UNIVERSIDAD NACIONAL DE TRUJILLO
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
ESCUELA ACADÉMICO PROFESIONAL DE
INGENIERÍA INFORMÁTICA

“DISEÑO DE UN ALGORITMO DE RECONOCIMIENTO EN LA IDENTIFICACIÓN DE TUMORES DE MAMA”

TESIS
PARA OPTAR EL TÍTULO PROFESIONAL DE INGENIERO INFORMATICO

AUTORES : Bach. BARDALES BRUNO, EDUARDO.

ASESOR : ING. ARAUJO GONZALEZ, CHRISTIAN.

TRUJILLO – PERÚ

2013
Dedicatoria

A mis padres,
Santos y María.
A mis abuelos
Adán y Rosa.
A mi Hermano
David Bardales.
Y a mis Tíos,
Amparo Bardales y Antonio Bazán.
AGRADECIMIENTO

En la presente tesis expreso eternamente el agradecimiento a Dios por sus bendiciones para llegar hasta donde he llegado y darme las fuerzas conocimientos necesarios y dirección en todo.

Agradezco a mis padres María Bruno y Santos Bardales, por su apoyo y confianza para cumplir los objetivos como persona y estudiante, a ellos por su especial trato y su total confianza depositada en mi persona.

A mi hermano David Bardales por estar siempre presente acompañándome en las etapas más difíciles. También agradezco a mis tíos y tías: Antonio Bazán, Melva, Marilú, Edita, Jaime; y de manera muy especial a mi tía Amparo Beatriz Bardales Alvarado quien ya partió a la presencia del Altísimo y que permanentemente me apoyo con espíritu alentador, contribuyendo incondicionalmente a lograr las metas y objetivos propuestos pues es quien me inicio en este trabajo de investigación ya mis abuelos Rosa Alvarado, Adán Bardales.

A mis profesores de la carrera de ingeniería informática, ya que durante toda mi carrera profesional con su esfuerzo y dedicación han aportado parte de sus conocimientos tanto a mi formación profesional como de también de la vida impulsándome hacia delante para abrirme vista hacia un mundo globalizado de hoy en día.

Debo agradecer también de manera muy especial y sincera al ING. CHRISTIAN ARAUJO GONZALEZ por aceptarme para realizar esta tesis de pregrado bajo su dirección. Su apoyo y confianza en mi trabajo y su capacidad para guiar mis ideas ya que han sido un aporte inapreciable, no solamente en el desarrollo de esta tesis, sino también en mi formación e iniciación para desarrollarme como investigador. Las ideas propias, siempre enmarcadas en su orientación y rigurosidad, han sido la clave del buen trabajo que hemos realizado juntos, el cual no se puede concebir sin su siempre oportuna participación. Le agradezco también el haberme facilitado siempre los medios suficientes para llevar a cabo todas las actividades propuestas durante el desarrollo de esta tesis.

Finalmente agradezco a todas las personas que contribuyeron directa o indirectamente en la realización de la presente tesis.
# Índice

AGRADECIMIENTO ..................................................................................................................... III

ÍNDICE ........................................................................................................................................ IV

ÍNDICE DE FIGURAS .............................................................................................................. VI

ÍNDICE DE TABLAS ................................................................................................................ VII

RESUMEN ........................................................................................................................................ VIII

ABSTRACT ....................................................................................................................................... IX

CAPÍTULO I: PLAN DE INVESTIGACIÓN .................................................................................. 1

1.1. DESCRIPCIÓN DEL PROBLEMA ....................................................................................... 1
1.2. FORMULACIÓN DE PROBLEMA ....................................................................................... 1
1.3. HIPÓTESIS ........................................................................................................................ 1
1.4. OBJETIVOS ....................................................................................................................... 2
  1.4.1. General ....................................................................................................................... 2
  1.4.2. Específico ................................................................................................................... 2
1.5. JUSTIFICACIÓN ................................................................................................................ 2
  1.5.1. Desde el punto de vista Computacional ................................................................. 2
  1.5.2. Desde el punto de vista Médico ............................................................................. 3
1.6. LIMITACIONES ............................................................................................................... 3
1.7. ORGANIZACIÓN DE ESTUDIO .................................................................................... 4

CAPÍTULO II: MARCO TEÓRICO .............................................................................................. 5

2.1. EL CÁNCER DE MAMA .................................................................................................... 5
  2.1.1. Definición .................................................................................................................. 5
  2.1.2. Tipos de Cáncer ....................................................................................................... 5
  2.1.3. Etapas del Cáncer de Mama .................................................................................. 6
  2.1.4. El Cáncer de Mama en el Perú y en el Departamento de la Libertad ...................... 8
  2.1.5. Tratamientos del Cáncer de Mama ....................................................................... 9
  2.1.6. Diagnósticos del Cáncer de Mama .......................................................................... 10
2.2. DISEÑO DE ALGORITMOS ............................................................................................. 11
  2.2.1. Diseño de Algoritmos ............................................................................................. 11
  2.2.2. Arquitecturas de los Sistemas de Identificación y Clasificación de Patrones o Formas 12
  2.2.3. El Procesamiento Digital de Imágenes en la extracción de información para la Identificación .......................................................... 16
  2.2.4. La Inteligencia Artificial en el Reconocimiento y Clasificación de Patrones o Formas 30

CAPÍTULO III: DISEÑO DE LA INVESTIGACIÓN ..................................................................... 49

3.1. TIPO DE LA INVESTIGACIÓN ......................................................................................... 49
3.2. DISEÑO DE LA INVESTIGACIÓN .................................................................................... 49
3.3. POBLACIÓN Y MUESTRA ................................................................. 49
3.4. VARIABLES DE ESTUDIO ............................................................. 50
3.5. TÉCNICAS E INSTRUMENTOS ...................................................... 50
3.5.1. Técnicas ................................................................................. 50
3.5.2. Instrumentos ......................................................................... 50

CAPÍTULO IV:RESULTADOS .......................................................... 51

4.1. ANÁLISIS DE MODELOS DE ALGORITMOS DE RECONOCIMIENTO Y CLASIFICACIÓN DE TUMORES DE MAMA ................................................................. 51
4.1.1. Algoritmo de Clasificación Difusa para Tumores de Mama ........................................ 53
4.1.2. Algoritmo de Selección de Rasgos de reconocimiento para el Diagnóstico de Tumores Mamarios ........................................................................................................... 55
4.1.3. Algoritmos Genéticos de Reconocimiento para Diagnósticos de Tumores de Mama ... 57
4.1.4. Algoritmo de Clasificación y Reconocimiento de Patrones de Gabor para Diagnósticos de Tumores Mamarios ................................................................. 59

4.2. ANÁLISIS COMPARATIVO DE MODELOS DE ALGORITMOS DE RECONOCIMIENTO Y CLASIFICACIÓN DE TUMORES DE MAMA ............................................ 62

4.3. DISEÑO DEL ALGORITMO DE RECONOCIMIENTO EN LA IDENTIFICACIÓN DE TUMORES DE MAMA ................................................................. 68
4.3.4. Descripción del Diseño del Algoritmo de Reconocimiento en la Identificación de Tumores de Mama ...................................................................................... 78
4.3.5. Propósito del Diseño del Algoritmo de Reconocimiento en la Identificación de Tumores de Mama ...................................................................................... 82

CAPÍTULO V: DISCUSIÓN DE RESULTADOS .................................. 88

5.1. ANÁLISIS Y DISCUSIÓN DE LA MEDICIÓN DE LA VARIABLE “MEJORA EN LA IDENTIFICACIÓN DE TUMORES DE MAMA” ................................................................. 88
5.1.1. DATOS GENERALES DEL ESTUDIO DE PRECIÓS DE CONTENIDOS AL RECUPERAR INFORMACIÓN ................................................................. 88
5.1.2. HÍPÓTESIS Y VARIABLE ESTADÍSTICAS ..................................... 88
5.1.3. RESULTADOS OBTENIDOS DEL ESTUDIO DE “MEJORA EN LA IDENTIFICACIÓN DE RECONOCIMIENTO DE TUMORES DE MAMA” ................................................................. 90
5.1.4. ................................................................................................. 91
5.1.5. DISCUSIÓN DE RESULTADO .................................................... 93

CAPÍTULO VI: CONCLUSIONES Y RECOMENDACIONES ............... 94

6.1. CONCLUSIONES ........................................................................ 94
6.2. RECOMENDACIONES ................................................................. 94
6.3. TRABAJOS FUTUROS ................................................................. 94
6.4. REFERENCIAS BIBLIOGRÁFICAS ............................................. 95
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Vecindad de un píxel</td>
<td>19</td>
</tr>
<tr>
<td>2</td>
<td>Peculiaridad de la implementación en paralelo</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>Plantillas para reconocimiento de píxeles esqueletales</td>
<td>22</td>
</tr>
<tr>
<td>4</td>
<td>Referencia Técnica Algoritmo para reconocimiento de píxeles esqueletales</td>
<td>24</td>
</tr>
<tr>
<td>5</td>
<td>Diseño Algorítmico de Clasificación Difusa</td>
<td>69</td>
</tr>
<tr>
<td>6</td>
<td>Diseño Algorítmico de Clasificación de Rasgos</td>
<td>70</td>
</tr>
<tr>
<td>7</td>
<td>Ejemplo de Wavelet de Gabor</td>
<td>71</td>
</tr>
<tr>
<td>8</td>
<td>Diseño de un proceso de filtro de Wavelet de Gabor</td>
<td>72</td>
</tr>
<tr>
<td>9</td>
<td>Diseño Algorítmico de Wavelet de Gabor</td>
<td>73</td>
</tr>
<tr>
<td>10</td>
<td>Reconocimiento de imagen en escala de filtros de grises</td>
<td>79</td>
</tr>
<tr>
<td>11</td>
<td>Arquitectura de entrenamiento para la identificación del reconocimiento del tumor de mama</td>
<td>84</td>
</tr>
<tr>
<td>12</td>
<td>Histograma de efectividad</td>
<td>91</td>
</tr>
<tr>
<td>13</td>
<td>Regiones de Rechazo y Aceptación obtenidas de las pruebas estadísticas realizadas</td>
<td>92</td>
</tr>
</tbody>
</table>
Índice de Tablas

Tabla 1: Conceptos básicos de la Inteligencia Artificial .............................................................. 30
Tabla 2: Variables de Estudio ..................................................................................................... 50
Tabla 3: Cuadro Comparativo de Modelos ................................................................................. 64
Tabla 4: Resultados obtenidos del estudio de efectividad de identificación de reconocimiento de tumores de mama ................................................................. 90
Tabla 5: Resultados obtenidos dentro del análisis estadísticos en la identificación de reconocimiento de tumores de mama ................................................................. 91
Tabla 6: Resultados obtenidos dentro del análisis Estadístico la identificación y reconocimiento de tumores de mama ................................................................. 92
RESUMEN

En la actualidad la principal causa de muerte por enfermedad es el cáncer de mama. El control integral del cáncer abarca la prevención, la detección precoz, el diagnóstico y tratamiento, la rehabilitación y los cuidados que se dan al mismo, según Burga y Valencia (2001).[25]

La sensibilización del público en general sobre el problema del cáncer de mama y los mecanismos de control, así como la promoción de políticas y programas adecuados, son estrategias fundamentales para el control poblacional del cáncer de mama. Éste tipo de cáncer (mama) puede afectar a distintos órganos como páncreas, mama, cuello uterino, próstata entre otros. Él cáncer mamario presenta mayor número de casos y esta enfermedad presenta estados o fases de desarrollo siendo una de ellas la metástasis que es la proliferación de células cancerígenas a órganos cercanos al del origen y causante de un gran porcentaje de muertes.

Para su diagnóstico de este tipo de cáncer existen diferentes pruebas y una de ellas es mediante el reconocimiento de imágenes tomadas en tiempo real y con una respuesta muy segura para su diagnóstico.

Ahora se requiere que estos sistemas sean sumamente rápidos para poder ser aplicados en tiempo real, para ello en un ambiente no controlado existen algunos inconvenientes como son: la condición de la luz de cierta escena a procesar, problemas como contraste de la imagen, presencia de ruido por la calidad de respuesta, entre otros.

Por esto, es necesario establecer una representación formal de los elementos básicos que intervienen en el desarrollo del diseño algorítmico, esto a través de un algoritmo de reconocimiento y lograr mejoras en la identificación de tumores de mama, además de estar a la vanguardia con los nuevos avances tecnológicos que se presentan en reconocimiento de algoritmos a través de imágenes.

El presente trabajo de investigación denominado “DISEÑO DE UN ALGORITMO DE RECONOCIMIENTO EN LA IDENTIFICACIÓN DE TUMORES DE MAMA” ayudará a mejorar la identificación de tumores de mama utilizando el diseño de un algoritmo de reconocimiento.

Este diseño algorítmico propuesto toma inicio en el ingreso de la mamografía inicial posteriormente se procede a reducir el área donde se está trabajando como posibles ruidos o problemas que dificultan el rápido diagnóstico y la acertación del grado con el que debe ser la respuesta del diagnóstico seguidamente el algoritmo propuesto toma internamente la capacidad de hacer un contraste procediendo a llevar a cabo el procesamiento de la imagen ingresada. Toma la detección de rasgos buscando alguna anomalía está a la vez nos presenta muestras que nos pueden dar una respuesta acertada a través del proceso de filtro que se dio a la imagen, todo ello es el proceso que nos servirá para la identificación rápida de un tumor de mama y determinar el grado de anomalía y etapa en la que se encuentra el tumor.Como resultado del estudio se concluye que es beneficioso llevar a cabo el desarrollo de un diseño algorítmico de reconocimiento, para poder mejorar la identificación y diagnóstico de tumores de mama, permitiendo una mayor aceptación en base a la rapidez del diagnóstico y a la seguridad de la respuesta.
ABSTRACT

Currently the leading cause of death by disease is breast cancer. Comprehensive cancer control includes prevention, early detection, diagnosis and treatment, rehabilitation and care given to it as Burga and Valencia (2001). [25]

The general public awareness about the problem of breast cancer and control mechanisms, and promoting policies and programs, are key strategies for controlling breast cancer population. This type of cancer (breast) can affect different organs such as pancreas, breast, cervix, prostate and others. Breast cancer has a greater number of cases and the disease presents states or developmental stages one of which is the metastasis of cancer cell proliferation in organs close to the source and cause of a large percentage of deaths.

For diagnosis of this cancer there are different tests and one of them is through the recognition of images taken in real time and with a very safe answer for diagnosis.

Now these systems are required to be extremely fast to be applied in real time to this in an uncontrolled environment there are some drawbacks such as: the condition of the light in a certain scene to process, problems such as image contrast, the presence of noise response quality, among others.

Therefore, it is necessary to establish a formal representation of the basic elements involved in the development of algorithmic design, that through algorithm recognition and improvements in identifying breast tumors, in addition to being at the forefront with new technological advances in recognition algorithms presented through images.

This research entitled "DESIGN OF A RECOGNITION ALGORITHM FOR IDENTIFICATION OF BREAST TUMORS" help improve the identification of breast tumors using the design of a recognition algorithm.

The proposed algorithmic design takes income starting in the initial mammogram then proceeds to reduce the area where you are working as any noise or problems that hinder the rapid diagnosis and aeration the degree to which the response should be below the diagnostic proposed algorithm takes the capacity internally to contrast proceeding to carry out the image processing entered. Take detecting abnormality looking features is both presents samples that can give us the right answer through the filter process that gave the picture, all this is the process that will serve for the rapid identification of a tumor breast and determining the degree of abnormality and stage where the tumor

As a result of the study concluded that it is beneficial to carryout the development fare cognition algorithm design, to improve the identification and diagnosis of breast tumors, allowing greater acceptance based on the speed of diagnosis andres ponsense security.
CAPÍTULO I: PLAN DE INVESTIGACIÓN

1.1. Descripción del Problema

A nivel mundial el principal problema del desarrollo del cáncer de mama en su totalidad es el tardío reconocimiento e identificación del tumor que se le da o la realización de diferentes pruebas que se le hace al paciente para determinar el grado, tipo y estado en el que se encuentra el tumor. Según Burga y Valencia (2001). [25]

Del reconocimiento e identificación de tumores mamarios por conceptos médicos se sabe que se debe realizar pruebas muy sencillas rápidas y de fácil diagnóstico, reconocimiento e identificación. Según Burga y Valencia (2001). [25]

Todo esto conlleva a las siguientes características problemáticas:

- Desconcierto en el paciente cuando se le hace la toma de diferentes exámenes de mama y no se le da una respuesta eficaz y adecuada.
- Exámenes mamo grafico (imágenes mamarias), recurrentes que se derivan hacia un segundo examen Biopcial (Biopsia) mucho más riguroso para el paciente al realizar la extracción de un trozo de mama; esto supondría una rebeldía y diseminación del tumor de mama por todo el cuerpo de la persona.
- Uso de métodos tediosos y complicados para el reconocimiento de tumores mamarios, dentro del paradigma científico y medico experimental pero en la práctica no se da y esto con lleva al desconcierto del paciente.
- La pérdida de tiempo en el reconocimiento de tumores mamarios, ocasionada por los tediosos métodos e infinidad pruebas que se hace en el diagnóstico de un cáncer mamario.

1.2. Formulación de Problema

¿Cómo podemos mejorar el reconocimiento en la identificación de tumores de mama?

1.3. Hipótesis

El diseño de un algoritmo de reconocimiento de tumores de mama permitirá mejorar la identificación de los tumores de mama.
1.4. Objetivos

1.4.1. General

Diseñar un algoritmo de reconocimiento de tumores de mama para mejorar la identificación de tumores cancerígenos.

1.4.2. Específico

- Recopilar y Analizar la información referente al cáncer de mama obtenida de diversos medios.
- Hacer el análisis de los algoritmos de reconocimiento de tumores de mama para la reducción de la complejidad.
- Diseñar un algoritmo de reconocimiento de un tumor de mama por medio de una imagen.
- Comprobar la mejora del reconocimiento en la identificación de un tumor de mama.

1.5. Justificación

1.5.1. Desde el punto de vista Computacional

Promueve la utilización de las Tecnologías de la Información y Comunicación (TIC), en la solución de problemas prácticos para mejorar la calidad de vida de la población, mediante la aplicación de un diseño de un “algoritmo de reconocimiento en la identificación de tumores de mama” permitiendo al ámbito médico mejorar el reconocimiento y la identificación de los tumores de mama, integrando de esta manera la ciencia informática y la tecnología computacional, en la solución de los problemas médicos.

Así mismo el diseño del algoritmo como parte del avance científico en el área de las ciencias de la computación de nuestro país va a ser de vital importancia dado que el estudio que se realiza en la presente investigación, apoya y fortalece el área de la inteligencia artificial, a la cual pocos profesionales de la ciencias de la computación brindan sus conocimientos para fortalecerla y hacer de ella un área que apoye el desarrollo económico de nuestro país, generando puestos de trabajo en el área mencionada, por eso la investigación de esta tesis, analiza diversos métodos de la inteligencia artificial, y propone una solución de clasificación de patrones para el reconocimiento de tumores
de mama los cuales, dependiendo de su gravedad causan el cáncer de mama una enfermedad que ocasiona muchas muertes en nuestro país.

Adicionalmente este algoritmo es una propuesta que pueda competir con el resto de algoritmos ya comerciales implantados en diversos equipos que cuentan las diferentes instituciones.

1.5.2. Desde el punto de vista Médico

La aplicación del diseño de un “algoritmo de reconocimiento en la identificación de tumores de mama” apoya en el ámbito médico en lo siguiente:

- Apoya en el diagnóstico e investigación médica mediante el diseño de un algoritmo de reconocimiento para la identificación de tumores de mama.

- Apoya al reconocimiento y clasificación de tumores de mama para su identificación médica.

El diseño de un algoritmo para el reconocimiento de tumores de mama es para incrementar la exactitud y utilidad del reconocimiento de tumores de mama resumiendo la tediosa tarea identificación de tumores en el ámbito médico.

1.6. Limitaciones

En el presente estudio se presentan las siguientes limitaciones:

- Bibliográficas: Las fuentes bibliográficas relacionadas con este tema son escasas y se tuvo que recurrir a bibliotecas de otras universidades en otras provincias o departamentos, así como el uso de fuentes electrónicas.

- Operativa: Para la presente investigación se requiere de ciertos permisos especiales de la institución que nos brinda la información necesaria para el desarrollo del mismo, como para obtener ciertas imágenes de su base de datos y así poder realizar las pruebas necesarias en el proyecto.

- Social: La falta de instituciones que apoyen a las personas con cáncer de mama, y no apostando por el desarrollo de nuevas técnicas, métodos, entre otros. Y permitan que la detección de dicho cáncer sea más rápido y económico.
1.7. Organización de Estudio

El trabajo cuenta con 6 capítulos y está organizado de la siguiente manera:

En el Capítulo 1, describe el planteamiento del problema a estudiar, partiendo desde la realidad problemática, el problema concreto, la hipótesis, los objetivos, la justificación y las limitaciones del estudio.

En el Capítulo 2, se presenta el marco teórico donde se aborda temas como procesamiento digital de imágenes, el conjunto de conocimientos teóricos y tecnológicos que sustentan el estudio realizado y Métodos de reconocimiento de tumores de mama.

En el Capítulo 3, se presenta los materiales y métodos utilizados para realizarla presente investigación y poder comprobar la hipótesis planteada.

En el Capítulo 4, se presenta los resultados donde se ven el diseño de algoritmos para reconocer los tumores de mama propuesta, desde la elección de un modelo de referencia utilizando un análisis comparativo, hasta la muestra a utilizar para la implementación del modelo.

En el Capítulo 5, comprende todos los resultados obtenidos luego de la implementación del algoritmo muestra.

En el Capítulo 6, está comprendido por las conclusiones obtenidas luego de la investigación y las recomendaciones que el estudio arroja como resultado de la investigación realizada, así como trabajos futuros.

Finalmente se presenta las referencias bibliográficas, referencias web usadas y los anexos.
CAPÍTULO II: MARCO TEÓRICO

2.1. El Cáncer de Mama

Para un mejor entendimiento se detallará a continuación algunas definiciones que intervienen en el proceso del cáncer mamario:

2.1.1. Definición

La mama está dividida en unas 20 partes (lóbulos), que a su vez se dividen en partes más pequeñas (lobulillos), unidas por unos conductos que terminan en el pezón. A la proliferación de células malignas en los lóbulos o en los conductos le llamamos cáncer de mama que en medicina le pondremos con el nombre científico de carcinoma mamario.

Si las células malignas se proliferan en los lobulillos le llamamos "cáncer de mama lobular" y si se proliferan en los conductos le llamaremos "cáncer de mama ductal".

La mayoría de los tumores que se producen en la mama son benignos, no cancerosos, y son debidos a formaciones fibroquísticas. El quiste es como una bolsa llena de líquido y la fibrosis es un desarrollo anormal del tejido conjuntivo. La fibrosis no aumenta el riesgo de desarrollar un tumor y no requiere de un tratamiento especial. Los quistes, si son grandes, pueden resultar dolorosos. La eliminación del líquido con una punción suele hacer desaparecer el dolor. La presencia de uno o más quistes no favorece la aparición de tumores malignos.

Los tumores benignos están relacionados en su mayoría con factores genéticos. Los síntomas que producen son dolor e inflamación pero ni se diseminan al resto del organismo ni son peligrosos.

Dentro de los tumores malignos, existen varios tipos en función del lugar de la mama donde se produzca el crecimiento anormal de las células y según su estadio.

Los tumores pueden ser localizados o haberse extendido, a través de los vasos sanguíneos o mediante los vasos linfáticos, y haber dado lugar a metástasis, es decir, a un cáncer en un órgano distante al originario. De todos los casos de cáncer de mama, sólo el 7-10% de ellos presenta metástasis de inicio (Tamblay, 1992).

2.1.2. Tipos de Cáncer

En el desarrollo del cáncer de mama encontramos también diferentes tipos y clasificaciones de tipos de tumores de mama. Según Tamblay y Bañados (1992), los tipos de cáncer de mama son los siguientes:
a. El carcinoma ductal in situ

Se origina en las células de las paredes de los conductos mamarios. Es un cáncer muy localizado, que no se ha extendido a otras zonas ni ha producido metástasis. Por este motivo esta enfermedad 'pre maligna' puede extirparse fácilmente. La tasa de curación ronda el 100%. Este tipo de tumor se puede detectar a través de una mamografía.

b. El carcinoma ductal infiltrante (o invasivo)

Es el que se inicia en el conducto mamario pero logra atravesarlo y pasa al tejido adiposo de la mama y luego puede extenderse a otras partes del cuerpo. Es el más frecuente de los carcinomas de mama, se da en el 80% de los casos.

c. El carcinoma lobular in situ

Se origina en las glándulas mamarias (o lóbulos) y, aunque no es un verdadero cáncer, aumenta el riesgo de que la mujer pueda desarrollar un tumor en el futuro. Se suele dar antes de la menopausia. Una vez que es detectado, es importante que la mujer se realice una mamografía de control al año y varios exámenes clínicos para vigilar el posible desarrollo de cáncer.

d. El carcinoma lobular infiltrante (o invasivo)

Comienza en las glándulas mamarias pero se puede extender y destruir otros tejidos del cuerpo. Entre el 10% y el 15% de los tumores de mama son de este tipo.

Este carcinoma es más difícil de detectar a través de una mamografía.

d. El carcinoma inflamatorio

Es un cáncer poco común, tan sólo representa un 1% del total de los tumores cancerosos de la mama. Es agresivo y de rápido crecimiento. Hace enrojecer la piel del seno y aumentar su temperatura. La apariencia de la piel se vuelve gruesa y ahuecada, como la de una naranja, y pueden aparecer arrugas y protuberancias. Estos síntomas se deben al bloqueo que producen las células cancerosas sobre los vasos linfáticos.

2.1.3. Etapas del Cáncer de Mama

Fernández (2003) [3], recomienda que después que se diagnostica el cáncer de mama, se deben realizar pruebas para determinar si las células cancerosas se han diseminado dentro de la mama o a otras partes del cuerpo. El proceso utilizado para determinar si el cáncer se ha esparcido dentro de la mama o a otras partes del cuerpo se denomina clasificación en etapas. La información que se reúne de los procesos de
estadios determina la etapa de la enfermedad. Es importante conocer la etapa de la enfermedad a fin de planear el mejor tratamiento.

Según este autor en el caso del cáncer de mama se utilizan las siguientes etapas:

a. **Etapa 0** (carcinoma in situ).

Hay dos tipos de carcinoma mamario in situ:

- **Carcinoma ductal in situ (CDIS)** es un cáncer de mama muy precoz que puede desarrollarse en un tipo invasivo de cáncer mamario (es decir, cáncer que se ha esparcido desde el conducto a los tejidos circundantes).

- **Carcinoma lobular in situ (CLIS)** no es cáncer, más bien un marcador o indicador que identifica a una mujer con un riesgo incrementado de padecer cáncer de mama invasivo (es decir, cáncer que se ha esparcido a los tejidos circundantes). Es común que ambas mamas estén afectadas.

b. **Etapa I**

En la etapa I, el cáncer tiene un tamaño no mayor a 2 centímetros (aproximadamente 1 pulgada) y no se ha diseminado fuera de la mama.

c. **Etapa II**

Esta etapa a la vez se subdivide en dos sub etapas:

Etapa IIA

El cáncer tiene un tamaño no mayor a 2 centímetros (aproximadamente 1 pulgada) pero se ha diseminado a los ganglios linfáticos axilares (los ganglios linfáticos bajo el brazo), o tiene entre 2 y 5 centímetros (1 a 2 pulgadas) pero no se ha diseminado a los ganglios linfáticos axilares.

Etapa IIB

El cáncer tiene un tamaño entre 2 y 5 centímetros (1 a 2 pulgadas) y se ha esparcido a los ganglios linfáticos axilares (los ganglios linfáticos bajo el brazo), o tiene un tamaño mayor a 5 centímetros (aproximadamente 2 pulgadas) pero no se ha esparcido a los ganglios linfáticos axilares.

d. **Etapa III**

Esta etapa a la vez se subdivide en dos sub etapas:
Etapa III A

El cáncer tiene un tamaño inferior a 5 centímetros (aproximadamente 2 pulgadas) y se ha diseminado a los ganglios linfáticos axilares (los ganglios linfáticos bajo el brazo), y los ganglios linfáticos están adheridos entre ellos o a otras estructuras, o tiene un tamaño mayor a 5 centímetros y se ha diseminado a los ganglios linfáticos axilares y los ganglios linfáticos pueden estar adheridos entre ellos o a otras estructuras.

Etapa III B

El cáncer se ha diseminado a tejidos cerca de la mama (la piel o la pared pectoral, incluidos las costillas y los músculos pectorales), o se ha esparcido a los ganglios linfáticos dentro de la pared pectoral a lo largo del esternón.

e. Etapa IV

El cáncer se ha diseminado a otros órganos del cuerpo, más a menudo los huesos, los pulmones, el hígado o el cerebro, o se ha diseminado a los ganglios linfáticos en el cuello, cerca de la clavícula.

2.1.4. El Cáncer de Mama en el Perú y en el Departamento de la Libertad

De acuerdo al reporte del Ministerio de Salud (2010) [16], en el Perú el cáncer de mama es la segunda causa de muerte por cáncer, mientras que en Lima Metropolitana ocupa el primer lugar. Representa el 25% de todos los tipos de cáncer en la mujer. Lamentablemente aún la mayor proporción de casos se diagnostican en estadios avanzados, inoperables (>50%), lo cual no ha cambiado en los últimos años.

La única forma que ha demostrado mejorar la sobrevida del paciente es el diagnóstico precoz y el mejor método para ello hasta el momento en el Perú es la Mamografía.

A nivel nacional en el Perú la Tasa de incidencia va en aumento, en Lima metropolitana: 32.9 por 100 mil. En el Perú: 1/20 mujeres tienen cáncer de mama y una de cada 200 mamografías realizadas dan diagnóstico de cáncer definido.

En el Perú, según el doctor Risco (2010) [16] el cáncer de mama representa el cáncer más frecuente en Lima, Arequipa y Trujillo. En los últimos 30 años su incidencia a 26.5 por 100,000 mujeres, y la tasa de mortalidad es de 9.3 por100 mil.

Burga (2011) [25], señala que en el departamento de La Libertad, la frecuencia es 100 veces mayor o más que en el hombre y cada vez más va en aumento. En promedio cada año 935 mujeres contraen la enfermedad a una edad media de 54 años. El cáncer de mama representa la segunda causa de mortalidad por cáncer en la mujer, aproximadamente
1 de cada 10 mujeres desarrolla este tipo de cáncer a lo largo de su vida. Según el Registro de Cáncer de Trujillo, la tasa de incidencia es de 30 por 100, ocupando el 2do lugar después del cáncer de cuello uterino.

2.1.5. Tratamientos del Cáncer de Mama

En el presente apartado se muestran las diversas formas de como actualmente se trata el clínicamente el cáncer de mama, las cuales se detallan a continuación:

a. Cirugías

De acuerdo a lo afirmado por Figueroa (2011) [28], existen diversas cirugías para el tratamiento del cáncer de mama entre ellas tenemos:

i. Cirugía Conservadora

La Lumpectomía (biopsia escisional o escisión amplia), es la extracción del tumor canceroso y en una zona de seguridad a su alrededor y en ocasiones algún ganglio axilar. La cual se asocia posteriormente con la radioterapia. Mastectomía parcial, en ella se extrae el cáncer con parte del tejido de alrededor del mismo y el recubrimiento de los músculos del tórax debajo del tumor, con los ganglios linfáticos axilares. También se aplica posteriormente la radioterapia.

ii. Cirugía Agresiva o Radical

La Mastectomía radical es la extracción de toda la mama y el recubrimiento de los músculos del tórax y de la pared torácica situados debajo del tumor, con los ganglios linfáticos axilares. En la mastectomía parcial modificada parte de los músculos de la pared del tórax se conservan.

b. La Radioterapia

En la radioterapia, la radiación de alta intensidad de los rayos X, rayos gama u otras fuentes, se utiliza para matar las células cancerosas y reducir el tamaño de los tumores.

c. La Quimioterapia

Dichos fármacos se denominan medicamento citotástico, citostáticos o citotóxicos. La terapia antineoplásica tiene una gran limitación, que es su escasa especificidad. El mecanismo de acción es provocar una alteración celular ya sea en la síntesis de ácidos nucleicos, división celular o síntesis de proteínas. La acción de los diferentes citostáticos varía según la dosis a la que se administre. Debido a su in especificidad afecta a otras células y tejidos normales del organismo, sobre todo si se encuentran en división activa. Por lo tanto, la quimioterapia es la utilización de diversos fármacos que tiene la propiedad de interferir con el ciclo celular, ocasionando la destrucción de células.
d. La Terapia Hormonal

Se aplica si las células cancerosas analizadas tienen receptores para estrógenos o progesterona. El tamoxifeno se utiliza como terapia hormonal, se administra durante 5 años en el cáncer de mama sin extensión a los ganglios linfáticos. El problema es que aumenta la probabilidad de aparición del cáncer de útero.

2.1.6. Diagnósticos del Cáncer de Mama

Según barragán (2010) [2], entre los diferentes tipos de diagnóstico de cáncer de mama que existe en la actualidad tenemos:

a. Palpitación

Tocar cuidadosamente el nudo y el tejido que le rodea su tamaño, su textura y si se mueve fácilmente. Los nudos benignos se sienten, con frecuencia, de una forma diferente a los que son cancerosos.

b. La mamografía de diagnóstico

La toma de un rayo X del seno.

c. Ultrasonografía (Ultrasonido)

Ondas sonoras de alta frecuencia, que los humanos no pueden oír. Las ondas sonoras entran en el seno y regresan a la máquina. Los patrones de sus ecos producen una foto llamada Sonograma, la cual aparece en una pantalla. Este examen se usa frecuentemente acompañado de un mamograma.

d. Examen de la secreción del pezón

Basándose en estos exámenes, su médico podría decidir que no es necesario hacer más exámenes y que no hay necesidad de un tratamiento. En dichos casos, su médico podría querer revisarle regularmente para observar cualquier cambio. Sin embargo y con frecuencia, el médico tiene que remover fluido o tejido del seno para enviarlo al laboratorio para su diagnóstico. A este procedimiento se le llama biopsia, y hay varios tipos:

i. Biopsia por aspiración con aguja fina

Una aguja muy fina se guía dentro del área sospechosa para obtener una pequeña muestra del tejido.

ii. Biopsia core

Una aguja gruesa se guía dentro del área sospechosa para obtener un pequeño cilindro de tejido.
iii. Biopsia quirúrgica

Un cirujano obtiene parte o la totalidad de un nudo o una zona sospechosa a través de una incisión en el seno.

iv. Biopsia guiada por imágenes

Aquellas que utilizan la ayuda de la ecografía o de los rayos X para localizar una zona sospechosa, las cuales incluyen:

**Biopsia estereostática** - Tipo de biopsia guiada por imágenes. La biopsia estereostática localiza la ubicación exacta de la zona sospechosa por medio de dispositivos computarizados y escáneres para crear una imagen tridimensional (3D) del seno. Se obtiene una muestra de tejido con una aguja.

**Sistema Mammotome para biopsia del seno** - Un tipo de biopsia guiada por imágenes que utiliza una guía ecográfica o estereostática para insertar una aguja en la ubicación exacta de la zona sospechosa. Luego, el sistema Mammotome se utiliza para aspirar cuidadosamente tejido del seno de la zona sospechosa. En 1999, la Administración de Alimentos y Medicamentos de Estados Unidos (US Food and Drug Administration, FDA) aprobó el uso del dispositivo manual Mammotome.

v. Biopsia del ganglio linfático centinela

Procedimiento que consiste en inyectar un líquido de contraste o una sustancia radioactiva cerca del tumor. Esta inyección ayuda a localizar al ganglio linfático más cercano al tumor (ganglio centinela), aquel que tiene más probabilidades de presentar células cancerosas si el cáncer se ha propagado. El cirujano extirpa el ganglio linfático que absorbe el líquido de contraste o la sustancia radioactiva y lo envía al patólogo para que lo analice rigurosamente y determine la presencia de células cancerosas. Es posible que las células cancerosas aparezcan primero en el ganglio centinela antes de propagarse a otras partes del cuerpo.

2.2. Diseño de Algoritmos

2.2.1. Diseño de Algoritmos

Palma (2008) [18], señala que en el diseño se determina como hace el programa la tarea solicitada. Los métodos más eficaces para el proceso de diseño se basan en el conocido divide y vencerás. Es decir, la resolución de un problema complejo se realiza dividiendo el problema en subproblemas y a continuación dividiendo estos subproblemas en otros de nivel más bajo, hasta que pueda ser implementada una solución en la computadora. Este método se conoce técnicamente como diseño
descendente (top – down) o modular. El proceso de romper el problema en cada etapa y expresar paso en forma más detallada se denomina refinamiento sucesivo.

Cada Subprograma es resuelto mediante un módulo (subprograma) que tiene un solo punto de entrada y un solo punto de salida.

2.2.2. Arquitecturas de los Sistemas de Identificación y Clasificación de Patrones o Formas

Ochoa (2005) ubica dentro de los sistemas de identificación y clasificación de arquitectura de patrones o formas obtendremos algunas definiciones para nuestro mejor entendido las cuales se presentan a continuación:

a. Definición de Arquitecturas de Patrones

En informática, término general que se aplica a la estructura de un sistema informático o de una parte del mismo. El término se aplica asimismo al diseño del software de sistema por ejemplo el sistema operativo (el programa que controla la computadora) y también se refiere a la combinación de hardware y software básico que comunica los aparatos de una red informática. La arquitectura de ordenadores se refiere a toda una estructura y a los detalles necesarios para que ésta sea funcional. Por tanto, la arquitectura de ordenadores cubre sistemas informáticos, microprocesadores, circuitos y programas del sistema. Generalmente, el término no suele referirse a los programas de aplicación, como hojas de cálculo o procesadores de textos, que son necesarios para realizar una tarea pero no para hacer funcionar el sistema.

b. Definición de Sistemas de Patrones

La definición del autor James A. Senn, citado por Ochoa(2005) señala que la definición de Sistemas de Información “Es el proceso de examinarla situación de una empresa con el propósito de mejorarla con métodos y procedimientos más adecuados.”

Ian Sommerville, citado por Ochoa (2005) define que sistemas es “la colección de componentes interrelacionados que trabajan conjuntamente para cumplir algunos objetivos”.

Asimismo Kendall & Kendall, citado por Ochoa (2005) menciona que “Es una colección de subsistemas interrelacionado se interdependientes que trabajan de manera conjunta para llevar a cabo metas y objetivos predeterminados.
c. **Definición de Patrón Informático:**

Se define básicamente el concepto de patrón. De la misma forma que sucede con otros conceptos de la informática (y como es el caso de la arquitectura), no es fácil establecer una definición taxativa y definitiva del concepto de patrón. John Vlissides (miembro del Golf), citado por Ochoa (2005) hace hincapié en la dificultad de esta tarea. En las diversas obras de referencia podemos encontrar diferentes definiciones. En esta sección intentaremos reunir a las más influyentes y significativas que nos permitan comprender la idea y establecer las bases para nuestro posterior estudio.

Así mismo el Arquitecto Christopher Alexander, citado por Ochoa (2005) define al patrón en la siguiente manera:

“Cada patrón es una regla de 3 partes, que expresa una relación entre un contexto, un problema y una solución. Como un elemento en el mundo, cada patrón es una relación entre un contexto, un sistema de fuerzas que ocurren repetidamente en ese contexto y una configuración espacial que permite que esas fuerzas se resuelvan entre sí.”

“Como elemento de un lenguaje, un patrón es una instrucción que muestra como puede ser usada esta configuración espacial una y otra vez para resolver el sistema de fuerzas, siempre que el contexto lo haga relevante.”

Continuando con la definición anterior, podemos agregar otro párrafo de Alexander, citado por Ochoa (2005).

“Cada patrón describe un problema que ocurre una y otra vez en nuestro entorno, para describir después el núcleo de la solución a ese problema, de tal manera que esa solución pueda ser usada más de un millón de veces sin hacerlo ni siquiera dos veces de la misma forma.”

Estos patrones son oblicuos y permiten alcanzar la “calidad sin nombre”. Para aclarar estas definiciones, se utilizara un ejemplo concreto aplicado a la construcción, donde tal como refiere Ochoa (2005) “Si nos fijamos en las construcciones de una determinada zona rural, observaremos que todas ellas poseen apariencias parejas (tejados de pizarra con gran pendiente, etc.), pese a que los requisitos personales por fuerza han debido ser distintos. De alguna manera la esencia del diseño se ha copiado de una construcción a otra, y a esta esencia se pliegan de forma natural los diversos requisitos. Diríase aquí que existe un patrón que soluciona de forma simple y efectiva los problemas de construcción en tal zona.”
A modo de complemento, se considera muy importante e interesante citar 2 principios postulados por Fowler, citado por Ochoa (2005) \cite{17} y que debemos tener en mente en todo momento al utilizar patrones:

- Los patrones son un punto de partida, no un destino.
- Los modelos no están bien o mal, sino que son más o menos útiles.

Como reflexión final de esta sección, podemos añadir que un patrón de diseño no es una solución en sí misma, sino la documentación de la forma en que construyeron soluciones a problemas similares en el pasado, lo cual permite una mejor gestión de la experiencia y transferencia de conocimientos.

d. **Arquitecturas de los Sistemas de Identificación y Clasificación de Patrones:**

Dentro de los sistemas de identificación y clasificación de patrones o formas tenemos:

i. **Definición de Arquitecturas de Identificación de Patrones**

Un patrón de arquitectura de software es un esquema genérico probado para solucionar un problema particular, el cual es recurrente dentro de un cierto contexto. Este esquema se especifica describiendo los componentes, con sus responsabilidades y relaciones.

ii. **Características de Arquitectura de Identificación de Patrones**

Para las características de las arquitecturas dispuestas tenemos algunas descripciones como las cuales son utilizados para problemas recurrentes que ocurren en situaciones específicas y dan una solución. Con ellos se puede hacer la variabilidad de las interfaces, documentar experiencias de diseño existentes y bien probadas, se puede obtener experiencia en el desarrollo de sistemas interactivos, identificar y especifican abstracciones de alto nivel, proveen un vocabulario común y comprensible para entendimiento. Con la forma de documentar la arquitectura muestra el funcionamiento del sistema software. Facilitan la construcción de software con propiedades definidas (propiedades particulares). Con ellos se puede desarrollan interfaces intercambiables y modelo utilizables. Como también Ayudan a construir arquitecturas heterogéneas y complejas dependiendo del sistema que queramos mostrar.
iii. Tipos De Arquitecturas de Patrones

Así también tenemos diferentes arquitecturas en el reconocimiento de identificación de patrones o formas que se diversifican con diferentes aéreas entre ellas tenemos:

**Arquitectura de Patrones Simples**

Entre la arquitectura de patrones simples tenemos: Layers, Tubos – y - Filtros, Pizarrón, Repositorio, que a continuación se definiremos.

- **Layers**: Ayuda a estructurar aplicaciones que pueden ser descompuestas en grupos de subtareas con distintos niveles de abstracción (granularidad).
- **Pizarrón**: Útil en problemas para los cuales no hay una solución conocida. Generalmente se provee aproximaciones a la solución final.
- **Repositorio**: Ayuda a estructurar sistemas centrados en los datos, haciéndolos más flexibles y adaptables.

**Arquitectura de Sistemas Distribuidos**

Entre las arquitecturas de sistemas distribuidos tenemos Broker (Microkernel, Pipes-y-Filtros), CAGS, Cliente-Servidor. Así definiremos a continuación el patrón arquitectónico de tuberías y filtros:

- **Patrón Arquitectónico de Tubos y Filtros**: Provee una estructura para procesar flujos de datos. Cada paso de procesamiento se encapsula en un filtro. Los datos se pasan usando los “pipes” entre filtros adyacentes. Recombinando los filtros pueden construirse distintas familias de sistemas relacionados (Ochoa, 2005) [17].

**Arquitectura de Sistemas Interactivos**

Entre las arquitecturas de sistemas interactivos tenemos Modelo-View-Controlador, Presentación-Abstracción-Control. Así definiremos:

- **Modelo-View-Controlador**: Divide a una aplicación interactiva en tres componentes: modelo, vistas y controladores.

- **Presentación-Abstracción-Control**: Define al sistema como una jerarquía de agentes que cooperan entre sí para implementar la funcionalidad de la aplicación.
Arquitectura de Patrones Adaptables

Entre las arquitecturas de patrones adaptables tenemos Microkernel, Reflexión:

- **MicroKernel**: Es usado en sistemas de software que deben adaptarse a los cambios en los requisitos. Este separa el núcleo funcional, la funcionalidad extendida y los aspectos relativos al cliente.
- **Reflexión**: Provee un mecanismo para cambiar la estructura y el comportamiento de un sistema de software, en forma dinámica. Este patrón divide a una aplicación en dos partes: un meta nivel que provee información acerca de las propiedades del subsistema seleccionado, y un nivel base que provee la lógica de la aplicación.

iv. Diferencias de Arquitecturas de Patrones

Los patrones arquitectónicos brindan una sugerencia (genérica) acerca de cómo resolver algunos de los problemas arquitectónicos más comunes dentro del reconocimiento de formas.

Se pueden utilizar mezclas de patrones, y adaptaciones de ellos pues la única diferencia que engloba totalmente a todas ellas es que hay arquitecturas que se interrelacionan entre ellas, dependiendo del escenario de trabajo (dominio + infraestructura previa), del tipo de software a desarrollar, y de lo que se pretende obtener.

2.2.3. El Procesamiento Digital de Imágenes en la extracción de información para la Identificación

En el procesamiento digital de imágenes obtenemos diferentes conceptos y aspectos que engloban a todo el procesamiento digital de una imagen. Descrito por Imbaquingo (2004) \(^{13}\), así tenemos los siguientes:

a. **Definición de Procesamiento Digital de Imágenes (DPI)**

El Campo del procesamiento digital de imágenes (DPI) se refiere se refiere a procesar las imágenes del mundo real de manera digital por medio de la computadora. Él interés del DPI se basa principalmente en dos áreas de aplicación:

El mejoramiento de la información pictórica para la interpretación humana y el procesamiento de datos de la imagen para su almacenamiento, trasmisión y representación para percepción autónoma.
de máquinas. Para ello describiremos a continuación la definición de imagen.

**La Imagen:**

Definimos a la imagen como uno de los puntos clave para el procesamiento digital de imágenes:

Una imagen puede ser definida como una función bidimensional \((x, y)\), donde \(x\), y son coordenadas espaciales en un (plano) y la amplitud \(f\) es llamada intensidad o nivel de gris en ese punto.

Cuando \(x\), y, \(f\) son todos finitos, cantidades discretas, llamamos a la imagen digital.

Una imagen digital está compuesta por un número finitos de elementos, cada uno de los cuales con un valor y una posición particular, llamados píxeles. (Álvarez, D. & Bello, R., 2007) [24].

b. **Definición de Procesamiento Digital de Imágenes en tonos de Grises y tonos RGB.**

Para el procesamiento digital de imágenes en tonos grises y RGB tenemos ciertos criterios que detallaremos a continuación.

- **Definición de Imágenes en tonos de Grises**

  En computación una escala de grises es una escala empleada en la imagen digital en la que el valor de cada pixel posee un valor equivalente a una graduación de gris. Las imágenes representadas de este tipo están compuestas de sombras de grises, que van desde el negro más profundo variando gradualmente en intensidad de grises hasta llegar al blanco.

- **Definición de Imágenes RGB.**

  Descripción RGB (del inglés Red, Green, Blue; "rojo, verde, azul") de un color hace referencia a la composición del color en términos de la intensidad de los colores primarios con que se forma: el rojo, el verde y el azul. Es un modelo de color basado en la síntesis aditiva, con el que es posible representar un color mediante la mezcla por adición de los tres colores luz primarios. El modelo de color RGB no define por sí mismo lo que significa exactamente rojo, verde o azul, por lo que los mismos valores RGB pueden mostrar colores notablemente diferentes en diferentes dispositivos que usen este modelo de color. Aunque utilicen un mismo modelo de color, sus espacios de color pueden variar considerablemente.
Para indicar con qué proporción mezclamos cada color, se asigna un valor a cada uno de los colores primarios, de manera, por ejemplo, el valor 0 significa que no interviene en la mezcla y, a medida que ese valor aumenta, se entiende que aporta más intensidad a la mezcla. Aunque el intervalo de valores podría ser cualquiera (valores reales entre 0 y 1, valores enteros entre 0 y 37, etc.), es frecuente que cada color primario se codifique con un byte (8 bits). Así, de manera usual, la intensidad de cada una de las componentes se mide según una escala que va del 0 al 255.

c. Algoritmos de Procesamiento Digital de Imágenes en tonos de Grises

Así tenemos algunos algoritmos de tonos de escala grises en el procesamiento de imágenes.

- Algoritmos de adelgazamiento sobre gráficas vecinas

Los algoritmos de adelgazamiento intentan obtener esqueletos que se aproximan al eje medio. Estos pueden agruparse en dos clases:

Algoritmos implementados en forma secuencial y Algoritmos implementados en forma paralela. Esta denominación se les asigna por la forma en cómo son analizados los píxeles en la imagen digital y no tienen ninguna relación con el hecho de usar computadoras que implementen computo secuencial o paralelo.

Ambas implementaciones efectúan tres pasos básicos:

Ubicación de píxeles de frontera sobre el objeto de interés y el análisis de células de frontera, preservación o borrado de células analizadas.

A continuación se proporciona una breve descripción de estas implementaciones, describiendo los pasos efectuados en una iteración sobre la frontera del objeto de interés.

- Algoritmos implementados en forma paralela

En este tipo de implementación se analizan simultáneamente en la imagen digital cada uno de los píxeles, para detectar a aquellos píxeles que pertenecen a la frontera del objeto de interés; es decir, todo aquel píxel del objeto que tenga en su vecindad un vecino directo que pertenezca al fondo será considerado un píxel de frontera. Un vecino directo es aquel píxel, que ocupa una posición 0, 2, 4 o 6 (Figura 1) en la vecindad del píxel de interés, o también se
puede hacer referencia a él como 4 – vecino podemos observar en la Figura 1.\textsuperscript{[13]}

![Figura 1: Vecindad de un píxel](Image)

Figura 1: Vecindad de un píxel


Una vez detectados estos píxeles sobre la frontera del objeto de interés, reciben una marca especial que los distingue como píxeles de frontera; posteriormente cada uno de ellos es comparado con un conjunto de plantillas (patrones que presentan vecindades específicas sobre el píxel analizado, para preservar condiciones de conectividad, píxeles finales.) con el propósito de encontrar a aquellos píxeles de frontera cuya vecindad sea similar con alguna de las plantillas utilizadas; de existir dicha similitud este píxel es marcado nuevamente, pero ahora como píxel no borrable, en caso contrario solo es considerado como píxel de frontera (Imbaquingo, 2007)\textsuperscript{[13]}.

Una vez finalizado el análisis de todos los píxeles sobre la frontera del objeto de interés, se procede a cambiar píxeles marcados como de frontera a píxeles de fondo; generándose así un nuevo objeto; al cual, se le vuelve a implementar el procedimiento anterior; este proceso se repite hasta obtener curvas y arcos digitales que representan al objeto adelgazado.

Un problema fundamental en este tipo de implementación se presenta en la Figura 2.\textsuperscript{[13]}
En ella se observa que, los píxeles marcados como ceros pertenecen al fondo de la imagen digital y los píxeles marcados con 1 pertenecen al objeto. Debido a su simetría, todo los píxeles del objeto son también píxeles de frontera; al buscar similitud con las plantillas estos píxeles no serían preservados y en el siguiente paso serían borrados, no siendo este el efecto deseado. Para evitar esta peculiaridad se pre asigna un orden de borrado. Este pre asignación no tiene un fundamento formal, por lo que el objeto adelgazado resultante puede ser diferente.

Pese a esta peculiaridad la implementación paralela proporciona cierta ventaja con respecto al tiempo de cómputo.

- **Algoritmos implementados en forma secuencial**

  En este tipo de implementación, se detectan uno a uno los píxeles pertenecientes a la frontera del objeto de interés, siguiendo una orientación predefinida (por ejemplo, búsqueda en sentido contrario a las manecillas del reloj del vecino directo que pertenezca al fondo de la imagen digital).

  Al ser detectado el primer píxel de la frontera del objeto, éste es analizado buscando la similitud con las plantillas empleadas. De existir similitud con alguna de las plantillas usadas éste píxel es preservado, en caso contrario, éste es borrado.

  A partir de éste píxel se busca al siguiente píxel de frontera, repitiéndose los pasos anteriores, es decir; una vez ubicado el siguiente píxel de frontera del objeto se busca similitud con alguna de las plantillas usadas, si existe esta similitud el píxel es preservado o en caso contrario es borrado. Este proceso es repetido hasta concluir con los píxeles de frontera en el caso del objeto de interés.
Al igual que en la implementación paralela este proceso es repetido sobre el nuevo objeto obtenido de la interacción anterior hasta obtener curvas y arcos digitales que representen al objeto adelgazado.

En éste tipo de implementación no presenta la peculiaridad que está presente en la implementación paralela; pero su tiempo de cómputo es grande y necesita un algoritmo de seguimiento de contorno para ubicar a cada uno de los píxeles subsecuentes de la frontera.

Debido a las características propias de cada implementación, los esqueletos obtenidos en una implementación secuencial de un algoritmo no son los mismos que los obtenidos en una implementación paralela.

En las secciones siguientes describo formalmente un algoritmo de adelgazamiento en estructuras vecinas y un procedimiento estándar implementado en un software de análisis de imagen, los cuales utilizaremos para tener una referencia al comparar los esqueletos obtenidos mediante el procedimiento de adelgazamiento propuesto por Kovalevsky.

- **Algoritmo clásico de adelgazamiento**

  A continuación se va a describir un algoritmo clásico de adelgazamiento, al cual hacemos referencia en esta tesis como algoritmo Pavlidis.

  Este se basa en la idea muy conocida en la literatura, de que los píxeles esqueletales son sólo aquellos píxeles que presentan similitud con alguna de las siguientes plantillas de la imagen (Figura 3). \[13\]
Figura 3: Plantillas para reconocimiento de píxeles esqueletales


Las plantillas de vecindad para píxeles esqueletales, utilizan la siguiente convención:

En las plantillas del inciso A, al menos un píxel de los grupos etiquetados como A o B debe ser diferente de cero. Los píxeles marcados con el valor de 0 son píxeles esqueletales.

En la plantilla del inciso B, al menos un píxel marcado debe ser diferente de cero. Si ambos píxeles etiquetados con C son diferentes de cero, entonces los píxeles etiquetados con C o A pueden tener cualquier valor. De otra manera, al menos un miembro de cada par marcados con B A y deben ser diferentes de cero.

Este algoritmo implementa una mezcla de procesamiento secuencial y paralelo; pues los píxeles de frontera no son analizados todos a la vez, ya que se preestablece una secuencia de análisis, es decir se analizan solo los píxeles cuyo vecino - n sea cero, donde los valores de \( n = 0, 2, 4, 6 \), y son analizados en esa secuencia (Imbaquingo, 2004).
El algoritmo se aplica sobre imágenes binarias; la vecindad del píxel que presenta la similitud con las plantillas, hace que el píxel analizado reciba una marca específica que evita que este sea borrado en iteraciones posteriores.

El algoritmo al cual se hace referencia como algoritmo Paulidises es el siguiente (Figura 4)\[^{13}\]:

Notación:

I es la imagen inicial.

P es el conjunto de plantillas de vecindad para encontrar píxeles esqueletales, incluyendo este conjunto también las rotaciones de 90° de la primera plantilla y las tres rotaciones de 90° de la segunda plantilla (Imbaquingo, 2004)\[^{13}\].

La bandera REMAIN, se utiliza para denotar la existencia de píxeles esqueletales.

La bandera SKEL, se utiliza cuando alguna vecindad de un píxel es similar a alguna de las plantillas existentes en P.
Figura 4: Referencia Técnica Algoritmo para reconocimiento de píxeles esqueletales

El algoritmo citado en las líneas anteriores fue implementado mediante un programa bajo la plataforma computacional DIAS. En la imagen se muestra una imagen y su esqueleto obtenido mediante la aplicación de este algoritmo, así como el esqueleto inscrito en el objeto. (Imbaquingo, 2004). [13]

d. Algoritmos de Procesamiento Digital de Imágenes en tonos de RGB.

Dentro de los algoritmos de procesamiento digital de imágenes en tonos de RGB tenemos:

- **Algoritmo K-Medias**

  El algoritmo de las K medias es probablemente el algoritmo de agrupamiento más conocido. Es un método de agrupamiento heurístico con número de clases conocido (K). El algoritmo está basado en la minimización de la distancia interna (la suma de las distancias de los patrones asignados a un agrupamiento al centro de dicho agrupamiento). De hecho, este algoritmo minimiza la suma de
las distancias al cuadrado de cada patrón al centro de su agrupamiento.

El algoritmo es sencillo y eficiente. Además, procesa los patrones secuenciales (por lo que requiere un almacenamiento mínimo). Sin embargo, está sesgado por el orden de presentación de los patrones (los primeros patrones determinan la configuración inicial de los agrupamientos) y su comportamiento depende enormemente del parámetro K.

- **Algoritmo Multi - ScaleRetinex**

Si quieres mejorar sustancialmente una imagen que ha sido tomada con poca luz, creo que deberías orientarte hacia la teoría Retinex y sus algoritmos de mejora de imágenes, en particular el MSRCR (Multi - ScaleRetinexwith Color Restoration).

Retinex no es más que una teoría descrita por Edwin H. Land (fue fundador y principal accionista de Polaroid Corporation) en la cual sugirió que tanto el ojo como el cerebro están involucrados en la percepción de los colores, de ahí la palabra "retinex" que está formada por "retina" y "córtex".

Las técnicas de mejora de imágenes con Retinex lo que hacen básicamente es distinguir la información lumínica de un punto a partir de su intensidad y la de sus vecinos para poder así aclarar las zonas oscuras de una imagen.

e. **Diferencias de los Algoritmos de DPI a escala de Grises y RGB**

En esencia, la diferencia entre usar un algoritmo para una imagen en escala de grises y una imagen en RGB es el número de “canales de color”: una imagen en escala de grises tiene uno; una imagen RGB tiene tres. Una imagen RGB se puede entender como tres imágenes en escala de grises superpuestas, una coloreada de rojo, otra de verde y otra de azul.

En realidad, tanto los algoritmos para imágenes RGB como en escala de grises tienen una función para un canal de color adicional, llamado el canal alfa, que representa la opacidad. Cuando el valor del canal alfa de una capa es cero en cierto lugar, la capa es completamente transparente en ese lugar (puede ver a través de ella), y el color en ese lugar lo determina lo que haya debajo. Cuando alfa es máximo (255), la capa es opaca (no se puede ver a través de ella), y el color lo determina el color de la capa. Valores intermedios de alfa corresponden a grados de
transparencia y opacidad: el color en el lugar es proporcional a la mezcla de color de la capa y del color que haya debajo.

f. Algoritmos de DPI en la extracción de Información para la Identificación del Cáncer de Mama

Así tenemos algunos algoritmos utilizados actualmente que en la actualidad la mayoría de ellos están en investigación, de ellos definiremos algunos a continuación:

i. Algoritmo de Clasificación Difusa en la Identificación de tumores de Mama

En general, las técnicas de clasificación (clustering) se basan en la minimización de una función objetivo que determina los prototipos de los grupos buscados. La función objetivo más utilizada en clustering, aunque no la única, es la que nos mide las desviaciones respecto de los prototipos como se muestra en la siguiente fórmula:

\[ J(m) = \sum_{k=1}^{c} \sum_{x \in X} u_{xk}^{3/2} \sqrt{g d_{xm}^2} \]

Fórmula 1: Formula básica para el algoritmo de clasificación difusa


Esta función será mínima cuando los prototipos estén situados de tal manera que representen agrupaciones de elementos alrededor de los prototipos. Cuando las probabilidades son cero o uno, \( U(x,k) = \{0, 1\} \), estamos realizando un clustering hard o crisp, y el algoritmo que encuentra los prototipos se denomina k-means. Cuando las probabilidades están comprendidas entre cero y uno, \( U(x,k) \in [0, 1] \), estamos realizando un clustering difuso (fuzzy clustering), y el algoritmo que encuentra los prototipos se denomina fuzzy c-means (FCM), que fue formulado por Dunn.

En Flores - Sintas (1998), citado por Soto y Egea (2003) se fundamenta que estudiando las propiedades de una métrica local derivada de la matriz de covarianza de la muestra, se consigue encontrar una función que mide la desviación de los elementos de la muestra.
$$J(m) = \sum_{k=1}^{c} \sum_{x \in X} u_{xm}^{3/2} g_{ma}^{2} x_{m}$$

Fórmula 2: Formula de Representación Matemática para el algoritmo de clasificación Difusa.


ii. Algoritmo Genético en la Identificación de Tumores de Mama
Álvarez y Bello (2007), [24] consideran que sea X el problema a resolver. Dada una representación de candidatas a soluciones en una cadena de bits, un algoritmo genético simple, tal y como se describe, trabajaría del siguiente modo:

- Comenzar con una población P generada aleatoriamente de n cromosomas del bit.

- Calcular la capacidad f(x) para cada cromosoma x de P.

- Repetir los siguientes pasos hasta que se hayan creado n descendiente:

  - Seleccionar un par de cromosomas padre de P, siendo la probabilidad de selección una función creciente de la capacidad. La selección se realiza “con reemplazamiento”, es decir, que el mismo cromosoma puede ser seleccionado en más de una ocasión para ser padre.

  - Con probabilidad P {c}, (probabilidad de cruce, o tasa de cruce), cruzar el par en un punto elegido aleatoriamente (con probabilidad uniforme) para formar dos descendientes. Si no tiene lugar ningún cruce, formar dos descendientes que sean copias exactas de sus respectivos padres. (Obsérvese que aquí la probabilidad de cruce se define como la probabilidad de que dos padres se crucen sobre un único punto. Hay otras versiones de algoritmos genéticos que son de “cruces en múltiples puntos”, en los que la tasa de cruce para una pareja de padres es el n° de puntos en los que tiene lugar un cruce).
- Mutar los dos descendientes en cada lugar con probabilidad \( P\{m\} \) (probabilidad de mutación, o tasa de mutación), y colocar los cromosomas resultantes en la nueva población \( P'\).

- Si \( n \) es impar, se puede rechazar aleatoriamente a un miembro de la nueva población.

- Remplazar la población actual \( P \) con la nueva \( P' \).

- Volver al paso 2.

iii. **Algoritmo de Selección RS - Reduct en la Identificación de tumores de Mama**

RS Reduct es un método que trata de encontrar un reducto de manera que éste sea lo suficientemente bueno para el análisis de datos en tiempos aceptables. Para ello se utiliza la búsqueda heurística como estrategia de búsqueda, debido a que se tomará en cuenta sus variantes para su búsqueda.

Encontrar un reducto significa encontrar un conjunto mínimo de atributos que cubran a \( B \), es decir, un subconjunto que satisfaga, como observamos en la siguiente fórmula:

\[
\forall (K, n) \exists i \in R : b((K, n), i) = 1 \cup b((K, n), N + 1) = 1
\]

Fórmula 3: Fórmula Básica para Algoritmo de Selección de Rasgos Rs-Reduct.


Según Tolmos (2002) [32], esto significa que para cada par de objetos que se tomen, o son de la misma clase o tienen algún atributo de \( R \) para el cual los objetos son similares “, es decir, cumplen la relación \( R \).
iv. **Algoritmo de selección a través de RM en la Identificación de Tumores de Mama**

Definiendo el algoritmo de selección en resonancia magnética para tumores mamarios, según García, Marín y Robles (2001) \[^{29}\], su base de datos es utilizada en los experimentos, en continua expansión, esta consta de 304 casos de diversos hospitales. Los casos se han agrupado en diferentes estirpes o agrupaciones histológicas (quiste, grasa, vascular, sinovial, muscular, fibrosa, nerviosa, metástasis, pseudo tumor, otros malignos, otros benignos, linfoides y mesenquimales). Los diferentes parámetros de cada caso se han extraído independientemente por diferentes radiólogos. Estas variables se obtienen a partir de las imágenes de RM (Resonancia Magnética) potenciadas en SE-T1 y SE-T2/STIR (homogeneidad, dependencia de estructura, hemorragia, T1, T2/STIR, diana, forma, márgenes, edema, T1Trac, única-múltiple, dependencia de estructura, atrofia, calcio, vasos.

Se presenta además hemorragia, grasa, fibrosis/hemosiderina, fascia y alteración del hueso y datos epidemiológicos (edad, presentación clínica) con lo que se obtienen 22 parámetros por cada registro. Estas características, junto con las etiquetas correspondientes de benignidad / malignidad y estirpe histológica son los que conforman la base de datos final.

v. **Algoritmo de Wavelet Gabor en la Identificación de tumores de Mama**

De La Cruz y Torres (2012) \[^{27}\], considera que esta técnica se fundamenta en la transformada wavelets - Gabor. Este último algoritmo fue diseñado para que los puntos extraídos correspondan a puntos característicos de la imagen tomada, los cuales se utilizan comúnmente en medidas reconocimiento y segmentación de imágenes. Las modificaciones consisten en un conjunto de restricciones y filtros geométricos para ajustar inicialmente la ubicación de los centros de búsqueda, así mismo del reconocimiento y extracción de sus puntos característicos a través de filtros de imagen.
2.2.4. La Inteligencia Artificial en el Reconocimiento y Clasificación de Patrones o Formas

En la inteligencia artificial encontramos desarrollos de conceptos que nos llevarán a un mejor entendimiento de esta rama de la ciencia de la informática. Así tenemos:

a) Definición de Inteligencia Artificial

A la inteligencia artificial La podemos definir de cuatro maneras, las cuales se detallan en el cuál se detallan en Tabla 1:

Tabla 1: Conceptos básicos de la Inteligencia Artificial

<table>
<thead>
<tr>
<th>Sistemas que piensa como Humanos</th>
<th>Sistemas que piensa racionalmente</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Es el nuevo y excitante esfuerzo de hacer que las computadoras piensen.</td>
<td>• Es el estudio de las facultades mentales mediante el uso de modelos computacionales.</td>
</tr>
<tr>
<td>• Maquinas con mentes, en el más amplio sentido literal.</td>
<td>• Es el estudio de los cálculos que hacen posible percibir, razonar y actuar.</td>
</tr>
<tr>
<td>• La automatización de actividades la vinculamos con procesos de pensamiento humano, actividades como la toma de decisiones, resolución de problemas de, aprendizaje entre otros.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sistemas que actúan como Humanos</th>
<th>Sistemas que actúan racionalmente</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Es el arte de desarrollar máquinas con capacidad para realizar funciones que cuando son realizadas por personas requieren de inteligencia.</td>
<td>• La inteligencia computacional es el estudio del diseño de agentes inteligentes IA es la que está relacionada con conductas inteligentes en artefactos</td>
</tr>
<tr>
<td>• El estudio de cómo lograra que los computadores realicen tareas que, por el momento, los humanos hacen mejor.</td>
<td></td>
</tr>
</tbody>
</table>

Algunas definiciones de inteligencia artificial, organizadas en cuatro categorías

Las que aparecen en la parte superior se refieren a los que refiere en procesos mentales y al razonamiento, mientras que la parte inferior aluden a la conducta. Las definiciones de la izquierda miden el éxito en términos de la fidelidad en la forma de actuar de los humanos, mientras que los de la derecha toman como referencia un concepto ideal de inteligencia, que llamaremos racionalidad. Un sistema es racional si hace “lo correcto”, en función de su conocimiento.

A lo largo de la historia se ha seguido los cuatro enfoques mencionados. Como es de esperar, existe un enfrentamiento entre los enfoques centrados en los humanos y los centrados en su entorno a la racionalidad. Él enfoque centrado en el comportamiento humano debe de ser una ciencia empírica, que incluya hipótesis y confirmaciones mediante experimentos. Él enfoque racional implica una combinación de matemáticas e ingeniería. Cada grupo al mismo tiempo ha ignorado y ayudado al mismo tiempo.

b) Definición de Patrón

Hill (2007)\(^{(12)}\), sostiene que patrón es sinónimo de objeto. En ocasiones se le llama así a los objetos ya clasificados.

Por su parte Alba (2006)\(^{(1)}\), explica que tras los procesos de segmentación, extracción de características y descripción, cada objeto queda representado por una colección (posiblemente ordenada y estructurada) de descriptores, denominada patrón. En los problemas de reconocimiento, cada patrón se supone perteneciente a una categoría o clase, existen patrones como vectores, cadenas y/o árboles.

c) Definición de Reconocimiento

De acuerdo a Hill (2007)\(^{(12)}\) es el Proceso de clasificación de un objeto en una o más clases.

Así mismo Alba (2006)\(^{(1)}\) asevera que es el proceso por el cual se asigna a una “etiqueta”, que representa una clase, a un patrón concreto.

d) Definición de Clasificador

Alba (2006)\(^{(1)}\) considera que un subsistema que utiliza un vector de características de la entidad cuantificable y lo asigna a una de M clases.
e) Paradigmas de la Inteligencia Artificial

En este orden de ideas, Yen (2009) considera que en su evolución en los primeros años la Inteligencia Artificial (IA) pretendía lograr un algoritmo general que fuese capaz de resolver cualquier tipo de problema, el famoso solucionador general de problemas, a través de métodos heurísticos. Se creía que todos los problemas se podían resolver recorriendo un árbol de objetivos y la solución del problema, consistía en encontrar el camino correcto, desechar todas aquellas vías poco fértiles. La dificultad estaba en cómo desechar esas ramas o caminos improductivos, quién podría garantizar que ese camino después no fuese el más adecuado, hasta qué profundidad había que recorrer el árbol para estar seguro de que esa vía era errónea, cómo evitar, qué producto de la ramificación creciente del árbol, un exceso de combinaciones incalculables (explosión combinatoria). Con el tiempo estos revolvedores generales fueron abandonados y se trató de darle solución a problemas más específicos, problemas que dependían de un dominio de aplicación y los cuales se solucionaban a través de los conocimientos que se poseían sobre ese dominio, lo que dio lugar a la explosión de los sistemas basados en conocimientos más conocidos como sistemas expertos, los cuales debido a la estrechez de su dominio en la solución de los problemas, no cumplían con las expectativas de los usuarios. Esto, entre otras cosas, trajo como consecuencia la crisis del paradigma simbolista dentro de la IA lo que originó nuevos paradigmas siendo los más importantes los basados en el conexionismo y los basados en la cooperación y es sobre estos últimos donde haremos hincapié dada su relación con la distribución inteligente de la información y la cooperación entre agentes inteligentes los cuales se ajustan a las necesidades de una sociedad basada en el conocimiento.

Por otra parte se pretendía comprender los estados mentales a través de "mundos de juguetes", ya que no se conocía lo suficiente sobre la naturaleza de la inteligencia y, por otro lado, existía aquello de que lo aprendido en esos dominios de juguete se podría extender sin dificultades a problemas más complejos.

La IA abandona los sistemas autistas (encerrados en sí mismos) para dar lugar a nuevos sistemas abiertos que comparten información de forma inteligente, a la vez que permite la combinación de diferentes paradigmas de representación del conocimiento como los basados en reglas, los marcos (frames), los guiones, entre otros. Entre ellas también tenemos el desarrollo de la inteligencia artificial como son los reconocedores de imágenes con patrones, o formas tratando de resolver nuevos problemas actuales.
f) Reconocimiento de Patrones (RP) o Formas (RF) en la IA

Rusell y Norving (2004) \( ^{[21]} \), sostienen que para el reconocimiento de Patrones o Formas en la inteligencia artificial encontramos diferentes conceptos para un mejor entendimiento de este punto clave en esta rama en la se detalla a continuación:

i. Definición de Reconocimiento de Formas

Carasco (2005) \( ^{[3]} \), sostiene que son los medios por los cuales se puede interpretar el mundo. Es la ciencia que se ocupa de los procesos sobre ingeniería, computación y matemáticas relacionados con objetos físicos y/o abstractos, con el propósito de extraer información que permita establecer propiedades de o entre conjuntos de dichos objetos.

ii. Estructura de Reconocimiento de Formas

El reconocimiento de formas está estructurada de la siguiente, manera:

- **Selección de variables:** Consiste en seleccionar cuál es el tipo de características o rasgos más adecuados para describir los objetos. Se deben localizar los rasgos que inciden en el problema de manera determinante.

- **Clasificación supervisada:** También es conocida como clasificación con aprendizaje, en este tipo de problemas ya se encuentran definidas las clases, y éstas cuentan con algunos objetos previamente clasificados.

- **Clasificación parcialmente supervisada:** También conocida como aprendizaje parcial, en éstos problemas existe una muestra de objetos sólo en algunas de las clases definidas.

- **Clasificación no supervisada:** Esta taxonomía, conocida como clasificación sin aprendizaje, en estos problemas no existe ninguna clasificación previa de objetos y en algunas ocasiones ni siquiera se han definido las clases.

iii. Tipos de Representación de Patrones (RP) o Formas (RF)

Entre los tipos de Patrones o Formas encontrados tenemos:

- **Geométrico (Clustering):** Los patrones deben ser gráficas, en éste enfoque se emplea el cálculo de distancias, geometría de formas, vectores numéricos y puntos de atracción.
• **Estadístico:** Se basa en la teoría de la probabilidad y la estadística, utiliza análisis de varianzas, covarianzas, dispersión y distribución.

• **Sintáctico-Estructural:** Estudia la estructura de los objetos, es decir, usa teoría de lenguajes formales, gramáticas y teoría de autómatas.

• **Neuro - Reticular:** También se puede definir a la representación de patrones como la utilización de redes neuronales que se “entrenan” para dar una cierta respuesta ante determinados valores.

• **Lógico - Combinatorio:** Se basa en la idea de que la modelación del problema debe ser lo más cercana posible a la realidad del mismo, sin hacer suposiciones que no estén fundamentadas. Se utiliza para conjuntos difusos y utiliza lógica simbólica, circuitos combinacionales y secuenciales.

• **Aproximación de Teoría de la Decisión:** Se considera cuando el sistema manipula formas simples, entendiéndose por tales aquellas que al tiempo son elementos primitivos de la representación, no planteándose descomposición estructural de las mismas. Se utilizan funciones de decisión para clasificar las formas, representadas como vectores de características.

• **Aproximación Estructural/Relacional:** Su utilidad se encuentra en problemas en los que las formas en estudio se consideran complejas, es decir, constituídas por formas simples y relaciones estructuradas entre las formas simples. En este caso, las formas están representadas como cadenas, gramáticas árboles, grafos, estableciéndose una analogía entre la estructura de la forma y la sintaxis de un lenguaje, o entre una red estructural. Estos procesos de reconocimiento resulta, en el caso sintáctico, uno de los análisis (parcing), o en el caso estructural, un problema de comparación entre grafos o de inferencia.
iv. Aproximaciones Metodológicas al diseño de etapas de Aprendizaje

Las aproximaciones metodológicas asociadas al diseño de las etapas de aprendizaje pueden ser:

- **Aprendizaje Supervisado.** En este caso se dispone de un conjunto de muestras (dataset) controladas de las cuales se conocen a priori la pertenencia a clases de todos y cada una de ellas. Él proceso de aprendizaje consiste pues en la determinación de cuáles son las reglas de clasificación a partir de las regularidades de las muestras.

- **Aprendizaje No Supervisado.** En este caso no se dispone de los valores de pertenencia a clase de los elementos de la muestra controlada, por lo que antes de obtener reglas de clasificación es preciso analizar el conjunto de datos para determinar el número de clases de formas que la constituyen, así como sus regularidades.

v. Esquemas de Identificación de Patrones

Para el esquema de identificación y clasificación de patrones tenemos algunos aspectos que a continuación daremos a conocer.

a. Esquema Geométrico

En este caso, la planificación se puede plantear como la asignación del vector de características incógnita, entendido este como un punto del espacio de características, a una de las particiones (regiones) mutuamente exclusivas previamente definidas en dicho espacio. Cada una de estas particiones se corresponde con una de las C clases de formas $\Omega_i$. Para efectuar el proceso de asignación, durante el proceso de aprendizaje se define un conjunto de funciones discriminantes $d_i(x)$ asociada cada una a una de las c clase de formas.

b. Esquema Estadístico

En este caso se considera que los elementos del vector de características son variables aleatorias, siendo una medida ruidosa de la característica. Para cada clase de formas se hace uso del conocimiento de su distribución de probabilidad y de la probabilidad a priori de ocurrencia de cada clase. Basándose en estos datos se construye una regla de decisión de naturaleza probabilística basada, por ejemplo en el objetivo de minimizar las probabilidades de reconocimiento erróneo.
c. Esquema de Redes Neuronales

En este caso se considera la utilización de elementos provenientes del paradigma de las redes neuronales artificiales, para resolver problemas de RF. Sin embargo, se pueden considerar este esquema como uno independiente a los dos anteriores, o bien como una consecuencia del esquema geométrico.

vi. Modelos Estadísticos en el RF.

Según Russell y Norving, (2004) en los modelos a detallar a continuación, consideraremos a los vectores de características como variables aleatorios n - dimensionales y a las clases de formas como distribuciones de densidades de probabilidad. En tal sentido será obtener reglas de clasificación óptimas en el sentido de minimizar determinadas tasas relacionadas con la clasificación errónea en el reconocimiento de formas.

- **Modelo Clasificador bayesiano de mínimo error**

  Este modelo define una regla de clasificación de formas entre clases Ω_i partiendo de un vector de medidas XεU, conociendo las probabilidades a priori, P (Ω_i), de que una muestra pertenezca a una de las clases para dichas probabilidades se cumple (Fórmula 4):\[ \sum_{i=1}^{c} P (\cap i) = 1 \]

  Fórmula 4: Fórmula Matemática de la Definición de Regla del Modelo de Clasificador Bayesiano de Mínimo de Error.

\[ \int u^{p(x)/p(\alpha_0)} \, dx = 1 \]

Fórmula 5: Representaciones de Densidades de Probabilidad de Clasificador Bayesiano Mínimo de Error.


Para el modelo de clasificador bayesiano mínimo de error se puede analizar la tasa de error probabilístico a partir de valores de X, en este caso se puede dar en función a dicha fórmula:

\[
P(\varepsilon/x) = \begin{cases} 
  p(P(\Omega_1)/X) & X \in \Omega_2 \\
  p(P(\Omega_2)/X) & X \in \Omega_1 
\end{cases}
\]

Fórmula 6: Representaciones de Casos de Valor Probabilístico de Clasificador Bayesiano Mínimo de Error.


A partir de dicha fórmula observaremos que el error viene definido por el menor de las probabilidades tomadas como un X dado. Por ello, a este esquema de clasificación se le denominara regla bayesiana de mínimos de error.

- **Modelo Clasificador bayesiano de mínimo riesgo**

  Este modelo requiere muchas veces de conceptos de teoría de juegos, podemos realizar una introducción cualitativa, no rigurosa. El clasificador bayesiano de mínimo Riesgo supone que todos los errores son idénticos en importancia y por tanto se suman de la misma forma. Este modelo surge en situaciones en las que los errores no se pueden agrupar homogéneamente, sino que son diferentes en relevancia o importancia para el sujeto o sistema que los sufre. Por ejemplo puede detenerla misma trascendencia tanto al clasificar un observable que sea un león como si fuese un antílope, como el caso contrario, clasificar un antílope como un león. En términos de dicho modelo, ambos errores son diferentes y comportan riesgos diferentes en conclusión a dicho modelo no todos los errores son idénticos ni comportan las mismas consecuencias para el sistema de decisión. La función objetivo que se desea optimizar no debe ser
homogénea en el tratamiento de los errores, para ello se debe introducir una valoración o intención propia del usuario y por tanto ajena a la estadística subyacente. Para modelar esta nueva función introduciremos el concepto de pérdida, por adoptar una decisión errónea. Sea que X pertenece a la clase \( \Omega_i \), si el clasificador decide que pertenece a la clase \( \Omega_j \), incurre en una pérdida \( L_{ij} \). Ahora bien, como existen \( c \) clases, el vector incógnita puede pertenecer a cualquiera de ellas, por lo que la pérdida promedio esperada o riesgo de asignar la muestra incógnita a la clase \( \Omega_j \) se definirá de la siguiente fórmula:

\[
r_j(x) \sum_{i=1}^{c} L_{ij} p(\Omega_i|x)
\]

Fórmula 7: Representaciones de Densidades de Probabilidad en el Modelo de Clasificador bayesiano Mínimo de Riesgo.


Dado \( P(x) \) como un término común obtendremos la regla del modelo de clasificador bayesiano mínimo riesgo expresándolo como (Fórmula 8):

\[
R(x) = \frac{1}{p(x)} \sum_{i=1}^{c} L_{ij} p\left(\frac{x}{\Omega_i}\right) p\Omega_i
\]

Fórmula 8: Regla Funcional del Modelo de Clasificador bayesiano Mínimo de Riesgo.

• **Modelo distribuida normal multivalente**

   A efectos prácticos el termino común p(X) puede eliminarse para realizar las comparaciones. En algunos problemas de reconocimiento de formas, la perdida de nula para decisiones correctas, y un valor fijo distinto de cero para decisiones erróneas. Así por ejemplo, podemos definir: Lij= 1 – δij . A esta función se le denomina función de perdida simétrica o cero – uno (Fórmula 9).

   \[ P(x) \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \]

   Fórmula 9: Representaciones de Formas de Densidades de Probabilidad en el Modelo Distribuido Normal Multivalente.


• **Modelo de estimaciones de distribuciones**

   El modelo de estimación de distribuciones puede clasificarse en función del conocimiento a priori disponible o asumido acerca de la naturaleza de las clases. En primer lugar tenemos la estimación paramétrica, aplicable cuando se conoce la forma analítica de las funciones de distribuciones de las muestras pero no así sus parámetros. En segundo lugar, la estimación no paramétrica cuando no se conoce tal naturaleza, en la que intentaremos aproximar una distribución de probabilidad consistente con las muestras. Una de las técnicas básicas de estimación paramétrica es la de máxima Verosimilitud. Dado un conjunto de datos D con m muestras Xj pertenecientes a una clase correcta, la probabilidad de que tal muestra sea obtenida en un muestreo es P(x, j, θ). Dado que en este modelo las muestras tomadas son independientes unas de otras, las probabilidades de obtener un conjunto de datos completo a partir de este modelo estadístico (formula 10),
\[ P\left( \frac{D}{\theta} \right) = \prod_{j=1}^{m} P(x, \theta) \]

Fórmula 10: Representaciones Matemática para Obtener el Modelo de Estimaciones de Distribuciones.


vii. Modelos geométricos de la Decisión en el RF

Si bien los modelos y técnicas de la decisión son mucho más rigurosos que la de los planteamientos geométricos, estos son más intuitivos y resultan mucho más precisos en muchas situaciones (Russell, 2004). Entre los modelos geométricos de la decisión de en reconocimientos de formas tenemos:

- Modelos de funciones discriminantes:

  Dado un espacio de representación donde se ha definido un conjunto de c clases \( \Omega_1, \Omega_2, \ldots, \Omega_n \), asociadas a cada clase se define un funcional \( d_i(X) \), donde \( X \in \mathbb{R}^n \)

  Representa el vector de características del espacio. Se puede establecer una regla de clasificación basada en estos funcionales de la siguiente forma (Fórmula 11):

  \[ x \in \Omega_i \iff d_i(x) = (\max_j (d_j(x))) \]

  Figura 51: Regla Matemática Discriminante


  Al conjunto de funcionales mencionados se les denomina funciones de decisión o funciones discriminantes y la podemos describir en primer lugar a la forma de la función de que debe estar directamente relacionada con el comportamiento geométrico de las clases en consideración. En segundo lugar, de la determinación de los parámetros de la función, qué se resuelve mediante esquemas de aprendizaje.
Modelos de clasificación por funciones de distancia

La manera más simple e intuitiva de realizar un proceso de clasificación de formas es probablemente la basada en los conceptos de distancia entre las formas y prototipos. El motivo de ello es que la proximidad o similaridad entre vectores de características es una manera obvia de determinar pertenencia a categorías. En este caso los vectores de características se consideran como puntos de un espacio con estructura tal que admite la definición de una distancia e incluso si es posible de una métrica.

Definiremos el concepto formal de distancia, que no siempre coincide con el concepto intuitivo de la misma. Para ello, sea \( U \subseteq R^n \) un conjunto finito o infinito de elementos, que se corresponda con los vectores de características asociados a las muestras de un problema dado. Se denomina función de distancia \( D \) a una correspondencia (Fórmula 12).

\[
D: U \times U \rightarrow R
\]

Para que \( D \) sea un funcional de distancia se debe cumplir que para un par arbitrario \( (x, y) \in U \):

\[
D(x, y) \geq D_0
\]

\[
D(x, x) = D_0
\]

\[
D(x, y) = D(Y, X)
\]

Fórmula 12: Fórmula Matemática en función al Modelo Clasificación por Funciones a Distancia.


Otra forma de definir el modelo de función a distancia sería el de distancia mínima donde tomas aspectos relacionados con los algoritmos de vecino más cercano y K vecino más cercanos.

La regla de clasificación del vecino más cercano

Es una extensión de la regla de clasificación por distancia mínima y parte de la suposición de que los prototipos de una misma clase están cerca unos de otros en el espacio de representación. Esta regla consiste en asignar a la muestra la clase de su vecino más cercano en el conjunto de entrenamiento.
La regla de clasificación de los k vecinos más cercanos

Es una extensión de la regla del vecino más cercano en la que se clásica la muestra \( x \) en la clase más frecuente de entre los \( k \) vecinos más cercanos. Estas dos reglas son muy sencillas de implementar y tienen unas propiedades estadísticas muy interesantes, como se ha comentado anteriormente.

Una regla de reciente aparición es la regla de clasificación de los \( k \) vecinos de centro más cercano (que utiliza para la clasificación aquellos \( k \) vecinos cuyo centro de se encuentre más cerca de la muestra. Es una regla especialmente adecuada cuando el conjunto entrenamiento es reducido y existe un cierto solapamiento entre las clases, pero necesita que los prototipos se representen como puntos en \( \mathbb{R}^n \).

Modelos de separación lineal y la dimensión Vapnik-Chervonenkis

Un ejemplo muy sencillo de conjunto de datos con dos clases es el conjunto de las funciones booleanas. En este caso los atributos o características son igualmente booleanas: \( X_i \in \{0,1\} \), donde se puede considerar a cualquier función booleana como un conjunto de datos con dos clases. Por ejemplo las funciones booleanas bidimensionales básicas and y or son conjuntos que demuestran linealidad separable. No es el caso de la función Or – Exclusive y su complementaria nor - Exclusive, que si son separables pero por discriminantes cuadráticos.

Estos ejemplos nos llevan a introducir el concepto de separabilidad lineal de un conjunto de muestras, o su relacionado de capacidad de separación de un discriminante lineal. En general un clasificador lineal en un espacio \( n \)-dimensional puede separar \( n+1 \) puntos no bolinéales.

La argumentación precedente sobre la separabilidad lineal nos lleva a presentar un concepto relacionado de carácter general sobre la capacidad discriminante de un clasificador. Él mismo es la denominada dimensionalidad de Vapnik – Chervonenkis de un clasificador.
viii. Aprendizaje de Clasificadores Lineales en el RF

La tarea del aprendizaje supervisado en reconocimiento de formas consiste en la determinación de los parámetros del clasificador a partir de conjuntos de muestras representativas. El aprendizaje supervisado en su modalidad de aprendizaje en la que la información aportada por el usuario, o supervisión, es de vital importancia en el desarrollo del procedimiento. En la práctica, tal supervisión suele reducirse a incorporar información de la clase a la que pertenece a cada muestra. Los clasificadores lineales constituyen la forma de clasificadores más simples de clasificadores en concreto la del modelo bi clásico bastante sólida como un conjunto de muestras constituidas por pares que incluyan las coordenadas es el caso más simple de todos.

Se ha desarrollado una teoría de aprendizaje lineal bi clásico bastante sólida. Representamos el conjunto de datos que se utiliza para el aprendizaje como un conjunto de m muestras constituidas por pares que incluyen las coordenadas \( X_i \in \mathbb{R}^n \), además la información de supervisión dada por la pertenencia a la clase.

a) Procedimiento del Perceptrón

En la práctica este procedimiento utiliza el método del gradiente o del descenso (método de interacciones lineales) para alcanzar la minimización de la función objetivo, está basada en múltiples muestras o batch perceptrón, porque en la corrección de los coeficientes se hacen intervenir todas las muestras del conjunto de datos incorrectamente clasificadas.

b) Procedimiento de mínimo error cuadrático

Los procedimientos de mínimo error cuadrático se basan en utilizar funciones de la naturaleza menos heurística, pero con mayor significado. De esta manera se recurre al concepto de significado de error. Teniendo dos grupos uno que se orienta a ver el desarrollo del discriminante lineal con márgenes de seguridad y el segundo que contiene y trabaja sobre la formulación de error. En conclusión la respuesta de error se obtiene con la diferencia entre el comportamiento verificado y el deseo que está definido por una variable de caso bioclásico.
c) Procedimiento Widrow–Hoff

También denominado procedimiento LMS (least - mean-square). Se basa en computar las correcciones en base a una gradiente definida anteriormente. La principal ventaja que no precisa computar la matriz pseudo - inversa y los problemas que pudieran implicar su singularidad. La principal dificultad de este procedimiento es que precisa definir como datos los márgenes para cada muestra, por lo que incluye en la decisión final, por lo que converge a una solución que no es correcta.

d) Procedimiento Ho-Kashyap

El procedimiento Ho- Kashyap considera que los márgenes son incógnitos que el procedimiento debe a su vez determinar, este procedimiento parte de la computación previa de la matriz pseudo-inversa y de una elección arbitraria y positiva del vector de márgenes .Una de las grandes ventajas de este procedimiento es que permite detectar en la ejecución del procedimiento cuando no existe solución. En los anteriores procedimientos no se da esto, dado que finalizaran sin convergencia solo por haber agotado un límite superior de limitaciones.

e) Maquina Lineal Mínimo de Error

En este caso pretendemos construir una máquina lineal que presente un error mínimo entre la respuesta que produce y la que debiera producir. En este caso no se trabaja con una sobre la ecuación de desigualdad del discriminante .Se traba con la respuesta de una máquina lineal. Para evitar problema de derivabilidad en el cálculo del gradiente, se utiliza una función suavizada de tipo sigmoide.

f) Procedimiento de Máquinas Vectoriales de Soporte

Todos los procedimientos presentados para obtener clasificadores tratan de proporcionar una solución válida de entre las muchas posibles .En algunos casos de forma no muy frecuente, la frontera de separación entre clases de encuentra muy cerca de alguna de las muestras fronterizas y simultáneamente muy alejada de las demás. En estos casos las márgenes de separación de muestras en el clasificador lineal son muy desiguales .Dado que en un sistema de reconocimiento de formas es frecuente la existencia del ruido, este puede producir un desplazamiento de alguna muestra y causar error que es más probable si existe un fuerte desequilibrio de los márgenes. Un objetivo adicional al obtener un clasificador valido, es el de obtener un clasificador con mayor margen de separación posible para
minimizar el efecto de ruidos en el sistema de reconocimiento de formas.

Este procedimiento de la máquina de máquina de soporte vectorial trata de determinar un clasificador que tenga una banda lo más ancha posible entre las muestras más próximas.

ix. Aprendizaje de Clasificadores no Lineales en el RF

Además de los clasificadores NN, KNN y bayesianos, es posible definir clasificadores para clases no linealmente separables utilizando perceptrones dispuestos en múltiples capas, que son redes sin realimentarse con una o más capas de nodos entre las entradas y salidas. Estas capas adicionalmente contienen capas ocultas, o nodos que no están conectados directamente a las entradas ni a las salidas. El perceptrón multicapa es el primer antecedente de red neuronal, pero en esta disciplina se estudian otras topologías más complejas que la considerada.

- **Topología de Perceptrones Multicapa**

  Los perceptrones tanto simples como en multicapa generan regiones de clasificación que tiene propiedades geométricas notables, conectando con conceptos de geometría computacional tales como son las regiones convexas. En principio para realizar una reflexión general nos referimos a perceptrones con funciones de activación de naturaleza booleana, bien sea con salidas \{-1,+1\} o bien \{0,1\}. Un perceptrón simple genera una estructura de clasificación definida por un hiperoano que separa el espacio multidimensional en dos regiones abiertas, siendo ambas convexas. Pará comprender la topología de la decisión de un perceptrón de dos capas, es necesario referirse a un modelo simplificado del mismo en el que la segunda capa realiza una función similares a un and booleano. En el marco de este modelo simplificado un perceptrón de tres capas puede asimilarse a uno en el que la tercera capa realiza una función booleana or. En los perceptrones multicapas se utilizan funciones de activación continuas que permitan la continuidad y derivabilidad de la función de trasferencia al efecto de poder generar procedimientos prácticos de entrenamiento. En una red neuronal en problemas de clasificación no lineal, hemos de destacar dos tipos de elementos:

  - **Los elementos visibles de la red**, que corresponden con las variables de entrada y salida de la misma que tiene un significado concreto en el ámbito de la **aplicación**.
- **Los elementos ocultos de la Red**, que son todas las unidades de proceso y capas de variables internas. Estos carecen de significados traducibles directamente a una explicación simbólica y son metodológicamente invisibles y carentes de significado concretos.

- **Redes defunciones de base Radial**

  Los perceptrones de tres capas proporcionan clasificadores universales que son capaces de adaptarse a un conjunto de datos arbitrario, pero su principal inconveniente lo constituye un procedimiento masivo, de fuerza bruta, con infinidad de elementos lineales y parámetros que ajustar y sin un conocimiento preciso sobre si tales elementos son o no necesarios para una aplicación dada. Las redes neuronales basadas en funciones de base radial proporcionan una alternativa conceptualmente más simple. Las redes de funciones de base radiales una forma de clasificación no lineal y consiste en cuanto mayor sea la dimensionalidad del espacio trasformado, mayor será la probabilidad de que un problema no linealmente separable en las coordenadas $x$ si lo sea en las coordenadas $z$.

  En conjunto sus parámetros han de ser obtenido en un procesos de aprendizaje que se basa en minimizar el error para un conjunto de $m$ muestras.

**x. Selección de Características en el RF**

La complejidad de los procedimientos de aprendizaje en general y de los clasificadores particularmente dependen de dos factores principales: el volumen de muestras de aprendizaje y la dimensionalidad del espacio de representación. Tanto como el volumen de muestras de aprendizaje como la dimensionalidad de espacio se puede reducir con técnicas como la se lección de atributos y la trasformación de atributos.

- **Selección de Atributos**

  En este caso los datos resultantes son extraídos desde el conjunto original formando un subconjunto del original. Él gran problema de este enfoque es que es necesario determinar cuáles son los más significativos respecto a las clases del conjunto de datos.
Transformaciones de Atributos

En este caso se genera una transformación de los datos de forma que los resultantes pueden ser ordenados sistemáticamente en cuanto a su relevancia. Existen técnicas que simplifican este problema al generar conjuntos significativos en el número deseado, los conjuntos sistemáticamente ordenados en cuanto a relevancia.

Se pueden considerar algunas técnicas para el reconocimiento de formas entre estas tenemos:

a. **Análisis de Componentes Principales (PCA)**

El procedimiento de análisis de componentes principales es uno de los más utilizados en tareas de análisis de datos por su simplicidad computacional y robustez matemática. Consiste en generar una trasformación lineal en los atributos, de forma que se dé lugar a un conjunto final que sea estable en cuanto a representar la variabilidad inicial de datos. El objetivo del análisis de componentes principales es determinar un nuevo sistema de coordenadas en las que expresar un conjunto de datos tal que exista independencia.

b. **Análisis de Componentes Independientes (ICA)**

El objetivo del análisis de componentes independientes es generar una trasformación de los datos tal que el resultado esté formado por atributos que sean estadísticamente independientes. En análisis de componentes principales (PCA) los atributos no son estadísticamente independientes, sino que solo lo son respecto a los estadísticos de segundo orden. Por esta razón ICA es más general que PCA, pero evidentemente más complejo. La forma práctica de realizar la convención es básicamente la de una maquina lineal.

c. **Escaleo Multidimensional**

Esta técnica de Escale o multidimensional aborda el problema de reducción de la dimensionalidad mediante un enfoque diferente. En este caso no existe trasformación de coordenadas, sino que el objetivo es que las muestras en el espacio transformado guarden entre sí unas distancias que sean lo más parecida a las distancias existentes entre ellas en el espacio original. Este enfoque permite que la técnica sea más abstracta que las anteriores y pueda ser aplicada a situaciones de análisis de datos más generales donde se parte de matrices...
de interdistancias entre muestras, aunque las mismas carezcan de base vectorial.

xi. **Validación y Comparación en el RF**

Una vez efectuado el aprendizaje del clasificador, es necesario realizar un conjunto de pruebas o test para poder reconocer sus prestaciones. Estas pueden hacer referencia a los dos procesos de reconocimiento de formas, por un lado tenemos la eficiencia del clasificador en línea, y por otro lado, la de los procesos de entrenamiento. Podemos dividir la eficiencia en dos tipos: la primera hace referencia a la eficiencia computacional, es decir al consumo de recursos en términos de tiempo y memoria que requiere. El segundo tipo, es relativo a la tasa de error que comete el clasificador.

Algunas de las cuestiones que deben de ser resueltas posteriormente del entrenamiento de un clasificador son las siguientes:

- Test y evaluación de las prestaciones de un clasificador. Para ello se suele utilizar un conjunto de datos, que mediante diversas técnicas permita entrenar y evaluar sus prestaciones.

- Comparación de las prestaciones de diferentes clasificadores para determinar el más adecuado o para validar comparativamente las prestaciones de un nuevo clasificador respecto a otros existentes.

- Combinación de los resultados de diferentes clasificadores para obtener una clasificación más precisa. También suele utilizarse para obtener clasificación de mayor calidad a partir de varios clasificadores de baja calidad pero bajo coste.
CAPÍTULO III: DISEÑO DE LA INVESTIGACIÓN

3.1. Tipo de la Investigación
Aplicada.

3.2. Diseño de la Investigación
Aplicando el diseño de Pre Prueba – Post Prueba (Sánchez, 2006), con un solo grupo de control, cuyo esquema es el siguiente:

<table>
<thead>
<tr>
<th>Pre prueba</th>
<th>Post prueba</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grupo de estudio</td>
<td>0₁</td>
</tr>
</tbody>
</table>

Dónde:

X: Es el algoritmo de reconocimiento en el diseño de identificación de tumores de mama de transformación propuesto.

O₁: Es la Observación antes de Aplicar el Método al Grupo Experimental.

O₂: Es la Observación al Grupo de Control.

De donde, se aplicará una comparación entre O₁ y O₂ para poder validar la Hipótesis.

Reconocimiento.

La comparación de los resultados de las observaciones, determinarán la validez de la hipótesis formulada, mediante el método estadístico T-Student para análisis de grupos en observaciones independientes. De esta manera se verá si hay mejora o no, a partir de los resultados obtenidos.

3.3. Población y Muestra

El Universo de la presente investigación está conformado por todas las imágenes de carcinoma mamario del Hospital “Almenor Aguinaga Asenjo” de la ciudad de Chiclayo que demandan el diagnóstico por resonancia magnética, en el año 2013.

La muestra estuvo conformada por varias imágenes de pacientes que fueron atendidas, con la técnica de la mamografía digital, donde se aplicó el algoritmo para la fase de reconocimiento en entorno tecnológico.
3.4. Variables de Estudio

A continuación en la Tabla 2 se muestran las variables del estudio tanto independiente como dependiente las cuales son producto de la hipótesis de la investigación, adicionalmente se mencionan los indicadores con las cuales se están midiendo estas variables.

<table>
<thead>
<tr>
<th>Tipo</th>
<th>Variables</th>
<th>Indicadores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independiente</td>
<td>Diseño de un algoritmo de reconocimiento de Tumores mamarios.</td>
<td>-Exhaustividad del algoritmo de reconocimiento de tumores de Mama.</td>
</tr>
<tr>
<td>Dependiente</td>
<td>Reconocimiento en la identificación de tumores de mama.</td>
<td>-Efectividad de Aciertos del Algoritmo. -Rapidéz en el reconocimiento.</td>
</tr>
</tbody>
</table>

Fuente: Elaboración propia

3.5. Técnicas e Instrumentos

En la presente investigación se utilizaran las siguientes técnicas e instrumentos que permitirán probar los objetivos planteados:

3.5.1. Técnicas

En las la presente investigación se utiliza la técnica de Observación Indirecta de algoritmos de reconocimiento diseñados tradicionalmente, además se hace uso del análisis comparativo.

La comparación de características técnicas de la aplicación (medición) en base al rendimiento su evaluación y Verificación, la Investigación y consulta bibliográfica y el desarrollo humano entre otros.

3.5.2. Instrumentos

También utilizaremos otros medios instrumentales como son:

Los Test, cuadros comparativos.
CAPÍTULO IV: RESULTADOS

4.1. Análisis de Modelos de Algoritmos de Reconocimiento y Clasificación de Tumores de Mama

Analizaremos diversos algoritmos para el diagnóstico del cáncer de mama. Se presenta una comparación con respecto a ciertas características que permiten observar el mayor beneficio de algunos diseños de algoritmos con respecto a otros; el diseño o diseños algorítmicos seleccionado servirá de referencia para el modelo de nuestro diseño algorítmico computacional.

Los algoritmo que estoy utilizando para la realización de este modelo que ayudaran a mejorar el reconocimiento en la identificación de tumores de mama son: Algoritmo de clasificación Difusa para Tumores de mama, Algoritmo de Selección de Rasgos para el diagnóstico de tumores mamarios, Algoritmos Genéticos para Reconocimiento de Tumores de mama, Algoritmos de Clasificación y reconocimiento de Patrones de Gabor para diagnósticos de tumores de todos los algoritmos, también se abala en esta parte ciertos criterios que se definirán a continuación y entre ellos tenemos:

a. Sincronización

Herrera (2004) [11], nos dice que la sincronización es un componente importante de cualquier sistema de comunicación, basado en el tiempo de respuesta que nos da toda tecnología de la información.

b. Técnicas Orientadas a Objetos (Programación en clases)

Duran (2007) [4], nos dice que se definen técnicas con atributos externos de un objeto para conseguir clases que tengan características fundamentales como modularidad, abstracción, encapsulación y ocultación de información.

c. Técnicas de Programación Lógica

Fernández (1994) [7], nos dice que la programación lógica pura se entiende, en el sentido más tradicional, como programación en el lenguaje de la lógica de cláusulas.

d. Técnicas Imperativa

Rodríguez (2003) [20], nos dice que las técnicas imperativas es una secuencia finita de instrucciones, que se ejecutan una tras otra.

e. Técnicas de clasificación de bases de datos

Rob (2005) [19], señala que la clasificación de una base de datos es importante por la demanda de procesamiento que posee la tarea realizada como reconocimiento de rostros, mama, iris, entre otros, debido que son distintos tipos de base de datos dependiendo de la envergadura del proyecto.
f. Técnica Dinámica (programación de algoritmo voraces)

Ziviani (2007) sostiene que la programación dinámica calcula la solución de todos los subproblemas, partiendo de los subproblemas menores hacia los mayores, es decir resolviendo el problema utilizando la técnica divide y vencerás.

g. Reconocimiento por color a escala grises

Esqueda (2005) nos indica que los algoritmos trabajan a escala de grises porque una imagen a color posee 3 matrices R G B y la escala de grises es el promedio de las 3 matrices.

h. Reconocimiento por formas

Gonzales y Woods (2002), nos indican que el reconocimiento de formas es esencial para extraer las características relevantes de la imagen.

i. Método de implementación rápida

Gonzales y Woods (2002) sostiene que la implementación rápida de este tipo de algoritmos es inusual, porque en la etapa de entrenamiento estos algoritmos demoran a ejecutarse dependiendo del número de imágenes analizadas en la etapa de entrenamiento.

j. Finitud Poblacional (Programación imperativa o estructurada)

Gonzales y Woods (2002) señala que la población de estudio de los algoritmos mientras más amplia mejor, pero usualmente las base de datos de imágenes de rostro, iris, cáncer mamario, formas, entre otros casos. Llevan un número de imágenes válidas para su análisis.

k. Opacidad

Gonzales y Woods (2002) sostiene que el procesamiento de imágenes la opacidad de dicha imagen es importante para ciertos algoritmos con entorno controlado.

El algoritmo para ayudar a la identificación del reconocimiento tumores de mama será, Algoritmo de características de Gabor en mamografías digitales, este algoritmo con una serie de algoritmos como también el de soporte de máquinas vectoriales (SMV) y el de creación propia ayudaran al reconocimiento en la identificación de tumores de mama.

Los criterios de evaluación para la comprobar del presente diseño de algoritmo fueron analizando la complejidad algorítmica y el porcentaje de casos en el que ayudo a mejorar la identificación y el reconocimiento.
Para el análisis proporcionado en este capítulo vamos a seguir la clasificación anterior, ya que vamos a primar, las que a nuestro entender son las características expresivas más fundamentales de los diseños algoritmos de reconocimiento en la identificación de tumores de mama.

4.1.1. Algoritmo de Clasificación Difusa para Tumores de Mama

El diseño algorítmico de clasificación difusa es una herramienta que nos servirá para la obtención de los resultados; por último, esta muestra aplicaciones del algoritmo sobre tres bases de datos WDBC, WBCD y WPBCC, y los resultados que se han obtenido.

En los trabajos citados se hace referencia al estudio de tres bases de datos: Wisconsin Diagnostic Breast Cáncer (WDBC), Wisconsin Breast Cáncer Data base (WBCD) y Wisconsin Prognostic Breast Cáncer Chemothera- py Data base (WPBCC). Aquí se pretende mostrar como mediante la clasificación difusa se puede encontrar resultados similares que en algunos casos puede sustituir a procedimientos más tediosos de diagnóstico de cáncer.

a. Sincronización

Parcialmente. El diseño algorítmico de clasificación difusa utiliza parcialmente su sincronización ya que usualmente la fase primaria es la que consume mayor tiempo de procesamiento, y por lo general, se aplican en ella modelos concurrentes (conjunto relacionados de archivos) de selección de los algoritmos difusos por lo que es parcialmente sincronizada en el tiempo.

b. Técnicas Orientada a Objetos (Programación en clases)

Sí. El diseño algorítmico de clasificación difusa si utiliza técnicas orientadas a objetos ya que los clasifica a través de su heurística (solución de problemas alternativos que se dan).

c. Técnicas de Programación Lógica

Sí. El diseño algorítmico de clasificación difusa por su mismo nombre difuso si utiliza las técnicas de programación lógica como en todo algoritmo difuso porque en ella las categoriza y clasifica lógicamente a través de tres bases de datos.

d. Técnicas Imperativa (programación imperativa o estructurada)

No. El diseño algorítmico de clasificación difusa no utiliza la técnica de programación estructurada o imperativa ya que utiliza la técnica orientada a objetos que difiere a esta técnica.
e. **Técnicas de clasificación de bases de datos**
   Sí. El diseño algorítmico de clasificación difusa si utiliza técnicas de clasificación de base de datos y en ella las utiliza muy avanzadamente ya que las clasifica en tres bases de datos distintos.

f. **Técnica Dinámica (programación de algoritmo voraces)**
   No. El diseño algorítmico de clasificación difusa no utiliza técnicas de programación dinámica ya que estos son propios de los algoritmos voraces o algoritmo glotón.

g. **Reconocimiento por color a escala grises**
   Parcial. El diseño algorítmico de clasificación difusa utiliza parcialmente la técnica de reconocimiento por color a escala grises ya que se evoca más a la clasificación de los datos obtenidos con diferenciación de su base de datos de ella misma y solo toma reconocimiento de colores en mínimas cantidades pero evocado a la cantidad de datos que pueda obtener o recibir del o de los múltiples usuarios.

h. **Reconocimiento por formas**
   Parcial. El diseño algorítmico de clasificación difusa utiliza parcialmente la técnica de reconocimiento por formas ya que se evoca más a la clasificación de los datos obtenidos con diferenciación de su base de datos de ella misma y solo toma reconocimiento por formas en mínimas cantidades pero evocado a la cantidad de datos que pueda obtener o recibir del o de los múltiples usuarios.

i. **Método de implementación rápida**
   No. El diseño algorítmico de clasificación difusa es más extensa de implementar y aun no tiene proporcionado un método de rápida de implementación.

j. **Finitud Poblacional**
   Parcialmente. El diseño algorítmico de clasificación difusa tienes solo una finitud poblacional parcial ya que ha sido probado solo en algunos casos.

k. **Opacidad**
   Parcial. El diseño algorítmico de clasificación difusa tiene parcial opacidad ya que se caracteriza por la presencia de opacidades.
múltiples pero no en su totalidad en la clasificación de sus bases de datos.

4.1.2. **Algoritmo de Selección de Rasgos de reconocimiento para el Diagnóstico de Tumores Mamarios**

El diseño algorítmico de selección de rasgos de reconocimiento nos habla sobre la teoría de los conjuntos aproximados ya que ha abierto nuevas tendencias en el desarrollo de las técnicas de análisis de rasgos. Dentro de estas es significativo el concepto de reducto través de imágenes de exámenes mamarios cancerígenos, cuya obtención en un sistema de decisión es un proceso computacionalmente costoso aunque importante en análisis de datos y nuevo conocimiento. Debido a esto, se ha hecho necesario desarrollar diferentes variantes para calcular reductos. El presente trabajo investiga la utilidad que ofrece el modelo de los conjuntos aproximados en selección de rasgos y partículas cancerígenas de células muertas y se presenta un nuevo método con el propósito de calcular un buen reducto. Este nuevo método consiste en un algoritmo glotón que usa heurísticas para encontrar un buen reducto en tiempos aceptables. Se presentan, además, los resultados experimentales obtenidos usando diferentes conjuntos de datos de cáncer de mama.

**a. Sincronización**

Sí. El diseño algorítmico de selección de rasgos desinteresadamente desarrolla diferentes técnicas de extracción y selección de características, y por tanto, también tiene que ver con la sincronización rápida que efectúa en el reconocimiento de las mismas.

**b. Técnicas Orientada a Objetos (Programación en clases)**

Parcialmente. El diseño algorítmico de selección de rasgos maneja una estructura con clases muy parcial ya que dichas clases las maneja pero de manera o forma más estructurada.

**c. Técnicas de Programación Lógica**

Sí. El diseño algorítmico de selección de rasgos si demuestra técnicas de programación lógica ya que se encuentra basado en diseños matemáticos y teoremas demostrativos.

**d. Técnicas Imperativa (programación imperativa o estructurada)**

Sí. El diseño algorítmico de selección de rasgos si utiliza técnicas imperativas porque de una manera otra está compuesta por un dominio abstracto de valores y un conjunto de operaciones definidas sobre ese dominio, con un comportamiento específico.
e. **Técnicas de clasificación de bases de datos**
   Parcial. El diseño algorítmico de selección de rasgos utiliza parcialmente a través de un proceso computacional de análisis muy particularmente pero basándose más a partir de ello en el reconocimiento de sus rasgos.

f. **Técnica Dinámica (programación de algoritmo voraces)**
   Sí. El diseño algorítmico de selección de rasgos si utiliza técnicas dinámica porque en ella emplea el algoritmo voraz, glotón o llamado también goloso propio de esta técnica de programación.

g. **Reconocimiento por color a escala grises**
   No. El diseño algorítmico de selección de rasgos no utiliza técnicas de reconocimiento por color ya que solo utiliza técnicas de reconocimientos por formas y entre otras técnicas que le son más elementales para realizar su objetivo.

h. **Reconocimiento por formas**
   Sí. El diseño algorítmico de selección de rasgos si utiliza técnicas de reconocimiento por formas ya que son propios de este diseño algorítmico para realizar su objetivo.

i. **Método de implementación rápida**
   No. El diseño algorítmico de selección de rasgos algoritmo no tiene un método rápido de implementación porque lo hace a través de rasgos particulares que se encuentre como un método más tedioso y dificultoso para implementarlo ya contiene un lenguaje alturado y personas que no están capacitadas en un nivel altamente avanzado no podrían desarrollarlo.

j. **Finitud Poblacional**
   No. El diseño algorítmico de selección de rasgos no tiene una finitud poblacional ya que solo fue desarrollado como un avance en las ciencias aplicadas y no de una manera extensa por la misma razón que es dificultoso y costoso en su manera de implementar.

k. **Opacidad**
   Sí. El diseño algorítmico de selección de rasgos si contiene procesos de opacidad en su estructura propia característica (oscura) de dicho algoritmo
4.1.3. Algoritmos Genéticos de Reconocimiento para Diagnósticos de Tumores de Mama

El diseño algorítmico genético de reconocimiento para diagnósticos de tumores de mama está basado en los diseños de algoritmos de J. Holland en el año 1970 se encuentra como un logro más de la Inteligencia Artificial en su intento de replicar comportamientos biológicos, con los avances científicos que ello implica, mediante la computación. Se trata de algoritmos de búsqueda basados en la mecánica de la selección natural y de la genética. Utilizan la información histórica para encontrar nuevos puntos de búsqueda de una solución óptima del problema planteados con esperanzas de mejorar los resultados. En el presente artículo se realizará una introducción a los Algoritmos Genéticos: qué son, de dónde proceden, y en qué difieren de otros métodos de búsqueda, analizándose asimismo, su aplicación principal que es la de reconocer y diagnosticar tumores mamarios.

a. Sincronización

Parcialmente. El diseño algorítmico genético de reconocimiento para diagnóstico de tumores de mama utiliza parcialmente la sincronización ya que su fase primaria consumen poco tiempo de procesamiento, y por lo general, se aplican en la evaluación genética de la búsqueda y resultados en dichos algoritmos en memoria compartida que contienen pocas instrucciones de sincronización solamente para las zonas de procesamiento y evaluación genética.

b. Técnicas Orientadas a Objetos (Programación en clases)

Sí. El diseño algorítmico genético de reconocimiento para diagnóstico de tumores de mama si utiliza técnicas orientadas a objetos ya que de una manera u otra trata de desarrollar métodos que tienen que ver con comportamientos biológicos y que se envuelven con la herencia a través de sus genes y que difieren de otros métodos de búsqueda.
c. Técnicas de Programación Lógica

Sí. El diseño algorítmico genético de reconocimiento para diagnóstico de tumores de mama si emplea técnicas de programación lógica porque trata de entrelazar conocimientos de las ciencias biológicas con la inteligencia artificial.

d. Técnicas Imperativa (programación imperativa o estructurada)

Sí. El diseño algorítmico genético de reconocimiento para diagnóstico de tumores de mama si utiliza la programación estructurada ya que este algoritmo resulta de los pasos y diseños de la estructura de un árbol muchos de estos resultados incluyeron en la réplica o desarrollo de varias invenciones en los años 2000 como por ejemplo en el GP también se ha aplicado al hardware evolable (hardware que puede cambiar su arquitectura, comportamiento dinámico y autónomamente relacionándose con su ambiente) y así como también a programas de ordenador.

e. Técnicas de clasificación de bases de datos

Sí. El diseño algorítmico genético si contiene técnicas de clasificación buscando resultados óptimos en los tipos de modelos de genos prototipos ya que en ella explora, analiza y en base a ello identifica el problema para un mejor reconocimiento de tumores de mama.

f. Técnica Dinámica (programación de algoritmo voraces)

No. El diseño algorítmico genético no contiene técnicas de programación dinámicas ya que estos son propios de otros diseños algorítmicos usados anteriormente.

g. Reconocimiento por color a escala de grises

Sí. El diseño algorítmico genético si utiliza técnicas de reconocimiento a escala grises para diferenciar el color de genes previamente computarizados y modelados en contrastes a tonos grises.

h. Reconocimiento por formas

Sí. El diseño algorítmico genético si utiliza técnicas de reconocimiento por formas para la diferenciación esquematizada de genes por formas y tamaños específicos previamente computarizados y modelados.
i. Método de implementación rápida

Parcial. El diseño algorítmico genético contiene una parcial implementación ya que es difícil implementarla cuando no se tiene la información suficiente y necesaria como son diseños matemáticos estructurales genéticos y algorítmicos y muchas veces esa información es relevante para su implementación.

j. Finitud Poblacional

Parcial. El diseño algorítmico genético sólo puede trabajar con poblaciones finitas no muy grandes.

k. Opacidad

No. El diseño algorítmico genético sólo está guiado por la aptitud de los individuos (comportamiento de sus genes), y no incorpora ningún otro conocimiento específico del problema en cuestión.

4.1.4. Algoritmo de Clasificación y Reconocimiento de Patrones de Gabor para Diagnósticos de Tumores Mamarios

El diseño algorítmico de clasificación de patrones Gabor trata sobre el diagnóstico histológico correcto en los tumores de partes blandas es muy importante como fase previa al abordaje de estas lesiones (tumores de mama). Las técnicas de reconocimiento de patrones pueden ser una herramienta muy útil tanto en el diagnóstico como en la extracción de la información relevante, utilizando imágenes de RM (Resonancia Magnética) y datos epidemiológicos. Se han obtenido muy buenos resultados, por encima del 90% de acierto, al discriminar entre estirpes. La exactitud al diferenciar entre una estirpe y las demás ha sido también muy elevada. Estos resultados, junto con información sobre benignidad/malignidad con modelos previamente descritos de tumores de mama, proporcionan clasificadores a integrar en una herramienta médica de apoyo al diagnóstico de los tumores de partes blandas.

a. Sincronización

Sí. El diseño algorítmico de clasificación de patrones Gabor si contiene sincronización ya que su diseño tiene que ver con la dependencia de cada patrón interconectándose entre sí que también puede dividirse como una central sincronizada de acuerdo a los procesos que realiza dentro de su estructura.
b. **Técnicas Orientada a Objetos (Programación en clases)**

Sí. El diseño algorítmico de clasificación de patrones Gabor si contiene técnicas orientada a objetos porque en ella hereda clases a través de un patrón de uno o varios objetos en este caso los objetos serán las imágenes RM los rasgos directos o indirectos de y entre otros elementos que son parte esta técnica que utiliza dicho algoritmo.

c. **Técnicas de Programación Lógica**

Sí. El diseño algorítmico de clasificación de patrones de Gabor si utiliza técnicas de programación lógica porque emplea de una manera u otra partiendo de su funcionalidad de reconocimiento en el acierto y diferenciación de estípes y tumores benignos o malignos.

d. **Técnicas Imperativas (programación imperativa o estructurada)**

Sí. El diseño algorítmico de clasificación de patrones Gabor si utiliza técnicas imperativas porque parte de acuerdo al diseño algorítmico estructural del scanner computacional de resonancias magnéticas instrumentó con el cual trabaja conjuntamente para realizar las imágenes RM.

e. **Técnicas de clasificación de bases de datos**

Sí. El diseño algorítmico de clasificación de patrones si contiene técnicas de clasificación de base de datos porque a través de su reconocimiento con patrones contiene los llamados patrones DAO (objeto de acceso a datos) dicha característica también es fundamental para sus reconocimientos acertados de este tipo de diseño algorítmico.

f. **Técnica Dinámica (programación de algoritmo voraces)**

Parcialmente. El diseño algorítmico de clasificación de patrones utiliza parcialmente técnicas de programación dinámica partiendo de su diseño matemático funcional que busca de una manera fraccionaria la forma más óptima de subdividir el problema de diagnóstico de patrones tratando de buscar en muchos de sus puntos determinantes la mejor solución posible, dicho diseño parte de la técnica dinámica basado en el método de divide y vencerás.
g. **Reconocimiento por color a escala grises**

Si. El diseño algorítmico de clasificación de patrones si utiliza técnicas de reconocimiento por color a escala grises para la diferenciación y precisión de imágenes que son obtenidos de la resonancia magnética los cuales son propios del diseño algorítmico Gabor para diagnóstico de tumores de mama.

h. **Reconocimiento por formas**

Si. El diseño algorítmico de clasificación de patrones si utiliza técnicas de reconocimiento por formas para la diferenciación del reconocimiento de formas tomadas de las imágenes de resonancia magnética.

i. **Método de implementación rápida**

Si el diseño algorítmico de clasificación de patrones si utiliza métodos de implementación sencillos y rápidos efectivos en la identificación de tumores de mama ya que utiliza diversos métodos y factores que ayudan a la implementación completa de dicho diseño algorítmico.

j. **Finitud Poblacional**

Parcial. El diseño algorítmico de clasificación de patrones utiliza parcialmente la finitud poblacional ya que es difícil de reconocer un tumor en zonas donde aún no se aplica el método de Gabor ya que es un factor clave para hacer este reconocimiento.

k. **Opacidad**

Sí. El diseño algorítmico de clasificación de patrones si utiliza métodos y técnicas de opacidad en sus formatos de imágenes RM.
4.2. Análisis Comparativo de Modelos de Algoritmos de Reconocimiento y Clasificación de Tumores de Mama

En este capítulo hemos analizado cuatro algoritmos de reconocimiento e identificación de tumores de mama de muy distinta naturaleza.

Todos presentan ventajas e inconvenientes, y debido a la heterogénea naturaleza de las aplicaciones de reconocimiento e identificación no podemos decir que unos sean mejores que otros. No olvidemos que hoy en día existe una gran cantidad de aplicaciones de algoritmos que se ajustan perfectamente al algoritmo de clasificación de patrones Gabor, la cual puede ampliarse también con la manera formal de aplicación de identificación descrita en el algoritmo de selección de rasgos. También hay aplicaciones como el del algoritmo de selección de rasgos de reconocimiento con un fuerte reconocimiento a través de selección de rasgos de imágenes, en cuyo caso necesitarían aplicar un sistema de reconocimiento capaz de identificar y representar totalmente estas capacidades.

Si abstraemos las principales características de cada sistema de representación, identificación y reconocimiento; de manera que haremos una comparación en formato tabular, obtendremos los resultados descritos que se presentan en la Tabla 3.

De lo antes expuesto podemos resumir en lo siguiente:

Las aplicaciones de algoritmos orientadas al reconocimiento e identificación tumores de mama ofrecen mucho más enriquecimiento de diagnóstico, pues desde la definición de reconocimiento e identificación se habla de no sólo diagnóstico de tumores de mama como único objetivo sino también de imágenes, pruebas de diagnóstico, entre otros.

Las propuestas de algoritmo reconocimiento no necesariamente deben brindar soporte para una arquitectura de identificación de tumores de mama eficiente por parte del usuario, puesto que el diseño de identificación nos permite el reconocimiento en su totalidad de las imágenes captadas para el diagnóstico de tumores de mama y trae como consecuencia la usabilidad del mismo.

Las propuestas de algoritmos de diagnóstico de tumores de mama deben tener en consideración algunos factores como la complejidad y la reutilización, que influyen en la viabilidad y la propia ejecución de proyectos de desarrollo, reconocimiento e identificación de imágenes, entre ellos para la diagnosticación de tumores de mama.
Por último, las propuestas de reconocimiento e identificación de tumores de mama deben tener en consideración a la sincronización, ya que es la sincronización el principal recurso de diagnosticación, reconocimiento e identificación de tumores de mama hacia el usuario.

Por tanto, en la presente investigación se considera que para el diseño de un Algoritmo de Reconocimiento en la Identificación de Tumores de Mama nos basaremos en tomar algunas definiciones formales requeridas de cada aplicación de reconocimiento e identificación para el diseño el mismo.

Basada en las características descritas anteriormente (numeral 4.1), se concluye obteniéndose el siguiente cuadro comparativo.
<table>
<thead>
<tr>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>F</th>
<th>G</th>
<th>h</th>
<th>i</th>
<th>j</th>
<th>k</th>
</tr>
</thead>
</table>

Fuente Creación Propia

- a. Sincronización
- b. Técnicas Orientada a Objetos (Programación en clases)
- c. Técnicas de Programación Lógica
- d. Técnicas Imperativa
  (Programación imperativa o estructurada)
- e. Técnicas de clasificación de bases de datos
- f. Técnica Dinámica (programación de algoritmo voraces)
- g. Reconocimiento por color a escala grises
- h. Reconocimiento por formas
- i. Método de implementación rápida
- j. Finitud Poblacional
- k. Opacidad

| ✔ | : Si |
| × | : No |
| ✔ ✔ | : Parcialmente |
4.3. Diseño del Algoritmo de Reconocimiento en la Identificación de Tumores de Mama

Dentro del diseño del algoritmo de reconocimiento en la identificación de tumores de mama definiremos y enfocaremos el algoritmo propuesto que estamos planteando basándonos en los siguientes puntos que a continuación detallaremos.

4.3.1. Enfoque del algoritmo propuesto

Para llegar a desarrollar el diseño del algoritmo propuesto de reconocimiento de tumores de mama hemos tomado en cuenta conceptos y partes fundamentales de los algoritmos mencionados como son el Algoritmo de Clasificación Difusa para tumores de mama, Algoritmo de Selección de Rasgos para el diagnóstico de tumores mamarios y el Algoritmo de Wavelet Gabor de nuestros algoritmo describiremos a continuación partes que se ha tomado en cuenta para el diseño del algoritmo propuesto.

El algoritmo de clasificación procesa los tipos de tumores de mama y los agrupa entre los distintos tipos de tumores y no tumores, este algoritmo se usa para determinar las clases y agruparlos en un espacio de puntos con características o puntos de interés similares.
Algoritmo de Clasificación Difusa para tumores de mama

1: Determinar el número de clusters \( K \) de imágenes de tumores de mama.
2: Seleccionar \( k \) centroides iniciales, definiendo como centroide el punto que está más equidistante a los clusters se ubica de una forma arbitraria y teniendo en cuenta que el número de centroides tiene que ser menor o igual al número de datos.

\[
\sum_{i=1}^{k} \frac{x_i}{n}
\]

3: Determinar las fronteras de cada cluster.
4: Hallar las distancias entre cada centroide y lo datos de cada cluster.

\[
\sqrt{(x_i - x_f)^2 + (y_i - y_f)^2}
\]

5: Asignar a cada cluster el centroide que esté más cercano.
6: Repetir los pasos 3 y 4 hasta que las fronteras de los clusters no cambien significativamente.

\[\text{dis}(C_i, C_{i+1}) \leq \varepsilon\]

Figura 5: Diseño Algorítmico de Clasificación Difusa.

Este algoritmo ha sido seleccionado para extraer las características más significativas de cada imagen, ha sido seleccionado por su gran utilidad con otros algoritmos.

Algoritmo de Selección de Rasgos para el diagnóstico de tumores mamarios

1: Se calcula el vector \( R(T) = (R(A_1), \ldots, R(N)) \). Para todos los atributos del problema se calcula su \( R(A) \) y así con todos los valores se forma el vector \( R(T) \).

2: Se determina los \( n \) mejores atributos por los cálculos del paso anterior. El valor de \( n \) se puede seleccionar por el usuario. Como resultado de este paso se obtiene el vector:

\[
RM = (R(A_i), R(A_j), \ldots) \quad \text{con} \quad n = |RM|
\]

3: Se determinan las combinaciones de \( n \) en \( p \) (valores seleccionados por el usuario) desde los atributos seleccionados en el paso 2 se obtiene el vector de combinaciones.

\[
Comb = \left\{ \left\{ a_1, a_3, a_5 \right\}, \left\{ a_1, a_3, a_8 \right\}, \left\{ a_3, a_5, a_8 \right\}, \left\{ a_1, a_5, a_8 \right\} \right\}
\]

4: Se calcula el grado de dependencia de las clases con respecto a cada una de las combinaciones obtenidas en el paso anterior. Como resultado de este paso se obtiene el vector de dependencias.

\[
DEP = (k(comb_1, d), \ldots, k(comb_n, d))
\]

5: Para cada atributo \( A \) se calcula \( H(A) \) según la siguiente ecuación.

\[
H(A) = \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{k(comb_i, d)}{}
\]

Nos proporciona una clasificación por atributos

Figura 6: Diseño Algorítmico de Clasificación de Rasgos

Desde el descubrimiento de la organización cristalina de la corteza visual en el cerebro de mamíferos hace aproximadamente 30 años, una enorme cantidad de investigaciones experimentales y teóricas han permitido un mejor entendimiento del cerebro y de las propiedades de respuesta de sus células.

Se han desarrollado avances bajo la premisa que las células simples en la corteza visual pueden ser modeladas por funciones de Gabor.

Las funciones bidimensionales de Gabor son filtros espaciales, que permiten alcanzar la resolución conjunta de información máxima en los espacios bidimensionales espacial y frecuencial.

Este fenómeno indica que existe una relación entre la resolución en el dominio temporal y frecuencial. Gabor descubrió que las funciones gaussianas moduladas por exponenciales complejas resultan tener mejor relación entre la resolución en los dominios temporal y frecuencial.

Gabor nos permitirá extraer las características con mayor precisión para ayudar en la identificación de reconocimiento de tumores de mama, este algoritmo fue usado en reconocimiento de voz, rostro, iris, entre otros. Para aplicar este algoritmo a los tumores de mama realizaremos unos ajustes a las imágenes como aconsejan los autores y algunas mejoras y ajustes a los algoritmos.

Figura 7: Ejemplo de Wavelet de Gabor.

Fuente: De La Cruz, S. & Torres, J. (2012) [27].
Figura 8: Diseño de un proceso de filtro del Algoritmo Wavelet de Gabor.

Fuente: De La Cruz, S.& Torres, J. (2012) [27].
Este algoritmo nos sirve para reducir el tamaño de los datos, sin perder las características más esenciales de la imagen o matriz bidimensional.

**Algoritmo de Gabor**

**Entrada:** Una Imagen de m x n

1: Se define una estructura de grafo sobre la cara, cuyos nodos son puntos de interés, que se puedan localizar fácilmente y que posean la misma estructura en todos los rostros.

2: Cada nodo del grafo es caracterizado utilizando un banco de filtros de Gabor de diferentes frecuencias y orientaciones. En cada nodo, se calcula la respuesta de todos los filtros, a lo que se da el nombre de Jet. Por consiguiente, cada nodo queda etiquetado con sus coordenadas y su Jet asociado.

3: Para una imagen nueva se busca en la base de datos el conjunto de Jets que sean más similares. Para esto se centra el grafo en los ojos de la nueva imagen, se calculan los jets de estos puntos y se comparan contra los jets de los ojos en la base de datos. Para la comparación de los jets se utiliza una función de similitud por fase, ésta es similar a una correlación, y está definida como se describe en la ecuación.

\[
S(j, J) = \frac{\sum_{i=1}^{n} a_i a_j' \cos(\phi_i - \phi_j)}{\left(\sqrt{\sum_{j=1}^{x} a_j^2 \sum_{i=1}^{y} a_i^2}\right)}
\]

4: Donde: \(a'\) es la magnitud del Jet de la base de datos, \(a\) la magnitud del Jet de la imagen nueva, \(\phi'\) la fase del Jet de la base de datos, \(\phi\) la fase del Jet de la imagen nueva y \(n\) el número de coeficientes del Jet.

**Salida:** Vector Unitario de la Imagen.

Figura 9: Diseño Algorítmico de Wavelet de Gabor

Fuente: De La Cruz, S. & Torres, J. (2012) [27].
La representación de la transformada de Wavelet de Gabor de una imagen, es la evolución de la imagen original con el banco de filtros Gabor.

En el desarrollo de esta investigación se utiliza la magnitud de los filtros (no la fase) ya que se ha determinado que es robusta y tolerante a distorsión para el reconocimiento de patrones.

La transformada de Gabor será implementada en el dominio de la frecuencia. Todas las imágenes de tumores de cáncer de mama deberán ser submuestreadas antes de ser procesadas a un tamaño de 180x160 para acelerar el proceso de extracción de características.

El tamaño de la imagen fue reducido para la reducción del vector de características ya que se está trabajando en un espacio altamente dimensional.

En esta investigación se manejarán 8 orientaciones y 4 frecuencias formando un banco de 32 (8x4) filtros de Gabor. Cuando se evoluciona una imagen con este banco de filtros se producen 32 imágenes filtradas o de las que se tiene cada una con diferentes características a diferentes frecuencias y orientaciones. Cada imagen filtrada es de tamaño 180 x 160 pixeles o menos previamente predefinida. Para agrupar en una sola matriz toda la información de las 32 imágenes filtradas.

La adaptación al algoritmo de Gabor viene en Algoritmo Completo Propuesto basado en Gabor donde observamos la adaptación de Gabor para imágenes en el paso 7 del diseño del algoritmo de reconocimiento de tumores de mama que veremos a continuación.

### 4.3.2. Diseño del algoritmo propuesto

El algoritmo propuesto se basara en algunas técnicas del algoritmo de clasificación difusa para tumores de mama y algoritmo de selección de rasgos para diagnóstico de tumores mamarios como también en la mayor parte de técnicas que utiliza el algoritmo de los wavelets de Gabor que nos permitirán las sincronizaciones determinadas, orientaciones y determinadas frecuencias. Estos a la vez están formados por un banco de filtros lineales e invariantes temporales, por lo que se los caracteriza por su respuesta al impulso. La misma se basa en un núcleo gaussiano modulado mediante una sinusoides.

Estos Algoritmos antes mencionados dieron por concluido el diseño del algoritmo propuesto.
4.3.3. Algoritmo de Reconocimiento e identificación de tumores de mama

Propuesto (Basado en Reducción de dimensiones de las imágenes y reducción del área de trabajo)

1: Ingresa la mamografía inicial.
2: Se procede a reducir el área de trabajo.
3: Se procede a reducir el ruido.
4: Se procede a mejorar el realce de contraste de la RM.
   //este paso es la modificación al algoritmo de Gabor aplicada a imágenes
5: Desde O = 1 hasta 8 (Indica el número de orientaciones que se maneja)
   Desde f = 1 hasta 4 (número de frecuencias)
   Se forma un filtro O a frecuencia f orientaciones O, la cual tiene un tamaño de 112 x 92 o predefinida
   Se submuestraa la matriz de 10304 x 1 por un factor de 4 formando ahora una matriz de 28 x 23
   Se concatenan las columnas de la imagen O formando una matriz de 644 x 1
   Se guarda ese resultado
   
   f = f + 1
   O = O ++ 1
   Se concatenan los 32 resultados obtenidos para formar una matriz de 1 x 20608.
6: Hasta ahora tenemos una mamografía procesada [resultado del paso 1 al 5].

   //este paso es para ajustar las imágenes de entrada e iniciar la cauterización.
7: Detección de regiones de interés con patrones de orientación asociados a la presencia de distorsiones de la arquitectura de la glándula mamaria.
4.3.3.1. Sub módulos del algoritmo propuesto

A continuación veremos los sub módulos del algoritmo propuesto descrito líneas anteriores.

A. Sub Modulo: Ingresar la mamografía Inicial

Modulo necesario para convertir las imágenes digitalizadas en matrices para poder realizar el procesamiento algorítmico propuesto.

1: Leer imagen en formato jpg (fopen).
2: IMG_ALTO = obtener alto de la imagen
3: IMG_ANCHO = obtener ancho de la imagen
4: Para pixelx = 0 hasta IMG_ALTO
   Para pixely = 0 hasta IMG_ANCHO
       MatrizRGB = rgb[pixelx][pixely]
   End
End
5: Return matrizrgb.

Fuente: Creación Propia.

B. Sub Módulo: Reducir el área de trabajo

Reducción del tamaño de la matriz bidimensional con el fin de reducir las dimensiones de dicha matriz para un rápido procesamiento computacional.

1: Suavizado de imagen en 1º derivada de la gaussiana.
2: Detección: máximos locales significativos…
   Supresión de no-máximos
       Poner a 0 los valores de magnitud que no sean máximo local en la dirección del gradiente.

   Histéresis de umbral (doble umbral)
       Dos umbrales: T1 < T2
       Los valores > T2 → bordes
       Los valores < T1 → no bordes
       Los intermedios, solo si están conectados a algún borde → bordes (algoritmo recursivo)
3: estimación con precisión subpixel (opcional).

Fuente: Creación Propia.
C. **Sub Módulo: Reducir ruido en imágenes**

Modulo para reducir el ruido en las imágenes digitalizadas a través de filtros.

1: Para pixely = 0 hasta IMG_ALTO
   Para pixelx = 0 hasta IMG_ANCHO
      Filtro = \[
                  \begin{bmatrix}
                    1 & 2 & 1 \\
                    2 & 4 & 2 \\
                    1 & 2 & 1 \\
                  \end{bmatrix}
              \]
      NUEVA_IMG(i, j, Filtro)
   Fin - para
Fin - para

Fuente: Creación Propia.

D. **Sub Módulo: Mejoramiento de realce (filtro de paso alto)**

Modulo necesario para realizar el realce de las imágenes a través del filtro de paso alto, paso necesario para mejorar la calidad de la imagen.

1: para i = 0 hasta filas
   Para j = 0 hasta columnas
      Si i>=1 AND i<filas-1 AND j>=1 AND j<columnas-1
         Filtro = \[
                    \begin{bmatrix}
                      -1 & -1 & -1 \\
                      -1 & 8 & -1 \\
                      -1 & -1 & -1 \\
                    \end{bmatrix}
              \]
         NUEVA_IMG(i, j, Filtro)
      Fin-si
   Fin - para
Fin - para

Fuente: Creación Propia.
E. Sub Módulo: Detección de regiones de interés (Segmentación por Crecimiento de Regiones)

Modulo necesario para obtener las regiones de interés y poder demarcar donde está la sección de tumor que vamos a procesar.

1: Generar el conjunto S de semillas
2: Sea I, una imagen, tal que Iᵢ = 0
3: Para todos ε S hacer
   Sean vᵢ, los 4-vecinos sin marca de s en I
   Si P(s U vᵢ) entonces
      Marcar s y vᵢ con su marca de región en Iᵢ
      Apilar vᵢ en S
   Fin-si
Fin-para

Fuente: Creación Propia.

4.3.4. Descripción del Diseño del Algoritmo de Reconocimiento en la Identificación de Tumores de Mama

El algoritmo computacional de reconocimiento en la identificación de tumores de mama parte de elementos que aportan características de reconocimiento de patrones al desarrollo de un sistema de identificación eficiente tomando en cuenta de que algunas de las aplicaciones descubiertas a través de algoritmos de reconocimiento e identificación donde la inmediatez y evolución continua son atributos principales, esto sugiere una arquitectura de aplicación que pueda ser altamente especializada y que enfatice un enfoque de desarrollo riguroso dando por finalizado al diseño del algoritmo propuesto.

De los pasos generales del Algoritmo de Reconocimiento en la identificación de tumores de mama Propuesto. (Basado en Reducción de dimensiones de las imágenes y reducción del área de trabajo) módulos centrales tomados en cuenta obtenemos la siguiente descripción:

El pasó1. Donde se inicia todo el proceso de reconocimiento e identificación ingresa la mamografía inicial.

El pasó 2. Se procede a reducir el área de trabajo que consiste en reducir el dimensionalidad del tamaño de la imagen ya que este factor es un paso muy importante para un mejor procesamiento de la misma.
El pasó 3: Se procede a reducir el ruido que consiste en reducir el distorsionamiento que puede darse al querer reconocer la imagen tomando en cuenta de los procesos de filtros que se le dan a la imagen.

En el paso 4: En este paso es el más fundamental y necesario para obtener las regiones de interés y poder demarcar donde está la sección de tumor que vamos a procesar.

El pasó 5: En este paso se da la detección de regiones de interés con patrones de orientación asociados a la presencia de distorsiones de la arquitectura de la glándula mamaria este paso se da la modificación al algoritmo de Gabor aplicada a imágenes y se procede al reconocimiento por regiones empleando métodos del algoritmo de Gabor modificando dichos proceso en la técnica del aprendizaje supervisado ya que es una técnica para deducir una función a partir de datos de entrenamiento.

Se toma el número de orientaciones que se da para la imagen o que se trata de manejar.

Desde $O = 1$ hasta 8 (Indica el número de orientaciones que se maneja).

Desde $f = 1$ hasta 4 (número de frecuencias).

Se forma un filtro $O$ a frecuencia $f$ orientaciones $O$, la cual tiene un tamaño de 112 x 92 o predefinida.

Se submueстра la matriz de 10304 x 1 por un factor de 4 formando ahora una matriz de 28 x 23.

Se concatenan las columnas de la imagen $O$ formando una matriz de 644 x 1

Se guarda ese resultado

$f = f + 1$

$O = O ++ 1$

Se concatenan los 32 resultados obtenidos o de las imágenes que se han tendido en cuenta a evaluar para formar una matriz de 1 x 20608 y así dar como resultado a la respuesta de reconocimiento e identificación de las imágenes de tumores mamarios que hayamos evaluado.

El pasó 6: En este paso se tiene una mamografía procesada de las tomadas o de las que se tiene procesando y es el resultado del paso 1 al 5. En este paso se ajustan las imágenes de entrada e inicia la cauterización o
resultado del procesado de filtrado que se le da a la imagen. Esto se ve en el filtro de imagen que ve a continuación:

![Filtros de grises](image)

Figura 10: Reconocimiento de imagen en escala de filtros de grises
Fuente: Creación Propia.

El paso 7: Se procede a mejorar el realce de contraste de la RM el realce de las imágenes a través del filtro alto y detección de la imagen más cercana con el algoritmo KNN, paso necesario para mejorar la calidad de la imagen

El diseño del algoritmo de reconocimiento de tumores de mama especifica nuevos elementos sobre los diseños algorítmicos computacionales de reconocimiento en la identificación de tumores de mama como son los algoritmos de selección de rasgos como también de algoritmos, genéticos, difusos y de patrones, vistos en el análisis comparativo realizado en el punto 4.2 descritos en el análisis que se hizo a dichos algoritmos, para aportar conceptos y elementos sobre los trabajos ya mencionados al entorno de reconocimiento de tumores de mama, así tenemos que proponer estos elementos de manera similar a la representación de los modelos elegidos, de tal manera que pueda generalizar los aspectos considerados para el diseño de nuestro algoritmo de reconocimiento en la identificación de tumores de mama, ya que esto nos permitirá un modelo propuesto desempeñándose en distintos sistemas de reconocimiento, más aun en identificaciones de tumores de mama, que es nuestro propósito.

El diseño del algoritmo de reconocimiento de tumores de mama separa claramente dos niveles, uno de ellos es el reconocimiento de tumores de mama clásico, y por otra parte la representación de conocimiento y ontología. Además aporta la idea de reconocimiento de
patrones o formas a escala de grises sobre el modelo; soporta explícitamente la sincronización, que ya vienen incorporadas en algunos modelos de los algorítmicos vistos; presenta organización relacional de contenidos gracias a los conocimientos previamente vistos, aporta eficiencia y seguridad en el reconocimiento preciso de tumores.

Los recursos sobre los cuales el algoritmo de reconocimiento en la identificación de tumores de mama actuará, estará involucrado en sistemas de reconocimiento de imágenes a través de patrones.

Las imágenes de reconocimiento a través de patrones de los cuales se recopila la información esencial para diseño de algoritmo del reconocimiento, así como las funciones que se esperan, son definidos formalmente de manera estructurada, separando rasgos, estructura y clasificaciones del formato de reconocimiento de tumores de mama.

Estos elementos posteriormente son instanciados como conocimiento bajo relaciones inherentes de su naturaleza lo cual es representado mediante una ontología que permite describir los rasgos o las formas partiendo de una clasificación de imagen de una manera computacionalmente viable.

El modelo también permitirá la posibilidad de identificación y reconocimiento en base al conocimiento, bajo inferencias lógicas esto debido a que los elementos del modelo están basados bajo ciertos axiomas formales sobre los cuales trabajan permitiendo así un eficiente razonamiento.

Así mismo el modelo posibilita las consultas para reconocimiento de tumores de mama sobre elementos formales bajo lenguajes de consultas de identificación dentro del enfoque del algoritmo de reconocimiento en la identificación de tumores de mama, permitiendo así la identificación, reconocimiento y optimización de resultados requeridos por parte de los usuarios.
4.3.5. Propósito del Diseño del Algoritmo de Reconocimiento en la Identificación de Tumores de Mama

El propósito del Diseño de un Algoritmo de Reconocimiento en la Identificación de Tumores de Mama es de proporcionar una representación formal de reconocimiento que permita el procesamiento de elementos que podamos encontrar en una imagen de tumores de mama, estos al ser formalmente definidos brindan la posibilidad de interoperabilidad con otros algoritmos (otros sistemas de reconocimiento), esto es la visión de algoritmo de reconocimiento.

Brindar a los usuarios potenciales reconocimientos en la identificación de tumores de mama específicamente, una mejor calidad de identificación en cuanto a sus consultas y de reconocimiento de imágenes de tumores de mama de una manera más rápida; y a los desarrolladores la posibilidad de construir agentes inteligentes que trabajen sobre recursos de identificación y reconocimiento para diferentes propósitos que se necesiten dentro del algoritmo especificado de una manera más óptima.

4.3.6. Elementos de algoritmo propuesto

A continuación explicaremos los elementos fundamentales del diseño computacional de reconocimiento de mama ya mencionados.

El modelo del algoritmo de reconocimiento de tumores de mama representa un diseño que ayuda en la identificación de tumores de mama, así tenemos:

\[ I_R = I_E \cap BD_{IP} \]

Dónde

\[ I_R = \text{Imagen reconocida} \]
\[ I_E = \text{Imagen a evaluar} \]
\[ BD_{IP} = \text{Base de datos de imágenes procesadas tras entrenamiento} \]

Estos componentes se detallan en los algoritmos antes mencionados, pero el resultado de esta intersección o de las imágenes más cercanas en vecindad a la imagen a evaluar proyectada son las imágenes reconocidas, clasificarla y así poder dar un resultado más preciso.
4.3.7. Arquitectura del algoritmo propuesto

El diseño del algoritmo de reconocimiento de tumores de mama se representa como un conjunto de elementos que permiten representar cualquier actividad computacional en un sistema de reconocimiento eficaz de esta patología tumoral ya mencionada con contenidos estructurados y relacionados explícita y formalmente, basándose en una combinación de algoritmos usados de este tema especificado en el cual encontraremos las relaciones y reglas de los diferentes conceptos a representar lo que permitirá al sistema obtener conclusiones y/o tomar decisiones procesando los reconocimientos extraídos del mismo sistema, en nuestro caso particular, el reconocimiento e identificación de tumores de mama.

La solución propuesta parte de un proceso adaptativo basado en la incorporación de elementos representativos sobre los algoritmos seleccionados, que permiten describir formalmente los recursos bajo el esquema de reconocimiento de tumores de mama.
Algoritmo propuesto para la extracción de características de tumores de cáncer de mama.

Figura 11: Arquitectura de entrenamiento para la identificación del reconocimiento de tumor de mama.

Fuente: Creación Propia.
4.3.8. Descripción de los componentes de la arquitectura

Entre los componentes de la arquitectura del algoritmo propuesto que tenemos describiremos a continuación:

**Nivel de Lectura de Imagen**

Este nivel incluye la lectura de las imágenes para su procesamiento, en este caso se leerán las imágenes de la base de datos de imágenes que servirán de entrenamiento en las otras etapas, en esta etapa se necesita de un médico experto en el tema para el correcto entrenamiento del diseño propuesto.

**Base de datos de imágenes RM**

Esta base de datos contendrá todas las imágenes que servirán de entrenamiento para nuestra arquitectura, pasando a formar una equivalencia en un espacio vectorial producido por segmentación.

**Nivel de pre - procesamiento de imágenes**

En este nivel incluye el mejoramiento de las imágenes con técnicas de procesamiento digital de imágenes como suavizamiento de la imagen, eliminación de ruido y realce de la imagen.

**Nivel de detección de regiones de interés**

En esta etapa se aplican algoritmos de detección de regiones de interés mediante segmentación de crecimiento de regiones para obtener puntos de interés que formaran un vector, que nos permitirá comparar por centroides.

**Nivel de Base de datos de regiones de interés (clasificación clustering)**

Este nivel provee de una clasificación de tumores y no tumores, esta clasificación se realizara con basado en clasificación vectorial por segmentación aplicada a crecimiento de regiones.

**Nivel de evaluación del clustering KNN**

Este nivel establece un nivel donde el experto doctor nos indicara si el resultado de la clasificación es correcta en caso contrario se procederá a ajustar desde la etapa de los filtros de Gabor.
4.4. Validez y verificación del diseño del algoritmo de Reconocimiento en la Identificación de Tumores de Mama

Para la verificación y validez del diseño del algoritmo de reconocimiento en la identificación de tumores de mama se realizó mediante métodos de muestreo de identificación de tumores y no tumores, con este diseño se hace viable la verificación del correcto funcionamiento de un identificador de tumores de mama.

4.4.1. Aspectos de contenido estándar del diseño del algoritmo de reconocimiento en la identificación de tumores de mama.

A continuación detallaremos los aspectos de contenido estándar del diseño algorítmico, nos servirá como referencia de que el diseño cumple con algunas normas que como ya se ha explicado además sirven para que las hospitales y doctores especialistas tengan un apoyo necesario para el uso de una mejor identificación de tumores de mama. Así tenemos:

Ámbito del Diseño del Algoritmo propuesto

Según concierne el diseño del algoritmo propuesto para mejora en la identificación de tumores de mama, se trata de una estructura algorítmica con manejo de un conjunto de clústeres de clasificados en el aprendizaje, y con conocimiento de experiencia de resultados para dar resultados como si fuera un experto, pero a la vez estará supervisada por un profesional especializado en el tema dotados de los conocimientos necesarios, así que el diseño propuesto tiene un ámbito en el cual se desempeña en su respectivo entorno de producción.

Imágenes de tumores nuevas

El diseño del algoritmo de reconocimiento de tumores de mama propuesto, para la mejora de identificación, hace referencia a que el diseño está probado con imágenes no ubicadas en los clúster, debido a que la posición de las imágenes para cierta persona siempre varía.

4.5. Caso de estudio.

El caso de estudio utilizado para la implementación de la presente investigación comprende realizar las siguientes funciones:

- Base de datos de radiografías de tumores de mama.
- Análisis de los resultados médicos emitidos por el doctor especialista.
- Manejo de base de datos multimedia de imágenes.
 Esto se realizó a la vez en un ambiente del Hospital “Almanzor Aguiñaga Asenjo” que se encuentra ubicado en la ciudad de Chiclayo, este hospital se encarga de evaluar casos de tumores de mama, con la finalidad de promover la aplicación del diseño para mejorar la identificación de tumores de mama, a la vez se ha tomado imágenes de casos reales cuyos nombres no serán reservados para no publicar su identidad, y que están publicadas en la base de datos pero que conjuntamente con el doctor especialista se puede analizar nuestro diseño los casos clínicos para probar mi diseño.

El hospital según las historias clínicas proporcionadas por el doctor del hospital antes mencionado nos proveerá las herramientas necesarias para realizar nuestro análisis.

Un total de 431 mamografías (digitalizado por el escáner Lumisys) se analizaron en realidad para validar el diseño con la detección de masas, que incluía 159 imágenes normales (de 80 individuos sanos) y 272 imágenes cancerosas (con al menos una masa maligna presente en cada imagen, a partir de 270 pacientes cancerosos).

Para utilizar suficientemente el conjunto de datos (de 431 muestras) y llevar a cabo unos resultados fiables, se utilizara un conjunto de datos en la formación independiente y los grupos de prueba. Por este procedimiento, cada uno de los diez subconjuntos, a su vez, se llevara a cabo a cabo para ser utilizado como un conjunto de pruebas independiente, mientras que los otros nueve subconjuntos se combinaron para formar un conjunto de entrenamiento.
CAPÍTULO V: DISCUSIÓN DE RESULTADOS

5.1. Análisis y discusión de la medición de la variable “Mejora en la identificación de tumores de mama”

Para poder medir la precisión de identificación de tumores de mama, usaremos la fórmula de la efectividad de Makhoul (1999) \cite{15}, que se define:

\[
F = 2 \cdot \frac{\text{Precision} \cdot \text{Exhaustividad}}{\text{Precision} + \text{Exhaustividad}}
\]

Dónde:

\[
\text{Precision} = \frac{|\text{Imágenes Evaluadas} \cap \text{Imágenes Reconocidas}|}{|\text{Imágenes Reconocidas}|}
\]

\[
\text{Exhaustividad} = \frac{|\text{Imágenes Evaluadas} \cap \text{Imágenes Reconocidas}|}{|\text{Imágenes Evaluadas totales}|}
\]

- Para medir necesitamos el factor complejidad, para indicar que nuestro algoritmo es mejor a otros algoritmos propuestos, así demostraremos la eficiencia en el desempeño del algoritmo.
- El factor más importante será el número de aciertos dados por el algoritmo propuesto comparado con resultados de otros algoritmos, con el mismo número de imágenes de entrenamiento.

5.1.1. Datos generales del estudio de precisión de contenidos al recuperar información

**Población:** Estuvo conformada por 500 imágenes de pacientes con carcinoma mamario que utilizaron para realizar estudios de diagnóstico.

**Muestra:** Lo conformaron 431 imágenes de pacientes que se practicaron la técnica del diagnóstico por mamografía digital.

5.1.2. Hipótesis y variable estadísticas

**Hipótesis Nula (H₀):** La efectividad de identificación en el reconocimiento de tumores de mama en mi algoritmo propuesto sin usar el diseño algorítmico propuesto es mayor o igual a la efectividad de reconocimiento en la identificación usando el diseño basado en el algoritmo propuesto (\(F_{atms}\)).

\[
H₀=F_{atm} \geq F_{atms}
\]
Hipótesis Alternativa (H₁): La efectividad en la identificación de reconocimiento de imágenes de tumores de mama sin usar el diseño algorítmico propuesto (F_{atm}) es menor a la efectividad de identificación en el reconocimiento usando el diseño basado en el algoritmo propuesta (F_{atms}).

\[ H₁ = F_{atm} < F_{atms} \]

En la que se debe de probar.

Variables Estadísticas:

- \( F_{atm} \): Efectividad Promedio del Algoritmo sin usar el Diseño Algorítmico propuesto.

- \( F_{atms} \): Efectividad Promedio del Algoritmo usando el Diseño basado en Algorítmico propuesto.

- \( S_{atm} \): Desviación estándar del Algoritmo sin usar Diseño basado en Algoritmo propuesto.

- \( S_{atms} \): Desviación estándar del Algoritmo usando Diseño basado en Algoritmo propuesto.

- \( N_{atm} \): Tamaño de muestra en grupos de 500 imágenes del algoritmo sin usar Diseño basado en Algoritmo propuesto.

- \( N_{atms} \): Tamaño de muestra en grupos de 500 imágenes del Algoritmo del Diseño basado en Algoritmo propuesto.

- \( T₁ \): Valor t “student”.

- \( GI \): Grado de Libertad
5.1.3. Resultados obtenidos del estudio de “Mejora en la identificación de reconocimiento de tumores de mama”

Los resultados obtenidos de los estudios realizados en el Hospital “Almanzor Aguiñaza Asenjo”, de Essalud, de la ciudad de Chiclayo en relación a la identificación del reconocimiento de los tumores de mama, son los siguientes:

Tabla 4: Resultados obtenidos del estudio de efectividad de identificación en reconocimiento de tumores de mama

<table>
<thead>
<tr>
<th>CONSULTAS</th>
<th>Análisis sin Algoritmo Propuesto</th>
<th>Análisis con Algoritmo Propuesto</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Precisión</td>
<td>Exhaustividad</td>
</tr>
<tr>
<td>Consulta 1</td>
<td>0.94</td>
<td>0.9325</td>
</tr>
<tr>
<td>Consulta 2</td>
<td>0.8045</td>
<td>0.807</td>
</tr>
<tr>
<td>Consulta 3</td>
<td>0.794</td>
<td>0.9165</td>
</tr>
<tr>
<td>Consulta 4</td>
<td>0.837</td>
<td>0.9485</td>
</tr>
<tr>
<td>Consulta 5</td>
<td>0.67</td>
<td>0.757</td>
</tr>
<tr>
<td>Consulta 6</td>
<td>0.745</td>
<td>0.9425</td>
</tr>
<tr>
<td>Consulta 7</td>
<td>0.569</td>
<td>0.788</td>
</tr>
<tr>
<td>Consulta 8</td>
<td>0.8325</td>
<td>0.9</td>
</tr>
<tr>
<td>SUMA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PROMEDIO</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fuente: Creación Propia.

Del cuadro anterior, las consultas numeradas para el desarrollo del experimento se encuentran en el Anexo, dentro del listado de Implementación de Funciones.
Figura 12: Histograma de Efectividad con ATM y sin ATM para las consultas realizadas

Fuente: Creación Propia.

Análisis La implementación de todos los métodos a evaluar fue realizada en Matlab, utilizando las librerías disponibles, a continuación se muestra la siguiente tabla.

5.1.4. 

Tabla 5: Resultados obtenidos dentro del análisis Estadístico en la identificación de reconocimiento de tumores de mama.

<table>
<thead>
<tr>
<th>Variable</th>
<th>$S_{atm}$</th>
<th>$S_{atms}$</th>
<th>$F_{atm}$</th>
<th>$F_{atms}$</th>
<th>$N_{atm}$</th>
<th>$N_{atms}$</th>
<th>GI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valor</td>
<td>0.0698</td>
<td>0.0145</td>
<td>0.8763</td>
<td>0.9737</td>
<td>8</td>
<td>8</td>
<td>14</td>
</tr>
</tbody>
</table>

Fuente: Creación Propia.
Estadística de Prueba

\[ t_i = \frac{F_{atm} - F_{atms}}{\sqrt{\frac{S_{atm}^2}{N_{atm}} + \frac{S_{atms}^2}{N_{atms}}}} \]

Fórmula 13: Prueba estadística t de Student, para grupos independientes.


Nivel de Significación

Considerando \( \alpha = 0.05 \) \( \Rightarrow \gamma = 1 - 0.05 = 0.95 \)

Lo que da un nivel de confiabilidad de 95%.

Región de Rechazo

Conforme a la forma de la distribución t,
Se rechazara la hipótesis nula \( (F_{atm} \geq F_{atms}) \), es decir si \( t_i < t_{0.05} \)

Sustituyendo valores en la estadística de prueba, tenemos

\[ t_i = \frac{-0.0974}{0.01841} = -5.29145 \]

Al acudir a la tabla de distribución “t” de Student, se ubica el grado de libertad correspondiente. La tabla contiene los siguientes valores.

<table>
<thead>
<tr>
<th>GI</th>
<th>( t_{0.05} )</th>
<th>( t_{0.01} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>1.761</td>
<td>2.624</td>
</tr>
</tbody>
</table>

Fuente: Creación Propia

Figura 13: Regiones de Rechazo y Aceptación obtenidas de las pruebas estadísticas realizadas

Fuente Creación Propia
5.1.5. Discusión de resultado

Como \( t_i = -5.29145 \) y \( t_i < 1.761 \) por lo tanto se rechaza la hipótesis nula y se acepta la hipótesis alternativa, la cual se pretendía probar.

Como algunos resultados mostrados en el cuadro son menores a los ejecutados por otros algoritmos, entonces se rechaza la hipótesis nula y se acepta la hipótesis alternativa, la cual se pretendía probar.

Considero los resultados experimentales en un escenario en el que realizamos la tarea de clasificación sobre 8 consultas de grupos de bases de datos.
CAPÍTULO VI: CONCLUSIONES Y RECOMENDACIONES

6.1. Conclusiones

- El diseño de un algoritmo de reconocimiento de tumores de mama elaborado, mejoró la identificación delos tumores de mama cancerígenos.
- La investigación bibliográfica realizada nos permitió el diseño del algoritmo.
- Se logró diseñar un algoritmo de reconocimiento de un tumor de mama por medio de una imagen.
- Además de las limitaciones es posible la creación de algoritmos basados en investigación científica en las ciencias de la computación.
- El algoritmo es preciso y efectivo de acuerdo a los métodos de muestreo estadístico

6.2. Recomendaciones

- Para un mejor aprovechamiento de la arquitectura, se recomienda antes de implementar consultar con un experto médico y verificar cuando el entrenamiento este en un porcentaje de error aceptable.
- Las imágenes para su base de datos, debe encontrase en formato JPG.
- Utilizar el algoritmo de reconocimiento de tumores de mama cancerígenos diseñado en el presente trabajo, para servir de apoyo en el reconocimiento de los tumores mama cancerígenos.

6.3. Trabajos futuros

Considerando que los algoritmos computacionales para la identificación de reconocimientos de tumores de mama, entre ellos el algoritmo SVM es un campo muy estudiado y aplicado a otros algoritmos. Entre los trabajos futuros se espera evaluar nuestra propuesta utilizando algoritmos eficientes para la selección de atributos y poder realizar experimentos en modelos basados en lógica difusa.
6.4. Referencias Bibliográficas


419&sa=X&ei=dCq6UeelCdTF4APlroCQBw&ved=0CGMQ6AEwBw#v=onep age&q=algoritmo%20de%20gabor&f=false, obtenido el 19/09/2012.


Anexo 1

IMPLEMENTACIÓN DEL DISEÑO ALGORÍTMICO EN LA IDENTIFICACIÓN DE TUMORES DE MAMA

La implementación del Diseño Algorítmico en la Identificación de Tumores de Mama lo vemos a continuación en el lenguaje de programación MATLAB.

```matlab
function modelo = svm_cancer(data,options)
% entrada de argumentos pasados
%-------------------------------------------
data = c2s(data);
[dim,num_data] = size(data.X);
if nargin< 2, options=[]; else options=c2s(options); end
if ~isfield(options,'ker'), options.ker = 'linear'; end
if ~isfield(options,'arg'), options.arg = 1; end
if ~isfield(options,'C'), options.C = inf; end
if ~isfield(options,'norm'), options.norm = 1; end
if ~isfield(options,'mu'), options.mu = 1e-12; end
if ~isfield(options,'eps'), options.eps = 1e-12; end
%----------------------------
y = data.y(:);
y(find(y==2)) = -1;
% Calculando la matriz de kernel del svm
H = kernel(data.X,data.X,options.ker,options.arg).*(y*y');
% añadir vector en diagonal
H = H + options.mu*eye(size(H));
Aeq = y';
beq = 0;

LB = zeros(num_data,1);% 0 <= Alpha
f = -ones(num_data,1);% Alpha
x0 = zeros(num_data,1);% Punto de partida
if options.norm==1,
%---------------------
if length(options.C) == 1,
UB = options.C*ones(num_data,1);
elseif length(options.C) == 2,
UB=zero(num_data,1);
UB(find(data.y==1))=options.C(1);
UB(find(data.y==2))=options.C(2);
else
UB=vectorC=zeros(num_data,1);
end
vectorC=zero(num_data,1);
else
%---------------------
```

Esta obra ha sido publicada bajo la licencia Creative Commons Reconocimiento-No Comercial-Compartir bajo la misma licencia 2.5 Perú. Para ver una copia de dicha licencia, visite http://creativecommons.org/licences/by-nc-sa/2.5/pe/
UB = ones(num_data,1)*inf;
vectorC = ones(num_data,1);
if length(options.C) == 1,
    vectorC = vectorC*options.C;
elseif length(options.C) == 2,
    inx1 = find(data.y == 1); inx2 = find(data.y == 2);
    vectorC(inx1) = options.C(1);
    vectorC(inx2) = options.C(2);
else
    vectorC = options.C(:);
end
vectorC = 1./(2*vectorC);
H = H + diag(vectorC);
end

% optimizando proceso
%-----------------------
qp_options = optimset('Display','off');
[Alpha,fval,exitflag] = quadprog(H, f, [], [], Aeq, beq, LB, UB, x0, qp_options);

inx_sv = find(Alpha > options.eps);
% calculando
%-----------------------
inx_bound = find(Alpha > options.eps & Alpha < (options.C - options.eps));

if length(inx_bound) ~= 0,
    model.b = sum(y(inx_bound) - H(inx_bound,inx_sv).*...
    Alpha(inx_sv).y(inx_bound))/length(inx_bound);
else
disp('Bias cannot be determined. ');model.b = 0;
end

% calculando margen del svm
%-----------------------
if options.norm == 1
    w2 = Alpha(inx_sv)'*H(inx_sv,inx_sv)*Alpha(inx_sv);
else
    w2 = Alpha(inx_sv)'*(H(inx_sv,inx_sv)-diag(vectorC(inx_sv)))*Alpha(inx_sv);
end
margin = 1/sqrt(w2);
% calculando error de entrenamiento de clasificación
%-------------------------------------------
%-------------------------------------------
%-------------------------------------------
Alpha = Alpha.*y;
model.Alpha = Alpha(inx_sv);
model.sv.X = data.X(:,inx_sv);
model.sv.y = data.y(inx_sv);
model.sv.inx = inx_sv;
model.nsv = length(inx_sv);
model.margin = margin;
model.exitflag = exitflag;
model.options = options;
model.kercnt = num_data*(num_data+1)/2;
model.trnerr = cerror(data.y,svmclass(data.X, model));
model.fun = 'svmclass';
ifstrcmpi(options.ker,'linear')
model.W = model.sv.X*model.Alpha
end
IMPLEMENTACIÓN DE FUNCIONES

La implementación de las funciones se hace a través de consultas, las cuales se listan a continuación.

Consulta 1:
500 imágenes de casos de tumor a evaluar
A continuación muestro la imagen no usada en el entrenamiento, para evaluar el reconocimiento

La precisión es: 0.9885, después de la ejecución de 111 casos para esta consulta.
Consulta 2:
500 imágenes de casos de tumor a evaluar
Las imágenes poseen 16 bits, con una resolución de 42 microns
A continuación muestro la imagen no usada en el entrenamiento, para evaluar el reconocimiento

![Imagen]

La precisión es: 0.7415, después de la ejecución de 117 casos para esta consulta.
Consulta 3:

500 imágenes de casos de tumor a evaluar

Las imágenes poseen 16 bits, con una resolución de 42 microns

A continuación muestro la imagen no usada en el entrenamiento, para evaluar el reconocimiento

La precisión es: 0.9, después de la ejecución de 38 casos para esta consulta.
Consulta 4:
500 imágenes de casos de tumor a evaluar
Las imágenes poseen 16 bits, con una resolución de 42 microns
A continuación muestro la imagen no usada en el entrenamiento, para evaluar el reconocimiento

La precisión es: 0.9485, después de la ejecución de 57 casos para esta consulta.
Consulta 5:
500 imágenes de casos de tumor a evaluar
Las imágenes poseen 16 bits, con una resolución de 42 microns
A continuación muestro la imagen no usada en el entrenamiento, para evaluar el reconocimiento

La precisión es: 0.7205, después de la ejecución de 47 casos para esta consulta.
Consulta 6:
500 imágenes de casos de tumor a evaluar
Las imágenes poseen 16 bits, con una resolución de 42 microns
A continuación muestro la imagen no usada en el entrenamiento, para evaluar el reconocimiento

La precisión es: 0.9115, después de la ejecución de 60 casos para esta consulta.
Consulta 7:
500 imágenes de casos de tumor a evaluar
Las imágenes poseen 16 bits, con una resolución de 42 microns
A continuación muestro la imagen no usada en el entrenamiento, para evaluar el reconocimiento

La precisión es: 0.6065, después de la ejecución de 78 casos para esta consulta.
Consulta 8:
500 imágenes de casos de tumor a evaluar
Las imágenes poseen 16 bits, con una resolución de 42 microns
A continuación muestro la imagen no usada en el entrenamiento, para evaluar el reconocimiento

![Imagen]

La precisión es: 0.905, después de la ejecución de 59 casos para esta consulta.