

UNIVERSIDAD MAYOR DE SAN ANDRES
POSTGRADO EN INFORMÁTICA
MAESTRIA EN SOFTWARE LIBRE Y ESTANDARES ABIERTOS



TESIS DE MAESTRIA
MÉTODO HEURÍSTICO PARA EL DIAGNÓSTICO DE
CÁNCER DE MAMA BASADO EN MINERÍA DE DATOS.

Tesis de postgrado presentada para la obtención del Grado de Magister

POR: Lic. SANDRO SAUL CAMACHO CENTELLAS

TUTOR: M. Sc. CARLA XIMENA OVANDO POLO

LA PAZ-BOLIVIA

Diciembre, 2016



**UNIVERSIDAD MAYOR DE SAN ANDRÉS
FACULTAD DE CIENCIAS PURAS Y NATURALES
CARRERA DE INFORMÁTICA**



LA CARRERA DE INFORMÁTICA DE LA FACULTAD DE CIENCIAS PURAS Y NATURALES PERTENECIENTE A LA UNIVERSIDAD MAYOR DE SAN ANDRÉS AUTORIZA EL USO DE LA INFORMACIÓN CONTENIDA EN ESTE DOCUMENTO SI LOS PROPÓSITOS SON ESTRICTAMENTE ACADÉMICOS.

LICENCIA DE USO

El usuario está autorizado a:

- a) visualizar el documento mediante el uso de un ordenador o dispositivo móvil.
- b) copiar, almacenar o imprimir si ha de ser de uso exclusivamente personal y privado.
- c) copiar textualmente parte(s) de su contenido mencionando la fuente y/o haciendo la referencia correspondiente respetando normas de redacción e investigación.

El usuario no puede publicar, distribuir o realizar emisión o exhibición alguna de este material, sin la autorización correspondiente.

TODOS LOS DERECHOS RESERVADOS. EL USO NO AUTORIZADO DE LOS CONTENIDOS PUBLICADOS EN ESTE SITIO DERIVARA EN EL INICIO DE ACCIONES LEGALES CONTEMPLADOS EN LA LEY DE DERECHOS DE AUTOR.

Dedicatoria

Esta tesis está dedicada a mi amada esposa Norka, a mis dos amadas hijas: Estefani y Luciana, que a partir de este gran paso en mi vida, tomen el ejemplo de constante superación y siempre anhelar a crecer y no conformarse con un solo logro; sino más bien varios; a pesar de los contratiempos y adversidades que puedan tener.

AGRADECIMIENTO

Ante todo agradecer a Dios, por brindarme vida y salud, a mí tutora M. Sc Carla Ovando Polo por su paciencia y conocimiento, a mí querida Madre Julia Centellas por su apoyo incondicional, a mis hermanos Jhonny, Nelva y Rubén por su apoyo moral y finalmente, a la Dra. Carmen Ovando Polo por brindarme su ayuda.

Índice General

<i>Introducción</i>	11
<i>Descripción del problema</i>	13
<i>Importancia para afrontar el problema</i>	14
<i>Capítulo I MARCO DEL PROBLEMA</i>	15
1.1 Estado del Arte.....	15
1.1.2 Heurística.	17
1.1.3 Mamografía.	18
1.1.4 Minería de datos.....	22
1.2 Planteamiento del problema.....	23
1.3 Formulación del problema de investigación.....	24
1.4 Planteamiento de objetivo.....	24
1.4.1 Objetivo general.	24
1.4.2 Objetivos específicos.	25
1.5 Planteamiento de Hipótesis.....	25
1.5.1 Variable dependiente.....	25
1.5.2 Variable Independiente.	25
<i>CAPITULO II MARCO METODOLÓGICO</i>	26
2.1 Diseño metodológico.	26
2.2 Clasificación de la Investigación.....	27
2.2.1 Según el Propósito o Razón.	27
2.2.3 Según la Estrategia Empleada.....	27
2.2.4 Según la Ubicación de los Datos en el Tiempo.	27

2.3	<i>Método de investigación</i>	28
2.4	Técnicas de investigación	29
2.5	Universo o Población de referencia	29
2.6	Delimitación geográfica.....	30
<i>Capitulo III Marco Teórico</i>		31
3.1	<i>Referencia Teórica.</i>	31
3.1.1	Cáncer de Mama.	31
3.1.2	Identificación de tonos claro, oscuro y gris	32
3.1.3	Radiografía convencional (placa simple).....	33
3.1.4	Tomografía Computarizada (TC).....	34
3.1.5	Ecografía	35
3.1.6	Mamografía	36
3.1.7	Resonancia Magnética (RM)	37
3.1.8	Las cinco densidades básicas	38
3.1.9	Minería de datos.....	41
3.1.10	Heurística.....	47
3.1.11	Estudios realizados por otros investigadores.....	50
3.2	Marco Legal o institucional.	51
3.2.1	Constitución Política del Estado.	51
<i>Capitulo III Marco ingeniera de prototipo</i>		53
3.1	Análisis y Diseño del prototipo.	53
3.1.1	Método Heurístico.....	53
3.1.2	Una visión general del método propuesto.....	53
3.2	Arquitectura tecnológica.....	63
3.2.1	Transformada de Wavelet Discreta.....	63
3.2.2	Preparación de la imagen mamográfica	69

3.2.3	Análisis de la imagen mamográfica	71
3.2.4	Supresión del ruido de la imagen mamográfica.....	72
3.2.5	Extracción de Patrones	73
3.2.6	Clasificación de Patrones	75
3.2.7	Optimización basada en el Algoritmo Genético	75
<i>Capítulo IV Marco de Resultados.....</i>		<i>79</i>
4.1	Estado de los Objetivos.....	79
4.1.1	Objetivo General	79
4.1.2	Objetivos Específicos.....	79
4.2	Estado de la Hipótesis.....	80
4.2.1	Aplicación del Método Heurístico	80
<i>Conclusiones.....</i>		<i>90</i>
4.3	<i>Recomendaciones.....</i>	<i>91</i>
<i>Primera Recomendación</i>		<i>91</i>
<i>Bibliografía.....</i>		<i>93</i>
<i>Anexos.....</i>		<i>99</i>
ANEXO A.....		100
“OPERACIONALIZACION DE LAS VARIABLES”		100
ANEXO “B“		101
GUÍA DE ENTREVISTA.....		101
<i>Apéndices.....</i>		<i>105</i>
<i>Apéndice A Terminología clínica</i>		<i>106</i>
<i>Glosario.....</i>		<i>108</i>

Masa: 108

Calcificaciones 108

Índice de Figuras

<i>Figura.- 1.Fases del Método Heurístico</i>	<i>54</i>
<i>Figura.- 2.Fases del proceso de extracción de Patrones de imágenes mamográficas con la minería de datos</i>	<i>56</i>
<i>Figura.- 3.Imagen Mamografía –ejemplo</i>	<i>62</i>
<i>Figura.- 4.Disecion y conformación de familias de patrones (FP).....</i>	<i>62</i>
<i>Figura.- 5.Componentes de alta y baja frecuencia separadas, muestreo completo y muestreo reducido con los coeficientes wavelet donde 2 significa muestreo reducido a la mitad</i>	<i>66</i>
<i>Figura.- 6.Descomposición de la señal</i>	<i>67</i>
<i>Figura.- 7.Descomposición en tres niveles de la misma señal.....</i>	<i>67</i>
<i>Figura.- 8.Superposición de las señales original y recuperada con reducción por coeficientes wavelets.....</i>	<i>68</i>
<i>Figura.- 9.Filtros de wavelet.....</i>	<i>72</i>
<i>Figura.- 10. Representación del Algoritmo Genético para la maximizar el carácter común</i>	<i>77</i>
<i>Figura.- 11.Imagen Mamográfica de aplicación</i>	<i>82</i>
<i>Figura.- 12 Familia de Patrones 1,2, 3 (FP1, FP2, FP3).....</i>	<i>82</i>
<i>Figura.- 13Familia de Patrones 4, 5, 6 (FP4, FP5, FP6).....</i>	<i>83</i>
<i>Figura.- 14 Diseción de la imagen mamográfica.....</i>	<i>83</i>
<i>Figura.- 15 Comparación de la ejecución del AG</i>	<i>86</i>
<i>Figura.- 16 FP 4 con detección.....</i>	<i>87</i>
<i>Figura.- 17 FP56 Con detección.....</i>	<i>88</i>
<i>Figura.- 18 FP6 Con detección.....</i>	<i>89</i>

Índice de Tablas

<i>Tabla 1. Las cinco densidades básicas que se observan en la radiografía convencional</i>	<i>40</i>
<i>Tabla 2 Blanco y negro: términos utilizados en cada modalidad.....</i>	<i>40</i>
<i>Tabla 3. Lista de similitud de índices.....</i>	<i>58</i>
<i>Tabla 4 Resultados de la experimentación de los resultados obtenidos del AG.....</i>	<i>85</i>
<i>Tabla 5 BI - RADS / Categorías.....</i>	<i>107</i>

Resumen

El cáncer de mama es una de las principales causas de muerte en las mujeres; por esta razón, existe una amplia evidencia de que la detección temprana juega un papel importante en la reducción de la mortalidad de este cáncer, muchos métodos han sido investigados para mejorar una detección temprana; actualmente se utiliza la mamografía, mamografía 3d y la Tomografía axial; entre otros. Sin embargo, la mamografía es el medio más accesible y menos dañino para proporcionar información útil sobre la presencia de cáncer de mama. Debido a la dificultad en la interpretación de las mamografías, son necesarias múltiples lecturas de un único examen; de este modo, en este trabajo se propone aumentar la fiabilidad de un diagnóstico con la utilización de procesamiento de imágenes digitales y análisis de las mismas; mediante un Método Heurístico basado en Minería de Datos para extraer información esencial de las imágenes mamográficas y transformarlas en patrones. Posteriormente, se clasificará en sub-grupos de patrones para la conformación de familias mediante la homogeneidad y maximización de índices de coincidencia. El enfoque propuesto reduce la intervención humana y mejora la precisión de los resultados computacionales y es relativamente fácil de adquirir los datos.

Palabras claves: Heurística, Minería de Datos, patrones, imágenes mamográficas, cáncer

Summary

Breast cancer is one of the leading causes of death in women; For this reason, there is ample evidence that early detection plays an important role in reducing the mortality of this cancer, many methods have been investigated to improve early detection; Mammography, 3d mammography and axial tomography are currently used; among others. However, mammography is the most accessible and least harmful way to provide useful information about the presence of breast cancer. Because of the difficulty in interpreting mammograms, multiple readings of a single examination are required; In this way, in this work it is proposed to increase the reliability of a diagnosis with the use of digital image processing and analysis of the same; Using a Data Mining-based Heuristic Method to extract essential information from mammographic images and transform them into patterns. Subsequently, it will be classified into sub-groups of patterns for the conformation of families through the homogeneity and maximization of coincidence indexes. The proposed approach reduces human intervention and improves the accuracy of computational results and is relatively easy to acquire data.

Keywords: Heuristics, Data Mining, patterns, mammographic images, cancer

Introducción

El cáncer de mama es el más frecuente en las mujeres; según el Instituto Nacional del Cáncer Americano, una de cada ocho mujeres que viva hasta los 70 años, desarrollará cáncer de mama a lo largo de su vida y, entre 25 a 30 mujeres mueren por esta enfermedad. La incidencia del cáncer mamario ha aumentado de un 5% en 1940 a un 12% en 1995. También en España es el tumor de mayor incidencia en la mujer (entre 40 a 70 mujeres por cada 100.000), según el Registro Español de Cáncer. En las mujeres, el cáncer en general, es la causa de muerte en el 19,9% de los casos, de los cuales, el más frecuente (16,6%) es el de mama, Se ha determinado que el cáncer de mamas es la segunda causa de muerte de mujeres con cáncer, se estima que 22 de cada 100.000 mujeres poseen dicha enfermedad (Abalo, 2003).

El Ministerio de Salud de Bolivia en el año 2010 informó que en Bolivia, 26,57 por cada 100.000 mujeres desarrollaron Cáncer de Mama, entre tanto la mortalidad por ese mal alcanzó a 8,71 de cada 100.000. Por esa razón, el Ministerio de Salud elaboró y aplicó líneas estratégicas dentro del Plan Nacional de Prevención, Control y Seguimiento del Cáncer de Mama que tiene como fin reducir la alta incidencia y la mortalidad por esta dolencia.

El cáncer de mama es una de las pocas enfermedades cancerosas que se pueden diagnosticar precozmente; antes de que se note algún síntoma; la mamografía es una prueba de imagen por rayos X que detecta la presencia del tumor en la mama, antes de que sea perceptible al tacto, los especialistas recomiendan que todas las mujeres se realicen esta prueba cada año, a partir de los 50 a 55 años.

La mamografía permite detectar lesiones en la mama hasta dos años antes de que sean palpables y cuando aún no han invadido en profundidad ni se han diseminado a los ganglios

ni a otros órganos. Cuando el tumor se detecta en estas etapas precoces es posible aplicar tratamientos menos agresivos; los cuales dejan menos secuelas físicas y psicológicas en la mujer (Antonie M., 2001) .

Descripción del problema

El diagnóstico clínico es un proceso cognitivo y complejo que precisa de: capacitación, experiencia, reconocimiento de patrones y cálculo de probabilidad condicional.

Muchos médicos han adquirido en el tiempo una intuición que facilita la tarea; la misma es producto de la experiencia de haber visto cientos de casos similares a lo largo de su carrera.

Los médicos menos experimentados pueden reducir sus errores al momento de diagnosticar a un paciente, realizando un proceso consciente y sistemático; tomando en cuenta la confiabilidad de sus apoyos diagnósticos y las fuentes más frecuentes de error.

Un médico puede sentirse fatigado o abordar de manera superficial a un paciente, puede no estar familiarizado con su enfermedad o estar influido por la opinión diagnóstica del colega que lo evaluó antes; y como consecuencia, no considerar un diagnóstico con un abordaje apropiado. Otro error común es la asociación de toda la sintomatología al diagnóstico inicial; utilizándolos solo para confirmar lo que se pensó en un inicio y descartando aquellos síntomas que no estén asociados.

En el cáncer de mama existen diagnósticos errados por factores humanos, debido a que la detección de la totalidad de los carcinomas visibles a través de análisis

retrospectivos de las imágenes, muchas veces resulta complicado; las lesiones varían desde alteraciones en partes blandas de distintas formas y márgenes, hasta calcificaciones de diferente morfología, tamaño y distribución, que pueden ser representativas de malignidad; por lo que son necesarias múltiples lecturas de un mismo examen.

A través de la aplicación de diferentes investigaciones sobre minería de datos para el procesamiento de imágenes médicas han obtenidos muy buenos resultados mediante las redes neuronales para tareas de clasificación y agrupamiento (Antonie M., 2001). Pero, el presupuesto de los hospitales en Bolivia, no alcanza a cubrir el alto costo de dichas herramientas por lo cual se dificulta el acceso a esta tecnología.

Importancia para afrontar el problema

La interacción de diferentes disciplinas permite resolver problemas complejos, algunas de ellas son bases de datos, estadística, aprendizaje automático y procesamiento de imágenes (Han, 2001).

La Heurística es un procedimiento para resolver un problema de optimización bien definido mediante una aproximación intuitiva, en la que la estructura del problema se utiliza de forma inteligente para obtener una buena solución.

La Minería de Datos es el conjunto de técnicas y herramientas aplicadas al proceso no trivial de extraer y presentar conocimiento implícito, previamente desconocido, potencialmente útil y humanamente comprensible, a partir de grandes conjuntos de datos, con objeto de predecir de forma automatizada tendencias y comportamientos; y describir de forma automatizada modelos previamente desconocidos (Piatetski-Shapiro, 1996).

Capítulo I MARCO DEL PROBLEMA

1.1 Estado del Arte

1.1.1 Cáncer de Mama

Se ha encontrado varios grupos de investigación en el diagnóstico del cáncer de mama, la mayor parte se encuentra en una fase experimental, con respecto a la explotación de datos aplicadas al procesamiento de imágenes, las investigaciones se viene realizado a través de grupos de investigación de grandes organizaciones gubernamentales como la Agencia Nacional Aéreo Espacial del gobierno de los Estados Unidos de Norteamérica (NASA) (Simoff S., 2002). Actualmente la resolución de tareas de explotación de datos se está realizando con herramientas propietarias ofrecidas por las diferentes empresas líderes en tecnología de bases de datos, como ORACLE o MICROSOFT (Liebstein, 2002).

Para la detección de microcalcificaciones, se ha encontrado distintos criterios y enfoques usando la transformada wavelet, que se analiza la forma, el tamaño y el número de microcalcificaciones por unidad de área (Strickland, 1994) (Karssemeijer B. a., 1997), seguidamente de técnicas de ecualización para ajustar la distribución de los datos, lo cual es determinante para aumentar la fracción de verdaderos positivos y minimizar la fracción de falsos positivos. (Karssemeijer W. V., 2000), varios estudios realizan la detección de anomalías combinando desde la detección de contornos hasta la extracción de características por técnicas multiresolución. Otros trabajos proponen un sistema basado en redes neuronales (Verma., 1790-1793). (Giger, 1997) Propone como los algoritmos genéticos para extraer las mejores características de cada píxel, para posteriormente clasificarlos como microcalcificaciones empleando una red neuronal basada en la función de error entropía cruzada. Mientras que (A. Hojjatoleslami, 1997) realiza un comparación de varios métodos de clasificación tras aplicar una etapa de pre procesado a las

mamografías, proporcionando mejores resultados el clasificador. Por otro lado, (Schiabel, 2001) tras una etapa de segmentación aplicada a la mamografía y la extracción y selección de características, emplea un red neuronal entrenada.

Otros estudios se basan en el filtrado para eliminar el fondo y aumentando la relación señal a ruido que permita el realzado de la región de interés, para identificar las posibles anomalías mediante las técnicas de umbralización local y global de los niveles de gris, mientras que la detección de microcalcificaciones se hace mediante la extracción de características como tamaño, distribución espacial de las microcalcificaciones, análisis de la textura. (R.M. Nishikawa Y. J., 1992) (K. Doi, 1993) (Y. Jiang, 1997), distintos trabajos se han centrado en clasificadores basados en redes neuronales, algoritmos genéticos, funciones de base radial,

(Karayiannis, 1998) Propone que a partir de la mamografía digital, se realiza un descomposición multiresolución basada en filtros de análisis wavelet, a continuación elimina las subbandas de baja frecuencia y por último reconstruye la mamografía que únicamente contiene componentes de alta frecuencia, incluyendo microcalcificaciones. (G.Ling, 2000) Combina las prestaciones entre el análisis de características de multiresolución de la imagen con el uso de la transformada wavelet y las características de la estructura de niveles de gris y las redes neuronales para la detección de microcalcificaciones.

Trabajos para la detección de microcalcificaciones se han basado en el uso de Support Vector Machine (A. Bazzani A. B., 2000) Combina dos métodos distintos, el primero para descubrir microcalcificaciones gruesas y el segundo de ellos es capaz de descubrir microcalcificaciones mas finas empleando técnicas de multiresolución mediante transformada wavelet. Mientras, (I. El-Naqa, A support vector machine approach for

detection of microcalcifications, 2002) propone un método para detectar microcalcificaciones y lo compara con otras técnicas empleadas por otros investigadores.

1.1.2 Heurística.

La evolución cronológica que ha llevado el trasplante renal en España. Apoyados en la heurística (Albacete, 2006).

Las personas ciegas, su cuerpo, el espacio y la representación mental, el planteo metodológico está centrado en la heurística, es decir, en el conocimiento, maduración y proyección de las fuerzas del individuo, que busca su genuina expresión, como miembro de una sociedad (Carla Beatriz, 2010)

Un enfoque heurístico para la programación de la producción en la industria de la fundición mediante lógica difusa, En este artículo se describe la concepción, el desarrollo y aplicación de una metodología para la programación de la producción en la industria de la fundición (Raul Landmann, 2011).

Diseño de una hiperheurística para la programación de la producción en ambientes job shop, El objetivo del trabajo es disminuir el tiempo de proceso (Makespan) e incrementar el tiempo de trabajo de las maquinas, disminuyendo el tiempo de ocio en ambientes de job shop, a través del diseño de una hiperheurística basada en colonia de hormigas y algoritmos genéticos (Omar Danilo Castrillón, 2010).

1.1.3 Mamografía.

En 1913 el cirujano alemán Albert Salomón, publicó la importancia de realizar radiografías a las piezas de mastectomía, para demostrar la extensión del tumor a los ganglios axilares, así como, la diferencia radiológica entre los márgenes del carcinoma infiltrante versus el circunscrito, describiendo por primera vez, los cambios radiográficos correspondientes a la presencia de microcalcificaciones en estos tumores (A.Salomon, 1913).

La siguiente publicación sobre radiología mamaria data de 1927 cuando otro cirujano, Otto Kleinschmidt, menciona por primera vez las indicaciones de la mamografía en un capítulo de un libro de texto (Keinssmidt O, 1927).

En 1930 Stafford L. Warren, un radiólogo del Rochester Memorial Hospital, Rochester, Nueva York, publicó la utilización de la técnica estereoscópica para las mamografías en vivo, usó película de grano fino, pantallas de refuerzo de grano fino, parrilla móvil para disminuir la radiación dispersa, 50-60 kV y 70 mA, con una distancia del tubo a la placa de 25 pulgadas y un tiempo de exposición de 2,5 segundos. (Warren SL, 1930).

Al año siguiente (1931), Walter Vogel en Leipzig (Vogel, 1932) y Paul Seabold en América (Seabold PS, 1931), publicaron de forma independiente sus investigaciones sobre mamografía, incluyendo la diferenciación entre enfermedad benigna y carcinoma.

En 1937 Hicken publicó la utilización del contraste en mamografía, realizando magnificas radiografías que mostraban los ductos lactofóricos normales y anómalos, los quistes, papilomas y carcinomas (Hicken NH Mammography, 1937). En ese mismo año Chester S. Carlson, desarrolló los principios básicos que permitían mediante la colaboración entre la industria y la medicina la creación de la Xeromamografía (Gold RH,

1992). Los principios físicos básicos de producción de imágenes electrostáticas ya eran conocidos en 1777 (Kossel F, 1967). Siendo en 1903 cuando Righi realizó su primera radiografía sobre una imagen electrostática (Righi A, 1907)

En 1938, Jacob Gershon-Cohen y Albert Strickler (Gershon-Cohen J, 1938), publican la variedad de imágenes radiográficas normales en función del momento del ciclo menstrual y de la edad de la paciente. Gershon-Cohen constató los hallazgos radiográficos con los macroscópicos y microscópicos, realizando la mayoría de sus trabajos en 1950 con la renombrada patóloga Helen Ingleby (Ingleby H, 1960).

En 1949 Raúl Leborgne en Uruguay, publicó la existencia de microcalcificaciones en el 30% de los carcinomas de sus series, describiéndolas como “finos granos de sal” y resaltando la necesidad de realizar las mamografías con una cuidada técnica y una adecuada colocación de la paciente (Leborgne R, 1951) (Leborgne R., 1953).

Durante la década de los 50, Gershon-Cohen et al., publicaron las anomalías radiográficas por ellos encontradas en la enfermedad benigna y maligna de la mama, permitiendo establecer un criterio diagnóstico para el cáncer de mama (GERS52, GERS56, GERS58). En Estrasburgo, en 1951, Gros y Sigrist confirmaron los criterios mamográficos usados para el diagnóstico de lesiones benignas y malignas, algunos años más tarde, crearon la primera unidad radiológica dedicada exclusivamente a la patología mamaria (GROS67).

Cuatro años más tarde, en 1955, John Roach y Herman Hilleboe realizan un estudio del potencial médico de la xeroradiografía general (ROAC55).

Siendo en 1960 cuando Howard Gould et al. (GOUL60) describen, con mayor detalle, las imágenes xeroradiográficas mamarias comparadas con las mamografías convencionales.

En 1960 Robert L. Egan describió la técnica mamográfica de alto mili amperaje-bajo kilo voltaje con placa industrial, que junto a la adecuada colocación de la paciente, dieron tan buenos resultados en sus primeras 1.000 pacientes, por lo que se renovó el interés por la mamografía convencional al desaparecer las barreras técnicas que dificultaban la difusión de la misma (EGAN60).

En 1963, el programa de control del cáncer del servicio de salud pública de Estados Unidos (Cáncer Control Program of the US Public Health Service), comunica la utilidad y reproducibilidad de la mamografía utilizando la técnica y colocación de las pacientes descritas por Egan, la mamografía puede ser utilizada para la detección de carcinoma en mujeres asintomáticas (CLAR65).

En 1965, el Colegio Americano de Radiología (American College of Radiology) (ACR), llegó a un acuerdo en cuanto a la técnica y dosis recomendadas, responsabilizándose de la enseñanza de radiólogos y técnicos en mamografía,

En 1963 Gerald Dodd, fue probablemente la primera persona en realizar una localización con aguja previa a la biopsia de una lesión no palpable y visible mamográficamente, publicando su método dos años después (DODD65).

En 1965 Charles Gros en colaboración con la "Compagnie Générale de Radiologie" (CGR), desarrollaron el Senógrafo, la primera unidad de Rayos x específica para mamografía.

En 1968 H. Stehen Gallager y Martin, publican los resultados de la correlación mamográfica e histológica del estudio de la totalidad de los cortes de las piezas de

mastectomía, viendo que $\frac{3}{4}$ de las 38 piezas tenían carcinoma difuso o múltiple y que más de los $\frac{3}{4}$ presentaba carcinoma intraductal o hiperplasia atípica (GALL69).

En 1971 hacen público su concepto de carcinoma "mínimo", que definían como, una lesión muy probablemente curable, consistente bien en un carcinoma lobular in situ, carcinoma intraductal no invasivo ó carcinoma invasivo (lobular o ductal) en una masa no mayor de 0,5 cm de diámetro (GALL71).

En la siguiente década (1970), en un esfuerzo por disminuir la radiación y el tiempo de exposición, J. L. Price y P.D. Butler experimentaron con pantallas intensificadoras de alta definición en íntimo contacto con las películas (PRICE70) Se dispuso del primer sistema comercial de xeromamografía al año siguiente (1971), Wolf et al. Publicaron los signos xeromamográficos del cáncer, (WOLF87) e investigaron la relación entre los patrones parenquimatosos mamarios y el cáncer (WOLF82).

En 1976, Howard Frank et al. Describieron como método de localización preoperatorio de lesiones no palpables encontradas en la mamografía, una aguja con una guía en forma de arpón en su interior [FRAN76]. Edward A. Sickles y cols, publican al año siguiente los resultados de su investigación sobre la magnificación en mamografía (SICK77), adquiriendo tanta importancia, que la mayoría de las unidades de mamografía llevan incorporadas un mecanismo de magnificación para poder estudiar mejor determinadas áreas de interés.

En 1980 Daniel Kopans y Salvatore de Luca (KOPA80), publican su experiencia utilizando como método de localización de lesiones no palpables y visibles en la mamografía.

En 1985, Marc Homer publica su trabajo (HOME85) con una guía curva retráctil, terminada en forma de "J" y realizada con un metal pseudoelástico que permitía reintroducir la guía una vez sacada para facilitar el reemplazamiento de la aguja. Mientras

que algunos hallazgos mamográficos son característicos de lesiones benignas o malignas, otros no nos permiten la diferenciación entre benignidad y malignidad, el diagnóstico citológico o histológico de estas lesiones solo puede alcanzarse mediante métodos intervencionistas mamarios (TARD02). Los más importantes son: la Punción Aspiración con Aguja Fina (PAAF), la Biopsia con Aguja Gruesa (BAG) y la colocación de guía para biopsia quirúrgica comentado anteriormente.

1.1.4 Minería de datos

La idea de la minería de datos viene desde los años 60, cuando los estadísticos de esa época manejaban términos como data fishing, data mining o data archeology, más tarde en los años 80, Rakesh Agrawal, Gio Wiederhold, Robert Blum y Gregory Piatetsky-Shapiro, entre otros empezaron a fortalecer los términos de data miningy KDD.

A finales delos años 80 solo existían un par de empresas quienes se dedicaban a esta tecnología; para el 2.002 este número se multiplicó considerablemente, ya que existían más de 100 empresas en el mundo con un portafolio de más de 300 soluciones que utilizaban la tecnología.

En la actualidad se ha incursionado con mayor fuerza en el desarrollo de aplicaciones que utilizan la minería de datos. Existen un conjunto de técnicas y herramientas capaces de ayudar a la toma de decisiones de los expertos. A pesar de ser relativamente joven, la minería de datos presenta aplicaciones en casi todos los sectores de la sociedad. En la salud, a nivel internacional se destaca la "Aplicación de técnicas de minería de datos para el diagnóstico prematuro del cáncer de mamas". Este sistema se encarga de realizar un diagnóstico del cáncer de mama a partir de una base de datos de imágenes de mamografías (Vallejo Delgado N, 2012).

En Cuba se han desarrollado investigaciones como por ejemplo "Aplicaciones de la minería de datos para el análisis de la Información Clínica". Este estudio se basa en el apoyo a la toma de decisiones a partir de coronariografías realizadas a pacientes que padecen cardiopatías isquémicas (Rosete Suárez A, 2009).

La UCI tampoco ha estado ajena al desarrollo de aplicaciones que emplean la minería de datos, y en ese sentido se destaca el "Diagnóstico de enfermedades de transmisión sexual mediante técnicas de inteligencia artificial (Bañobre Corpas Y).

1.2 Planteamiento del problema

La medicina es uno de los campos de la ciencia que resulta más beneficiada con la interacción de la informática y las matemáticas, habida cuenta que de esta manera pueden optimizarse problemas complejos; el aprendizaje automático y la minería de datos son una realidad en varias áreas del conocimiento como es el diagnóstico de enfermedades como el cáncer.

El diagnóstico clínico es un proceso cognitivo y complejo que precisa de: capacitación, experiencia, reconocimiento de patrones y cálculo de probabilidad condicional.

Muchos médicos han adquirido en el tiempo una intuición que facilita la tarea; la misma es producto de la experiencia de haber visto cientos de casos similares a lo largo de su carrera.

Los médicos menos experimentados pueden reducir sus errores al momento de diagnosticar a un paciente, realizando un proceso consciente y sistemático; tomando en cuenta la confiabilidad de sus apoyos diagnósticos y las fuentes más frecuentes de error.

Un médico puede sentirse fatigado o abordar de manera superficial a un paciente, puede no estar familiarizado con su enfermedad o estar influido por la opinión diagnóstica del colega que lo evaluó antes; y como consecuencia, no considerar un diagnóstico con un abordaje apropiado. Otro error común es la asociación de toda la sintomatología al diagnóstico inicial; utilizándolos solo para confirmar lo que se pensó en un inicio y descartando aquellos síntomas que no estén asociados.

En el cáncer de mama existen diagnósticos errados por factores humanos, debido a que la detección de la totalidad de los carcinomas visibles a través de análisis retrospectivos de las imágenes, muchas veces resulta complicado; las lesiones varían desde alteraciones en partes blandas de distintas formas y márgenes, hasta calcificaciones de diferente morfología, tamaño y distribución, que pueden ser representativas de malignidad; por lo que son necesarias múltiples lecturas de un mismo examen.

1.3 Formulación del problema de investigación.

¿De qué manera se podrá coadyuvar al diagnóstico de cáncer de mama, simplificando la selección de patrones obtenidos a través de la minería de datos?

1.4 Planteamiento de objetivo.

1.4.1 Objetivo general.

Elaborar un método heurístico que permita la selección de patrones obtenidos a través de la minería de datos para facilitar el diagnóstico oportuno del cáncer de mama.

1.4.2 Objetivos específicos.

- Dividir la mamografía y asociarlos en familias de patrones.
- Recoger toda la información de los patrones en una base de datos relacional.
- Evaluar el nivel de uniformidad en las familias de patrones, a través del cálculo de los índices de homogeneidad,
- Aplicar el algoritmo genético para maximizar el nivel de coincidencia.
- Generar las recomendaciones para la toma de decisiones, basadas en los resultados obtenidos a través de la minería de datos.

1.5 Planteamiento de Hipótesis.

La aplicación de un método heurístico para la selección de patrones obtenidos a través de la minería de datos facilitará el diagnóstico oportuno del cáncer de mama.

1.5.1 Variable dependiente.

Diagnóstico oportuno del cáncer de mama.

1.5.2 Variable Independiente.

Método heurístico para la selección de patrones obtenidos a través de la minería de datos.

CAPITULO II MARCO METODOLÓGICO

2.1 Diseño metodológico.

La investigación científica se concibe como un proceso dinámico compuesto por varias etapas que se encuentran relacionadas entre sí, la relación que exista entre las etapas identificadas puede ser secuencial o continua, dependiendo del tipo de investigación que se desea llevar adelante.

Para iniciar la investigación tal como se menciona (Hernandez, Fernández, & Baptista, 2006) es importante definir una idea y un planteamiento del problema que despierta el interés; en esta investigación se ha considerado de sumo interés aportar a la seguridad humana en el campo de la salud; siendo el “**cáncer de mama**” la segunda causa de la mortalidad en mujeres en Bolivia, es importante coadyuvar a su diagnóstico oportuno, reduciendo de esta manera los casos que derivan en el deceso de las pacientes, aportando así al fin supremo del estado que es el “vivir bien”

Considerando que el positivismo busca hechos de fenómenos sociales sin que prime para ello los estados subjetivos de los individuos, para llevar adelante esta investigación se adoptará el paradigma positivista.

El enfoque que se dará es cuantitativo por que se trabajará con datos estadísticos (patrones) obtenidos mediante la aplicación de la minería de datos y se buscará probar una hipótesis que plantea que mediante un método heurístico orientado a la selección de patrones obtenidos de mayor interés se podrá detectar y diagnosticar el cáncer de mama con mayor rapidez y precisión.

2.2 Clasificación de la Investigación.

Esta clasificación está hecha en base al texto (Raúl Mejía Ibáñez)

2.2.1 Según el Propósito o Razón.

Se desarrollará una investigación aplicada, porque se pretende desarrollar la investigación con fines prácticos y para su aplicación inmediata.

2.2.2 Según el nivel de Conocimientos.

En la fase la investigación será descriptiva y explicativa, por cuanto está dirigida a identificar la relación de causa-efecto entre las variables del problema.

2.2.3 Según la Estrategia Empleada.

En la primera fase será documental, a partir de un análisis bibliográfico de datos contenidos en fuentes de consulta.

Posteriormente en una segunda fase se hará una investigación de campo para obtener información de manera directa.

2.2.4 Según la Ubicación de los Datos en el Tiempo.

En la primera etapa será retrospectiva, porque se realizará una revisión de antecedentes.

En la segunda fase será prospectiva, porque en base a lo investigado se pretende diseñar un método heurístico que trabaje sobre los patrones obtenidos en la minería de datos, analizando e interpretando los mismos para luego hacer una selección de manera que se pueda coadyuvar al diagnóstico oportuno del cáncer de mama.

2.3 Método de investigación

2.3.1 Fases Metodológicas

El enfoque propuesto, para la elaboración del método heurístico que permita la selección de patrones, se basara en la optimización de selección de índices de homogeneidad, utilizando datos simples como entradas, conformación de familia de patrones mediante índices de homogeneidad que serán utilizados para evaluar el carácter en común de todo el conjunto de patrones, y la heurística se implementara para maximizar el valor de estos índices en común, de este modo se optimizara la selección de familia de patrones.

Paso 1: Familia de patrones

Seleccionar y clasificar los diferentes patrones de interés en un sub-grupos conformando familias de patrones, para obtener los datos para su posterior evaluación del carácter en común de los mismos. Cada patrón dentro de estas familias se clasificara hasta el nivel más bajo.

Paso 2: Recolección de datos

Almacenar los datos básicos necesarios para la conformación de cada familia.

Paso 3: Evaluación

Medir la similitud dentro de una familia de patrones, encontrando varios índices de homogeneidad para su posterior evaluación del grado de coincidencia dentro de la familia de patrones, basándose en diferentes parámetros; tales como, el número común de componentes. Estos índices serán el punto de partida para la conformación de un nuevo sub-grupo de patrones o el análisis de un sub-grupo existente.

2.4 Técnicas de investigación

Para nutrir la investigación se utilizará:

- La investigación documental contenida en la bibliografía, informes y otros documentos relacionados al tema.
- Investigación de campo para recolectar información de expertos en el manejo y selección de patrones obtenidos.

2.5 Universo o Población de referencia

La unidad de análisis para llevar adelante el presente trabajo de investigación contemplará las siguientes Instituciones Públicas:

- Unidades de Oncología de nosocomios públicos y privados

Hospital general de Clínicas, Unidad de Imagenología, La Paz – Bolivia

- Médicos de la Especialidad de Oncología

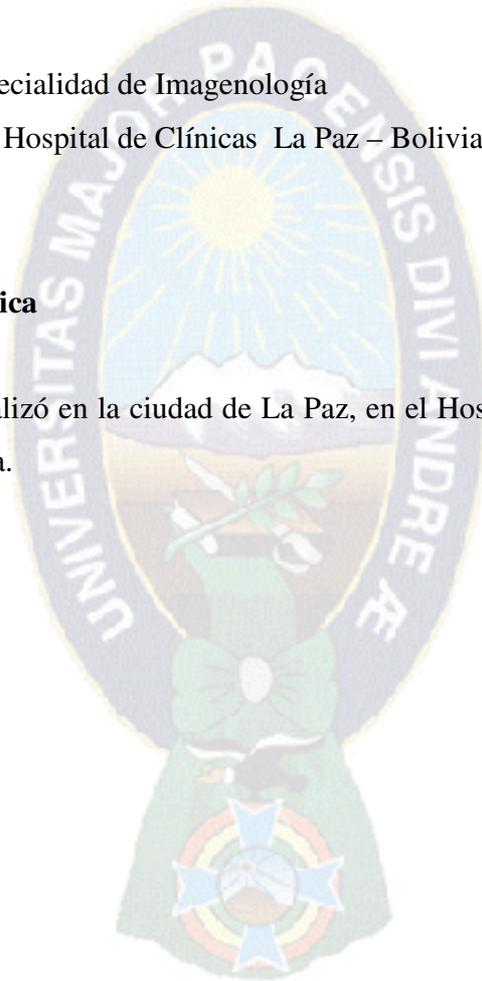
Dra., Martha Lucia Aguirre de Delgado, Jefa de la Unidad de Imagenología
Del Hospital de Clínicas La Paz – Bolivia

- Médicos de la Especialidad de Imagenología

Dra. Paola Castro. Hospital de Clínicas La Paz – Bolivia

2.6 Delimitación geográfica

La investigación se realizó en la ciudad de La Paz, en el Hospital de Clínicas el área de Oncología e Imagenología.



Capítulo III Marco Teórico

3.1 Referencia Teórica.

3.1.1 Cáncer de Mama.

Es una enfermedad por la que se forman células malignas (cancerosas) en los tejidos de la mama, la mama está compuesta por lóbulos y conductos. Cada mama tiene entre 15 y 20 secciones que se llaman lobulillos. Los lobulillos terminan en docenas de bulbos minúsculos que pueden producir leche. Los lóbulos, los lobulillos y los bulbos están conectados por tubos delgados que se llaman conductos.

Cada mama tiene también vasos sanguíneos y vasos linfáticos. Los vasos linfáticos transportan un líquido casi incoloro que se llama linfa. Los vasos linfáticos van a órganos que se llaman ganglios linfáticos. Estos son estructuras pequeñas con forma de frijol que se encuentran en todo el cuerpo. Filtran sustancias de un líquido que se llama linfa y ayudan a combatir infecciones y enfermedades. Hay racimos de ganglios linfáticos cerca de la mama en las axilas (debajo de los brazos), por encima de la clavícula y en el pecho.

El tipo más común de cáncer de mama es el carcinoma ductal, que empieza en las células de los conductos. El cáncer que empieza en los lóbulos o los lobulillos se llama carcinoma lobulillar y se encuentra con mayor frecuencia en ambas mamas que otros tipos de cáncer de mama. El cáncer de mama inflamatorio es un tipo de cáncer poco común en el que la mama está caliente, enrojecida e hinchada.

3.1.2 Identificación de tonos claro, oscuro y gris

Según (García, 2012) Todas las imágenes radiológicas se visualizaban en un soporte plástico, la placa. En algunos lugares todavía se utilizan, pero son cada vez menos frecuentes.

Las imágenes se obtenían a través de una combinación de rayos X y de luz que incidían sobre una película fotográfica, lo que, a su vez, generaba una imagen latente que después se procesaba en un cuarto oscuro mediante la aplicación de productos químicos y después se colgaba, literalmente, hasta su secado.

Cuando se solicitaba un estudio de carácter inmediato, las placas se interpretaban mientras todavía estaban mojadas por los productos químicos; de ahí nació el concepto de **lectura húmeda** para la interpretación «inmediata».

Las placas se visualizaban sobre cajas con iluminación denominadas negatoscopios (casi siempre mal colocadas cuando aparecen en escenas de alguna película o serie de televisión).

Este método de trabajo se mantuvo durante decenios, pero tenía dos desventajas principales:

- Requería una gran cantidad de espacio para almacenar cada vez un mayor número de placas.
- Las placas en sí mismas solamente podían estar en un lugar cada vez, y ese lugar no era necesariamente el requerido para atender al paciente.

Entonces apareció la **radiografía digital**, en la que la placa fotográfica fue sustituida por un **chasis o lámina fotosensible** que podía ser procesada por un dispositivo lector electrónico y esto permitía el almacenamiento digital de la imagen.

El disco duro de un ordenador utilizado como servidor podía almacenar incontables imágenes.

Todavía más importante que lo anterior, cualquier persona que tuviera permiso de acceso, en cualquier parte del mundo y en cualquier momento podía visualizar las imágenes.

Los estudios realizados se mantenían en los servidores en los que era posible archivar, transmitir y almacenar las imágenes. Este sistema se denominó **PACS (Picture Archiving, Communications, and Storage)**, sistema de archivo, comunicación y almacenamiento de imágenes).

Con los sistemas PACS es posible almacenar y recuperar todo tipo de imágenes, incluyendo las **radiografías convencionales (RC)**, las **imágenes de tomografía computarizada (TC)**, imágenes de ecografía (ultrasonidos [US]) y los estudios de resonancia magnética (RM).

3.1.3 Radiografía convencional (placa simple)

Las imágenes producidas mediante el uso de radiaciones ionizantes (es decir, rayos X) pero sin añadir material de contraste como bario o yodo se denominan radiografías convencionales o bien, con mayor frecuencia, «radiografías o placas simples».

La producción de estas imágenes es relativamente barata y se puede llevar a cabo casi en cualquier lugar mediante aparatos portátiles o móviles; en la actualidad sigue siendo el estudio de imagen que se realiza con mayor frecuencia.

Requieren una fuente de rayos X (el «aparato de rayos X»), un método para obtener la imagen (una placa, un chasis o una lámina) y un sistema para procesar la imagen obtenida (los productos químicos o el dispositivo lector digital).

Las indicaciones más habituales de la radiografía convencional son la radiografía torácica, la radiografía simple de abdomen y virtualmente cualquier imagen inicial del sistema esquelético para descubrir la presencia de fracturas o de problemas articulares.

Sabemos que la aplicación de dosis elevadas de radiaciones ionizantes (dosis sustancialmente mayores que las utilizadas en cualquier procedimiento radiográfico médico) causa mutaciones celulares que pueden dar lugar a distintas formas de cáncer y malformaciones. Las estimaciones epidemiológicas relativas a los niveles bajos de radiación varían en cuanto a su valoración del riesgo, pero suele aceptarse que solamente hay que llevar a cabo aquellos estudios radiológicos diagnósticos necesarios y que los estudios en los que se utilizan rayos X deben evitarse en las situaciones potencialmente teratógenos, como el embarazo (García, 2012).

3.1.4 Tomografía Computarizada (TC)

Fue introducida inicialmente en la década de 1970, representó una vuelta de tuerca en lo relativo a los estudios de imagen médica.

Mediante el uso de un armazón o caballete (gantry) en el que se colocan un haz rotatorio de rayos X y múltiples detectores en diferentes configuraciones (los cuales también rotan de

manera continua alrededor del paciente), junto con complicados algoritmos informáticos para el procesamiento de los datos, es posible formatear un elevado número de imágenes bidimensionales en forma de cortes y en múltiples planos.

Después de su obtención, las imágenes de TC también pueden ser «procesadas en diferentes ventanas» para optimizar la visualización de distintos tipos de procesos patológicos, una utilidad que se denomina post-proceso y que, en términos generales, ha hecho avanzar todo lo relativo a la imagen digital. El post-proceso permite la manipulación adicional de los «datos en bruto» (raw data) para una visualización más detallada de las alteraciones observadas, sin necesidad de repetir el estudio y así evitar la re exposición del paciente.

La obtención de las imágenes de TC requiere el uso de un escáner de precio muy elevado, un espacio considerable para su instalación y una capacidad notable para un complejo procesamiento informático.

A pesar de ello, la TC es la técnica de imagen clave de los estudios de imagen seccional y está disponible en numerosos centros asistenciales a pesar de que todavía no se ha conseguido que sea realmente un método portátil. Los escáneres de TC son más caros y complejos que las unidades radiográficas convencionales y, al igual que estas últimas, todavía utilizan radiaciones ionizantes (rayos X) para la producción de las imágenes.

3.1.5 Ecografía

Para producir las imágenes, la ecografía utiliza energía acústica con una frecuencia superior a la audible por el ser humano, en lugar de rayos X como la radiografía convencional y la TC.

Utiliza un transductor que genera los US y registra las señales ecográficas. Un ordenador incorporado en el propio equipo procesa la señal en función de sus características. Las imágenes ecográficas se generan digitalmente y pueden ser almacenadas con facilidad en sistemas PACS.

Los ecógrafos son relativamente baratos comparados con los aparatos de TC y RM. Se encuentran en la mayor parte de los centros asistenciales y pueden ser fácilmente transportables, incluso a mano.

Dado que en la ecografía no se utiliza radiación ionizante, este método es especialmente útil para la realización de estudios de imagen en mujeres en edad fértil, en mujeres embarazadas y en niños.

Asimismo, la ecografía es muy oportuna para la visualización de los tejidos blandos y para la diferenciación entre las estructuras sólidas y quísticas. También se utiliza de forma genérica para obtener biopsias guiadas mediante estudios de imagen y es un método no invasivo para el estudio del flujo sanguíneo.

La ecografía se considera, en términos generales, una modalidad de imagen muy segura que no produce efectos adversos importantes cuando se emplea para el diagnóstico médico (García, 2012).

3.1.6 Mamografía

Es el diagnóstico de imagen por rayos-X de la glándula mamaria, mediante el mamógrafo que disponen de tubos de emisión de rayos X especialmente adaptados para conseguir la mayor resolución posible en la visualización de las estructuras internas de la glándula mamaria.

El año 1913, el médico alemán Albert Salomón, hizo una publicación en la que especificaba la importancia de realizar radiografías a las mamas para verificar la extensión del tumor de los ganglios axilares (Amparo Vilarrasa Andrés, 2006).

Una mamografía periódica ayuda a disminuir la mortalidad del cáncer de mama, ya que se puede realizar una detección precoz. La prevención y la periodicidad de los exámenes son esenciales para el tratamiento y erradicación de la enfermedad. Estos pueden detectar problemas mucho antes de que se hagan palpables.

Una mamografía no es dolorosa. La prueba trata generalmente de cuatro radiografías (dos en cada mama) para captar la mayor parte del tejido. El estudio consistirá en situar el seno sobre la placa del mamógrafo; posteriormente, se aplastará con otra placa, esto se realiza para igualar densidades y para que el tejido sea más homogéneo para el haz de RX. La radiación utilizada es mínima. Este momento suele ser el más incómodo, pero no se demora mucho. Se recomienda evitar hacérsela durante la menstruación o una semana previa, ya que las mamas estarán más sensibles.

3.1.7 Resonancia Magnética (RM)

La RM utiliza la energía potencial almacenada en los átomos de hidrógeno del organismo. Los átomos son manipulados por campos magnéticos muy intensos y por pulsos de radiofrecuencia que producen los niveles necesarios de energía de localización y de energía con especificidad tisular para que sofisticados programas informáticos generen imágenes bi-dimensionales o tridimensionales.

Los escáneres de RM no son tan habituales como los de TC; son caros y es preciso ubicarlos en áreas arquitectónicamente específicas para su funcionamiento adecuado. En general, también conllevan costes de mantenimiento elevados.

No obstante, la RM no utiliza radiación ionizante y permite obtener un contraste mucho mayor entre los distintos tipos de tejidos blandos, en comparación con la TC.

La RM se usa con mucha frecuencia en el ámbito de la neurología y es especialmente útil para la visualización de los tejidos blandos como los músculos, los tendones y los ligamentos.

Existen problemas de seguridad relacionados con los campos magnéticos extremadamente intensos de los escáneres de RM, tanto en lo que se refiere a los objetos que puede haber en el cuerpo (p. ej., marcapasos) como en lo relativo a los proyectiles ferromagnéticos ubicados en el entorno del escáner de RM (p. ej., las bombonas metálicas de oxígeno). También son conocidos los efectos adversos de las ondas de radiofrecuencia que producen estos escáneres y siempre hay que tener en cuenta los posibles efectos adversos asociados a algunos de los agentes de contraste que se utilizan en los estudios de RM (García, 2012).

3.1.8 Las cinco densidades básicas

La radiografía convencional se limita a la demostración de cinco densidades básicas indicadas a continuación desde la menos densa hasta la más densa (tabla 1):

- Aire, que aparece con el color más oscuro en la radiografía.
- Grasa, que aparece con un tono de gris ligeramente más claro que el del aire.
- Tejidos blandos o líquidos (muestran una densidad similar en las radiografías convencionales, de manera que, por ejemplo, en una radiografía convencional no es posible diferenciar el miocardio de la sangre en el interior del corazón).
- Calcio (generalmente contenido en los huesos).
- Metal, que aparece con el tono más blanco en una radiografía.

- En el cuerpo no suele haber estructuras con densidad metálica. Los medios de contraste radiológicos y las prótesis de rodilla o cadera son ejemplos de estructuras con densidad metálica colocadas artificialmente en el cuerpo.

Una de las utilidades principales de la TC es su capacidad para expandir la escala de grises, lo que nos permite diferenciar muchas más de estas cinco densidades básicas.

El lector debe recordar que, cuanto mayor es la densidad de un objeto, más rayos X absorbe y más blanco aparece en las imágenes radiográficas.

Cuanto menos denso es un objeto, menos rayos X absorbe y más oscuro aparece en las radiografías.

Por desgracia, los términos específicos utilizados para describir lo que aparece como blanco en una imagen y lo que aparece como negro son diferentes en las distintas modalidades de imagen.

La tabla 2 es una guía sencilla que explica los términos utilizados para describir lo blanco y lo negro en las distintas modalidades de imagen (García, 2012).

Densidad	Aspecto
Aire	Absorbe la menor cantidad de rayos X y aparece con el color «más oscuro» en las radiografías convencionales
Grasa	Aparece con un color gris, algo más oscuro que los tejidos blandos
Líquido o tejidos blandos	El líquido (p. ej., la sangre) y los tejidos blandos (p. ej., el músculo) tienen la misma densidad en las radiografías convencionales
Calcio	Es la estructura más densa y el material natural (p. ej., los huesos) que absorbe la mayor cantidad de rayos X
Metal	Generalmente absorbe todos los rayos X y aparece con el color «más blanco» (p. ej., balas, bario)

Tabla 1. Las cinco densidades básicas que se observan en la radiografía convencional

Modalidad	Términos utilizados para «blanco»	
Radiografías convencionales	Aumento de la densidad; opaco	Disminución de la densidad; translúcido
TC	Atenuación aumentada (elevada); hiperdenso	Atenuación disminuida; hipodenso
RM	Intensidad de señal aumentada (elevada); brillante	Intensidad de señal disminuida (baja); oscuro
Ecografía	Ecogenicidad aumentada; ecodenso	Ecogenicidad disminuida; ecolúcido
Medicina nuclear	Aumento de la captación del marcador	Disminución de la captación del marcador
Estudios con bario	Radiopaco	No opaco; radiolúcido

Tabla 2 Blanco y negro: términos utilizados en cada modalidad

3.1.9 Minería de datos.

Es la extracción de conocimiento en bases de datos (Knowledge Discovery in Databases o KDD), mediante un proceso no trivial de identificación válida, novedosa, potencialmente útil y entendible de patrones comprensibles que se encuentran ocultos en los datos (Fayyad, Piatetsky-Shapiro, Smyth, & Uthurusamy, 1996), para reunir las ventajas de varias áreas para el procesamiento masivo, como la Estadística, la Inteligencia Artificial, la Computación Gráfica, principalmente usando como materia prima las bases de datos, también se define como la integración de un conjunto de áreas que tienen como propósito la identificación de un conocimiento obtenido a partir de las bases de datos que aporten un sesgo hacia la toma de decisión (MOLINA, 1998).

Se compone de cuatro etapas principales:

- Determinación de los objetivos. Trata de la delimitación de los objetivos que el *cliente* desea bajo la orientación del especialista en *data mining*.
- Pre procesamiento de los datos. Se refiere a la selección, la limpieza, el enriquecimiento, la reducción y la transformación de las bases de datos. Esta etapa consume generalmente alrededor del setenta por ciento del tiempo total de un proyecto de *data mining*.
- Determinación del modelo. Se comienza realizando unos análisis estadísticos de los datos, y después se lleva a cabo una visualización gráfica de los mismos para tener una primera aproximación. Según los objetivos planteados y la tarea que debe llevarse a cabo, pueden utilizarse algoritmos desarrollados en diferentes áreas de la Inteligencia Artificial.

- Análisis de los resultados. Verifica si los resultados obtenidos son coherentes y los coteja con los obtenidos por los análisis estadísticos y de visualización gráfica. El *cliente* determina si son novedosos y si le aportan un nuevo conocimiento que le permita considerar sus decisiones.

El término Minería de Datos Inteligente (Evangelos & Han & Michalski et al.) Refiere específicamente a la aplicación de métodos de aprendizaje automático (Michalski), para descubrir y enumerar patrones presentes en los datos, para estos, se desarrollaron un gran número de métodos de análisis de datos basados en la estadística (Michalski et al). En la medida en que se incrementaba la cantidad de información almacenada en las bases de datos, estos métodos empezaron a enfrentar problemas de eficiencia y escalabilidad y es aquí donde aparece el concepto de minería de datos. Una de las diferencias entre el análisis de datos tradicional y la minería de datos es que el primero supone que las hipótesis ya están construidas y validadas contra los datos, mientras que el segundo supone que los patrones e hipótesis son automáticamente extraídos de los datos (Hernández Orallo).

(Morales, 2003) Dentro la minería de datos existe un proceso completo de descubrimiento de conocimiento que involucra varios pasos:

- Entendimiento del dominio de aplicación, el conocimiento relevante a utilizar y las metas del usuario.
- Seleccionar un conjunto de datos en donde realizar el proceso de descubrimiento.
- Limpieza y pre procesamiento de los datos, diseñando una estrategia adecuada para manejar ruido, valores incompletos, valores fuera de rango, valores inconsistentes.
- Selección de la tarea de descubrimiento a realizar, por ejemplo, clasificación, agrupamiento o clustering, reglas de asociación.
- Selección de los algoritmos a utilizar.

- Transformación de los datos al formato requerido por el algoritmo específico de explotación de datos, hallando los atributos útiles, reduciendo las dimensiones de los datos.

3.1.9.1 Agrupación de datos.

La agrupación o *clustering* consiste en agrupar un conjunto de datos basándose en la similitud de los valores de sus atributos. El *clustering* identifica regiones densamente pobladas, denominadas clusters, de acuerdo a alguna medida de distancia establecida (Chen, 1996). De esta manera se busca maximizar la similitud de las instancias en cada cluster y minimizar la similitud entre clusters (Kamber, 2006). La técnica de *clustering* ha sido estudiada en las áreas de la estadística (Cheeseman, 1996), (Avrim L. Blum, 1996), base de datos espaciales y minería de datos (Cheeseman & Stutz, 1996; Ester *et al.*, 1995; Ng & Han, 1994; Zhang *et al.*, 1996). Dos de los algoritmos de clustering más utilizados son Self Organizing Maps (SOM) y *K-means*. *SOM*, también denominado redes de Kohonen, fue creado por Teuvo Kohonen en 1982. Se trata de un modelo de red neuronal con capacidad para formar mapas de características de manera similar a como ocurre en el cerebro. *SOM* está basado en el aprendizaje no supervisado y competitivo, lo cual quiere decir que no se necesita intervención humana durante el mismo y que se necesita saber muy poco sobre las características de la información de entrada. *SOM* provee un mapa topológico de datos, que se representan en varias dimensiones, utilizando unidades de mapa (las neuronas) para simplificar la representación [Kohonen, 1995]. Las neuronas usualmente forman un mapa bidimensional, por lo que el mapeo transforma un problema de muchas dimensiones en el espacio, a un plano. La propiedad de preservar la topología significa que el mapeo preserva las distancias relativas entre puntos. Los puntos que están cerca unos de los otros en el espacio original de entrada son mapeados a neuronas cercanas en *SOM*. Por esta razón, *SOM* es muy útil como herramienta de análisis de clases de datos de muchas dimensiones (Vesanto & Alhoniemi, 2000), y además tiene la capacidad de

generalizar (Essenreiter *et al.*, 1999), lo que implica que la red puede reconocer o caracterizar entradas que nunca antes ha encontrado.

K-means es un método iterativo que busca formar k clusters, con k predeterminado antes del inicio del proceso. *K-means* comienza particionando los datos en k subconjuntos no vacíos, calcula el centroide de cada partición como el punto medio del cluster y asigna cada dato al cluster cuyo centroide sea el más próximo. Luego vuelve a particionar los datos iterativamente, hasta que no haya más datos que cambien de cluster de una iteración a la otra. *K-means* se explica en mayor detalle en el capítulo 4. Otros algoritmos de *clustering* son *K-medoids* o *PAM* (*Partition around medoids*) y *CLARA* (*Clustering LARge Applications*) (Kaufman & Rousseeuw, 1990). Este último permite manejar conjuntos de datos más grandes que el primero. *CLARANS* (Ng & Han, 1994) integra los algoritmos *PAM* y *CLARA* en uno.

3.1.9.2 Clasificación de datos.

La clasificación se utiliza para clasificar un conjunto de datos basado en los valores de sus atributos. Por ejemplo, se podría clasificar a distintas personas para la otorgación de un préstamo en riesgo bajo, medio y alto, teniendo en cuenta información histórica de las mismas.

La clasificación encuentra las propiedades comunes entre un conjunto de objetos y los clasifica en diferentes clases, de acuerdo a un modelo de clasificación. Para construir este modelo, se utiliza un conjunto de entrenamiento, en el que cada instancia consiste en un conjunto de atributos y el valor de la clase a la cual pertenece. El objetivo de la clasificación es analizar los datos de entrenamiento y, mediante un método supervisado, desarrollar una descripción o un modelo para cada clase utilizando las características disponibles en los datos. Esta descripción o modelo permite clasificar otras instancias, cuya

clase es desconocida. El método se conoce como supervisado debido a que, para el conjunto de entrenamiento, se conoce la clase de pertenencia y se le indica al modelo si la clasificación que realiza es correcta o no. La construcción del modelo se realimenta de estas indicaciones del supervisor (Chen, 1996).

Los algoritmos mayormente utilizados para las tareas de clasificación son los algoritmos de inducción. En la actualidad existen numerosos enfoques de algoritmos de inducción y variedad en cada enfoque, el presente trabajo hará hincapié en aquellos orientados a generar árboles de decisión.

La clasificación basada en árboles de decisión es un método de aprendizaje supervisado que construye árboles de decisión a partir de un conjunto de entrenamiento.

Un sistema típico de construcción de árboles de decisión es *ID3*, que utiliza la teoría de la información para minimizar la cantidad de pruebas para clasificar un objeto. Al utilizar métodos heurísticos, *ID3* garantiza un árbol simple, pero no necesariamente el más simple. Una extensión de *ID3* es *C4.5* (Quinlan, 1993), que extiende el dominio de clasificación de atributos categóricos a numéricos. Un paso importante en la construcción del árbol de decisión es la poda, la cual elimina las ramas no necesarias, resultando en una clasificación más rápida y una mejora en la precisión de la clasificación de datos (Han, 2001).

Existen muchos otros algoritmos de clasificación de datos, incluyendo métodos estadísticos, como el análisis de regresión lineal (Elder IV, 1996); algoritmos de *machine learning* (Cheeseman P., 1996); redes neuronales (Lu H., 1995), algoritmos genéticos y lógica difusa.

3.1.9.3 Reglas de asociación.

La minería de reglas de asociación consiste en encontrar reglas de la forma $(A_1 \text{ y } A_2 \text{ y...y } A_m) \Rightarrow (B_1 \text{ y } B_2 \text{ y...y } B_n)$, donde A_i y B_j son valores de atributos del conjunto de datos. Por ejemplo, se podría encontrar en un gran repositorio de datos de compras en un supermercado, la regla de asociación correspondiente a que si un cliente compra leche, entonces compra pan. Una regla de asociación es una sentencia probabilística acerca de la concurrencia de ciertos eventos en una base de datos, y es particularmente aplicable a grandes conjuntos de datos (Hand, 1997).

Existen varios algoritmos que realizan el descubrimiento de reglas de asociación, uno de los más utilizados es *Apriori*.

3.1.9.4 Aplicación de la Minería de datos.

En Internet

- E-bussines: Perfiles de clientes, publicidad dirigida, fraude.
- Buscadores Inteligentes: Generación de jerarquías, bases de conocimiento web.
- Gestión del Tráfico de la Red: Control de e>ciencia y errores.

El Mundo de los Negocios:

- Banca: Grupos de clientes, préstamos, oferta de productos.
- Compañías de Seguros: Detección de fraude, administración de recursos.
- Marketing: Publicidad dirigida, estudios de competencia.

En el mundo de las ciencias:

- Meteorología: Tele conexiones (asociaciones espaciales), predicción.
- Física: Altas energías, datos de colisiones de partículas (búsqueda de patrones).
- Bio-Informática: Búsqueda de patrones en ADN, proyectos científicos como genoma humano, datos geofísicos, altas energías, etc.

Minería de Texto

Examina una colección de documentos y descubre información no contenida en ningún documento individual de la colección; en otras palabras, trata de obtener información sin haber partido de algo (Nasukawa & Nagano, 2001).

3.1.10 Heurística.

Un método heurístico es un procedimiento para resolver un problema de optimización bien definido mediante una aproximación intuitiva, en la que la estructura del problema se utiliza de forma inteligente para obtener una buena solución, los procedimientos heurísticos son acciones que comportan un cierto grado de variabilidad y su ejecución no garantiza la consecución de un resultado óptimo como, por ejemplo, reducir el espacio de un problema complejo a la identificación de sus principales elementos (Monero ,1995). Mientras que Duhalde y González (1997) señalan que un heurístico es “un procedimiento que ofrece la posibilidad de seleccionar estrategias que nos acercan a una solución”.

Los métodos heurísticos específicos están relacionados con el conocimiento de un área en particular. Este incluye estructuras cognoscitivas más amplias para reconocer los

problemas, algoritmos más complejos y una gran variedad de procesos heurísticos específicos.

El conocimiento que tienen los expertos solucionadores de problemas están los “esquemas de problemas” (Chi 1981, 1982). Estos consisten en conocimiento estrechamente relacionado con un tipo de problema en particular y que contiene:

- Conocimiento declarativo: principios, fórmulas y conceptos.
- Conocimiento procedimental: conocimiento acerca de las acciones necesarias para resolver un tipo de problema en particular.

Conocimiento estratégico: conocimiento que permite, al individuo solucionador del problema, decidir sobre las etapas o fases que debe seguir en el proceso de solución.

La resolución de problemas está relacionada con el conocimiento específico del área en cuestión (Mayer, 1992; Stenberg, 1987). En este sentido, estos autores coinciden en señalar que los tipos de conocimiento necesarios para resolver problemas incluyen:

- Conocimiento declarativo: por ejemplo, saber que un kilómetro tiene mil metros.
- Conocimiento lingüístico: conocimiento de palabras, frases, oraciones.
- Conocimiento semántico: dominio del área relevante al problema, por ejemplo, saber que si Álvaro tiene 5 bolívares más que Javier, ésto implica que Javier tiene menos bolívares que Álvaro.

- Conocimiento esquemático: conocimiento de los tipos de problema.
- Conocimiento procedimental: conocimiento del o de los algoritmos necesarios para resolver el problema.
- Conocimiento estratégico: conocimiento de los tipos de conocimiento y de los procedimientos heurísticos.

Para solucionar un problema hay que hacer algo que ataque las causas del problema, que vaya al fondo del asunto, de esta manera para resolver un problema hay que seguir un método al que podemos denominar método heurístico, este consiste en cuatro pasos que se pueden seguir para resolver muchos tipos de problemas.

- Analizar y definir el problema.
- Definir la estrategia a seguir para llegar a la solución y llevarla a la práctica.
- Definir alternativas de solución al problema y seleccionar la mejor.
- Comprobar la pertinencia de la solución seleccionada

El primer paso consiste en definir exactamente donde está el problema, para lo cual hay que analizarlo con detenimiento, en este sentido se encuentran dos tipos de problemas, los problemas que ya se presentan completa y correctamente planteados, como algunos de razonamiento que resolvemos en las clases de física y matemáticas, el planteamiento de un problema de razonamiento debe llenar las siguientes características:

- Estar expresados con claridad, de manera que se entienda lo que se está diciendo.
- Incluir todos los datos necesarios para su resolución.

- Expresar claramente las condiciones en las cuales se da la situación o el hecho que se va a resolver.
- Expresar claramente las reglas con las cuales se debe buscar la solución.
- Terminar con una pregunta en la que se indique con claridad el tipo de solución que se espera.

3.1.11 Estudios realizados por otros investigadores

En los Trabajos (R.N. Strickland, 1994), (G. te Brake, 1997), se detecta las microcalcificaciones usando la transformada wavelet en combinación con un criterio donde se analiza la forma, el tamaño y el número de microcalcificaciones por unidad de área.

Se aplicaron técnicas de ecualización para ajustar la distribución de los datos (W.J.H. Veldkamp, 2000), de anomalías combinando desde la detección de contornos hasta la extracción de características por técnicas multiresolución.

Se realizaron, tratamiento de la mamografía digitalizada, como el filtrado para eliminar el fondo y aumentando la relación señal a ruido que permita el realzado de la región de interés, para después identificar las posibles anomalías realizadas a través técnicas de umbralización local y global de los niveles de gris (R.M. Nishikawa Y. J., 1992).

Trabajos basados en redes neuronales, algoritmos genéticos, funciones de base radial se propone un sistema basado en redes neuronales consiguiendo detectar un 87 % de verdaderos positivos (M.A. Kupinski, 1997).

Trabajos con técnicas de multiresolución, propone la detección de microcalcificaciones usando wavelets (T.C. Wang, 1998).

Descubrimientos de microcalcificaciones gruesas y el segundo de ellos es capaz de descubrir microcalcificaciones más finas empleando técnicas de multiresolución mediante transformada wavelet (A. Bazzaní A. B., 2000).

Se realizó la clasificación empleando un clasificador SVM (I. El-Naqa, A support vector machine approach for detection of microcalcifications, 2002).

Sistema inteligente para la detección y diagnóstico de patología mamaria (Amparo Vilarrasa Andrés, 2006).

3.2 Marco Legal o institucional.

3.2.1 Constitución Política del Estado.

Sección II "Derecho de la salud y a la Seguridad Social" se tomó en cuenta los siguientes artículos: Art. N° 35. Inc. I y II, Art. N° 36 Inc. I y II, Art. N° 37, referente a que "El estado tiene la obligación indeclinable de garantizar y sostener el derecho a la salud, que se constituye en una función suprema y primera responsabilidad financiera. Se priorizará la promoción de la salud y la prevención de las enfermedades"

En el capítulo segundo principios, valores y fines del estado, inciso 5. Garantizar el acceso de las personas a la educación, a la salud y al trabajo.

En el capítulo segundo derechos fundamentales Artículo 18.

I. Todas las personas tienen derecho a la salud.

II. El Estado garantiza la inclusión y el acceso a la salud de todas las personas, sin exclusión ni discriminación alguna.

III. El sistema único de salud será universal, gratuito, equitativo, intracultural, intercultural, participativo, con calidad, calidez y control social. El sistema se basa en los principios de solidaridad, eficiencia y corresponsabilidad y se desarrolla mediante políticas públicas en todos los niveles de gobierno

En el capítulo cuarto derechos de las naciones y pueblos indígena originario campesinos, Artículo 30 inciso 13. Al sistema de salud universal y gratuito que respete su cosmovisión y prácticas tradicionales.



Capítulo III Marco ingeniera de prototipo

3.1 Análisis y Diseño del prototipo.

3.1.1 Método Heurístico

En este trabajo, se introdujo el enfoque basado en el algoritmo de optimización heurística con índices de homogeneidad para el diagnóstico de cáncer de mama.

El enfoque propuesto empleó la minería de datos como una entrada, obtenidos a partir de la disección de la mamografía dentro de una familia de patrones almacenados dentro de una base de datos relacional, estos datos se utilizaron para evaluar el carácter común de toda la familia mediante los índices de homogeneidad, y el algoritmo genético adaptado como algoritmo heurístico se implementó para maximizar el valor de estos índices en común.

3.1.2 Una visión general del método propuesto

Los detalles del método fueron los siguientes:

El primer paso fue el procesamiento de imágenes mamográficas y su almacenamiento en una base de datos relacional para obtener los datos necesarios.

El segundo paso fue la aplicación la minería de datos para la obtención de patrones y diseccionar los mismos para conformar las familias de patrones. Los datos básicos y la información se recogieron y se almacenaron en otra base de datos relacional.

En el tercer paso, se realizó una evaluación del nivel de uniformidad en los patrones, mediante el cálculo de los índices de homogeneidad, para la conformación de familias de patrones.

El cuarto paso fue el uso de la heurística para maximizar el nivel de coincidencia en la familia con caracteres específicos.

El paso final fue la generación de recomendaciones, basadas en los resultados experimentales, ver figura 1.

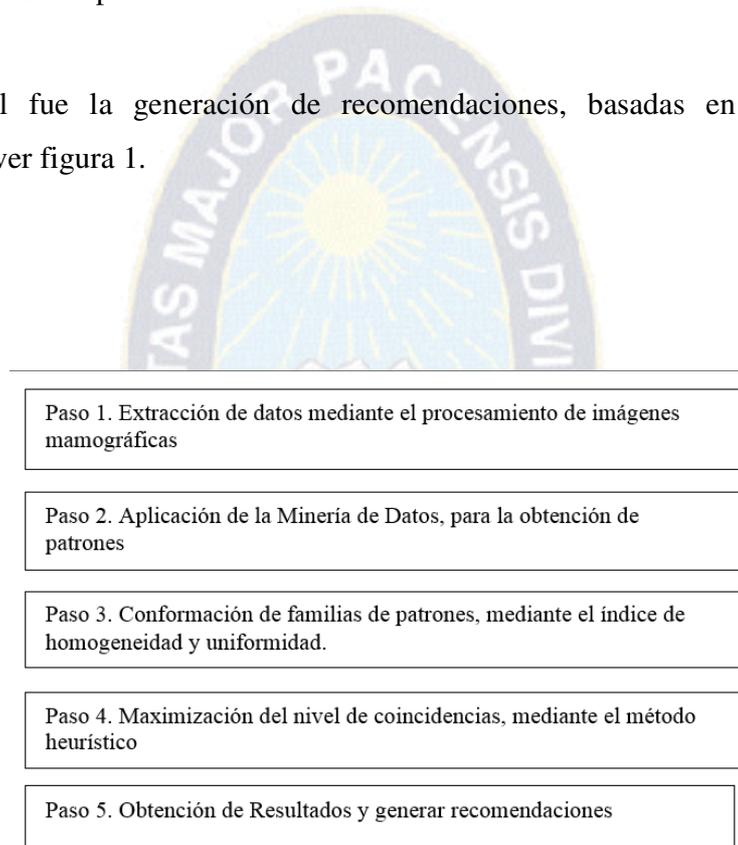


Figura.- 1.Fases del Método Heurístico

Paso 1: Disección de la mamografía y conformación de familia de patrones.

Se realizó la extracción de los datos mediante el procesamiento de imagen mamográfica, la mamografía se dividió hasta el nivel más bajo posible y se almacenó dentro de una base de datos relacional, mediante las siguientes fases.

- Fase 1: Se realizó la preparación de la imagen mamográfica y se analizó el contraste en diferentes niveles de gris, mediante re-escalado de la región de interés segmentado.
- Fase 2: Se aplicó la singularidad de la región de interés, mediante técnicas de procesamiento a diferentes escalas, y se aplicó la transformada de wavelet para la supresión del ruido y refuerzo de las singularidades de las regiones de interés; por otra parte se analizó la estructura de niveles de gris.
- Fase 3: Se realizó la agrupación y marcado de patrones de acuerdo a las características por pixel. En este caso, cada patrón obtenido en el procesamiento de imágenes de la segunda fase, contribuyó con una característica por píxel, y se agrupó en FP. Posteriormente, se etiquetó y clasificó por medio de la minería de datos, y se logró segmentar el tejido sano las microcalcificaciones.

Paso 2: Recolección de datos

Se aplicó la minería de datos (figura 2), al paso 1, para la selección, limpieza y transformación de los datos; de esta forma, se obtuvo una vista minable de un conjunto de patrones para determinar las mejores características del patrón, mediante una técnica de selección de características contextuales y realizar un análisis de correlación, entre los conjuntos de características de los patrones contextuales según su relevancia, reduciendo el error de clasificación.

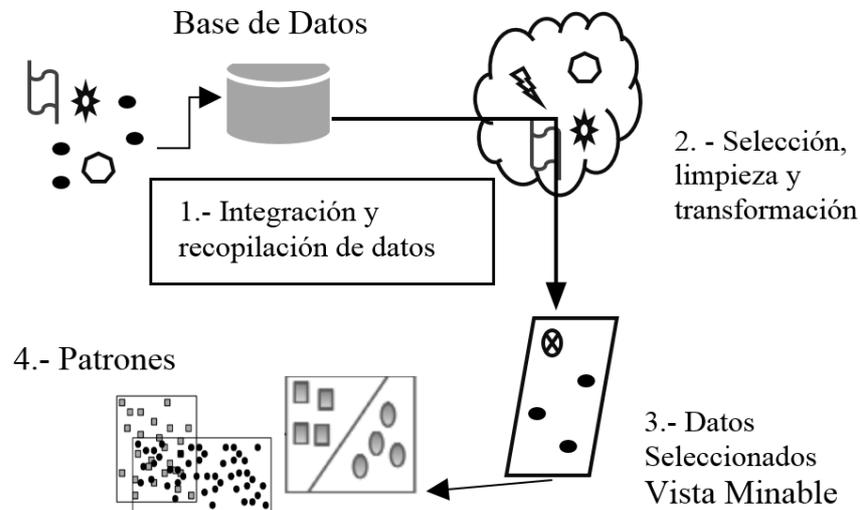


Figura.- 2.Fases del proceso de extracción de Patrones de imágenes mamográficas con la minería de datos

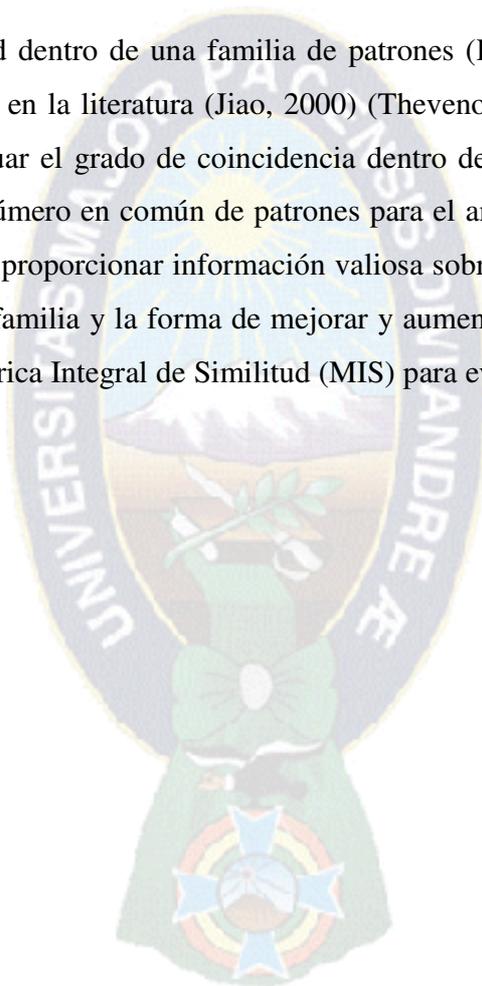
Para la integración y recopilación de datos, se transformaron todos los datos a un formato común de la base de datos para unificar de manera operativa toda la información recogida, detectando y resolviendo las inconsistencias y facilitar enormemente la navegación y visualización previa de los datos, para discernir los aspectos que podían interesar para su estudio.

Dado que los datos provienen del procesamiento de imágenes, pueden contener valores erróneos o faltantes. Estas situaciones se trataron en la fase de selección, limpieza y transformación, en la que se eliminaron o corrigieron los datos incorrectos y se decidió la estrategia a seguir con los datos incompletos.

Además, se proyectaron los datos para considerar únicamente aquellas variables o atributos que iban a ser relevantes con el objetivo de hacer más fácil la tarea propia de minería y para que los resultados de la misma fueran más útiles.

Paso 3: Evaluación de Similitud

Para medir la similitud dentro de una familia de patrones (FP), se propusieron varios índices de homogeneidad en la literatura (Jiao, 2000) (Thevenot, 2006). Un índice común es una métrica para evaluar el grado de coincidencia dentro de una FP. La evaluación se basó en parámetros del número en común de patrones para el análisis de una FP existente. Ellos fueron destinados a proporcionar información valiosa sobre el grado de coincidencia, alcanzado dentro de una familia y la forma de mejorar y aumentar la uniformidad. En este trabajo, utilizamos la Métrica Integral de Similitud (MIS) para evaluar el carácter común de FP.



Nombre		Desarrollado por	Medida de la similitud	Cero Similitud	Similitud Completa
GSI	Grado de similitud del índice	Collier (1981)	Toda la Familia	1	
IDCT	Índice de similitud de la constante	Wacher and Trleva (1986)	Toda la Familia	0	1
ISLP	Índice de similitud de la línea de producto	Kota, Sethuraman and Miller(2000)	Toda la Familia	0	100
PIDI	Porcentaje de índice de similitud	Siddike, Rose and Wang (1998)	individual	0	100
IS	Índice de similitud	Martin and ishii(1997)	Toda la Familia	0	1
PCS	Parte componente de similitud	Jiao and Tseng (2000)	Toda la Familia	1	
MCS	Métrica comprensiva para la similitud	Thevenot and Simpsom(2007)	Toda la Familia	0	1

Tabla 3. Lista de similitud de índices

Paso 4: Optimización mediante el Método Heurístico :

En este trabajo, se empleó el Algoritmo Genético (AG) para maximizar la (MIS). Un AG es un algoritmo de optimización estocástica de adaptación, que implican la búsqueda y optimización (Goldberg, D. E., 1989).

En este trabajo, cada atributo de un patrón se codificó como un entero, que más tarde se convirtió en una representación binaria de la AG.

El AG maximiza la MIS, con sujeción a las siguientes limitaciones adicionales para facilitar la selección de los patrones a ser encontrados.

- Microcalcificaciones
- Masas circunscritas
- Masas espiculadas
- Distorsiones y Asimetrías

Para poder determinar:

- CDIS: carcinoma ductal in situ
- CDI: carcinoma ductal invasivo
- CLI: carcinoma lobular invasivo
- Cáncer de mama inflamatorio
- CLIS: carcinoma lobular in situ
- Carcinoma tubular de la mama
- Carcinoma medular de la mama
- Carcinoma mucinoso de la mama
- Carcinoma papilar de la mama
- Carcinoma cribiforme de la mama
- Cáncer de mama en hombres
- Enfermedad de Paget en el pezón
- Tumor filoides de la mama

Mediante la adición de estas características, se especificó un número máximo de análisis. De ahí que el AG ofreció recomendaciones que más influyeron en el carácter común, ayudando al enfoque de análisis en patrones críticos para analizar. Actualmente, no hubo pautas para elegir el valor adecuado para este análisis. Sin embargo, se pudieron tomar un determinado porcentaje del número total de parámetros para este análisis.

Sobre la base de estas características, se eligieron los patrones. Dentro de este conjunto de patrones, se consideraron cuatro atributos:

- Microcalcificaciones
- Masas circunscritas
- Masas especuladas
- Distorsiones y Asimetrías.

Para una familia dada, si un atributo es común entre todos los patrones que utiliza esta familia, entonces este atributo no se considera durante la optimización.

Formulación:

Maximización (MIS) (Jiao, 2000):

Sujeto a:

$$P_{ijk} \in (V_i)_{ijk} \quad i = 1,2,3..,n \quad j = 1,2,3 \dots, t \quad k = 1,2,3,4 \dots c \quad (1.1)$$

$$\sum_{i=1}^n \sum_{j=1}^t \sum_{k=1}^c II(P_{ijk} \neq P_{ijk}^{ini}) \leq m \quad (1.2)$$

Donde:

P_{ijk} =Valor del parámetro k por el patrón i en FP j (decisión variable)

C_{ijk}^{ini} =Valor inicial del parámetro k del patrón i en la FP j

V_1 =Posible valor 1

$(V_l)_{ijk}$ =Asignación de posibles valores en k por el patrón i en el FP j

n =Número total de patrones en todas las FP.

t =Total número de patrones en FP.

m =Número máximo de parámetros asignados para el cambio.

$II(P_{ijk} \neq P_{ijk}^{ini}) = 0$ si $P_{ijk} = P_{ijk}^{ini}$ Caso contrario 1

Para entender la formulación, se consideró lo siguiente: Para la imagen mamográfica dada con varias FP, se estableció una lista de los patrones n.

Para cada patrón i en cada FP j, se consideraron:

$P_{ij1}, P_{ij2}, P_{ij3}, P_{ij4} \dots P_{ijc}$, que corresponden respectivamente a los valores de tejido sano o microcalcificaciones:

- Contraste de Niveles de Gris
- Estructura de niveles de Gris
- Entorno del pixel

El AG maximizó la Métrica Integral de Similitud (MIS) mediante la modificación de los valores de estos P_{ijk} bajo la restricción en análisis de lesiones, (es decir, el P_{ijk} puede tomar un conjunto particular de valores $(V_l)_{ijk}$ de todos los valores posibles de V_j).

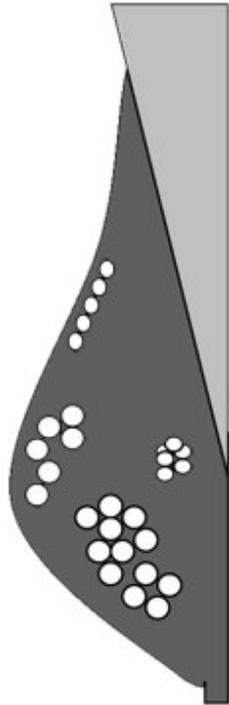


Figura.- 3.Imagen Mamografía –ejemplo



Figura.- 4.Disección y conformación de familias de patrones (FP)

Paso 5: Obtención de los resultados y recomendaciones

Una vez que la optimización fue completa, el AG propuso una nueva secuencia de análisis de patrones, que pudo ser comparado con los patrones originalmente extraídos con la Minería de Datos.

El AG no comprobó actualmente la viabilidad de la solución en cuenta; más bien, ofreció una lista clasificada de los parámetros que más influyeron en el grado de coincidencia en la FP. Esto pudo ser visto como una reducción del análisis de patrones, donde se comprobó la viabilidad de la solución a posteriori en la lista de recomendaciones propuestas, en lugar de comprobar la viabilidad de una solución de volver analizar los patrones.

3.2 Arquitectura tecnológica.

3.2.1 Transformada de Wavelet Discreta

Según (Walker, 1999) (Percival, 2002) , se considera una función discreta

$$f(n) = \{123456437892\} \quad (2.1)$$

Se Define

$$A(n) = \frac{F_{2n-1} + F_{2n}}{\sqrt{2}} \quad (2.2) \quad \text{y} \quad D(n) = \frac{F_{2n-1} - F_{2n}}{\sqrt{2}} \quad (2.3)$$

Se obtendrá las sub-funciones A^1 y D^1

$$A^1 = [1.5 \sqrt{2}, 3.5 \sqrt{2}, 5.5 \sqrt{2}, 3.5 \sqrt{2}, 7.5 \sqrt{2}, 5.5 \sqrt{2}] \quad (2.5)$$

$$D^1 = [-.5 \sqrt{2}, -.5 \sqrt{2}, -.5 \sqrt{2}, .5 \sqrt{2}, -.5 \sqrt{2}, 3.5 \sqrt{2}] \quad (2.6)$$

Ambas sub-funciones son de la mitad de tamaño que la función original y ésta se podrán recuperar de:

$$F(n) = \frac{A_1 + D_1}{\sqrt{2}}, \frac{A_1 - D_1}{\sqrt{2}}, \dots, \frac{\frac{A_n}{2} + \frac{D_n}{2}}{\sqrt{2}} \quad (2.7)$$

$$F(n) = \frac{1.5 \sqrt{2} - 0.5 \sqrt{2}}{\sqrt{2}}, \frac{1.5 \sqrt{2} + 0.5 \sqrt{2}}{\sqrt{2}}, \dots, \frac{5.5 \sqrt{2} - 3.5 \sqrt{2}}{\sqrt{2}}, \frac{5.5 \sqrt{2} + 3.5 \sqrt{2}}{\sqrt{2}}$$

Generando la señal original que finalmente se pudo escribir como:

$$F(n) = A_n^1 + D_n^1 \quad (2.8)$$

Donde

$$A_n^1 = \left[\frac{A_1}{\sqrt{2}}, \frac{A_1}{\sqrt{2}}, \frac{A_2}{\sqrt{2}}, \frac{A_2}{\sqrt{2}}, \dots, \frac{A_{n/2}}{\sqrt{2}}, \frac{A_{n/2}}{\sqrt{2}} \right]$$

$$D_n^1 = \left[\frac{D_1}{\sqrt{2}}, \frac{-D_1}{\sqrt{2}}, \frac{D_2}{\sqrt{2}}, \frac{-D_2}{\sqrt{2}}, \dots, \frac{D_{n/2}}{\sqrt{2}}, \frac{D_{n/2}}{\sqrt{2}} \right]$$

Si se define una matriz V_n^1 como:

$$\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

Y una matriz W_n^1 como:

$$\begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

Se obtuvo:

$$A_n^1 = [A^1][V_n^1] \quad y \quad D_n^1 = [D^1][W_n^1]$$

Se analizó en varios niveles como se muestra en (2.5)

Consideramos el nivel 2, donde se aplicó

$$A^2 = \frac{A_{2n-1}^1 + A_{2n}^1}{\sqrt{2}} \quad y \quad D^2 = \frac{D_{2n-1}^1 - D_{2n}^1}{\sqrt{2}}$$

$$A^2 = [5 \quad 9 \quad 12]$$

$$D^2 = [0 \quad -1 \quad -4]$$

De tal manera que $F_{(n)}$ se pudo recuperar de iteraciones sucesivas de $A^2: D^2:D^1$ pues de $A^2:D^2$ se recuperó A^1 y con $A^1:D^1$ como se vio anteriormente se recuperó $F_{(n)}$ (Walker, 1999).

La señal que se iba a analizar se descompuso como se muestra en la figura, separando los componentes de baja frecuencia de las de alta frecuencia se obtuvieron dos señales que en su conjunto produjeron el doble de muestras de la señal original (Misiti, Misiti, Oppenheim, & Poggi, 1997).

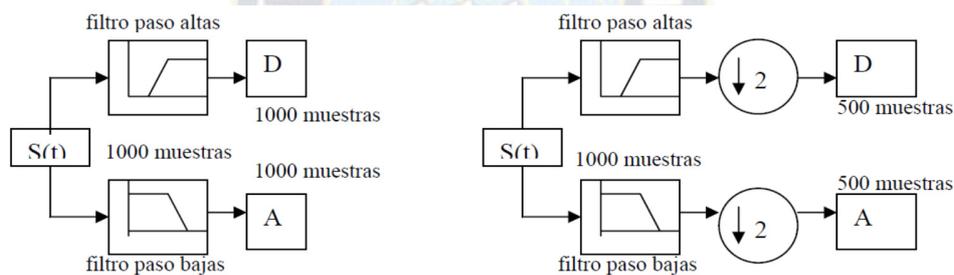


Figura.- 5. Componentes de alta y baja frecuencia separadas, muestreo completo y muestreo reducido con los coeficientes wavelet donde 2 significa muestreo reducido a la mitad

En algunos casos se requiere disminuir la cantidad de datos necesarios para el manejo de la señal, por lo que se hace el muestreo reducido, en el que se involucran los coeficientes wavelets obteniendo con ello una cantidad de muestras de la mitad que en el caso anterior como muestra la figura.

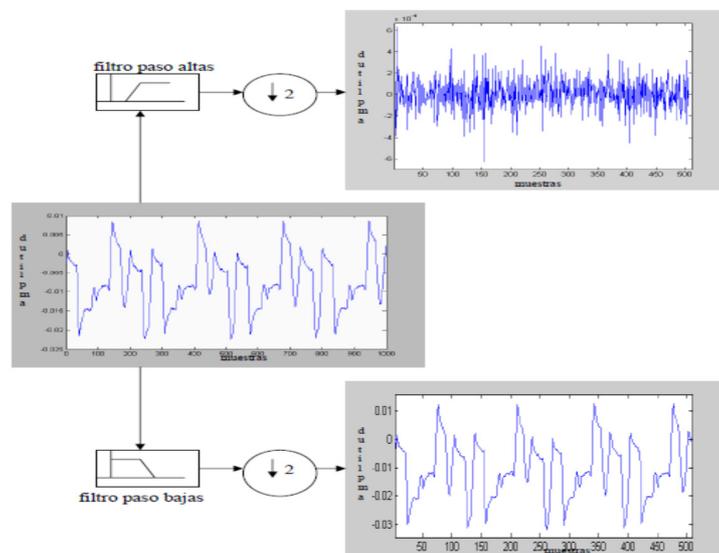


Figura.- 6.Descomposición de la señal

Este procedimiento se pudo repetir para varios niveles como se observa en la figura siguiente:

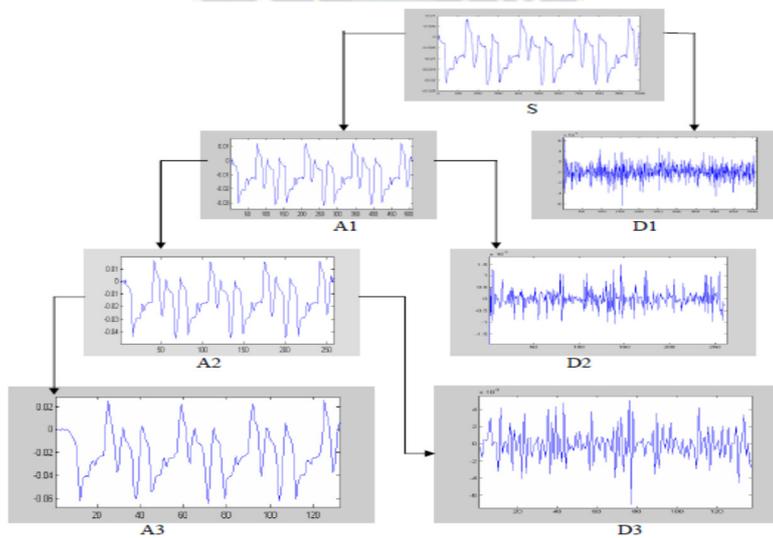


Figura.- 7.Descomposición en tres niveles de la misma señal

De tal manera que para el tercer nivel la señal tendrá un número de muestras en A_3 de 125 y D_3 igualmente 125, dado que en la componente de baja frecuencia la reducción en el número de muestras no distorsiona la señal de una manera significativa es posible la reconstrucción de la señal original de la ecuación (2.9), seleccionando el nivel dependiendo de la señal en serie de tiempo, de la wavelet utilizada y de la energía de los componentes de frecuencia de la wavelet de tal forma que la distorsión de la señal recuperada no afecte significativamente la información (Misiti, Misiti, Oppenheim, & Poggi, 1997).

$$S = A_3 + D_1 + D_2 + D_3 \quad (2.9)$$

Por lo general, las componentes de alta frecuencia son señales de ruido y fue conveniente deshacerse de ellas (Misiti, Misiti, Oppenheim, & Poggi, 1997), (Walker, 1999) por lo que la señal fue notablemente reducida, lo que es sumamente útil en procesamiento de señales. De cualquier manera, en la Figura 6 se observó la señal recuperada dada por la ecuación (2)

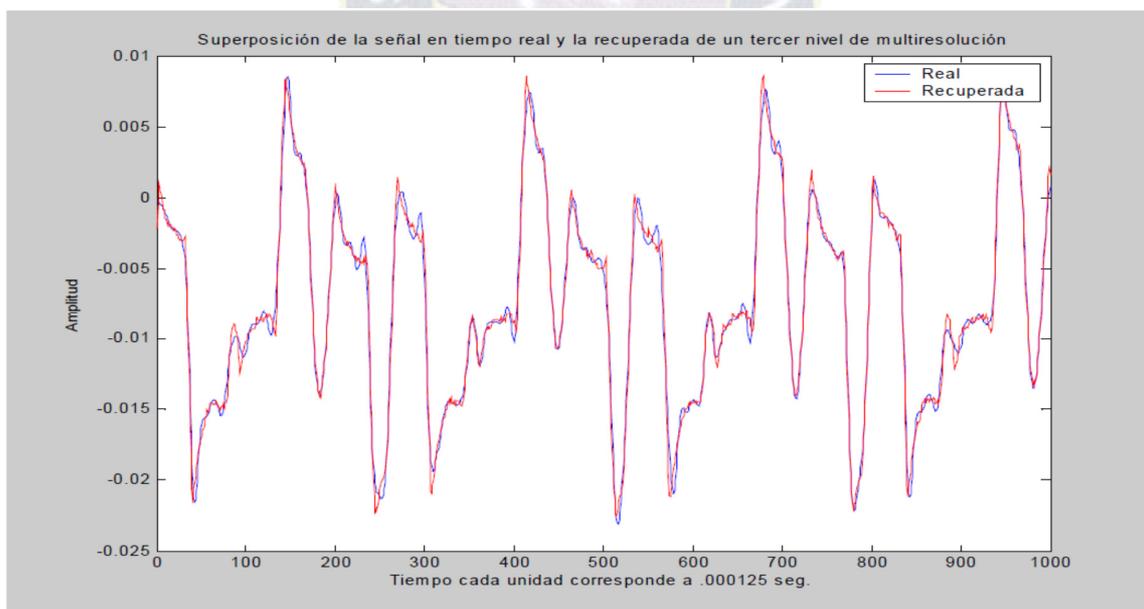


Figura.- 8. Superposición de las señales original y recuperada con reducción por coeficientes wavelets

3.2.2 Preparación de la imagen mamográfica

Los patrones en la imagen mamográfica, su tamaño estuvieron relacionados con la resolución y los niveles de gris, éstos pudieron ser tan pequeños como 50 μm o incluso más pequeñas, por tanto este valor mínimo que debió presentar la resolución de imagen mamográfica, para evaluar el efecto de los niveles de gris se realizó un escalado de éstos desde una resolución alta a otra más baja, realizando un escalado espacial, promediando los valores que cubre un píxel de la más alta resolución sobre un píxel a baja resolución por medio de la siguiente expresión (Lampinen J., 1995):

$$P_n = \sum_i [V_{n_i} \times I] \quad (3.1)$$

Dónde:

P_n =Valor del pixel gris

V_{n_i} =Es el valor que se asigna a cada pixel en función al patrón

I =Es el nivel de intensidad de los pixeles que tienen cada patrón.

Los niveles de gris son una función de distribución que muestra el número de píxeles para cada nivel de gris en un histograma, los cuales ayudan a conseguir información relevante de los patrones, a su vez es empleado para eliminar el fondo de la mamografía, pues se encuentra en los niveles más bajos de los niveles de grises, los píxeles que tienen un valor de amplitud pequeña y realzar aquellos con un valor de nivel de gris elevado. Esto lo hace mediante la aplicación de una transformación no-lineal que nos permite controlar de

manera eficiente la razón de contraste y el umbral de forma adaptativa, mediante la siguiente expresión.

$$F(f_{(x,y)}) = \alpha [s(k(f - \beta)) - s(k(f + \beta))] \quad (3.2)$$

$$\alpha = \frac{1}{s(s(k(1 - \beta)) - s(-k(1 + \beta)))} \quad (3.3)$$

$$s(x) = \frac{1}{1 + e^{-x}} \quad (3.4)$$

Dónde:

$f_{(x,y)}$ = Es el nivel de gris de un píxel de la imagen de entrada.

$F(f)$ = Es la función de transformación no-lineal por puntos

β y k = Son parámetros que representan el control del umbral y la proporción del contraste de la imagen respectivamente, valores óptimos para $\beta = 0.3$ y $k=20$.

Una vez preparada la imagen se realizó un análisis a través de la transformada wavelet en 2D, de lo cual se obtuvieron cuatro características distintas para cada píxel, se obtuvieron dos pares de imágenes del mismo tamaño a distintas frecuencias, teniendo el píxel i -ésimo cuatro características procedentes del valor del píxel i -ésimo de los dos pares de imágenes, luego se realzaron las singularidades y se suprimió el ruido, esto se hizo durante la etapa de etiquetado del patrón, para la pertenencia de un píxel a una clase.

3.2.3 Análisis de la imagen mamográfica

El método heurístico usó la transformada wavelet para la extracción de características, como es la transición de los niveles de gris de una lesión a los del fondo o al tejido sano de la mama, las wavelets se comportaron como un sistema de enfoque.

Este método heurístico empleó la detección de contornos a diferentes escalas para extraer las características que representan a las lesiones desde los máximos wavelet.

Este modelo se basó en la detección de extremos locales de la transformada wavelet discreta, descrito en el punto 3.2.1, aplicando un muestreo adaptativo de la transformada wavelet basado en su máximo local, el cual es invariante a la traslación, y en este caso el máximo indicó la posición de las variaciones rápidas en la amplitud de los niveles de gris, detección de contornos multi-escala, se obtuvieron las características wavelet o no contextuales.

Para las sub-bandas de frecuencia, se aplicó una combinación de filtros bajos y altas B-A, para la reconstrucción o síntesis desde las imágenes sub-banda se aplicó un proceso inverso, usando unos filtros distintos B y A, así se obtuvo una representación invariante a la traslación como se muestra en la figura 9 (Lampinen J., 1995).

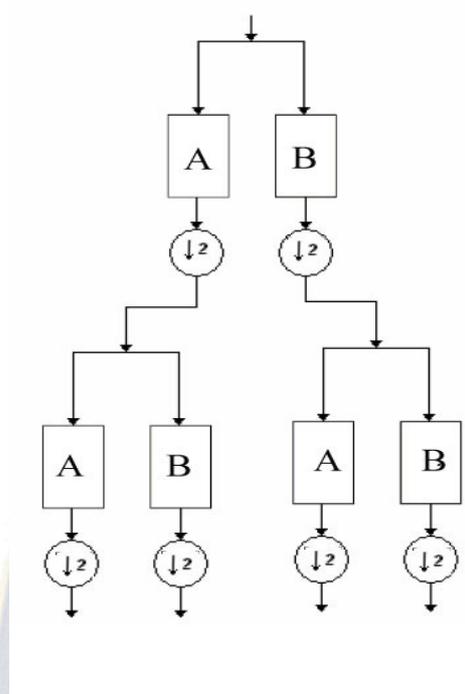


Figura.- 9.Filtros de wavelet

3.2.4 Supresión del ruido de la imagen mamográfica

El Método Heurístico usó un método de umbralización suave para eliminar el ruido a partir del umbral calculado, mediante la combinación del método de supresión de ruido Gradiente-Laplaciano y el proceso de realce no lineal, presenta buenos resultados, sobre todo para las características que se encuentran ocultas entre el fondo de la imagen y que de forma visual, no se perciben o se confunden con tejidos o vasos, debido a un bajo contraste.

De esta manera el método heurístico se logró separar el fondo de las características de interés, y es el que utiliza en el extractor de características, la adición de un píxel a la

imagen bajo estudio, es una técnica que sirve de referencia para verificar la respuesta de los filtros a diferentes escalas y diferentes frecuencias, de esta manera la singularidad no se pierde, esta técnica resulta muy adecuada para aumentar el contraste entre el fondo y las singularidades.

3.2.5 Extracción de Patrones

La extracción de patrones wavelet y de los niveles de gris la mamografía, se basa en cada patrón o FP es construida a partir de un conjunto de variables extraídas de los patrones wavelet y de los patrones de la estructura de los niveles de gris de la imagen.

La correspondencia de un patrón es para cada píxel, dando a conocer que en cada una de las existen píxeles si pertenecen a una lesión y píxeles que pertenecen a tejido sano, por tanto, podrán pertenecer a una clase u a otra en función de las patrones de cada uno de estos patrones, quedando la FP inicial de patrones dividido en dos subconjuntos FP0 (clase normal) y FP1 (clase anormal), de esta forma el Método Heurístico emplea dos tipos de patrones.

El primer tipo son extraídas de la disección de la imagen mamográfica inicial, a través de la transformada wavelet en dos dimensiones, por cuatro sub-imágenes correspondientes a cada nivel de disección wavelet, en concreto, se extrae el nivel de gris de cada píxel perteneciente a cada sub-imagen, de este modo las FP del primer tipo se denota como FP_c , son denotadas por que contienen información relevante sobre los niveles de gris del entorno de cada píxel analizado la información extraída desde cada píxel y es usada para formar un nueva FP que contiene información local sobre los niveles de gris del entorno, el conjunto de FP se denota como FP_ω ,

De esta manera se hace frente al problema de la detección de lesiones de una manera más realista, es necesario agregar otras que tienen que ver con la estructura de los niveles de gris del entorno de los píxeles, el contraste local se define como (Lampinen J., 1995).

$$C_L(x, y) = N(x, y) - P(x, y) \quad (4.1)$$

Dónde:

$C_L(x, y)$ =Contraste local del pixel

$N(x, y)$ =Nivel de gris del pixel en la posición (x, y)

$P(x, y)$ =Promedio de los niveles de grises sobre una región determinada.

El contraste normalizado se define como (Lampinen J., 1995):

$$C_N(x, y) = \frac{C_L(x, y)}{D(x, y)} \quad (4.2)$$

Dónde:

$C_L(x, y)$ =Contraste local

$D(x, y)$ =Es la Desviación estándar de los niveles de grises.

3.2.6 Clasificación de Patrones

Para determinar a qué clase pertenece cada patrón, se emplea la supresión de ruido explicado en el punto 3.2.4, los píxeles que son sensibles a pertenecer a la clase anormal son los que se encuentran por encima del umbral, mientras que aquellos que son sensibles a pertenecer a la clase normal son atenuados, para asignar las etiquetas a las FP no contextuales.

Los píxeles pertenecientes a lesiones se encuentran en la parte alta del histograma de los niveles de gris, se agrupan las FP en (n) clusters agregando aquellos que representan un nivel de gris elevado a la clase FP1 normal, el resto son asignados a FP0 (clase normal).

Como existen un elevado número de FP, y teniendo en cuenta además que el número de píxeles que pertenecen a lesiones es mucho menor que el total de los píxeles, se realiza el agrupamiento de una partición aleatoria de los píxeles que pertenecen a FP0, y consiguiendo así el conjunto definitivo, por tanto el número óptimo de clusters, se emplea un criterio de separación, basado en el análisis del discriminante de Fisher, que mide el nivel de separación entre cluster mediante las dos matrices de dispersión, matriz de dispersión intraclase y matriz de dispersión entreclase, por tanto, el subconjunto que genere una más grande, es considerado como el mejor subconjunto para el proceso de agrupamiento, y se escogera aquel valor mínimo de K que maximice la separabilidad.

3.2.7 Optimización basada en el Algoritmo Genético

La optimización de la familia de patrones basados en el algoritmo genético se basa en maximizar el carácter común de la familia de patrones mediante el uso de la MIS. El enfoque propuesto utiliza un conjunto de genes que controla la homogeneidad de un algoritmo genético que evalúe el carácter común de la familia de patrones.

Como se muestra en la (figura 7), la cadena de cromosomas en el algoritmo genético concatena los cromosomas de las cadenas individuales para cada patrón en una larga cadena y luego aumenta esta cadena con los genes que controlan el (nk) que controlan la entidad dentro de las cadenas cromosómicas individuales.

La longitud resultante de la cadena cromosómica es $(nk + nkp)$, donde:

- n es el número de variables de diseño.
- k es el valor de las masas o micro-calcificaciones, tales como forma, margen, densidad.
- n es el número de patrones.

Si cualquiera de estos primeros genes (nk) toma el valor de 1, entonces esa variable de coincidencia particular se hace común entre todos los patrones de una familia;

Un valor de 0 hace que la variable sea única dentro de la familia. Se sigue entonces que si estos primeros genes (nk) son todos los 1, son un cien por ciento común entre los patrones, mientras que una cadena de todos los 0 no indica ninguna comunidad entre los productos dentro de la familia. La variable de diseño C_{1p1} en (Figura 12) significa el primer valor de un parámetro como la forma, margen y la densidad de la imagen mamográfica en la familia de patrones, como tal, se consideran diferentes niveles de uniformidad, donde los resultados de la optimización indican:

1. Qué variables deben ser comunes.
2. Los valores que se deben tomar.
3. Los valores que las variables únicas deben tomar.



(Thevenot,

2006)

Figura - 10. Representación del Algoritmo Genético para la maximizar el carácter común

Dado que los parámetros para el AG son caso-dependiente, los valores que dan los mejores resultados no se conocen a priori. Antes de la optimización, los valores de los cuatro parámetros se determinan: es decir, de cruce (P_c), mutación (P_m), número máximo de generaciones, y de la población. Dimensionamiento de una población AG para asegurar la máxima influencia de cálculo y muestreo preciso se ha considerado empíricamente en varios estudios. (Goldberg, D. E., 1989), muestra cómo configurar el tamaño de la población en el contexto de la mezcla combi nativa. En este Método Heurístico, se consideró mínimo de 50 y un máximo de 200 como el tamaño de la población, Los ajustes de mutación se obtuvieron como se mostró en la literatura AG (Jong, 1975). Para el experimento, el más bajo (0.001) y el valor más alto (0.01) (Grefensette, 1986) de la tasa de mutación recomendado P_m .

Interpretación de Resultados.

Con el fin de lograr los objetivos planteados al inicio de esta tesis, se realizó el trabajo de campo mediante entrevistas (ver anexo B), se puede deducir que el Ministerio de Salud no cuenta con ninguna política de manejo y distribución de la información en cuanto a la detección de Cáncer de Mama; por otro lado, el hospital de Clínicas en el Área de Imagenología; si bien cuenta con la información de los pacientes de Cáncer de Mama, éstos son guardados en CD's y no son compartidos con otras entidades.

De cada 10 mujeres atendidas, 2 son detectadas con alguna lesión.

El 50 % sufre la pérdida de una mama.

Del 100% de casos atendidos con alguna lesión, el 7% culmina el tratamiento y seguimiento adecuado.

La Jefa de Unidad del área de Imagenología del Hospital de Clínicas de la Ciudad de la Paz – Bolivia, si bien cuenta con un alto prontuario de actualizaciones y capacitaciones, en congresos internacionales realizados en otros países, además de ejercer dentro el área más de 40 años, realiza pequeñas campañas de detección de Cáncer de Mama; pero debido al escaso presupuesto y una pobre área tecnológica en equipos no puede difundir dicho conocimiento y ampliar estas campañas a nivel nacional.

Estas son las razones por las cuales se debe implementar este Método Heurístico mediante un Sistema Informático con el principio de Software Libre y Estándares Abiertos, el cual permitirá la distribución libre y gratuita, para la estandarización del manejo de la información de pacientes con Cáncer de Mama.

Capítulo IV Marco de Resultados

4.1 Estado de los Objetivos.

4.1.1 Objetivo General

Se alcanzó el objetivo general propuesto, ya que el método desarrollado permitió el procesamiento de imágenes con la finalidad de obtener información necesaria para que esta sea manejada a través de la minería de datos para la obtención y selección de patrones que coadyuven a un diagnóstico oportuno en casos de cáncer de mama, aportando de esta manera a incrementar la esperanza de vida de muchas mujeres que padecen de esta enfermedad.

4.1.2 Objetivos Específicos

Se realizó una revisión bibliográfica y un trabajo de campo para tener el conocimiento necesario en la elaboración del método propuesto, por lo que se pudo alcanzar los objetivos planteados en el proyecto de acuerdo al siguiente detalle:

El primer objetivo fue logrado con el procesamiento de imágenes mamográficas y su almacenamiento en una base de datos relacional para obtener la información que sea necesaria y requerida; posteriormente se aplicó la minería de datos para la obtención de patrones y posterior disección de los mismos conformando familias con similares características.

En el segundo objetivo específico, los datos básicos y la información coleccionada fueron almacenados en otra base de datos relacional.

Para el logro del tercer objetivo específico, se realizó una evaluación del nivel de uniformidad en los patrones, mediante el cálculo de los índices de homogeneidad, para la conformación de familias de patrones.

Para alcanzar el cuarto objetivo específico se utilizó la heurística para maximizar el nivel de coincidencia en cada familia con caracteres específicos.

Las recomendaciones elaboradas en base a los resultados obtenidos después del procesamiento de toda la información se constituyen en el logro del quinto objetivo específico, hago notar que las mismas pueden variar de acuerdo a cada paciente.

4.2 Estado de la Hipótesis.

4.2.1 Aplicación del Método Heurístico

Para la aplicación del Método Heurístico, se analizó una Imagen mamográfica, previamente diagnosticada (Figura 11), de una paciente de sexo femenino con 65 años de edad.

Paso 1: Disección de la mamografía

Se procedió al diseccionado de la imagen mamográfica, como una familia de patrones (FP); del cual se obtuvo y evaluaron los datos de caracteres comunes de la imagen mamográfica (Figura 11).

Se logró diseccionar la imagen mamográfica hasta el nivel más bajo mediante la Transformada de Wavelet Discreta, la preparación de la imagen mamográfica, el análisis de la imagen mamográfica, la supresión del ruido de la imagen mamográfica, la extracción de

patrones y la clasificación de Patrones, descritos en el apartado de arquitectura tecnológica, y finalmente, se obtuvieron seis Familias de Patrones (FP) (ver Figura 14).



Figura.- 11.Imagen Mamográfica de aplicación



Figura.- 12 Familia de Patrones 1,2, 3 (FP1, FP2, FP3)



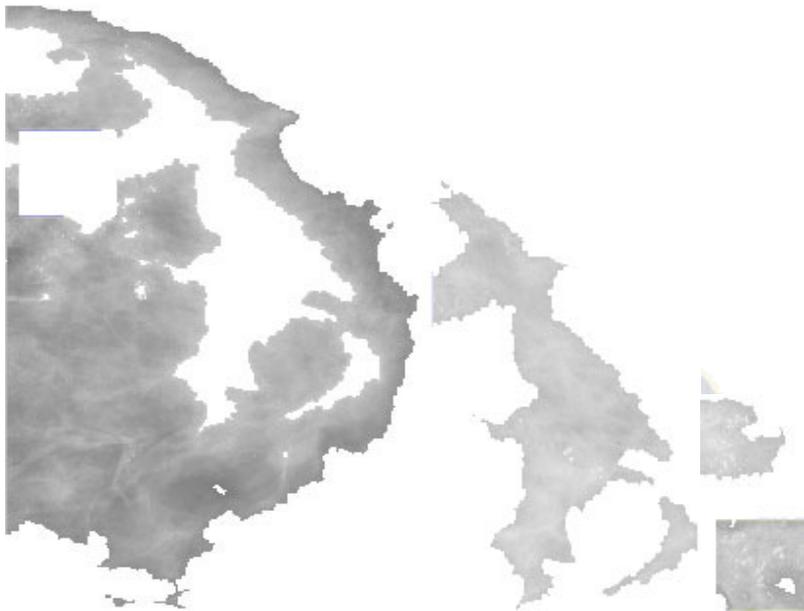


Figura.- 13 Familia de Patrones 4, 5, 6 (FP4, FP5, FP6)

Figura.- 14 Disección de la imagen mamográfica

Paso 2: Recolección de datos

La imagen mamográfica se diseccionó, y los datos fueron almacenados en una base de datos relacional, como se muestra en la Figura 14, luego se recolectaron los datos para calcular la Métrica Integral de Similitud (MIS), correspondiente a las familias (FP1, FP2..., FP6), como se muestran en las Figuras 12 y 13.

Se detectó un patrón en común que fue el mismo patrón exacto compartido en algunas de las FP; a su vez, se detectó un patrón variante que tenía la misma función entre algunas FP, que difería ligeramente uno de otro. También se detectó, un patrón único dentro de las FP, y por último, se detectó un patrón final, que hacía referencia a las características obtenidas después del nivel más bajo de disección posible, es decir, este patrón no se podía dividir aún más.

Se procedió a clasificar algunos patrones, como diferenciadores, los cuales excluyeron las regiones sin hallazgos, y otros como patrón con hallazgo de lesión.

Posteriormente, a la clasificación de los patrones se procedió al cálculo de la Métrica Integral de Similitud (MIS), como estrategia para reflejar con precisión la similitud de estos patrones.

$$MIS = \frac{\sum_i^p n_i * f_{1i} * f_{2i} * f_{3i} * f_{4i} * (C_i^{max} - C_i)}{\sum_i^p n_i * f_{1i}^{max} * f_{2i}^{max} * f_{3i}^{max} * f_{4i}^{max} * (C_i^{max} - C_i^{min})}$$

(5.1) (Thevenot, 2006)

Donde

p = Número total de componentes.

n =Número de familias de FP, patrón i .

f_{ji} = Relación entre el mayor número de modelos que compartan componente i con idéntico característica que tienen el patrón j .

Los valores obtenidos a través de la MIS en los conjuntos de FP, fueron seleccionados en función de su forma, margen y densidad, variando entre 0 y 1; posteriormente, se asignó el valor de 1, cuando todos los patrones no diferenciadores fueron comunes; y por último se asignó el valor de 0 para todos los patrones no diferenciadores que fueron diferentes.

Los resultados de la MIS se muestra en la Tabla 4.

Paso 4: Optimización basada en el Algoritmo Genético

Se realizó la optimización basada el AG como se describe en el apartado 3.2.7, con el objetivo de maximizar la Métrica Integral de Similitud (MIS) usando la ecuación 5.1, para todas las combinaciones posibles, los cuales se probaron con diferentes valores iniciales. Los valores más altos alcanzados por las MIS fueron las combinaciones resumidas en la Tabla 4 y los mejores resultados fueron las combinaciones de las filas 7, 8, 14 y 16, con un valor de 0.884 en la MIS; este valor fue el valor ideal, debido a que fue potencialmente común entre todos los patrones de la FP.

Nro.	Prob.	Mutación	Tamaño de la muestra	Nro. de Generaciones	Valores de la MIS	Porcentaje %
1	0.4	0.001	50	500	0.781	16.20
2	0.6	0.001	50	500	0.778	15.80
3	0.4	0.01	50	500	0.753	12.00
4	0.6	0.01	50	500	0.757	12.60
5	0.4	0.001	200	500	0.818	21.70
6	0.6	0.001	200	500	0.823	22.50
7	0.4	0.01	200	500	0.884	31.60
8	0.6	0.01	200	5000	0.884	31.60
9	0.4	0.001	50	5000	0.749	11.40
10	0.6	0.001	50	5000	0.792	17.80
11	0.4	0.01	50	5000	0.791	17.70
12	0.6	0.01	50	5000	0.804	19.70
13	0.4	0.001	200	5000	0.876	30.30
14	0.6	0.001	200	5000	0.884	31.60
15	0.4	0.01	200	5000	0.861	28.10
16	0.6	0.01	200	5000	0.884	31.60

Tabla 4 Resultados de la experimentación de los resultados obtenidos del AG

La comparación de estos parámetros: número de generaciones para converger y el número de llamadas de función, se muestra en la Figura 15.

Se normalizaron los valores entre 0 y 1, para indicar al valor más alto como un mejor rendimiento. La combinación más resaltante se mostró en la fila 16, donde el valor más alto MIS se obtuvo con el tiempo mínimo del cálculo.

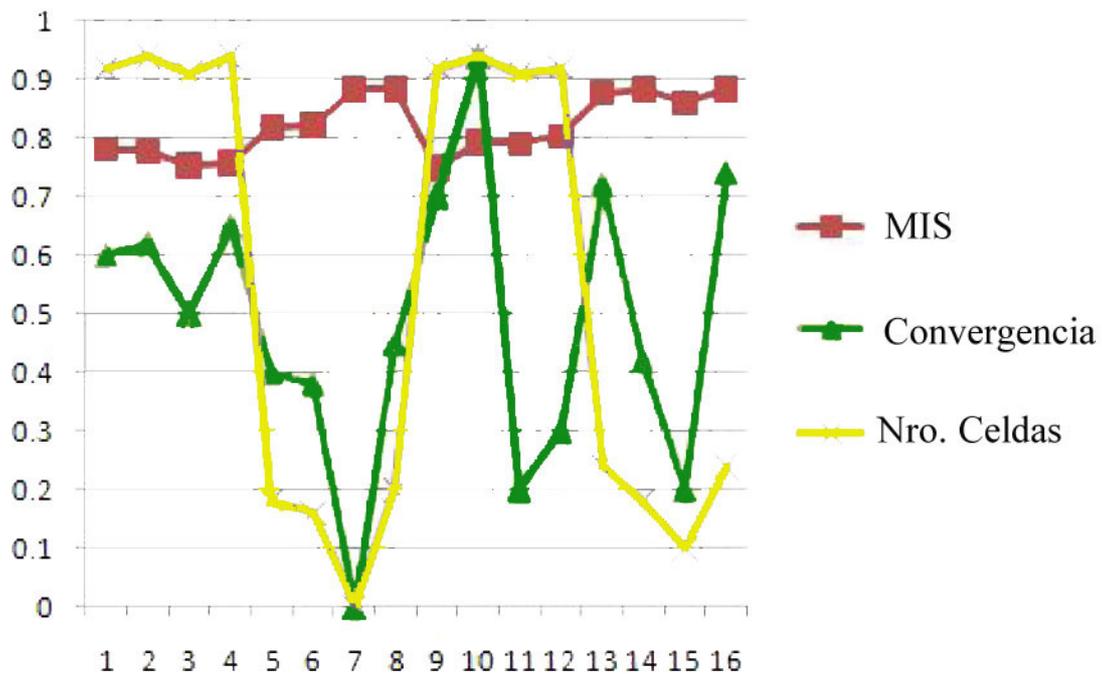


Figura.- 15 Comparación de la ejecución del AG

Paso 5: Obtención de los Resultados y Recomendaciones

Una vez analizada la imagen y realizada la optimización completa; mediante el optimizador AG, se obtuvo:

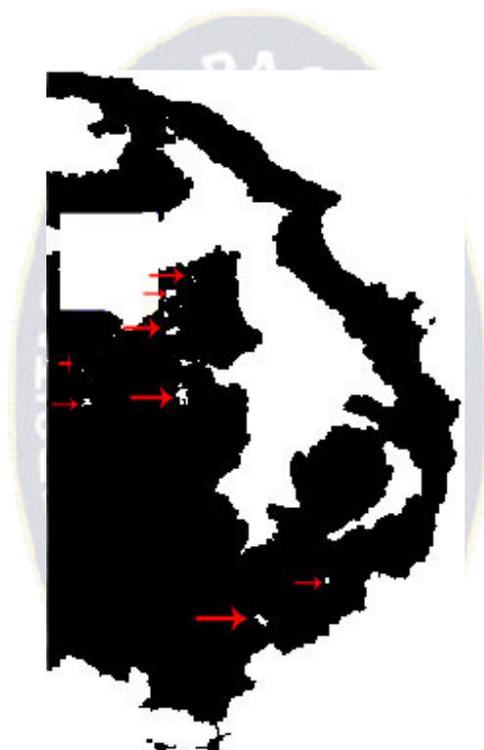


Figura.- 16 FP 4 con detección



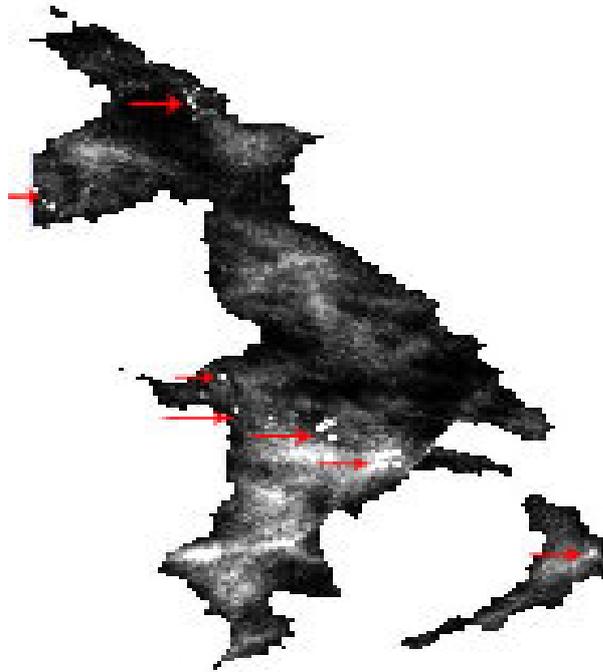


Figura.- 17 FP56 Con detección



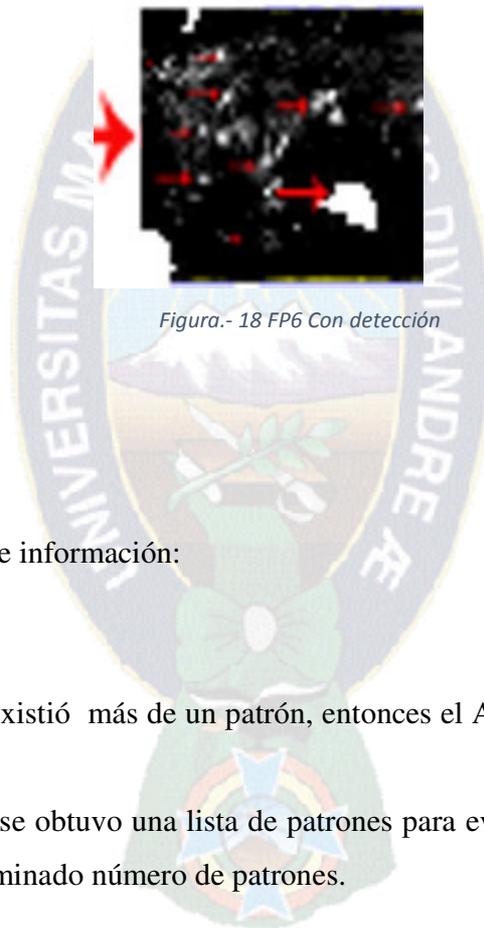
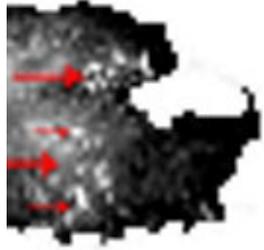


Figura.- 18 FP6 Con detección

Y dos tipos principales de información:

1. A nivel de la FP, si existió más de un patrón, entonces el AG evaluó cada patrón y lo clasificó.
2. A nivel de patrones, se obtuvo una lista de patrones para evaluar el más alto índice en común para un determinado número de patrones.

Conclusiones

La aplicación del Método Heurístico enfocado en el Algoritmo Genético y la Minería de Datos; está basada en la optimización de los índices de homogeneidad usando la MIS, permiten un óptimo diagnóstico de cáncer de mama.

El enfoque propuesto es un método sistemático y consistente para proporcionar recomendaciones y ayudar a la detección de patrones y conformación de familias de patrones.

El uso combinado de la Métrica Integral de Similitud (MIS) puede proporcionar información útil para el diagnóstico de una familia de patrones, tanto a nivel de la familia de patrones; así como, a nivel de patrón.

El enfoque propuesto se demostró y se validó a través de la aplicación del Método Heurístico en una mamografía previamente diagnosticada, realizando de este modo, la detección de lesiones. A su vez, mediante una lista clasificada de los patrones se ayudó a los profesionales a centrarse en los patrones detectados con lesiones, que podrían no haber sido identificados visualmente y sin tener un enfoque sistemático.

Al completar el optimizador Algoritmo Genético y el procesamiento de imagen mamográfica se propusieron una secuencia de detección de dos tipos principales de información:

1. A nivel de la familia de patrones; el cual indica que si existe más de una similitud para una familia en particular; entonces el Algoritmo Genético evalúa cada patrón y lo clasifica.

2. A nivel de patrones; se propone una lista de patrones individuales para clasificar y lograr el más alto índice en común; y así, determinar un número de patrones.

Mediante la entrevista, se determinó que en el Hospital de Clínicas, dentro la Unidad de Imagenología, no se cuenta con ningún tipo de Software para el procesamiento de imágenes mamográficas. Asimismo, tampoco poseen un estándar de informes de patologías sobre el Cáncer de Mama. Sin embargo, existe un repositorio de imágenes mamográficas, almacenadas en CD's, las cuales no son posibles intercambiarlas a través de ningún medio o sistema informático.

4.3 Recomendaciones.

Primera Recomendación

Recomendaciones a nivel de la familia de patrones. Si el profesional en el área, desea evaluar más de un patrón dentro de una familia de patrones, el algoritmo también se ejecuta, por lo tanto, una vez que el patrón está optimizado, se alcanza la similitud ideal, es decir, todos los patrones no son diferenciadores en común en la familia de patrones.

Recomendaciones a nivel de patrón. El Algoritmo Genético ofrece un conjunto de posibles similitudes; las cuales se podrían implementar para maximizar las características comunes dentro de cada familia de patrones.

Segunda Recomendación

Se recomienda al Ministerio de Salud, adoptar como política de salud pública, el Método Heurístico, desarrollado en esta investigación, mediante un Sistema Informático basado en Software libre y Estándares Abiertos. Dicha implementación, permitirá la distribución libre

y gratuita de este Sistema, para el beneficio de todas las unidades de Imagenología del País, Oncólogos y pacientes con el fin de optimizar el tiempo y la calidad de atención.



Bibliografía

- A. Bazzaní, A. B. (2000). *Automatic detection of clustered microcalcifications in digital mammograms using an SVM classifier*. European Symposium on Artificial Neural Networks. Bruges.
- A. Bazzaní, A. B. (2000). Automatic detection of clustered microcalcifications in digital mammograms using an SVM classifier. En *European Symposium on Artificial Neural Networks* (págs. 195-200). ESANN2000.
- A. Hojjatoleslami, L. S. (1997). An RBF based classifier for detection of microcalcifications in mammograms with outlier rejection capability. IEEE.
- A.Salomon. (1913). *Beiträge zur pathologie und klinik der mammaerzinome*. Arch Klin Chir.
- Abalo, E. (17 de 8 de 2003). *Información para el público en general*. Obtenido de www.samas.org.ar
- Albacete, M. P. (2006). Evolución cronológica del trasplante renal en España.
- Amparo Vilarrasa Andrés. (2006). SISTEMA INTELIGENTE PARA LA DETECCIÓN Y DIAGNÓSTICO DE PATOLOGÍA MAMARIA. En Amparo Vilarrasa Andrés, *SISTEMA INTELIGENTE PARA LA DETECCIÓN Y DIAGNÓSTICO DE PATOLOGÍA MAMARIA*. Madrid,.
- Antonie M., Z. O. (2001). *Application of Data Mining Techniques for Medical Image Classification* (segunda ed.). (S. Explorations, Ed.) International Workshop on Multimedia Data Minig.
- Avrim L. Blum, P. L. (1996). Selection of relevant features and examples in machine learning. *Artificial Intelligence* 97 (1997) 245-271.
- Bañobre Corpas Y, B. G. (s.f.). *Diagnóstico de Enfermedades de Transmisión Sexual mediante técnicas de Inteligencia Artificial*. . Universidad de las Ciencias Informáticas, La Habana.

- Beiträge, S. A. (1913). *zur pathologie und klinik der mammaerzinome* (101 ed.). Arch Klin Chir.
- Carla Beatriz, G. (2010). LAS PERSONAS CIEGAS, SU CUERPO, EL ESPACIO Y LA REPRESENTACIÓN MENTAL.
- Cheeseman P., J. S. (1996). Bayesian classification (AutoClass):. En G. P.-S. U.M. Fayyad, *Advances in Knowledge Discovery and Data Mining* (págs. 153-180). California: AAAI/MIT Press.
- Cheeseman, P. a. (1996). Bayesian classification (autoclass): Theory and results. In U.M. Fayyad,. En P. S. G. Piatetsky-Shapiro, *Advances in Knowledge Discovery and Data Mining* (págs. 153–180). AAAI/MIT Press.
- Chen. (1996). En I.-C. K. Chen, *Analysis of Branch Prediction Via Data Compression* (págs. 128-137). Cambridge, Massachusetts: ASPLOS VII.
- Corona, A. V. (2004). *Contribución a la diagnosis asistida sobre mamografía digitalizada mediante la aplicación sinérgica de la transformada wavelet y clasificadores neuronales*. U. Politécnica de Madrid, Tesis Doctoral.
- Elder IV, D. P. (1996). A statistical perspective on knowledge discovery in. En G. P.-S. U.M. Fayyad, *Advances in Knowledge Discovery and Data Mining* (págs. 83-115). California, EEUU: AAAI/MIT Press.
- Evangelos & Han, I., & Michalski et al., I. (s.f.).
- Fayyad, R., Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R. (1996). *Advances in knowledge and data mining*. Massachusetts: AAAI/MIT Press.
- G. te Brake, N. K. (1997). Detection criteria for evaluation of computer aided diagnosis systems. *Engineering in Medicine and Biology Society, 1996. Bridging Disciplines for Biomedicine, 18th Annual International* (págs. 1157-1159). IEEE,.
- G.Ling, Y. a. (2000). A cad system for automatic detection of clustered microcalcifications in digitized mammograms films. *IEEE Transactions on Medical Imaging*.
- García, S. M. (2012). Radiología básica - Aspectos fundamentales (Segunda Edición). En W. Herring, *Learning radiology: recognizing the basics* (págs. 1-7). Philadelphia, Pennsylvania: ELSEVIER.

- Gershon-Cohen J, S. A. (1938). Roentgenologic examination the normal breast: its evaluation in demonstrating early neoplastic changes. *AJR*.
- Giger, M. K. (1997). Feature selection and classifiers for the computerized detection of mass lesions in digital mammography. *IEEE*.
- Gold RH. (1992). The evolution of mammography. *Radiologic Clinics. North America: Radiologic Clinics*.
- Goldberg, D. E. (1989). *Genetic Algorithm in Search, Optimization and Machine Learning*, Addison-Wesley Publishing Company Inc. Reading, PA.
- Grefensette. (1986). Optimization of Control Parameters for Genetic Algorithms,. *IEEE Transactions on Systems*.
- Han, J. &. (2001). *Data Mining. Concepts and Techniques*. 548. Editorial Morgan Kaufmann.
- Hand, D. J. (1997). *Data Mining: Statistics and More*. The American Statistician.
- Hernández Orallo, 2. (s.f.).
- Hernandez, S., Fernández, C., & Baptista, P. (2006). *Metodología de la Investigación*. Mexico: Ed. Mc. Graw Hill.
- Hicken NH Mammography. (1937). *The roentgenographic diagnosis of breast tumours by means of contrast media*. Surgery, Gynaecology and Obstetrics.
- I. El-Naqa, Y. Y. (2002). A support vector machine approach for detection of microcalcifications. *IEEE*.
- I. El-Naqa, Y. Y. (2002). *A support vector machine approach for detection of microcalcifications*. *IEEE*.
- Ingleby H, G.-C. J. (1960). *Comparative anatomy, pathology and roentgenology of the breast*. Philadelphia: University of Pennsylvania.
- Jiao, J. a. (2000). *Understanding Product Family for Mass Customization by Developing Commonality Indices*, (11 ed., Vol. III). *Journal of Engineering Design*.
- Jong. (1975). Analysis of the Behavior of a Class of Genetic Adaptive Systems.

- K. Doi, M. G. (1993). Development of digital processing techniques for computer aided diagnosis in radiographic images. *The Third International Conference on Image Management and Communication in Patient Care* (págs. 110-115). .IMAC93.
- Kamber, J. H. (2006). Data Mining:Concepts and Techniques. *A M S T E R D A M B O S T O N*.
- Karayiannis, T. W. (1998). *Detection of microcalcifications in digital mammograms using wavelets*. IEEE.
- Karssemeijer, B. a. (1997). *Detection criteria for evaluation of computer aided diagnosis systems*. Engineering in Medicine and Biology Society.
- Karssemeijer, W. V. (2000). *Normalization of local contrast in mammograms*. IEEE Transactions on Medical Imaging.
- Keinssmidt O, B. Z. (1927). *Die Klinik der bösartigen geschwulste*.
- Kossel F. (1967). Physical aspects of xeroradiology as a tool in cancer. *Prog Clin Cancer*.
- Lampinen J., S. S. (1995). Wood surface inspection system based on generic visual features. *International Conference on Artificial Neural Networks* (págs. 9-13). Paris,: ICANN.
- Leborgne R. (1951). Diagnosis of tumours of the breast by simple roentgenography. *AJR*.
- Leborgne R. (1953). *The breast in roentgen diagnosis*. Montevideo Uruguay: Impresora.
- Liebstein, L. (2002). Data Mining. Teoría e Práctica. (16). Rio, Brasil.
- Lu H., R. S. (1995). NeuroRule: A Connectionist Approach to Data. *Proceedings of the 21th International Conference on Very Large Data*, (págs. 478-489).
- M.A. Kupinski, M. G. (1997). *Feature selection and classifiers for the computerized detection of mass lesions in digital mammography*. 2460-2463.
- Michalski et al, 1. (s.f.).
- Michalski, S. H. (s.f.). 1983.
- Misiti, M., Misiti, Y., Oppenheim, G., & Poggi, J. M. (1997). *Wavelet Toolbox For Use with MATLAB*. The MathWorks Inc.

- MOLINA, L. (1998). *Data mining no processo de extração de conhecimento de bases de Tesis de máster*. São Carlos (Brasil): Instituto de Ciências Matemáticas e Computação. Universidad de São Paulo.
- Morales, E. (2003). *Descubrimiento de Conocimiento en Bases de Datos*.
- Nasukawa, T., & Nagano, t. (2001). *Text analysis and knowledge mining system*. IBM Systems Journal, knowledge management.
- Omar Danilo Castrillón, W. A. (2010). DISEÑO DE UNA HIPERHEURÍSTICA PARA LA PROGRAMACIÓN DE LA PRODUCCIÓN EN AMBIENTES JOB SHOP. *Ingeniare. Revista chilena de ingeniería*, 203-214.
- Percival, D. B. (2002). *Wavelet Methods for Time Series Analysis*. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press.
- Piatetski-Shapiro et al., 1., Chen et al., 1., & Mannila, 1. (s.f.).
- Piatetski-Shapiro, G. U. (1996). *From data mining to Knowledge*. California, EEUU: AAAI Press/MIT Press.
- R.M. Nishikawa, Y. J. (1992). Computer aided detection of clustered microcalcifications. *IEEE International Conference on Man and Cybernetics Systems* (págs. 1375-1378). IEEE .
- R.M. Nishikawa, Y. J. (1992). *Computer aided detection of clustered microcalcifications”*. *IEEE International Conference on Man and Cybernetics Systems*. IEEE International Conference on Man and Cybernetics Systems.
- R.N. Strickland. (1994). Wavelet transforms for detecting microcalcifications in mamography. (págs. 402-406). IEEE.
- Raul Landmann, R. H. (2011). Un enfoque heurístico para la programación de la producción en la industria de la fundición mediante lógica difusa.
- Raúl Mejía Ibáñez. (s.f.). Metodología de la investigación.
- Righi A. (1907). *Die Bewegung der Ionen bei der Elektrischen*. Leipzig.
- Rosete Suárez A, R. D. (2009). Predicción de pacientes diabéticos. Preprocesado para Minería de Datos. *Informatica Médica*.

- Schiabel, A. P. (2001). Classifying clusters of microcalcifications in digitized mammograms by artificial neural network. IEEE.
- Seabold PS. (1931). *Roentgenographic diagnosis of diseases of the breast*. Surg Gynecol Obstet.
- Servente & García-Martínez, 2., Perichinsky & García-Martínez, 2., Perichinsky et al., 2., Perichinsky et al., 2., & Perichinsky et al., 2. (s.f.).
- Simoff S., D. C. (2002). *Multimedia Data Mining between Promise and Problems* (Tercera ed.). SIGKDD Explorations.
- Strickland, R. (1994). *Wavelet transforms for detecting microcalcifications in mamography*. IEEE International Conference.
- T.C. Wang, N. K. (1998). *Detection of microcalcifications in digital mammograms using wavelets*. IEE.
- Thevenot, H. J. (2006). *A Comprehensive Metric for Evaluating Commonality in a Product Family*, (17 ed., Vol. II). Journal of Engineering Design.
- Vallejo Delgado N, R. J. (2012). *Aplicación de técnicas de minería de datos para el diagnóstico prematuro de cáncer*.
- Verma., B. (1790-1793). *A neural network bse technique to locate and classify microcalcifications in digital mammograms*. IEEE 1998.
- Vogel, W. (1932). *Die roentgendarstellung der mammatumoren*. Arch Kin Chir.
- W.J.H. Veldkamp. (2000). Normalization of local contrast in mammograms. IEEE.
- Walker, J. S. (1999). *Primer on Wavelets and their Scientific Applications*. Chapman & Hall.
- Warren SL. (1930). *Roentgenologic study of the breast*.
- Y. Jiang, R. N. (1997). Computerized classification of malignant and benign clustered microcalcifications in mammograms. *Computerized classification of malignant and benign clustered microcalcifications in mammograms* (págs. 521-523). Engineeririg in Medicine and Biology Society.

Anexos



ANEXO A: OPERACIONALIZACION DE LAS VARIABLES

HIPOTESIS	Variables	Dimensión	Indicador	Instrumentos
La aplicación de un método heurístico para la selección de patrones obtenidos a través de la minería de datos facilitará el diagnóstico oportuno del cáncer de mama	Variable Independiente Aplicación de un método heurístico para la selección de patrones obtenidos a través de la minería de datos	Método Heurístico	Identificación del problema	Guía de Observación
			Definición de Estrategias	Guía de Observación
			Evaluación de resultados	Guía de Observación
	Selección de patrones con minería de datos	Recopilación de datos	Guía de Observación	
		Análisis automático de datos	Guía de Observación	
		Selección de Patrones	Cuestionario	
	Variable Dependiente: Diagnóstico oportuno del cáncer de mama	Diagnostico	Clasificación de Patrones	Guía de Observación Cuestionario
			Construcción de Familias de Patrones	Guía de Observación Cuestionario
			Evaluación e interpretación de patrones	Guía de Observación Cuestionario
	Cáncer de mama	Cáncer de mama	Identificación de Lesión	Guía de Observación Cuestionario
Identificación de Etapa			Guía de Observación Cuestionario	

ANEXO B: GUÍA DE ENTREVISTA

Nombre Completo: Dr, Martha Lucia Aguirre de Delgado, Jefa de la Unidad de Imagenología Del Hospital de Clínicas La Paz – Bolivia

OBJETO.

El objetivo de esta entrevista es conocer los procedimientos que sigue para la selección de patrones en las mamografías para detectar el cáncer de mama en todas sus etapas.

I. INSTRUCCIONES.

Por favor responda con absoluta sinceridad

II. PREGUNTAS.

1. ¿Qué tipo de Imagen es analizada comúnmente y para la detección de cáncer de mama? ¿por qué?

Mamografía – Ecografía – Tomografía - Resonancia Magnética - Radiología

Tomografía Axial

R: Generalmente se realizan, Rayos X mediante la Mamografía, y en algunos casos la digitalización de la Imagen Mamográfica

2. Considerando las lesiones benignas, malignas, dudosas o sospechosas, de manera muy general ¿qué pasos realiza para la detección de estas lesiones en los tejidos?

R: Inspección simple ocular y de tacto.

3. ¿Dentro del análisis de una o varias áreas o zonas, realiza alguna comparación de algún patrón en común en las lesiones encontradas?

Ninguna, se aplica el conocimiento directo a la detección.

4. ¿Realiza alguna clasificación en particular de estas lesiones?

La clasificación se la realiza de acuerdo a los BI-RADS(informe estándar, de clasificación de patologías).

5. ¿Qué confusiones se presentan en la interpretación de las imágenes?

De momento ninguno, pero las micro calcificaciones son tan pequeñas que no se logran realizar el análisis.

6. ¿De manera muy general que casos especiales se le presentaron? Explique

Lo más común es la extirpación de la mama.

7. ¿El formato de los informes que se presentan luego de un examen, se elaboran siguiendo un formato estándar en todos los centros de diagnóstico; o son elaborados a criterio del profesional que los realiza?

Son de manejo interno; así como los formatos de informes.

8. ¿Se cuenta con algún repositorio digital de estas imágenes e informes para hacer un seguimiento?

La información es guardada en CD's.

Muchas gracias por su colaboración

Nombre Completo: Dra. Paola Castro. Hospital de Clínicas La Paz – Bolivia

OBJETO.

El objetivo de esta entrevista es conocer los procedimientos que sigue para la selección de patrones en las mamografías para detectar el cáncer de mama en todas sus etapas.

III. INSTRUCCIONES.

Por favor responda con absoluta sinceridad

IV. PREGUNTAS.

1. ¿Qué tipo de Imagen es analizada comúnmente y para la detección de cáncer de mama? ¿por qué?

Mamografía – Ecografía – Tomografía - Resonancia Magnética - Radiología
Tomografía Axial

R: Mamografía

2. Considerando las lesiones benignas, malignas, dudosas o sospechosas, de manera muy general ¿qué pasos realiza para la detección de estas lesiones en los tejidos?

R: La paciente realiza el pago correspondiente, y de acuerdo a los análisis, se termina generalmente micro calcificaciones benignas.

3. ¿Dentro del análisis de una o varias áreas o zonas, realiza alguna comparación de algún patrón en común en las lesiones encontradas?

Ninguna.

4. ¿Realiza alguna clasificación en particular de estas lesiones?

No.

5. ¿Qué confusiones se presentan en la interpretación de las imágenes?

Algunas veces se consideró tejido sano y no se detectó el micro calcificaciones.

6. ¿De manera muy general que casos especiales se le presentaron? Explique

Una paciente asintomática, que se hizo la mamografía y se le detecto Cancero de Mama.

7. ¿El formato de los informes que se presentan luego de un examen, se elaboran siguiendo un formato estándar en todos los centros de diagnóstico; o son elaborados a criterio del profesional que los realiza?

Es de manejo interno.

¿Se cuenta con algún repositorio digital de estas imágenes e informes para hacer un seguimiento?

La información es guardada en CD's.

Muchas gracias por su colaboración

Apéndices



Apéndice A Terminología clínica

A.1. BI-RADS (Breast Imaging Reporting And Data System)

Fue creado por el Colegio Americano de Radiología (ACR) en 1993 (con ediciones posteriores en 1995 y 1998), es un sistema estandarizado de descripción de las lesiones mamográficas y de elaboración del informe radiológico.

Es la estandarizar de los informes, establecer un grado de sospecha de malignidad y permitir la toma de decisiones para así maximizar la detección del cáncer en estado precoz y reducir los falsos positivos en las biopsias. Esta clasificación otorga un valor pronóstico establecido por estudios que correlacionaron radiología y biopsias, determinando el porcentaje de malignidad para cada categoría BI-RADS.

Los objetivos son:

Objetivos:

- a. Estandarizar los informes mamarios (Hablar el mismo idioma)
- b. Disminuir la confusión en la interpretación de las imágenes.
- c. Facilitar el seguimiento.

BI - RADS / Categorías

BI-RADS	DESCRIPCION
0	Evaluación adicional
1	Negativa
2	Benigna
3	Probablemente benigna
4	Anormalidad sospechosa
5	Altamente sugestiva de malignidad
6	Malignidad conocida

Tabla 5 BI - RADS / Categorías

Objetivos de BI RADS

- **Herramienta útil que ayuda al control de calidad diseñada para estandarizar los informes mamográficos, reducir la confusión en la interpretación y facilitar el control posterior.**
- **Reportes claros, comprensibles y decisivos**
- **Unificar criterios entre radiólogos y clínicos (vigilancia o biopsia)**

Glosario

Masa:

Lesión ocupante de espacio en dos proyecciones, con un descriptor (morfología) y dos modificadora (margen y densidad)

Morfología: redonda ovalada, irregular

Margen: circunscrito, micro lobulado, mal definido, especulado, oscurecido por el parénquima adyacente

Densidad: (con relación al tejido mamario adyacente)

Calcificaciones

Se clasifican según su

Morfología, Típicamente benignas, Sospecha intermedia

Alta sospecha de malignidad

Distribución

Difusas, regionales, agrupadas, lineales o segmentarias