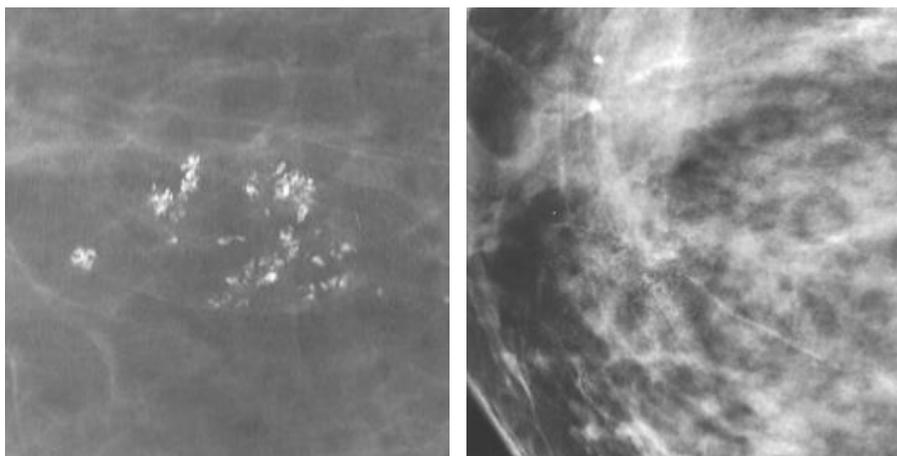


**Figura 16.** Calcificaciones de leche cálcica e hilos de sutura  
Fuente: (diagnosticorojas.com, 2018)

#### 2.12.5.1 Calcificaciones con morfología sospechosa

Según la forma de las calcificaciones se puede dar un diagnóstico ayudando al especialista a descubrir cualquier anomalía, dependiendo de la forma se puede definir si es benigno o maligno, a continuación, se muestra las formas más comunes de las calcificaciones (D'Orsi, 2016):

- **Amorfas:** Son partículas pequeñas asociadas de forma lineal, pero en general no se puede definir porque son diminutas y no son tan visibles para el ojo humano, se considera benigna si las partículas se distribuyen en el tejido mamario de una manera difusa caso contrario tiende a ser maligno.
- **Groseras heterogéneas:** Tienen una morfología irregular, se presentan como partículas asociadas, existe mayor probabilidad que se puede desarrollar cáncer si el origen de las calcificaciones no proviene por alguna intervención quirúrgica o traumatismo, si se encuentra varios casos en la misma mama entonces se puede considerar benignas.

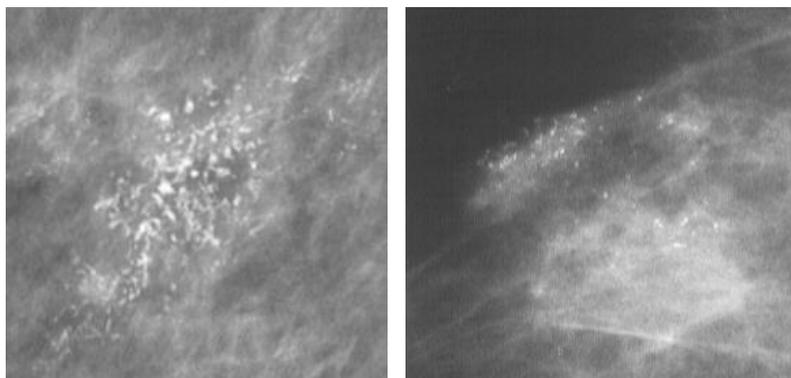


**Figura 17.** Calcificaciones con morfologías amorfas y heterogéneas

Fuente: (diagnosticorojas.com, 2018)

- **Finas pleomorfas:** Son más perceptibles para el especialista con una morfología más definida. Las calcificaciones finas pleomorfas no representan una forma lineal sino dispersas.

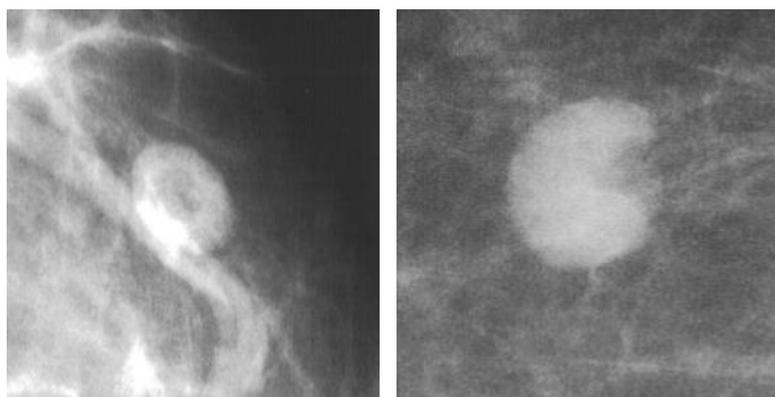
- **Calcificaciones lineales finas:** Tienen una morfología lineal discontinua sus bordes no están totalmente definidos, estas calcificaciones representan la mayor probabilidad de que desarrolle cáncer con un 70% de probabilidad.



**Figura 18.** Calcificaciones con morfología finas pleomorfas y lineales finas  
Fuente: (diagnosticorojas.com, 2018)

### 2.12.6 Ganglio linfático intramamario

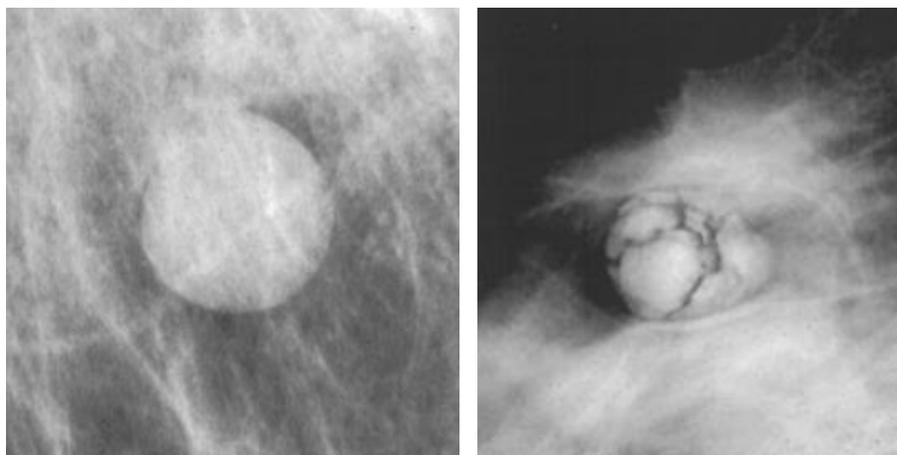
El ganglio contiene grasa y tiende a medir hasta 1 cm con una forma ovalada, la ubicación general donde se encuentra es la parte externa o superior de la mama muy cercanas a la axila, habitualmente se producen a lado de una vena ya que se necesita el drenaje venoso para su desarrollo.



**Figura 19.** Ganglio linfático intramamario  
Fuente: (diagnosticorojas.com, 2018)

### 2.12.7 Lesión cutánea

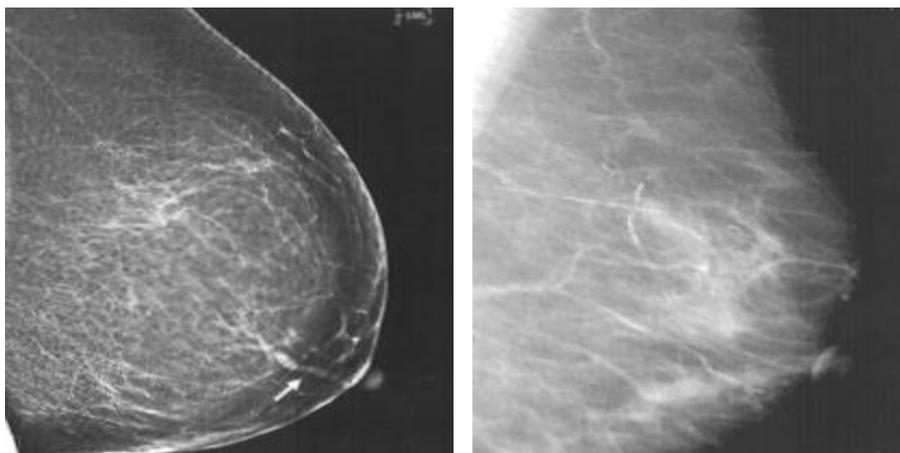
Es difícil de definir a simple vista se necesita más de una proyección para deducir que es una lesión cutánea, es probable tener errores al encontrar una lesión cutánea, el especialista para lograr detectar dichas lesiones debe señalar la lesión con un marcador radiopaco que permite detectar una lesión cutánea para que se refleje en la mamografía.



**Figura 20.** Lesión cutánea  
Fuente: (diagnosticorojas.com, 2018)

### 2.12.8 Conducto dilatado solitario

Son estructuras con una forma tubular se extienden ramificaciones a su alrededor esto representa la dilatación del conducto, según los especialistas sino existe otros aspectos inciertos en la misma mamografía se puede descartar que es una calcificación.



**Figura 21.** Conducto dilatado solitario

Fuente: (diagnosticorojas.com, 2018)

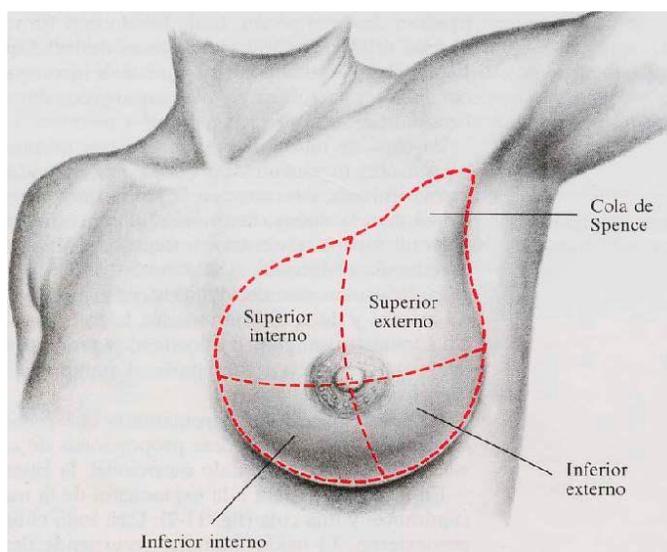
### 2.12.9 Ubicación de la lesión

Para un informe adecuado del diagnóstico del paciente se debe describir en que parte exacta de la mama tiene la lesión, se considera a la mamografía que se extrajo de manera frontal de la persona, con esas mamografías se traza un plano cartesiano invisible sobre el para establecer cuadrantes y posición horaria, se debe concretar en qué mama está la lesión si es izquierda o derecha, la manera de especificar la ubicación es en primer lugar en que mama ocurrió la lesión, luego el cuadrante, la posición horaria, la profundidad de la lesión, por último se debe definir la distancia que existe con relación al pezón, a continuación se indica que aspectos se debe considerar para definir la ubicación de una lesión (D'Orsi, 2016):

- **Lateralidad:** Indica en que mama se encuentra la lesión, puede ubicar en la mama derecha o izquierda.
- **Cuadrante y posición horaria:** Existe varias posiciones para definir el cuadrante que son inferointerno, inferoexterno, superointerno, superoexterno, preexiste la

posición central si la lesión está ubicada detrás del pezón, en cambio la posición horaria se refiere aun cuadrante en específico dependiendo de la mama si es derecha o la izquierda.

- **Profundidad:** Se toma a la mama desde un ángulo lateral separando a la mama de la pared torácica, los tres grados de profundidad son: posterior, medio y anterior.
- **Distancia desde el pezón:** Es la distancia que existe desde la lesión hasta el pezón.



**Figura 22.** Ubicación de la lesión de la mama  
Fuente: (diagnosticorojas.com, 2018)

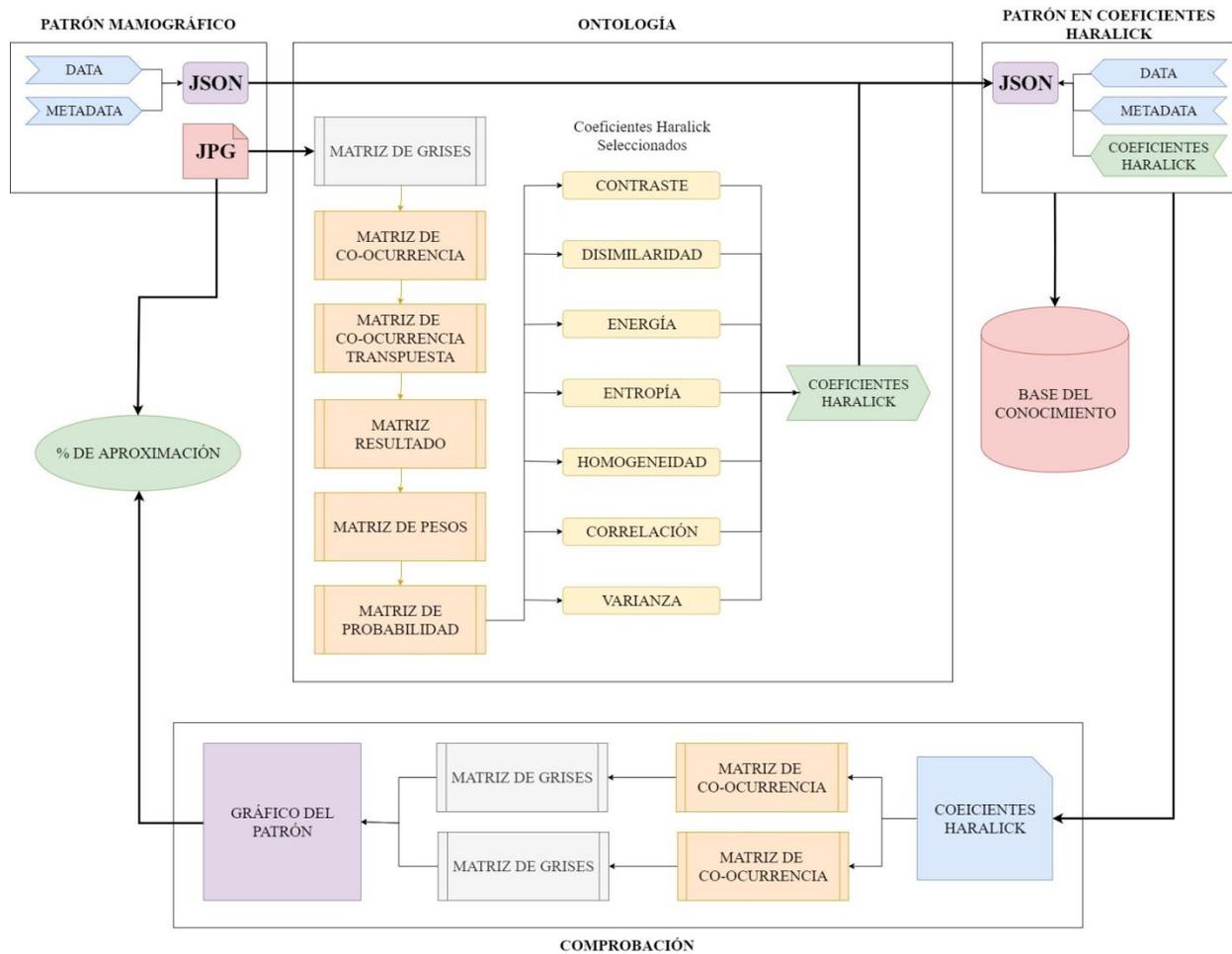
## CAPITULO III

### MODELO CONCEPTUAL DE LA ONTOLOGÍA

#### 3.1 Introducción

EL modelo está definido por un cluster conformado por los patrones de textura que poseen las mamografías, de cada patrón se obtiene una matriz de grises, posteriormente, se calcula la matriz de coocurrencia, finalmente se calculara los coeficientes de Haralick definidos, a partir de los coeficientes se hace inferencias de los resultados del diagnóstico de la mamografía. Para comprobar que la ontología funcione correctamente se genera la matriz de coocurrencia a partir de los coeficientes de Haralick, se forma la matriz de grises y a partir de la matriz de grises se puede generar la imagen.

A partir de la imagen creada se genera una nueva matriz de grises que se compara con la imagen inicial para medir el grado de efectividad de la ontología.

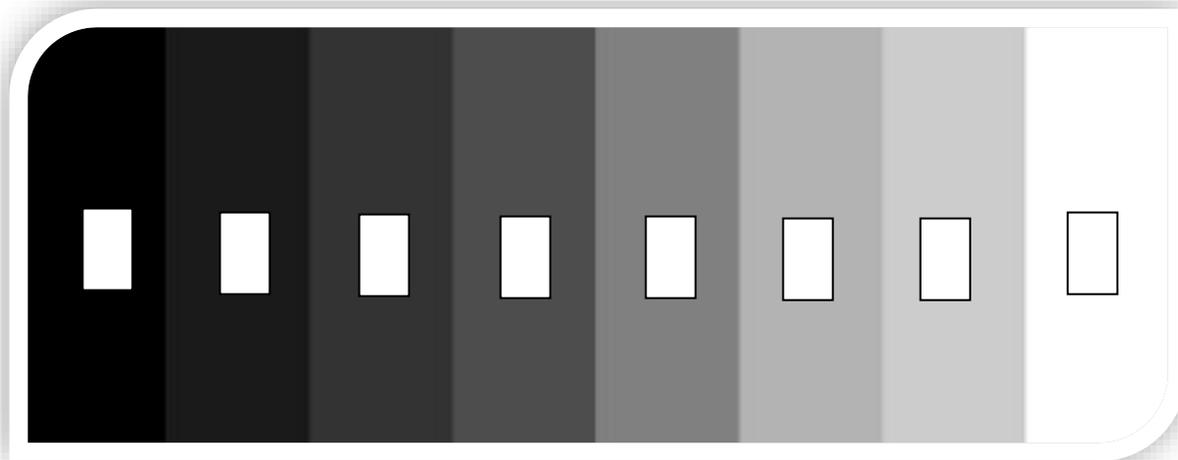


**Figura 23.** Modelo conceptual de la ontología - Detallado

## 3.2 Directrices para el diseño de la ontología.

### 3.2.1 Niveles de grises

Los niveles de grises se establecen de acuerdo al número de tonos de grises que se establezca, en este caso se tomará 256 tonos, para tener un mejor análisis de textura.

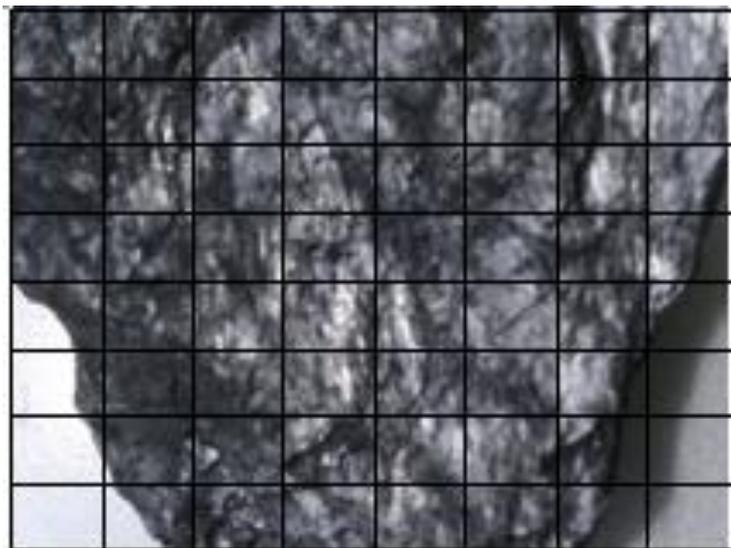


**Figura 24.** Distintos niveles de grises

Fuente: aboutespanol.com, 2019

### 3.2.2 Píxeles en una imagen

La imagen de los patrones ya predefinidos será dividida en una imagen de 8 píxeles, para realizar las primeras pruebas, para validar los resultados numéricos se procederá a realizar el mismo proceso con la misma imagen en 256 píxeles. La imagen será procesada en 256 píxeles dando un valor del tono de gris a cada celda, a partir de eso se generen las matrices y se pueda calcular los coeficientes de Haralick, cabe mencionar que la imagen debe ser cuadrada.



**Figura 25.** Imagen dividida en píxeles (8 x 8)

Fuente: Rodríguez, Morales, 2017

Como ejemplo, se tomó un patrón de un ganglio linfático donde se dividió la imagen en 16 celdas, la imagen será evaluada para determinar que valores de tono de gris debe ir en cada celda.

3	5	5	4
5	5	5	3
4	5	5	3
3	5	5	3

**Figura 26.** Imagen dividida en 16 celdas con valores de tonos de grises

Fuente: (tips.clip-studio.com, 2018)

### 3.2.3 Lista de coeficientes Haralick seleccionados

- Homogeneidad
- Contraste
- Disimilaridad
- Energía
- Correlación.
- Entropía.
- Varianza

### 3.2.4 Rangos de los coeficientes Haralick

#### Patrón 1

**Tabla 7**

*Rangos de los coeficientes Haralick Patrón 1*

	Rango 1	Rango 2	Rango 3	Rango 4	Rango 5
1.Homogeneidad		X			
2.Contraste			X		
3.Disimilaridad		X			
4.D. Standard					X
5.Correlación			X		
6.Entropía			X		
7.Varianza	X				

**Resultado Patrón 1** {(1,2), (2,3), (3,2), (4,5), (5,3), (6,3), (7,1)} **Nombre: Patrón n**

**Forma**

## Patrón 2

**Tabla 8**

*Rangos de los coeficientes Haralick Patrón 2*

	Rango 1	Rango 2	Rango 3	Rango 4	Rango 5
1.Homogeneidad	X				
2.Contraste		X			
3.Disimilaridad			X		
4.D. Standard				X	
5.Correlación					X
6.Entropía	X				
7.Varianza		X			

**Resultado Patrón 2** {(1,1), (2,2), (3,3), (4,4), (5,5), (6,1), (7,2)}

### Mamografía

Patrón 1 U Patrón 2

### 3.3 Diseño de la ontología con la metodología Methontology

Uno de los motivos para elegir esta metodología fue el esquema de plantillas que utiliza, los diagramas y las tablas empleadas en la conceptualización. En la elaboración de la ontología, se empleó la metodología “Methontology”, en la que se debe realizar los siguientes pasos para su construcción: especificación, conceptualización, formalización, implementación y mantenimiento que se mencionan a continuación:

### 3.3.1 Especificación de requisitos

Para realizar la ontología se ha planteado el siguiente modelo para explicar los requerimientos:

**Tabla 9**  
*Especificación de requisitos*

<b>Organismo</b>	Universidad de las Fuerzas Armadas ESPE		
<b>Proyecto</b>	ONTOLOGÍA PARA INTERPRETAR PATRONES DE IMÁGENES DE MAMOGRAFÍAS EN EL DOMINIO DE LOS COEFICIENTES HARALICK A TRAVÉS DEL ANÁLISIS DE LA TEXTURA .		
<b>Entregable</b>	Especificación de Requisitos		
<b>Autor</b>	Escorza Alcides, Carrillo Jaime		
<b>Versión/Edición</b>	1.0.0	<b>Fecha Versión</b>	20/03/2019
<b>Aprobado por</b>	Díaz Oswaldo	<b>Fecha Aprobación</b>	21/03/2019

### Introducción de la ontología

Se necesita el desarrollo de una ontología que interprete imágenes de mamografías mediante los coeficientes Haralick y que dicha ontología arroje resultados de los coeficientes en valores numéricos que previamente fueron seleccionados para el propósito de alimentar una base del conocimiento que determine juicios de valor que permita presentar resultados en porcentajes de la posible existencia de cáncer.

### Alcance

La ontología forma parte de un proyecto en el cual se encuentra la Universidad, por lo que el desarrollo de la ontología únicamente servirá como traductor de un corpus de imagen y los resultados deben ser enviados a una base de datos inteligente.

## Objetivos

- Lograr que la ontología obtenga todos los detalles (datos y metadatos) de un patrón de la imagen para su análisis.
- Mostrar una interfaz donde se pueda cargar las imágenes de los patrones antes de ir al análisis.
- Determinar las formas en 3D que se presenten en los patrones de imágenes.

## Descripción de la situación actual

El proyecto REDU tiene por objetivo, el desarrollo de una herramienta para la detección de cáncer de mama con la que se pretende mejorar el análisis tradicional de los médicos a través del análisis de la textura de la mamografía.

## Pros y Contras de la Situación Actual

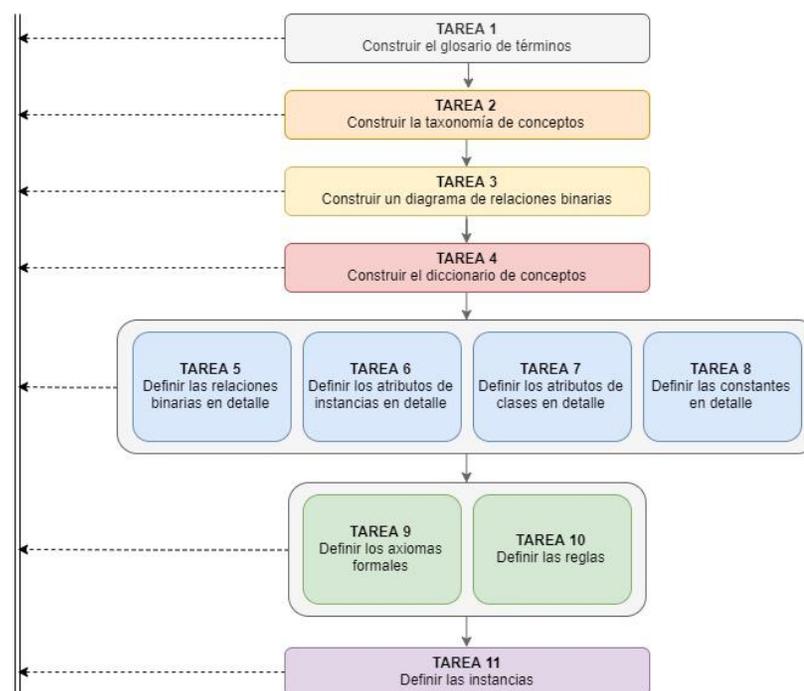
El proyecto está respaldado por varias universidades por lo que se cuenta con el apoyo en nuestro caso por las autoridades y los ingenieros de la Universidad de las Fuerzas Armadas ESPE, en contra se puede determinar que la coordinación con los demás integrantes del proyecto no es exacta por lo que la retroalimentación no es la mejor.

### 3.3.2 Conceptualización

En esta etapa consiste en estructurar los conocimientos en un modelo conceptual mediante tablas y diagramas que describirán el problema y la solución a través de las concepciones, propiedades, relaciones, constantes, proposiciones, pautas e demandas. En el desarrollo de la ontología es necesario que se tome en cuenta que

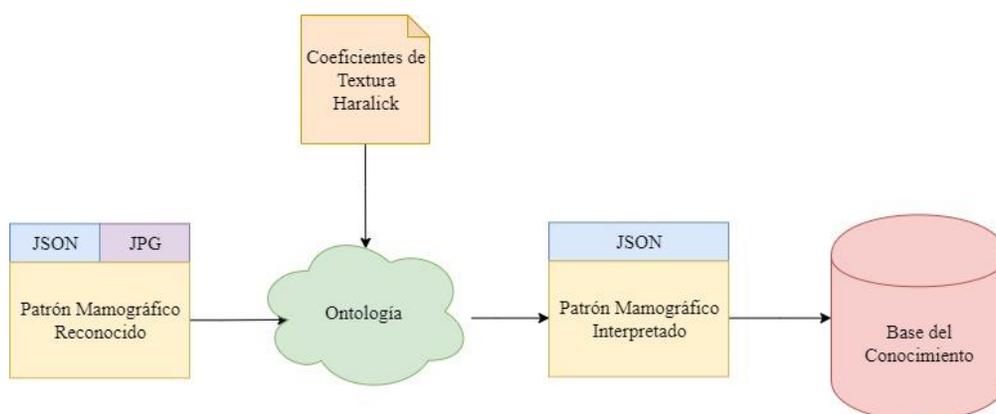
la mayoría de las tareas se ejecutan de forma simultánea junto al proceso de adquisición de conocimientos.

En la figura 27 se detallan las tareas necesarias para realizar la conceptualización de la ontología utilizando Methontology.



**Figura 27.** Tareas de la actividad de conceptualización de Methontology

En este caso no es un traductor léxico, ya que no se puede interpretar a simple vista las imágenes de los patrones de las mamografías, por lo cual en la imagen 28 se muestra un ejemplo de cómo se va a ir alimentando el conocimiento a la ontología, en este caso su conocimiento empieza en cero.



**Figura 28.** Proceso de almacenamiento de información en la base de datos

Para el proceso de almacenamiento de la ontología, se va guardando los coeficientes Haralick obtenidos de un patrón procesado de una mamografía, este se guarda en la base de datos y el proceso vuelve a realizarse, únicamente cambia que al volverse a repetir el proceso y al guardarse un nuevo grupo de coeficientes, se crean rangos en la base de datos, permitiendo establecer los niveles de BI-RADS que tiene cada caso.

### 3.3.3 Formalización e implementación

La implementación de la Ontología será realizada utilizando el lenguaje de programación PYTHON, utilizando la base de datos MongoDB para el almacenamiento de un JSON de cada patrón que se analice; para comprobar el funcionamiento se desplegará el resultado de cada paso que se ejecute en el programa.

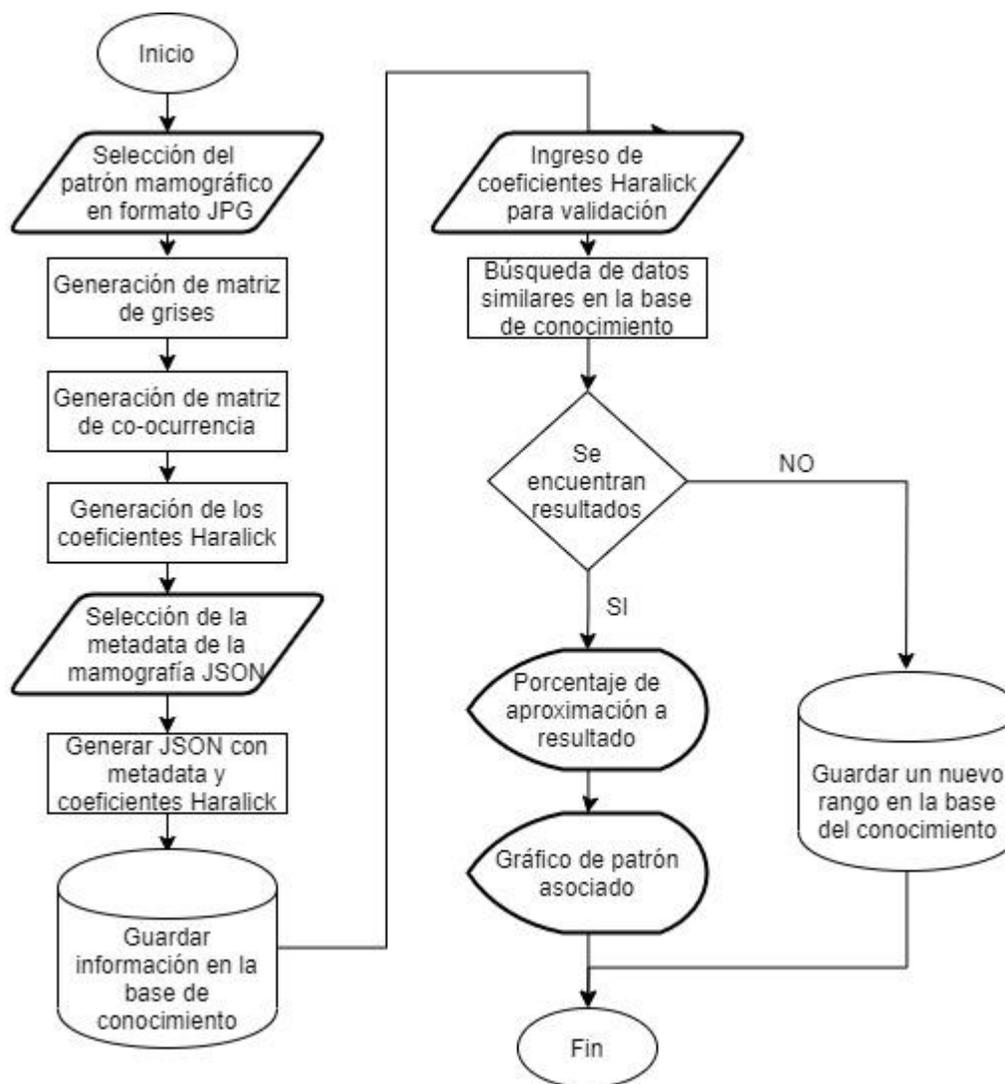
### 3.3.4 Mantenimiento

La Ontología a partir del conocimiento formado se podrá realizar actualizaciones e implementaciones de las instancias en caso de que se requiera perfeccionar los

resultados de la ontología; es decir, entre más casos que se agreguen en la base de datos, se tendrá resultados más exactos y con mejor precisión.

### **3.4 Diseño de la Ontología previo a la entrada al software**

- Imagen del patrón mamográfico en formato JPG.
- Data de la mamografía y Metadata del patrón en formato JSON.
- Definición de los grises de la imagen.
- Selección de las características de textura.
- Generar la matriz de grises del patrón mamográfico.
- Generar la matriz de co-ocurrencia del patrón mamográfico.
- Determinar los coeficientes Haralick del patrón mamográfico.
- Cuantificación de las características de textura.
- Generar archivo JSON con los Haralick obtenidos.
- Unir los JSON de la data, metadata y coeficientes Haralick.
- Guardar el JSON en la base de conocimiento.
- Proceso de verificación (Ingreso de los coeficientes Haralick obtenidos).
- Crear un nuevo resultado si no existe en la base del conocimiento.
- Clasificación de los resultados y verificar en qué nivel de BI-RADS se encuentra el caso ingresado.



**Figura 29.** Diagrama de flujo del proceso realizado por la ontología

## CAPITULO IV

### IMPLEMENTACIÓN DE LA ONTOLOGÍA PARA EL DOMINIO DE LOS COEFICIENTES HARALICK.

#### 4.1. Explicación de librerías utilizadas

##### **PyQt5.QtWidgets**

Es una librería que permite la elaboración de interfaces gráficas escrito en el lenguaje C++ aprovechando la velocidad que tiene este lenguaje, la versión es la 5 que ya tiene más compatibilidad con la mayoría de las aplicaciones; para realizar su instalación se debe ejecutar el siguiente comando ***pip install PyQt5*** (PyQt5 tutorial , 2019).

##### **matplotlib. pyplot**

Es una librería que permite realizar trazos para elaborar figuras simples o complejas, los gráficos se realizan a partir de los ejes preestablecidos por el matplotlib, las funciones de matplotlib se pueden llamar desde un Axes Object (Matplotlib, 2019), en nuestro caso se utilizará para la verificación del gráfico que se cargue a la aplicación.

##### **cv2**

Es una librería accede a que se realice una lectura y manejo de imágenes en escala de grises e imágenes a color (OpenCV, 2018).

## **NumPy**

Es una librería que contiene paquetes para procesar matrices multidimensionales desarrolladas en el lenguaje, además contiene características de álgebra lineal y números aleatorios, por lo que numpy se adapta a una amplia gama de base de datos (Numpy.org, 2016).

## **Skimage**

Es una colección que contiene algoritmos para procesar imágenes (Scikit-image, 2019).

## **EasyGUI**

Es una librería que proporciona el manejo de interfaces, este módulo se puede ejecutar en Python 2 y 3, y no posee dependencias (EasyGUI, 2014).

## **PIL**

Este módulo da paso a la edición de imágenes en especial imágenes con formato .gif, .jpeg, .png. En la actualidad se recomienda utilizar la librería Pillow en lugar de PIL (Recursos Python, 2014).

## **Tkinter**

Es una librería que proporciona herramientas robusto independiente de la plataforma, además es rápido, pero con limitada documentación (docs.python.org, 2019).

## PathLib

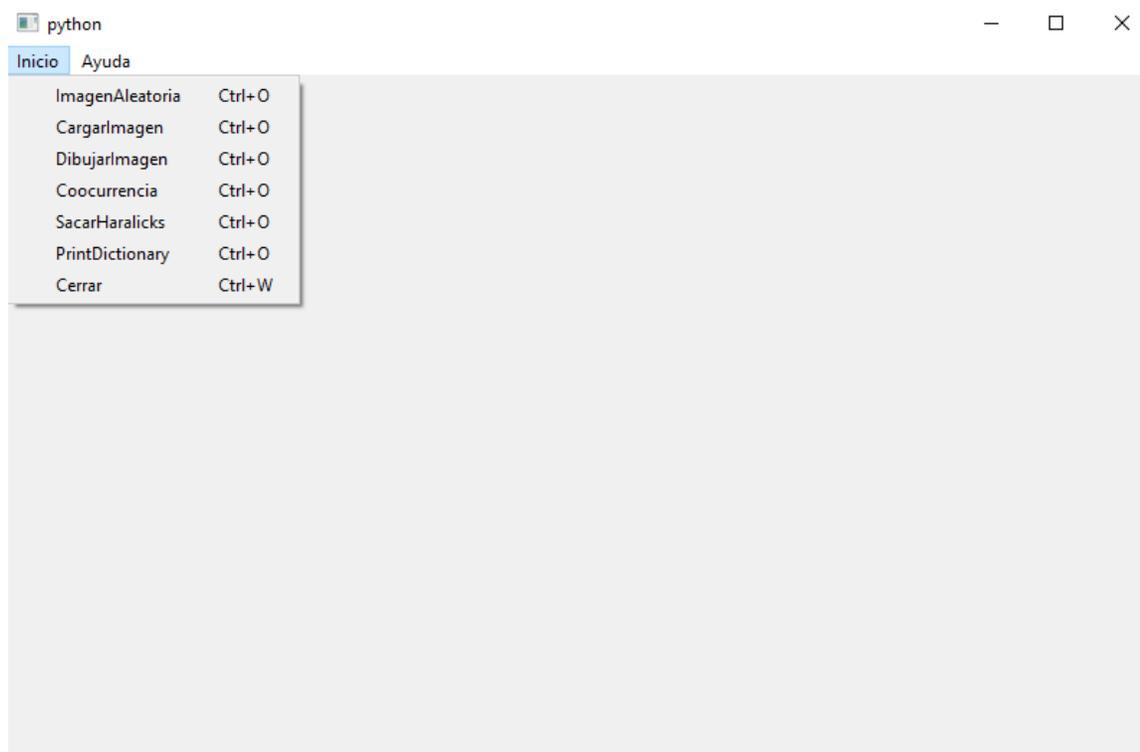
El módulo de pathlib ofrece clases que presenta un sistema de archivos con la semántica respectiva dependiendo del sistema operativo (docs.python.org, 2019).

## Json

El módulo Json maneja objetos tipo string no en formato bytes, por lo que permite string como entrada (https://docs.python.org, 2019).

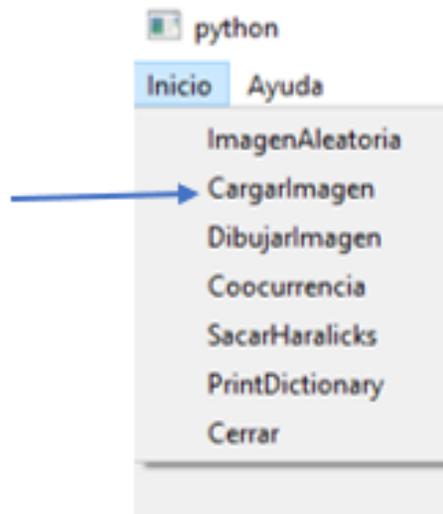
### 4.2. Descripción inicial del programa

Al iniciar el programa se carga la interfaz la cual tendrá las opciones para sacar los coeficientes Haralick de las imágenes de las mamografías.



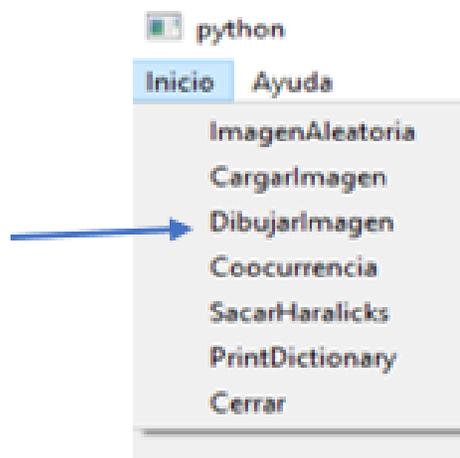
**Figura 30.** Interfaz gráfica de la ontología

Las opciones con las que cuenta el menú de la interfaz es cargar la imagen, dibujar la imagen, matriz de coocurrencia, sacar Haralicks.



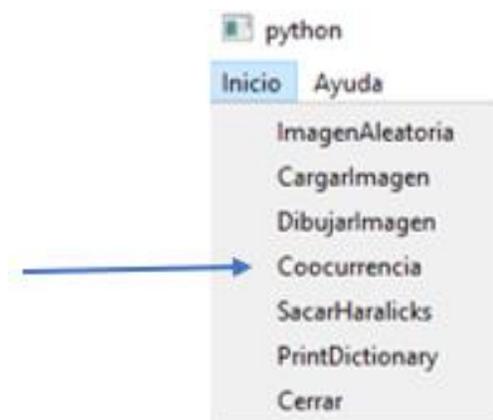
**Figura 31.** Opción Cargar Imagen

La opción cargar imagen permite seleccionar una imagen de cualquier directorio, para que procese dicha imagen.



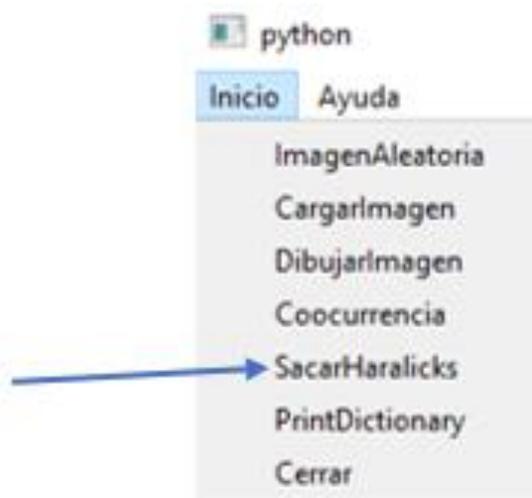
**Figura 32.** Opción Dibujar Imagen

La opción dibujar imagen se usa para verificar que la matriz de grises que se consigue al cargar la imagen sea lo más aproximado a la realidad, se dibuja la imagen con la matriz de grises y se puede comparar con la imagen general.



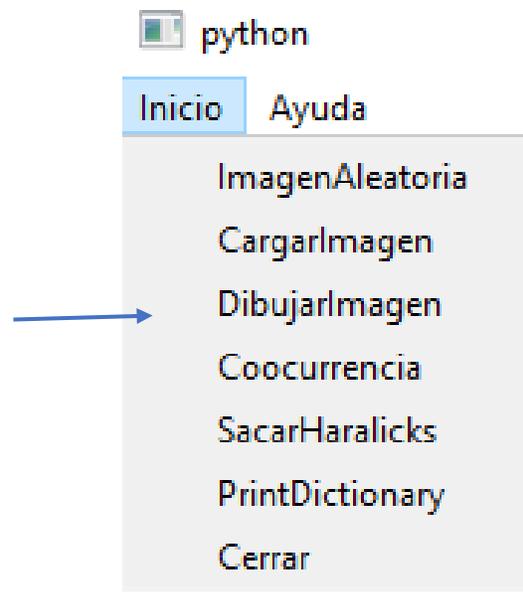
**Figura 33.** Opción coocurrencia

La opción coocurrencia es para procesar la matriz de grises que se obtuvo de la imagen para tener la matriz final que es la de coocurrencia.



**Figura 34.** Opción sacar Haralicks

La opción Haralicks accede a sacar todos los coeficientes Haralick a partir de la imagen que se cargue al comienzo de la ejecución del programa con el proceso respectivo.



**Figura 35.** Opción Cerrar

La opción cerrar consiente a dar por terminado la aplicación.

### 4.3. Explicación del código fuente

#### Class Window

En esta parte del código se crea la interfaz gráfica donde se permite el procesamiento de la imagen hasta presentar los coeficientes Haralick.

```
class Window(QMainWindow):  
  
    def __init__(self):  
        QMainWindow.__init__(self)  
        w = QMainWindow()  
        w.setStyleSheet("QMainWindow{background-image:url(C:/Users/ALCI/Desktop/lena.png)}")  
        self.resize(800, 500) #Tamaño inicial de la ventana 800x500  
        #Barra de estado  
  
        self.statusBar().showMessage("Bienvenid@")  
        self.font  
        #Objeto menuBar  
        menu = self.menuBar()  
        #Menú padre  
        menu_archivo = menu.addMenu("&Inicio")  
        #Menú padre  
        menu_editar = menu.addMenu("&Ayuda")  
  
        #Agregar un elemento acción al menu_archivo  
        menu_archivo_abrir = QAction(QIcon(), "&ImagenAleatoria", self)  
        menu_archivo_abrir.setShortcut("Ctrl+o") #Atajo de teclado
```

**Figura 36.** Clase inicial (Interfaz gráfica)

## Class LoadImage

Accede a cargar una imagen para obtener los coeficientes Haralick del directorio que se seleccione.

```

class LoadImage:
    def CargarImagen(self):
        matProbabilidad=[np.zeros((range1,range2))]
        matcoocurrencia=[np.zeros((range1,range2))]
        matcoocurrenciatranspuesta=[np.zeros((range1,range2))]
        matResultado=[np.zeros((range1,range2))]
        matPesos=[np.zeros((range1,range2))]

        Tk().withdraw()
        uri = askopenfilename()
        namearchivo=uri
        print(namearchivo)
        #print(uri)
        image=io.imread(uri)/255.0 # imread lee las imagenes con los pixel
        #uri.close()
        print("- Dimensiones de la imagen:")
        #print(image.shape)
        plt.imshow(image,vmin=0,vmax=1)
        I = Image.open(uri).convert("L")
        plt.imshow(np.asarray(I))
        coc=MatrizCoocurrenciaDef()

```

**Figura 37.** Coeficientes Haralick

### **Class MatrizCoocurenciaDef**

Primero se obtiene la matriz de grises que surge a partir de la imagen cargada, la matriz de grises se forma a partir del número de píxeles de la imagen, se sugiere de manera obligatoria que la imagen sea cuadrada y para mayor exactitud se recomienda una matriz de 256 x 256 px.

```

class MatrizCoocurrenciaDef:
    def Coocurrencia(self):
        lista=[]
        im = Image.open(uri)
        pixels = list(im.getdata())
        #obtener matriz de grises
        pix=np.array(pixels).reshape(dim1,dim2)
        print(pix)
        listCoocurrencia=[]
        #sacar la matriz de coocurrencia
        a=0

        for i in range(range1):
            for j in range(range2):
                for k in range(range1):
                    for l in range(range2):
                        if(l<range2):
                            if i==pix[k][l]:
                                if j==pix[k][l+1]:
                                    matcoocurrencia[0][i][j]=matcoocurrencia[0][i][j]+1

```

**Figura 38.** Clase para obtener la matriz de coocurrencia.

### **Función Dibujar**

Accede a recrear la imagen que se cargó a partir de la matriz de grises que se obtuvo, para que se pueda verificar si las dos imágenes son iguales.

```
def Dibujar():
    lista=[]
    listCoocurrencia=[]
    for elemento in pixels:
        lista.append(elemento/255.0)
    pix=np.array(lista).reshape(dim1,dim2)
    print("Iniciando")
    img=plt.imshow(pix,vmin=0,vmax=1)
    pixels1 = list(im.getdata())
    pix1=np.array(pixels1).reshape(dim1,dim2)
    print("Matriz de imagen generada")
    numFallas=0
    print(pix1)
    for i in range(32):
        for j in range(32):
            if(pix[i][j]==pix1[i][j]):
                numFallas+=1
    #Compara las dos matrices
    print("COMPARACION ENTRE LAS DOS MATRICES")
```

**Figura 39.** Función para dibujar una imagen.

### Función Haralicks

Se realiza el procesamiento de imágenes para obtener los coeficientes Haralick a partir de la matriz de coocurrencia que se obtiene previamente.

```

def Haralicks():
    listaHaralicks=[]
    #suma
    sumaHomogeneidad=0
    print("Test coocurrencia")
    print(matcoocurrencia)
    #Homogeneidad
    for p in range(range1):
        for q in range(range2):
            if q<range2:
                matProbabilidad[0][p][q]=(matResultado[0][p][q]/(1+pow((p-q),2)))
                sumaHomogeneidad=sumaHomogeneidad+matProbabilidad[0][p][q]

    #Suma de homogeneidad
    print("SUMA DE HOMOGENEIDAD")
    print(sumaHomogeneidad)
    ja="Homogeneidad: "
    ja=ja+str(sumaHomogeneidad)
    listaHaralicks.append(ja)
    print("Matriz de probabilidades")
    print(matProbabilidad)
    sumaContraste=0

```

**Figura 40.** Función Haralicks

### Proceso de almacenamiento de datos

Se realiza el almacenamiento en 2 sitios, primero se guarda en un archivo Json para que a continuación serán almacenados en una base de datos en Mongo.

```

data = {}
data['contraste'] = sumaContraste
data['disimilaridad'] = sumaDisimilaridad
data['energia'] = sumaEnergia
data['entropia'] = sumaEntropia
data['homogeneidad'] = sumaHomogeneidad
dir = WindowsPath('C:\\\\Haralicks')
fichero = dir / namefile

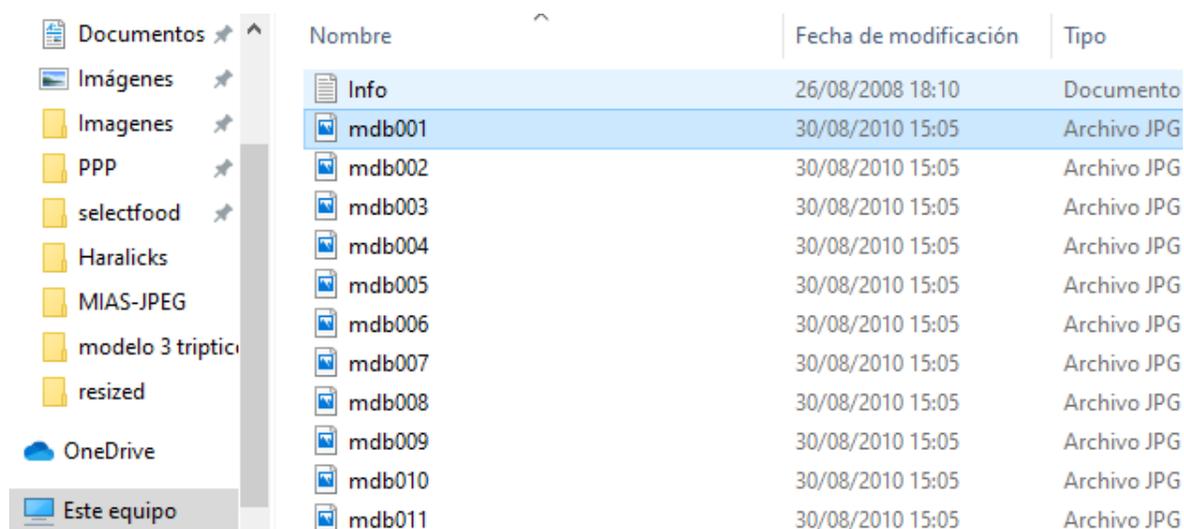
with fichero.open("w") as file:
    json.dump(data, file)
insertarDB(sumaHomogeneidad,sumaContraste,sumaDisimilaridad,sumaEntropia,sumaEnergia,sumaMed

```

**Figura 41.** Proceso de almacenamiento

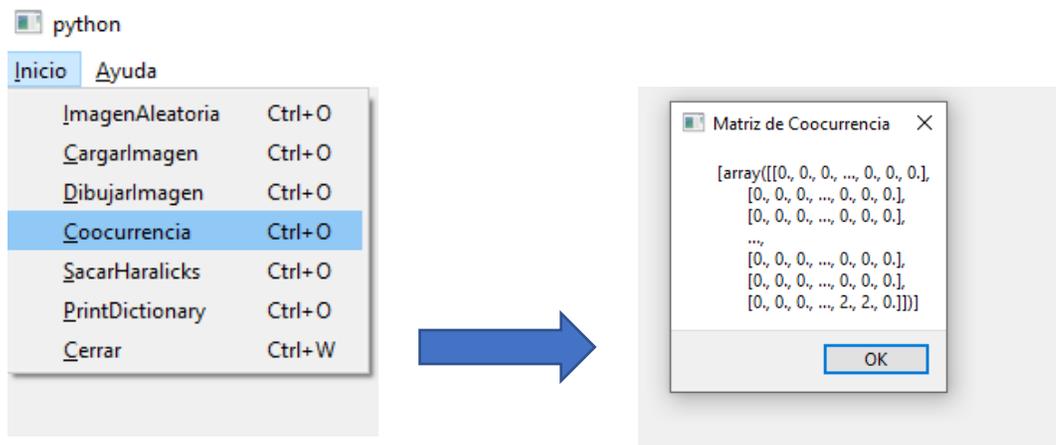
## 4.4. Ejecución del programa

### 1. Cargar Imagen: Seleccionar la imagen para procesarla



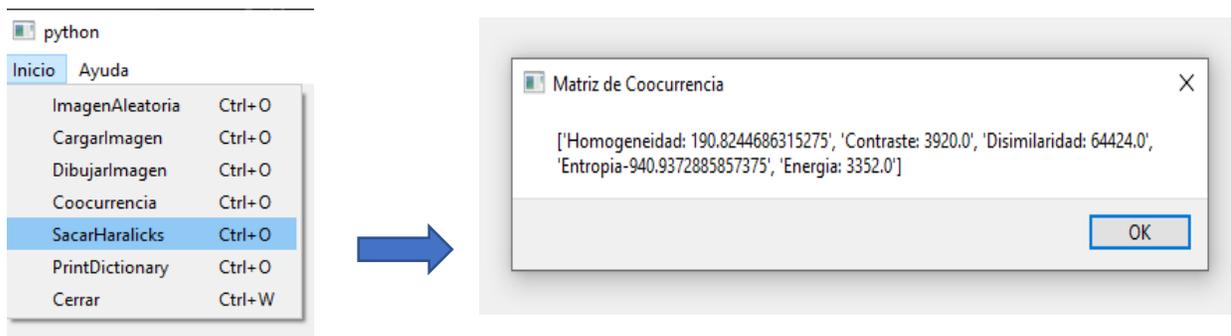
**Figura 42.** Seleccionar la imagen para procesarla

2. **Obtener la matriz de coocurrencia:** Permite la obtención de la matriz a partir de la matriz de grises que se generó con la imagen.



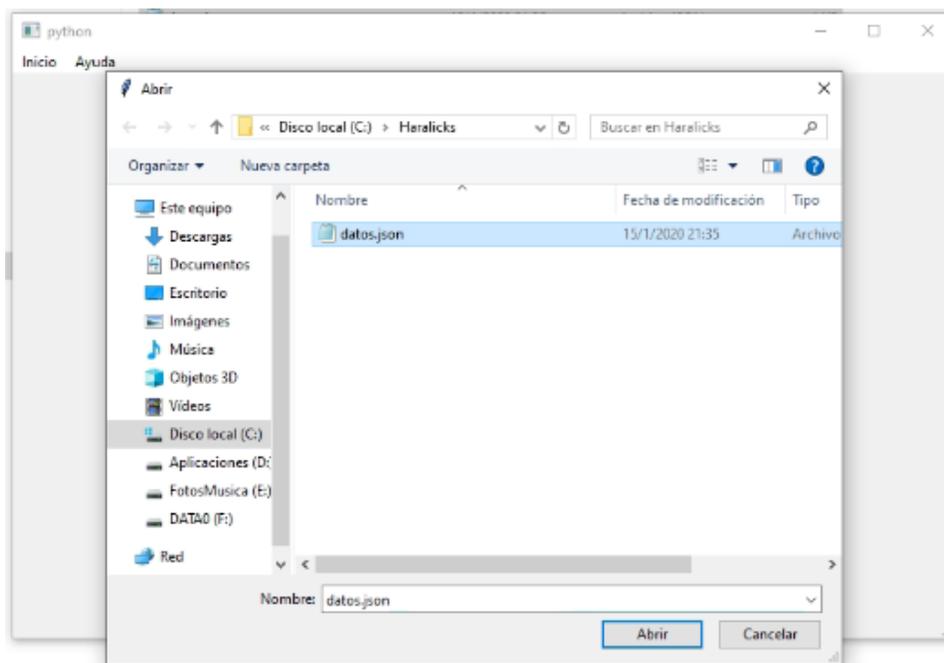
**Figura 43.** Obtener la matriz de coocurrencia

3. **Obtener Haralicks:** Es el proceso final para obtener los coeficientes Haralick a partir de la matriz de coocurrencia.



**Figura 44.** Obtener Haralicks

4. **Unión de archivos JSON:** Se solicita la selección del archivo JSON de la metadata con el que se va a unir los coeficientes Haralick generados.



**Figura 45.** Selección de archivo JSON con metadata de mamografía.

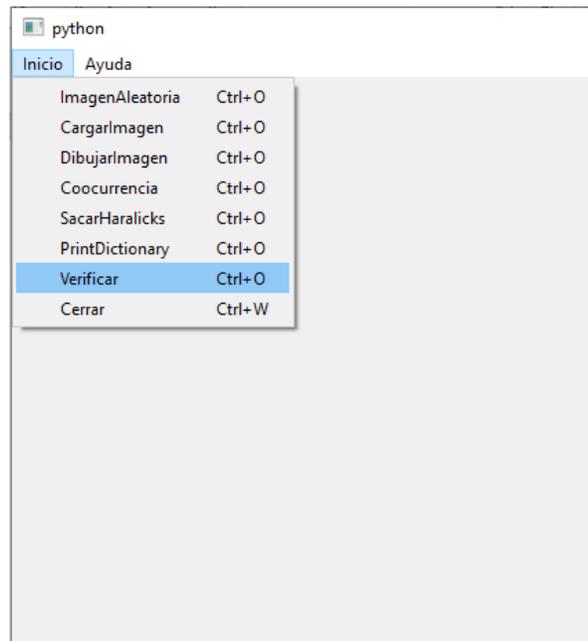
5. **Almacenamiento de los datos:** Se guarda en un archivo Json para un siguiente procesamiento.

Nombre	Fecha de modificación	Tipo	Tamaño
2019.12.23.HORA.19.59	23/12/2019 19:59	Archivo JSON	1 KB
2019.12.30.HORA.16.44	30/12/2019 16:44	Archivo JSON	1 KB

**Figura 46.** Almacenamiento de los datos

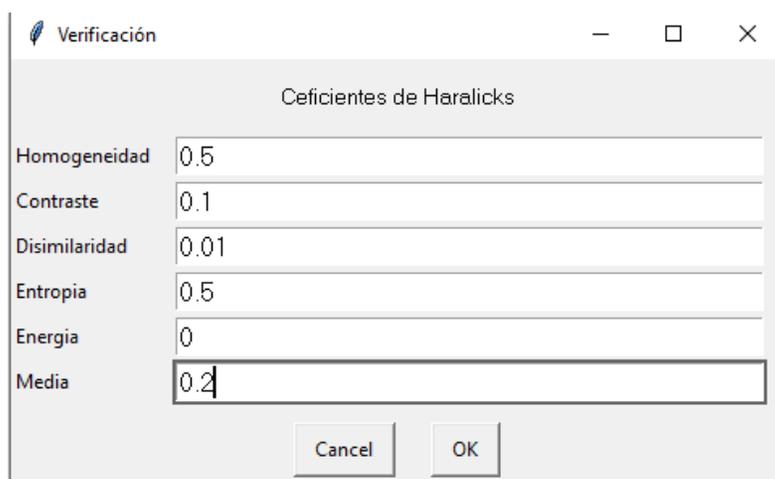
#### 4.5. Verificación de los coeficientes Haralick

Al momento, la aplicación se ha encargado de determinar los coeficientes, pero ahora en la sección de verificar se podrá ingresar los coeficientes y asociar con el caso que más se aproxime más la imagen asociada a dichos coeficientes.



**Figura 47.** Menú verificar

Al momento de seleccionar la opción verificar se permite el ingreso de los coeficientes Haralick según la Figura 46.

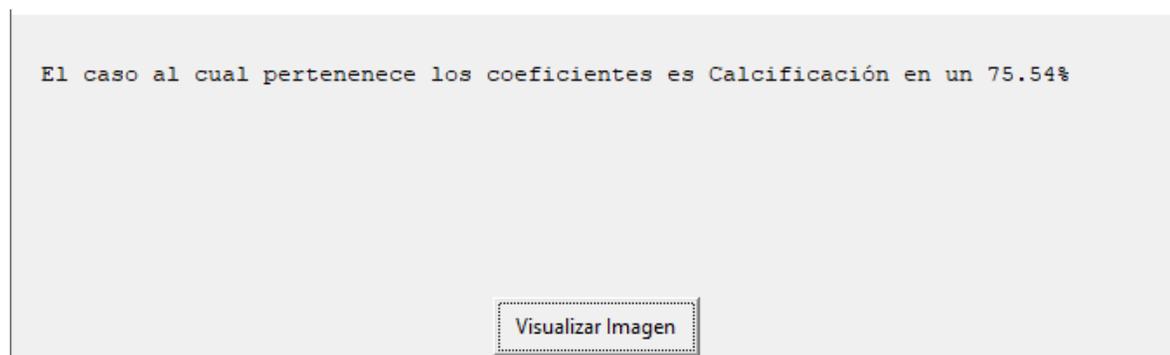


The image shows a software dialog box titled 'Verificación' with a subtitle 'Ceficientes de Haralicks'. It contains six input fields for different Haralick coefficients: Homogeneidad (0.5), Contraste (0.1), Disimilaridad (0.01), Entropia (0.5), Energia (0), and Media (0.2). At the bottom, there are 'Cancel' and 'OK' buttons.

Coefficient	Value
Homogeneidad	0.5
Contraste	0.1
Disimilaridad	0.01
Entropia	0.5
Energia	0
Media	0.2

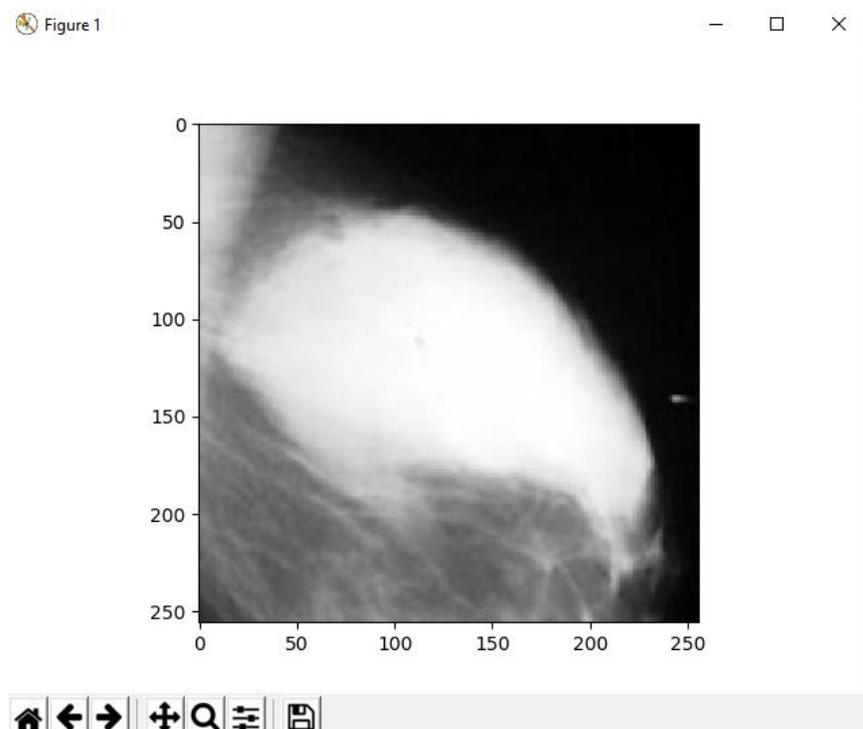
**Figura 48.** Ingreso de coeficientes Haralick

Una vez ingresado todos los coeficientes se hace la comparación con todos los casos ingresados para verificar que porcentaje es coincidente con alguno de los casos, se realiza la comparación con cada uno de los coeficientes y se descarta los casos que tiene poco parecido, de todos los casos seleccionados se selecciona el caso con mayor porcentaje de coincidencia, el resultado se refleja de la siguiente manera:



**Figura 49.** Resultado de coincidencia con los coeficientes Haralick ingresados.

Una vez que se ha reflejado el resultado se presenta la opción de visualizar la imagen que más se asemeje, esta se la dibuja con la matriz de grises que se ha guardado en la base de datos:



**Figura 50.** Imagen dibujada basada en la matriz de grises.

## CAPITULO V

### CONCLUSIONES Y RECOMENDACIONES

#### 5.1. Conclusiones

- El aporte significativo que representa la implementación de la ontología es el diagnóstico temprano del cáncer de mama a través del análisis e interpretación de los patrones de una mamografía utilizando como base los niveles del BI-RADS y los coeficientes de textura de Haralick.
- El problema de trabajar con las imágenes de patrones de una mamografías, es que no siempre cumplen con las mismas características para su manipulación (tamaño, resolución, ...) y genera muchos problemas en la definición de los coeficientes Haralick, produciendo que los valores finales no sean los reales.
- Al ser un estudio relativamente nuevo en el área del análisis de textura de mamografías utilizando los coeficientes de Haralick, ocasiono que la implementación de la ontología tardase más de lo planificado, por la limitada información que se puede hallar del tema.

## 5.2. Recomendaciones

- Se puede obtener mejores resultados en el análisis de los patrones mamográficos, si se realizan estimaciones o estudios de cuáles son los coeficientes Haralick que mejor discriminen una textura, o a su vez se puede añadir otras variables de análisis de textura para conseguir más información que se puedan analizar.
- Para trabajar con las imágenes de los patrones, se debe parametrizar las características que se utilizarán previo al análisis de cada caso; esto reducirá el tiempo de procesamiento y aumentará la eficacia de la ontología ya que los casos ingresados tendrán el mismo tipo de análisis.
- Para comenzar a realizar un estudio relacionado con el área médica, en este caso de las mamografías; es importante tener el apoyo de los expertos en el área para que solventen todas las inquietudes que se produzcan y así no tener mucho tiempo de demora en el desarrollo de la investigación.

### **5.3. Trabajos Futuros**

Se busca incrementar el campo de estudio, no solo limitando al análisis de mamografías, sino aumentando a otros tipos de exámenes médicos como las resonancias magnéticas, tomografías o ecografías de distintas partes del cuerpo (pulmones, abdomen, etc.) o incluso a tejidos del cuerpo (piel, tejidos musculares, etc.).

De la mayoría de los exámenes médicos se puede obtener un patrón gráfico que puede ser procesado y cuantificado utilizando los distintos patrones de textura, es decir que no solo se puede emplear los coeficientes Haralick, sino que existe la opción de que se use otros coeficientes como los de Gabor o Wavelet.

## BIBLIOGRAFÍA

- AEGO. (2010). *Un nuevo concepto lactancia y estática corporal*. Obtenido de:  
<https://www.aego.es/otra-informacion/generar-automaticamente-desde-el-titulo>
- Argentina.gob.ar. (2010). *Manual operativo para el uso de mamografía en tamizaje*.  
Obtenido de: <http://www.msal.gob.ar/images/stories/bes/graficos/0000000013cnt-10-Manual-operativo-de-uso-de-mamografia.pdf>
- Arija, C. (2012). *Taxonomía, Sistemática y Nomenclatura, herramientas esenciales en Zoología y Veterinaria*. REDVET.
- Asociación Española Contra el Cáncer. (2014). AECC. Obtenido de AECC:  
<https://www.aecc.es/sites/default/files/migration/actualidad/publicaciones/documentos/cancer-mama-2014.pdf>
- Bernaras, A., Laresgoiti, I., & Correa, J. (1996). *Building and Reusing Ontologies for Electrical Network Applications*. Semantic Scholar.
- Bru, D. (1992). *Principios básicos de la ecografía*. Patología Animal II.
- Buffa, R., Gamarra, S., & Viniegra, M. (2011). *Manual Operativo para el uso de mamografías de tamizaje*. Obtenido de:  
<http://www.msal.gob.ar/images/stories/bes/graficos/0000000013cnt-10-Manual-operativo-de-uso-de-mamografia.pdf>
- Caeiro, M. (2017). *Formación en artes plásticas y visuales para maestros de primaria*. Logroño.
- Chantal, P. (2002). *Explotación de los corpórea textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento*. Málaga.

- Domelles , C., Santos da Silva, P., Oliveira, C., Goldim, J., & Asthon-Prolla, P. (2015). *Conocimiento del cáncer de mama y cáncer de mama hereditario en el personal de enfermería de un hospital público.*
- D'Orsi, C. (2016). *BI-RADS. Atlas.*
- Febles, G. (2012). *Falsos negativos en mamografías.* Congreso Internacional de Radiología.
- Fernández , J. (Marzo de 2003). *Un Entorno de Integración de Ontologías para el desarrollo de sistemas de Gestión de Conocimiento.* Tesis Doctoral. Murcia, España.
- Gallego, G. (2005). *Nódulo palpable de mama.* Revista Colombiana de Obstetricia y Ginecología, 82-91.
- García, M. (Febrero de 2006). *Método de adquisición de modelos a partir de corpus textuales y su aplicación en la integración de bases de datos y fuentes de información.* Tesis Doctoral. Madrid, España.
- González, R. (Mayo de 2007). *Python para todos.* Obtenido de Python para todos: [http://www.utic.edu.py/citil/images/Manuales/Python\\_para\\_todos.pdf](http://www.utic.edu.py/citil/images/Manuales/Python_para_todos.pdf)
- Gruber, T. (1993). A Translation Approach to Portable Ontology Specifications. *Knowledge Systems Laboratory*, 71-92.
- Gruninger, M. (1995). Methodology for the Design and Evaluation of Ontologies.
- Haralick, R. (1973). Textural Features for Image Classification. *IEEE Transactions On Systems*, 610,621.
- Hernández, H., & Saiz, M. (2007). Ontologías mixtas para la representación conceptual de objetos de aprendizaje. *Red de Revistas Científicas de América Latina y el Caribe, España y Portugal.*
- Icaza, G., Nuñez, L., & Bugueño , H. (2017). Descripción epidemiológica de la mortalidad por cáncer de mama en mujeres en Chile. *Rev Med Chile*, 106-114.

- INEC. (2017). *El cáncer de mama en Ecuador*. Quito.
- Martinez Villaseñor, D., & Gerson Cwilich, R. (2006). *La tomografía por emisión de positrones*. Obtenido de La tomografía por emisión de positrones: <https://www.medigraphic.com/pdfs/circir/cc-2006/cc064n.pdf>
- Martinez, J., & Castro, C. (2018). Inmunohistoquímica en el cáncer de mama. Herramienta necesaria en la actualidad. *Medisur*.
- Marzal, A., & Gracia, I. (2009). *Introducción a la programación con python*. Obtenido de Introducción a la programación con python: <http://www1.herrera.unt.edu.ar/biblcet/wp-content/uploads/2014/12/ippython.pdf>
- MAYO CLINIC. (2018). *Tejido mamario denso: qué significa tener mamas densas*. Obtenido de Tejido mamario denso: qué significa tener mamas densas: <https://www.mayoclinic.org/es-es/tests-procedures/mammogram/in-depth/dense-breast-tissue/art-20123968>
- MIERCOLA. (Agosto de 2012). *Las Mamografías Tienen Un "Efecto Limitado o Ningún Efecto" en las Muertes por Cáncer de Mama: Sugiere un Estudio*. Obtenido de Las Mamografías Tienen Un "Efecto Limitado o Ningún Efecto" en las Muertes por Cáncer de Mama: Sugiere un Estudio: <https://espanol.mercola.com/boletin-de-salud/mamografia-y-la-tasa-de-mortalidad-de-cancer-de-mama.aspx>
- Neches, R., Fikes, R., & Finin, T. (1991). Enabling Technology for Knowledge Sharing. *AI Magazine* 12, 35-56.
- Noel, L., Aguilar, L., & Ramírez, J. (2004). Importancia de las calcificaciones en la mamografía. *Acta Médica Grupo Ángeles Mg*.
- OIML. (2011). *GUÍA OIML G – 14: Medición de densidad*.
- Peralta, O. (2011). Cáncer de mama: Estrategias de prevención y vigilancia Según nivel de riesgo. *Med.Clin.Condes*.
- Pinto, A. (2006). *Segmentación De Imágenes Por Textura*. Concepción.

Presutti, M. (2004). La matriz de co-ocurrencia en la clasificación multiespectral: tutorial para la enseñanza de mediadas texturales en cursos de grados universitarios. *Sensoriamento Remoto no Âmbito do Mercosul*.

Programacion.net. (2011). *Procesamiento de imágenes utilizando Python*. Obtenido de Procesamiento de imágenes utilizando Python: [https://programacion.net/articulo/procesamiento\\_de\\_imagenes\\_utilizando\\_python\\_1451](https://programacion.net/articulo/procesamiento_de_imagenes_utilizando_python_1451)

RadiologyInfo.org. (2017). *Exploración por tomografía por emisión de positrones tomografía computada*. Obtenido de Exploración por tomografía por emisión de positrones tomografía computada: <https://www.radiologyinfo.org/sp/pdf/pet.pdf>

RadiologyInfo.org. (Abril de 2018). *Detección temprana del cáncer de seno*. Obtenido de Detección temprana del cáncer de seno: <https://www.radiologyinfo.org/sp/info.cfm?pg=screening-breast>

Rivas, J., Urbina, R., Serpas, M., & García, E. (Marzo de 2006). *Guía técnica de prevención y control del cáncer de mama*. obtenido de guía técnica de prevención y control del cáncer de mama: [http://asp.salud.gob.sv/regulacion/pdf/guia/Guia\\_Mama\\_Mujer.pdf](http://asp.salud.gob.sv/regulacion/pdf/guia/Guia_Mama_Mujer.pdf)

Sanchez, N., Comas, R., & García, M. (2017). Propuesta de ontología para el desarrollo de un sistema de información geográfico para la energía eléctrica en cuba. *PUCESA*.

Sánchez, S. (14 de Mayo de 2007). Modelo de indexación de formas en sistemas VIR basado en ontologías. *Tesis profesional*. Puebla, México.

Soriano, G. (30 de Junio de 2009). *Introducción a Python*. Obtenido de Introducción a Python: [http://web.fi.uba.ar/~bortega/algoritmos\\_l/descargas/apunte\\_python.pdf](http://web.fi.uba.ar/~bortega/algoritmos_l/descargas/apunte_python.pdf)

Staab, S., Studer, R., Schnurr, & Sure, Y. (2001). Knowledge processes and ontologies. *IEEE*, 26-34.

- Swartout, B., Patil, R., Knight, K., & Rus, T. (1997). Toward Distributed Use of Large-Scale Ontologies. *AAAI*.
- Universidad Estatal Península De Santa Elena. (2017). *Prevalencia Y Medidas Preventivas En Cáncer De Mama*. Obtenido de: <https://www.upse.edu.ec/rcpi/index.php/revistaupse/article/viewFile/183/169>.
- Valencia, R. (Enero de 2005). Un Entorno para la Extracción Incremental de Conocimiento desde Texto en Lenguaje Natural. *Tesis Doctoral*. Murcia, España.
- Vega, B. (2011). Intervencionismo diagnóstico en patología de mama. *Radiología*, 531-543.
- Vergara, O. (Abril de 2003). Reconocimiento de texturas artificiales, aplicación a la inspección visual. *tesis para obtener el grado de maestro en ciencias computacionales*. Cuernavaca.
- Villa Veásquez, H., Pabón Romero, J. C., Jaramillo Botero, N., & Mantilla Suárez, J. C. (2006). *Ultrasonografía de mama*. Obtenido de Ultrasonografía de mama: <http://132.248.9.34/hevila/Medunab/2006/vol9/no2/4.pdf>
- Walker, F., & Cartwright, M. (2012). *Ecografía Neuromuscular*. España.
- docs.python.org. (2019). *Graphical User Interfaces with Tk — Python 3.8.1 documentation*. Obtenido de Graphical User Interfaces with Tk — Python 3.8.1 documentation: <https://docs.python.org/3/library/tk.html>
- docs.python.org. (2019). *pathlib- Rutas de sistema de archivos orientadas a objetos*. Obtenido de pathlib- Rutas de sistema de archivos orientadas a objetos: <https://docs.python.org/3/library/pathlib.html>
- EasyGUI. (2014). *EasyGUI - documentación de easygui*. Obtenido de EasyGUI - documentación de easygui : <http://easygui.sourceforge.net/>

<https://docs.python.org/3/library/json.html#module-json>. (2019). *json- Codificador y decodificador JSON*. Obtenido de json- Codificador y decodificador JSON: <https://docs.python.org/3/library/json.html#module-json>

Matplotlib. (2019). *Pyplot tutorial - Matplotlib 3.1.1 documentation*. Obtenido de Pyplot tutorial - Matplotlib 3.1.1 documentation: <https://matplotlib.org/3.1.1/tutorials/introductory/pyplot.html>

Numpy.org. (2016). *NumPy - NumPy*. Obtenido de NumPy - NumPy: <https://numpy.org/>

OpenCV. (2018). *Introducción a las imágenes: documentación de OpenCV-Python Tutorials 1*. Obtenido de Introducción a las imágenes: documentación de OpenCV-Python Tutorials 1: [https://opencv-python-tutroals.readthedocs.io/en/latest/py\\_tutorials/py\\_gui/py\\_image\\_display/py\\_image\\_display.html](https://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_gui/py_image_display/py_image_display.html)

PyQt5 tutorial . (2019). *PyQt5 tutorial 2019: crear una GUI con Python y Qt*. Obtenido de PyQt5 tutorial 2019: crear una GUI con Python y Qt: <https://build-system.fman.io/pyqt5-tutorial>

Recursos Python. (2014). *Instalar PIL / Pillow y aplicar efectos visuales*. Obtenido de Instalar PIL / Pillow y aplicar efectos visuales: <https://recursospython.com/guias-y-manuales/instalar-pil-pillow-efectos/>

Scikit-image. (2019). *Procesamiento de imagen en Python*. Obtenido de Procesamiento de imagen en Python: <https://scikit-image.org/>