



**INSTITUTO POLITÉCNICO NACIONAL**

---

**ESCUELA SUPERIOR DE INGENIERÍA  
MECÁNICA Y ELÉCTRICA**

**CLASIFICADOR AUTOMÁTICO DE RECURRENCIA DE  
CÁNCER**

# Tesis

**QUE PARA OBTENER EL TÍTULO DE:  
INGENIERO EN COMUNICACIONES Y ELECTRÓNICA**

**P R E S E N T A N**

**ABREU SANCHEZ JORGE ANTONIO  
FAJARDO OLIVARES SALVADOR**

**A S E S O R E S D E T E S I S**

**DRA. ACEVEDO MOSQUEDA MARIA ELENA  
DR. YAÑEZ MARQUEZ CORNELIO**



México D.F.

Noviembre 2009

**INSTITUTO POLITÉCNICO NACIONAL**  
**ESCUELA SUPERIOR DE INGENIERÍA MECÁNICA Y ELECTRICA**  
**UNIDAD PROFESIONAL “ADOLFO LÓPEZ MATEOS”**

**TEMA DE TESIS**

QUE PARA OBTENER EL TÍTULO DE INGENIERO EN COMUNICACIONES Y ELECTRÓNICA  
POR LA OPCIÓN DE TITULACIÓN TESIS COLECTIVA Y EXAMEN ORAL INDIVIDUAL  
DEBERA(N) DESARROLLAR  
C. JORGE ANTONIO ABREU SÁNCHEZ  
C. SALVADOR FAJARDO OLIVARES

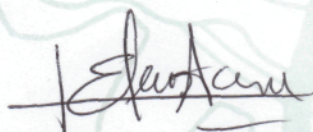
**“CLASIFICACIÓN AUTOMÁTICO DE RECURRENCIA DE CÁNCER”**

IMPLEMENTAR EL MODELO ASOCIATIVO GAMMA GENERALIZADO PARA CLASIFICAR LA  
RECURRENCIA DE CÁNCER

- ❖ ANTECEDENTES GENERALES DEL CÁNCER.
- ❖ MEMORIAS ASOCIATIVAS.
- ❖ MÉTODOS.
- ❖ PRUEBAS.
- ❖ CONCLUSIONES Y TRABAJO FUTURO.

MÉXICO D. F., A 20 DE OCTUBRE DE 2010.

**ASESORES**



**DRA. MARIA ELENA ACEVEDO MOSQUEDA**



**DR. CORNELIO YÁÑEZ MÁRQUEZ**



**ING. FEDERICO FELIPE DURÁN**



**M. EN C. SALVADOR RICARDO MENESES GONZÁLEZ**  
**JEFE DEL DEPARTAMENTO ACADÉMICO DE**  
**INGENIERÍA EN COMUNICACIONES Y ELECTRÓNICA**



## Resumen

En la presente tesis se propone una nueva aplicación para el Clasificador Gamma Generalizado ( $\gamma_g$ ), el cual, se vale del uso de memorias asociativas. Con dicho algoritmo se desarrolló un Clasificador Automático de Recurrencia de Cáncer.

El Clasificador utiliza la información de la Base de Datos “*Wisconsin Prognostic Breast Cancer*” la cual se encuentra disponible en la página web de la Universidad de Irvine California.

La Base de Datos se constituye por 194 registros: 148 No recurrentes y 46 Recurrentes, cada registro de la Base representa un caso de cáncer de mama. Son pacientes consecutivos revisados por el Dr. Wolberg desde 1984.

Se presentan los resultados obtenidos al someter al clasificador a tres pruebas distintas:

- *Hold Out*
- *Take One Out*
- *K-Fold Cross Validation*

En los resultados de las anteriores pruebas se demuestra que el Algoritmo Gamma Generalizado actúa eficientemente como Clasificador, y, en especial como Clasificador de Recurrencia de Cáncer llegando en algunos casos a superar los resultados obtenidos con otros tipos de Clasificadores.

Los resultados de las pruebas fueron los siguientes:

- *Hold Out* 90.2%
- *Take One Out* 78%
- *K-Fold Cross Validation* 91.2%

De los cuales el más relevante es el resultado de la prueba de *Hold Out*, 90.2% ya que es la prueba más difundida.

Con lo anterior se demuestra el buen funcionamiento del enfoque asociativo en el área de los clasificadores.



## Planteamiento del Problema

El cáncer de mama consiste en un crecimiento anormal y desordenado de las células de este tejido. Este tipo de cáncer es el más frecuente en la mujer, le siguen el de colon, pulmón, útero y ovarios. Es el cáncer más temido debido a su elevada frecuencia, la incidencia está creciendo ya que 1 de cada 9 mujeres desarrolla un cáncer de mama a lo largo de su vida. Es raro en hombres (aunque posible) y en mujeres por debajo de los 30 años, con mayor incidencia entre los 45 y los 65, aumenta después de la menopausia.

Es muy frecuente que aún después de una cirugía para remover el cáncer éste vuelva a aparecer, a esto se le llama recurrencia y, en algunos casos resulta aun más dañina que el cáncer original.

Es por todo esto que se debe contar de una herramienta capaz de realizar una predicción de recurrencia de cáncer de mama de forma eficaz y confiable.



## Justificación

El cáncer es en la actualidad una de las principales causas de muerte en el mundo. Se presenta actualmente como la segunda causa de muerte en nuestro país y aproximadamente el 6% de las defunciones por cáncer corresponden al mamario, ocupando el sexto lugar dentro de las causas de muerte por cáncer, pero, el segundo lugar de muerte por cáncer en mujeres.

Se considera hoy que al año 2010 habrá una tasa de mortalidad de 13 por 100,000 mujeres adultas y cerca de 4500 defunciones por año por esta causa en México. [1]

En el mundo se presentan alrededor de 570,000 casos nuevos por año y se estiman 1,000,000 de casos nuevos para el año 2,010, sin embargo, no es México un país con la mayor incidencia de cáncer mamario. Solamente en los Estados Unidos se calculan 200,000 casos nuevos por año y 40,000 defunciones anuales por esta causa. Actualmente la AMERICAN CANCERSOCIETY predice que una de cada 8 mujeres padecerá cáncer mamario durante su vida.

El pasado 26 de febrero de 2009 representantes de la Fundación Cima Organización no Gubernamental de México y Estados Unidos que luchan contra el cáncer de mama advirtieron que los últimos 10 años la incidencia de cáncer de mama ha crecido en 50 por ciento en México y cobra la vida de 12 mujeres al día [2].

La detección temprana de la recurrencia del cáncer brinda a los pacientes la posibilidad de recibir un tratamiento adecuado a tiempo y con ello salvar la vida, es por esto que se debe de contar con medios eficaces para su pronta detección.

De acuerdo con la Secretaría de Salud 46% de las mujeres que sufren esta enfermedad la padecerán antes de cumplir 50 años y una de cada ocho mexicanas corre riesgo de desarrollarla [3].

En la actualidad se tiene conocimiento de diversos enfoques de reconocimiento de patrones como lo son los basados en métricas, el enfoque neuronal, el enfoque estadístico –probabilístico y el enfoque asociativo.

La mayoría de los trabajos investigados están basados en redes neuronales o en un modelo matemático como el de la universidad de Craiova, el cual se basa en el teorema de Bayes y tiene una eficiencia del 74.24%.

Para el caso de este trabajo, el enfoque que más nos interesa es el enfoque asociativo. Este enfoque ha tenido un gran auge en el Centro de Investigaciones en Computación en el IPN, el enfoque asociativo en base al cual se han realizado otros trabajos de tesis, es el punto de partida para el presente trabajo.



En el trabajo se pretende aplicar el Clasificador automático de alto rendimiento basado en memorias asociativas para obtener un Clasificador de recurrencia de cáncer.

La razón de elegir el enfoque asociativo es que puede realizar clasificaciones con alto grado de exactitud y que puede ser implementado en algoritmos relativamente rápidos.



## Objetivos

### Objetivo General:

Implementar el modelo asociativo Gamma Generalizado para clasificar la recurrencia de Cáncer.

### Objetivos Particulares:

- Implementar el Algoritmo  $\gamma_g$ . (Gama Generalizado) para clasificar la recurrencia de cáncer.
- Encontrar el correcto escalamiento para cambiar los rasgos categóricos (características que tienen los elementos de la base de datos en forma de texto) en la base por interpretaciones numéricas.
- Codificar la base de datos mediante el código Johnson-Möbius Modificado.
- Utilizar el Método  $\gamma_g$  y la base de datos codificada para probar el Clasificador y, en caso de ser necesario, hacer las correcciones pertinentes.

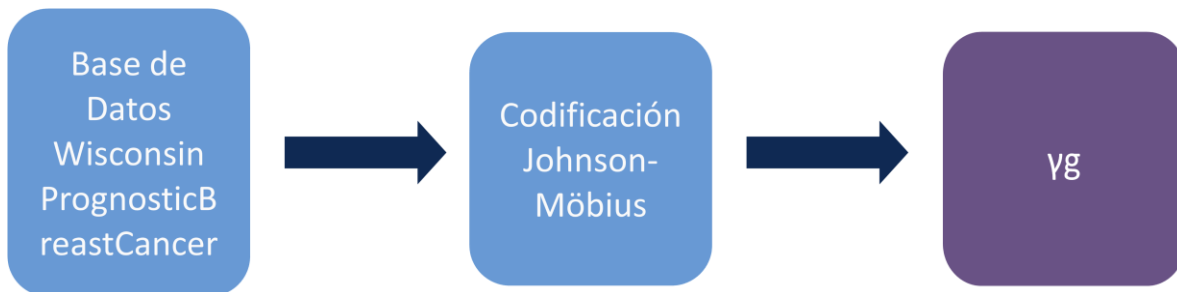


## Modelo propuesto

A continuación se presenta el modelo propuesto para la clasificación de Recurrencia de cáncer.

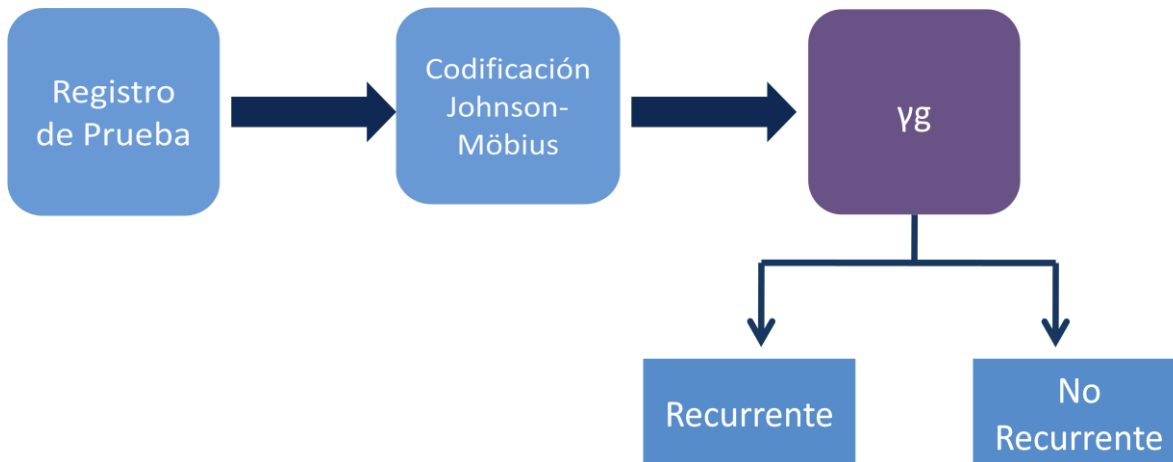
- *Fase de Aprendizaje*

En esta fase se toman los registros de la Base de Datos Wisconsin *Breast Cancer* que servirán como conjunto de aprendizaje, se codifican mediante el código Johnson-Möbius modificado y se almacenan según el algoritmo Gamma Generalizado ( $\gamma g$ ).



- *Fase de Recuperación*

En esta fase se toman los registros de la Base de Datos Wisconsin *Breast Cancer* que servirán como conjunto de prueba, se codifican mediante el código Johnson-Möbius modificado y se clasifican con el algoritmo Gamma Generalizado ( $\gamma g$ ), obteniendo como resultado “Recurrente” ó “No Recurrente” según sea el caso.







## INDICE GENERAL

1.1	Antecedentes .....	1
1.2	Cáncer .....	2
1.2.1	<i>Cáncer de mama</i> .....	2
1.2.2	<i>Estructura del órgano</i> .....	3
1.2.3	<i>Tipos de Cáncer de mama</i> .....	4
1.2.4	<i>Factores de Riesgo</i> .....	5
1.2.5	<i>Síntomas</i> .....	6
1.2.6	<i>Detección</i> .....	7
1.2.7	<i>Diagnóstico</i> .....	7
1.2.8	<i>Tratamientos</i> .....	9
1.2.8.1	Tratamiento de Cirugía.....	9
1.2.8.2	Tratamiento de Terapia Hormonal.....	9
1.2.8.3	Tratamiento de Quimioterapia .....	9
1.2.8.4	Tratamiento de Radioterapia.....	9
1.2.9	<i>Recurrencia</i> .....	10
1.3	Estado del arte .....	10
CAPITULO 2	Memorias Asociativas.....	13
2.1	Introducción .....	13



2.2 Memoria asociativa Lernmatrix, Correlograph, Linear Associator y Hopfield. ....	13
2.3 Conceptos básicos.....	15
2.4 Memorias Asociativas Alfa-Beta .....	17
CAPÍTULO 3 Métodos.....	21
3.1 Algoritmo.....	21
3.2. Operador $u_{\beta}$ .....	22
3.3 Código Binario Johnson-Möbius Modificado .....	22
3.4 Algoritmo del Código Johnson-Möbius Modificado: .....	23
3.5. Operadores de Cadena Binaria Mínima y k-Binario de Expansión .....	25
3.6. Operador Gama de Similitud.....	26
3.7 Algoritmo Clasificador de Patrones.....	28
CAPITULO 4 Pruebas .....	33
4.1 Pruebas de Eficiencia para clasificadores.....	33
4.2 Descripción de la Base de Datos .....	33
4.3 Programación del Algoritmo.....	35
4.4 Prueba Hold-Out .....	35
4.5 Prueba K-Fold Cross Validation.....	38
4.6 Take One Out .....	39
4.7 Resultados Generales.....	40
CAPITULO 5 Conclusiones y Trabajo Futuro .....	41



---

5.1 Conclusiones .....	41
5.2 Dificultades Durante las Pruebas.....	42
5.3 Trabajo Futuro .....	42
Anexo 1: Código Johnson-Möbius Modificado.....	43
Anexo 2: Código Gamma Generalizado .....	44
GLOSARIO.....	47
Referencias .....	49



## INDICE DE FIGURAS

Figura 1.1 Estructura del órgano	3
Figura 2.1 Representación de una Memoria Asociativa	15
Gráfica 4.1 Prueba Hold Out	36
Gráfica 4.2 Prueba Hold Out (2)	37
Gráfica 4.3 Resultados Prueba K-Fold	39



## INDICE DE TABLAS

Tabla 2.1 $A \times A \rightarrow B$	18
Tabla 2.2 $B \times A \rightarrow A$	18
Tabla 2.3 Definición del Operador Alfa	19
Tabla 2.4 Definición del Operador Beta	19
Tabla 2.5 Propiedades de la Operación Binaria $\alpha$	20
Tabla 3.1 Código Johnson-Möbius Modificado para 3, 7, 11, 17 y 19	24
Tabla 3.2 Ejemplo Código Johnson-Möbius Modificado	24
Tabla 4.1 Prueba Hold Out	36
Tabla 4.2 Prueba Hold Out (2)	37
Tabla 4.3 Resultados Prueba K-Fold	38
Tabla 4.4 Take One Out	39
Tabla 4.5 Resultados Generales	40



## CAPÍTULO 1 Introducción

### 1.1 Antecedentes

El Cáncer es un agente causante de muertes a nivel mundial, en su mayoría son tumores malignos que se caracterizan por una multiplicación anormal y desordenada de células, las cuales tienen la característica de invadir los tejidos adyacentes (metástasis).

De entre los tipos de cáncer, existe uno que es muy peligroso y ataca principalmente a las mujeres: el Cáncer de Mama.

El cáncer de mama también puede presentarse en los hombres. Sin embargo, por ser una enfermedad que afecta a miles de mujeres alrededor del mundo, se le ha asociado, por error, exclusivamente al sexo femenino.

El cáncer de mama es la segunda causa de muerte entre las mujeres en el mundo [4]. Diversos estudios demuestran que una de cada diez desarrollará este mal en su vida, pero que más del 90% puede curarse si la enfermedad se detecta en su fase inicial.

Una de las formas más efectivas de acabar con el Cáncer de mama es mediante cirugía. En la cirugía se remueve lo máximo posible del cáncer de mama. Sin embargo, después de la operación, es posible que algunas células cancerosas permanezcan en el cuerpo y se multipliquen. Esto no les ocurre a todas las mujeres, pero cuando sucede se le llama recurrencia. No existe un modo definitivo de identificar estas células porque no siempre pueden verse en pruebas médicas.

El desarrollo de un método efectivo para determinar la recurrencia de cáncer ayudará a salvar muchas vidas.

Una de las principales tendencias es desarrollar métodos de diagnóstico en base a los datos del paciente como: tamaño del tumor, edad, localización del tumor entre otros.

Ya que el diagnóstico se basa en una serie de datos almacenados se tiene la posibilidad de desarrollar un programa para computadora capaz de analizar los datos, comparar y determinar una posible recurrencia de forma automática. [2]



## 1.2 Cáncer

El cáncer es un trastorno en los mecanismos que controlan la proliferación y la diferenciación en las células de los organismos pluricelulares superiores. Un rasgo fundamental de todos los seres vivos es la existencia de un tiempo de vida finito de cada organismo, ésta propiedad también se extiende a las células de cada individuo, cuyos proceso de crecimiento y división están sutilísimamente controlados. La excepción a ésta regla la constituyen las células cancerosas, que han surgido como variantes y que han perdido control.

La proliferación y la diferenciación son dos procesos interdependientes, con características del crecimiento y desarrollo de seres pluricelulares. Estos procesos son cada vez más finamente controlados y regulados conforme se asciende en la cadena evolutiva.

En un organismo adulto debe existir una perfecta coordinación entre los procesos de proliferación y diferenciación. Este es un rasgo esencial del desarrollo cabal y la renovación normal de los tejidos. Es precisamente ésta coordinación la que se pierde en un tejido cuando en éste surge un cáncer.

En un tumor se producen más células nuevas que las que mueren al final del proceso de diferenciación.

Las células cancerosas tienen una vida más larga, es decir, tienen crecimiento indefinido, además de que su proliferación es descontrolada y la diferenciación es errónea.

De tal manera, podemos concluir que el cáncer se caracteriza por una proliferación celular descontrolada unida a una pérdida de rasgos de diferenciación. La coordinación precisa entre estos dos procesos es vital para el desarrollo normal y la renovación de los tejidos. [3]

### 1.2.1 Cáncer de mama

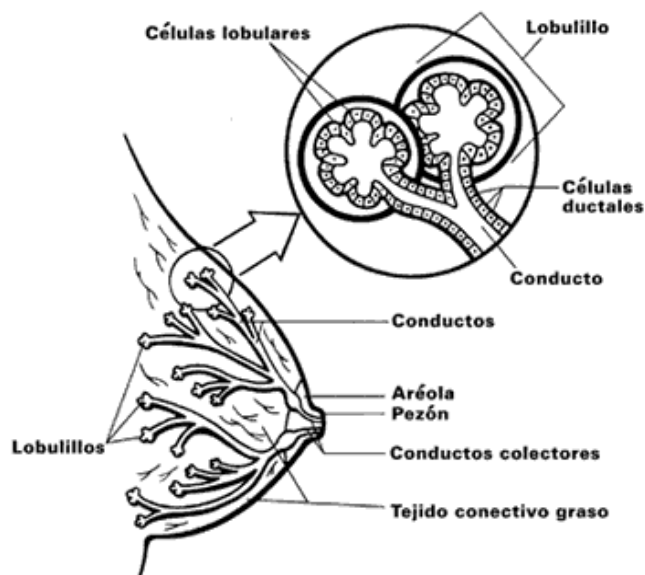
El cáncer de mama, junto con el cervicouterino, son las más comunes entre las mujeres mexicanas. Este se origina en el tejido mamario y al igual que otros cánceres, puede invadir y crecer hacia los tejidos que rodean el seno.

El cáncer de mama puede llamarse también cáncer de seno y es el crecimiento desordenado de células en las glándulas mamarias que generalmente se detecta como un bulto en el seno. [5]

### 1.2.2 Estructura del órgano

La apariencia exterior de la mama es variada, pero todas están conformadas por un pezón, que es el lugar donde se encuentran las aberturas del conducto de leche. Internamente, se constituye de tejido que va desde la clavícula hasta las costillas inferiores.

La mama se apoya en los músculos del pecho. En su interior se encuentran los lobulillos que producen la leche y conductos que la transportan hasta el pezón. Cada pezón tiene entre seis y nueve aberturas de conducto. Estas aberturas dirigen a un sistema de conductos que va desde el pezón hasta la parte de atrás de la mama. Dichos conductos se propagan en una serie de ramificaciones que terminan en lóbulos y estos a su vez en lobulillos más pequeños. Estos son lugares comunes para el desarrollo del cáncer de mama.



**Figura 1.1:** Estructura del órgano

Dividiendo el seno en cuatro partes, observamos que la mayor cantidad de conductos está localizada en el cuadrante superior externo de la mama. (Próximo a la axila)

Es en esta localización que tiene lugar la mayoría de los tumores malignos de mama. (Aproximadamente 50%). [6]





### 1.2.3 Tipos de Cáncer de mama

El cáncer de mama puede clasificarse de acuerdo al sitio en el que se originó el crecimiento irregular de células, ya sea en los conductos o en los lobulillos. Además se considera de acuerdo al grado de invasión y a la apariencia de las células vistas bajo un microscopio.

- *Carcinoma In situ*. El carcinoma es también conocido como “tumor”, *In situ* significa “en su sitio”, esto quiere decir que el cáncer está ubicado en los conductos de los lobulillos y no se ha extendido al tejido adiposo que se encuentra cerca del seno ni a otros órganos del cuerpo.
- *Carcinoma lobular*. También se conoce como neoplasia (tumor) lobular. Se origina en los llamados lobulillos pero no atraviesa las paredes de éstos. Usualmente no se convierte en cáncer invasivo sin embargo, hay casos en los que se puede llegar hacer invasivo.
- *Carcinoma ductal*. Es el tipo más común de cáncer no invasivo de seno. En este caso las células cancerosas que se encuentran dentro de los conductos no se propagan a través de las paredes hacia el tejido adiposo del seno. Este tipo de cáncer es tratado con cirugía o radiación, métodos que generalmente son favorables a la cura del padecimiento, pero si no se tratan a tiempo pueden convertirse en invasivos.
- *Carcinoma ductal infiltrante (o invasivo)*. Este tipo de cáncer tiene su origen en las glándulas que producen la leche. En este proceso el cáncer puede extenderse hacia los canales linfáticos o los vasos sanguíneos del seno y distribuirse a otras partes del cuerpo. Este tipo de tumor es el más común en el cáncer de mama.
- *Carcinoma lobular infiltrante (o invasivo)*. El origen de este tipo de cáncer se encuentra también en las glándulas productoras de leche y puede extenderse a otras partes del cuerpo. Se considera que entre 10 y 15 por ciento de los cánceres invasivos pertenecen a esta clasificación.
- *Carcinoma medular*. En este tipo de carcinoma las células cancerosas se encuentran agrupadas y en los bordes del tumor, hay células del sistema inmunitario que sirven para atacar y destruir las células anormales, así como a otros agentes extraños como bacterias o virus. El carcinoma medular es responsable de aproximadamente 5 por ciento de todos los casos de cáncer de seno.
- *Carcinoma coloide*. Pertenece al tipo de cáncer ductal invasivo y el término médico que se le da es carcinoma mucinoso ya que está formado por células que producen mucosidad. El pronóstico para este tipo de cáncer es favorable, ya que tiene menores probabilidades de propagación que el cáncer ductal invasivo o el lobular invasivo.



- *Carcinoma tubular.* Es un tipo especial de carcinoma ductal infiltrante. Las mujeres que son diagnosticadas con este tipo de cáncer, tienen menos probabilidades de que se propague fuera del seno en comparación con el cáncer ductal invasivo o el lobular invasivo. Este carcinoma es responsable de 2% de todos los casos de cáncer de mama.
- *Cáncer inflamatorio de mama.* Este tipo de cáncer no es común. La piel afectada del seno se siente caliente, está enrojecida y tiene la apariencia de una cáscara de naranja. Ahora se sabe que estos cambios no se deben a una simple inflamación, sino a las células cancerosas que bloquean los vasos linfáticos de la piel.

El cáncer inflamatorio tiene mayores probabilidades de propagación y un pronóstico menos alentador que otros tipos de cáncer [7].

### 1.2.4 Factores de Riesgo

A ciencia cierta, aún no se sabe la causa que lo origina, sin embargo, hay algunos factores que debe considerar y que sirven para prevenir esta enfermedad.

- *Sexo:* El cáncer de seno ataca principalmente a mujeres y solo en un 1% a hombres.
- *Edad:* El riesgo de padecer cáncer de seno aumenta con la edad. Cerca de 18% de los casos de cáncer de mama se diagnostican en las mujeres de 40 a 49 años de edad, mientras que alrededor de 77% de las mujeres con este tipo de cáncer son mayores de 50 años en el momento del diagnóstico. Aunque se han presentado casos en mujeres entre 20 y 30 años.
- *Los factores genéticos:* Los que se heredan, se presentan cuando hay cambios en los genes del cáncer (BRCA o BRCA2). Éstos son dos genes, que hasta el momento los especialistas han identificado, tienen una relación muy estrecha con el padecimiento.
- *Antecedentes familiares de cáncer de mama:* El riesgo de cáncer de seno es más alto entre las mujeres cuyos familiares consanguíneos cercanos tienen esta enfermedad, por lo que el riesgo podría aumentar sí:
  - 1.- Se tienen dos o más familiares con cáncer de seno o de ovario.
  - 2.- Algún familiar padeció cáncer antes de los 50 años de edad, ya sea en la rama materna o paterna. El riesgo es mayor si la madre o hermana tiene un historial de cáncer de seno.



3.- Se tiene uno o más familiares con uno o dos cánceres (seno y ovario, o dos tipos diferentes de cáncer de seno).

- *Antecedentes personales de cáncer de seno:* el riesgo de que aparezca un nuevo cáncer en otra parte del seno o en el otro seno aumenta de tres a cuatro veces en una mujer que ya tiene cáncer en un seno.
- *Períodos menstruales:* El riesgo aumenta si los periodos menstruales comenzaron a una edad temprana (antes de los 12 años) o la menopausia se presentó tardíamente (después de los 50 años).
- *Mujeres que no han tenido hijos:* El riesgo es ligeramente mayor en mujeres que no han tenido hijos o aquellas que tuvieron su primer hijo después de los 30.
- *Uso de anticonceptivos orales:* Las mujeres que usan anticonceptivos orales tienen un riesgo ligeramente mayor de padecer cáncer de mama que las que no.
- *Alcohol:* La ingestión de bebidas alcohólicas está asociada a un ligero aumento en la posibilidad de padecer esta enfermedad.
- *Obesidad y dietas con alto contenido de grasas:* La obesidad (exceso de peso) aumenta el riesgo, especialmente en las mujeres que han pasado por la menopausia. El exceso de tejido adiposo aumenta los niveles de estrógeno e incrementa la probabilidad de padecer cáncer de mama. [8]

### 1.2.5 Síntomas

Durante las primeras fases de la enfermedad se presenta una nudosidad o endurecimiento en una zona del pecho acompañada de una sensación de tensión o pesadez.

La mayoría de las mujeres no presentan dolor, o algún otro síntoma de alteración del estado de salud y se encuentra bien en el momento del diagnóstico. Raramente, hay retracción y salida del líquido del pezón, pero es importante considerar este factor.

En fases más avanzadas de la enfermedad, los síntomas son muy variados y dependen del tamaño y extensión del tumor, el cual en esta etapa será claramente palpable en la zona afectada e incluso los ganglios de las axilas pueden estar aumentando de tamaño.

Las señales de alerta son:

- Una masa, bolita dura o engrosamiento en cualquier parte del seno o en el área de la axila.
- Cambio en el tamaño o la forma del seno, hoyuelos o arrugas en la piel del seno.



- Hinchazón, enrojecimiento o calentamiento del seno que no desaparece
- Dolor en una parte del seno que no cambia con su ciclo menstrual.
- Retracción en el pezón o cualquier otra parte del seno.
- Secreción del pezón que comienza repentinamente y aparece solo en uno de los senos.
- Picazón, llaga o área escamosa en uno de los pezones.[9]

### 1.2.6 Detección

Sociedades médicas nacionales e internacionales recomiendan que las mujeres comiencen a examinar sus senos a partir de los 20 años de edad. Se ha demostrado que una de cada diez mujeres podría presentar cáncer de mama en algún momento de su vida, siendo más frecuente en edades comprendidas entre 40 y 65 años 95% de los casos de cáncer de mama pueden ser curados, siempre y cuando la enfermedad sea detectada en estadios tempranos.

La mamografía es el único método por excelencia para la detección del cáncer de mama, siempre y cuando un médico después de explorarte lo sugiera [10].

### 1.2.7 Diagnóstico

Una vez que se descubre una masa o abultamiento en el seno, es importante asegurarse que no sea maligno o canceroso, ya que ocho de cada diez tumores no lo son. Para esto, el médico iniciará con la elaboración de una historia clínica y con una serie de preguntas que aportarán información necesaria para determinar cada caso.

El médico reconoce una masa o abultamiento por su tamaño, textura y la forma como se mueve dentro del seno, generalmente las masas benignas se sienten diferentes a las cancerosas. En ocasiones una masa anormal es sólo un quiste que es una especie de saco lleno de líquido.

En estos casos, el médico puede deshacerlos por medio de un procedimiento sencillo y nada doloroso que consiste en introducir una aguja muy fina para sacar el líquido y examinarlo en el microscopio. Aunque pocas veces los quistes contienen células cancerosas, los resultados de éste análisis determinarán si es necesario pasar a una siguiente etapa que consiste en que la paciente se tome una mamografía, una ecografía o una biopsia.

#### ➤ *Mamografía*

La mamografía es una radiografía del seno y se realiza con una máquina de rayos X especialmente diseñada para esta zona del cuerpo. Es necesario saber que el estudio consiste en aplanar el seno, lo cual puede causar incomodidad, pero sólo dura unos



minutos. La mamografía puede detectar muchos cánceres antes de que sea posible sentirlos o palparlos, por eso es recomendable la práctica de dicha prueba por lo menos una vez al año, sobre todo en mujeres de 40 años en adelante o en aquellas mujeres que tienen antecedente familiares de cáncer.

Aunque la mamografía es muy precisa, no es perfecta puesto que en ocasiones no detectan el cáncer.

### ➤ *Resonancia magnética*

Es el siguiente paso después de la mamografía y la autoexploración constante, el médico podrá indicar la realización de un estudio de resonancia magnética denominado ecografía. Para realizar esta prueba primero se aplica una capa delgada de gel sobre los senos, luego el médico dirige una especie de dispositivo que se pasa alrededor de los senos. Este dispositivo emite ondas de sonido que no se pueden oír ni sentir pero producen diferentes clases de ecos cuando chocan con distintos tejidos en el cuerpo. Posteriormente una computadora convierte los ecos en imágenes que se visualizan en una pantalla de video.

Este examen puede distinguir entre diferentes tipos de abultamientos que pueden ser quistes llenos de líquido y masa sólida. El médico usa la ecografía para determinar el tamaño, la forma, la textura y la densidad de una masa en el seno.

### ➤ *Biopsia*

Este es otro procedimiento para verificar si la masa que ha aparecido en el seno es cancerosa. Consiste en tomar una muestra de las células o tejido del seno para examinarla bajo el microscopio en busca de signos de cáncer.

### ➤ *Prueba HER2*

HER2, por sus siglas en inglés (***Human Epidermal Growth Factor Receptor 2*** o Receptor Humano para el Factor de Crecimiento Epidérmico tipo 2) es una proteína que produce un gen que al acumularse en grandes cantidades tiene el potencial de producir cáncer.

Se realiza una prueba en un laboratorio de patología donde utilizando una muestra de tejido de cáncer de mama y a través de diversos médicos, se busca identificar el grado de presencia de la proteína HER2.

La prueba HER2 puede realizarse a pacientes que recientemente hayan sido diagnosticadas con cáncer de mama invasor o hayan tenido o actualmente tengan cáncer de mama y quieran saber si existe un tratamiento alternativo a la quimioterapia o a la terapia hormonal.



## 1.2.8 Tratamientos

### 1.2.8.1 Tratamiento de Cirugía

Es la primera opción de tratamiento para muchos casos de cáncer. Si está localizado y restringido a una zona, es recomendable la cirugía para extirparlo junto con el tejido circundante que pudiera contener células cancerosas.

El éxito de la cirugía depende de si el tumor se ha extendido o no a otras áreas. Este método sigue ofreciendo las mejores posibilidades de cura para muchos tipos de cáncer. [11]

### 1.2.8.2 Tratamiento de Terapia Hormonal

Involucra el uso de medicamentos o la extracción quirúrgica de los ovarios para prevenir que las hormonas estimulen el crecimiento de las células cancerosas que pudieran permanecer en el cuerpo después de la cirugía.

Consiste en consumir medicamentos orales que combaten el crecimiento o la extensión de algunos tipos de cáncer, bloqueando la acción del estrógeno, la hormona femenina. [12]

### 1.2.8.3 Tratamiento de Quimioterapia

Se utiliza para tratar las células cancerosas que se han propagado o se ha presentado el fenómeno de metástasis, es decir, a otras partes del cuerpo. Dependiendo del tipo y de su etapa de desarrollo, este método se puede utilizar para detener la enfermedad, evitar que se extienda, hacer más lento su crecimiento, eliminar las células cancerosas que se hayan extendido a otras partes del cuerpo o aliviar los síntomas ocasionados por el padecimiento.

Los efectos secundarios pueden ser cansancio, cambios en la piel en el área tratada y cierta pérdida del apetito. Otros por lo general se relacionan con el tratamiento en áreas específicas, tales como pérdida del cabello. Por fortuna, muchos de los efectos secundarios desaparecen con el tiempo. [13]

### 1.2.8.4 Tratamiento de Radioterapia

Se utiliza para tratar el cáncer localizado y mediante este método se destruyen o se dañan las células cancerosas para que no puedan reproducirse.



Esta se administra con rayos externos de alta energía o mediante implantes internos. La radioterapia externa no es dolorosa y es muy similar a la toma de una radiografía con rayos X. El paciente no necesita ser internado y el tratamiento se administra unos 5 días a la semana durante 5 u 8 semanas, lo cual depende del tamaño, ubicación y el tipo de cáncer.

### 1.2.9 Recurrencia

La recurrencia del cáncer es cuando éste vuelve a aparecer luego de haber terminado el tratamiento inicial. Existen 2 tipos de recurrencia:

- Recurrencia Local: Es cuando vuelve a aparecer en el sitio de la operación, y puede ser detectada en la mamografía. Es por ello la importancia del monitoreo constante. El tratamiento es el mismo que al inicio.
- Recurrencia Distal: Es la metástasis o la dispersión del cáncer a otras partes del cuerpo, como pulmones, hueso, cerebro o hígado. Ésta se detecta cuando la paciente presenta síntomas como disminución de peso, cansancio, dolor de huesos. Para su tratamiento se deben incluir quimio y radioterapia, que ayudarán a mejorar la calidad de vida.

Para reducir la posibilidad de recurrencia en algunas pacientes, éstas son tratadas con terapia hormonal o quimioterapia. El problema son los efectos secundarios tanto a corto como largo plazo de estos tratamientos, por lo que se debe evaluar con el médico y la paciente el costo-beneficio.

Es muy importante un monitoreo constante con el médico para evitar la recurrencia o bien, tratarlas a tiempo. [14]

### 1.3 Estado del arte

Los valores económicos y sociales para el diagnóstico del cáncer de mama son muy altos. Como resultado, muchos investigadores del área de Inteligencia Artificial (IA) han sido atraídos hacia este problema y han utilizado varias técnicas de IA para la detección y predicción del cáncer de mama.

Muchos de los métodos de IA han mostrado mejores resultados que los métodos ordinarios de laboratorio de los cuales se encontraron siete tipos de software semejantes al de nosotros en cuestión a la detección de recurrencia de cáncer de mama. Los cuales utilizan diferentes técnicas de clasificación las cuales no utilizan



memorias bidireccionales así como las Alfa-Beta. A continuación se explican los trabajos sobre el tema:

1. **2007 Chih-Lin Chia, W. Nick Street b, William H. Wolberg**, Universidad de Iowa Estados Unidos “ *Application of Artificial Neural Network-Based Survival Analysis on Two Breast Cancer Datasets*”. Este trabajo aplica las redes neuronales artificiales (RNA) para el problema de análisis de supervivencia, debido a que las Redes Neuronales Artificiales fácilmente pueden considerar la interacción de variables y crear un modelo de predicción no lineal, que ofrece una predicción más flexible que los métodos tradicionales. Este trabajo compara los resultados de Redes Neuronales Artificiales en dos diferentes conjuntos de datos de cáncer de mama, ya que ambos utilizan características nucleares morfo métricas. Los resultados muestran que la Red Neuronal Artificial puede predecir con éxito la probabilidad de recurrencia. Los resultados obtenidos arrojan una efectividad de: Clase No-Recurrente: 63.49%, Clase Recurrente: 47.92%.[15]
2. **2006 I. Maglogiannis, K. Karpouzis y M. Bramer**, New York Estados Unidos “ *Automated Diagnosis of Breast Cancer Based on the WDBC Patient Data*”. Este trabajo se basa en el uso de del algoritmo Support Vector Machine el cual es un nuevo algoritmo para la clasificación y regresión que permite la expansión de la información proporcionada por un conjunto de datos de entrenamiento por medio de la combinación lineal de un subconjunto de datos dentro del conjunto de entrenamiento (vectores de apoyo). Estos vectores localizan un hiper plano que separa los datos de entrada con un buen grado de generalización. La efectividad del clasificador está alrededor del 90%, fue probado en una Computadora personal con Procesador Pentium a 2.6 GHz con 512 MB en RAM, el tiempo de ejecución varía entre 70 y 74 segundos.[16]
3. **2003 Hng, L. Miyasaka, K. Inoue, H. Tsukamoto, M.** Universidad de Fuikui Japón “ *Automatic generation of fuzzy classification systems using hyper-cone membership functions*”. En este trabajo, se proponen los métodos de generación automática de reglas de clasificación difusas con los algoritmos genéticos (GA) para obtener sistemas compactos difusos. Se propone un enfoque de la función de miembros de Híper- cono para la construcción de normas para la parte de antecedentes. Entonces, este método se utiliza para determinar la ubicación y la forma de la función de miembros de Híper- cono en la parte antecedente, la clase de salida y el número de entradas necesarias de cada regla de los Algoritmos Genéticos. Además, mediante la regla del método en los procesos de los Algoritmos Genéticos, se obtienen los sistemas compactos de clasificación difusa. Este método se aplicó a los conjuntos de datos de WINE y Wisconsin Pronóstico Cáncer de Mama (WPBC) los conjuntos de datos. El conjunto WINE de datos constará de 13 entradas y





tres salidas, mientras que los datos WPBC establece en 33 entradas y dos salidas. [17]

4. **2006 Springer Berlin / Heidelberg**, Department of Information and Communication Systems Engineering, University of the Aegean, Karlovassi. Este artículo trata de las estimaciones del cáncer de mama así como diagnóstico y pronóstico en la implementación de redes neuronales utilizando la base de datos Wisconsin *Breast Cancer*. Para el problema de diagnóstico, la exactitud de la red neuronal en términos de sensibilidad y especificidad se midió en 98,6 y 97,5% respectivamente. En lo que se refiere al problema de pronóstico, la exactitud de la red neuronal se midió mediante un enfoque estratificado de validación cruzada por diez. La sensibilidad osciló entre 80,5 y 91,8%, mientras que la especificidad osciló entre el 91,9 y 97,9%, dependiendo de la prueba doble y la división del período previsto. [18]
  
5. **2000 Annals of University of Craiova, Math. Comp. Sci. Ser.** En este trabajo utiliza un clasificador bayesiano que tiene como nombre Naive Bayes, este es un clasificador probabilístico basado en el teorema de Bayes, considerando una fuerte (Naive) que es una suposición de independencia. Este clasificador Naive Bayes considera a todos los atributos de forma independiente y ayudan a contribuir a la probabilidad de una determinada decisión. El clasificador Naive Bayes se puede entrenar de manera muy eficiente en un aprendizaje supervisado. Éste clasificador se ha aplicado a los pronósticos del cáncer de mama de la base de datos de la universidad de Wisconsin (WPBC). Las pruebas de precisión de diagnóstico, oscilo entre 74. 24%. [19]
  
6. **1997 Burkeef** Compararon el desempeño del sistema de escenarios TNM (TNM staging system) con una red neuronal multicapa entrenada con *back propagation*. Burke demostró que el uso de la red neuronal aumentó en un 10% la capacidad de predicción, es decir, la cifra se elevó de un 44% a un 54%. [20]
  
7. **2001 Street, W. N., Kim, Y, Streaming ensemble algorithm (SEA)**. Utiliza una combinación de varios clasificadores para crear un clasificador de gran escala. Este método lo utiliza para clasificar a aquellos pacientes que morirán en los 5 años siguientes a la cirugía y a aquellos que sobrevivirán en ese mismo lapso. Con el mismo propósito, Wang y Witten. [21]



## CAPITULO 2 Memorias Asociativas

### 2.1 Introducción

En la actualidad las Memorias Asociativas han cobrado gran importancia en temas relacionados con la teoría y aplicaciones del reconocimiento y clasificación de patrones.

El propósito fundamental de una memoria asociativa es recuperar correctamente patrones completos a partir de patrones de entrada, los cuales pueden estar alterados con ruido aditivo, sustractivo o combinado. En el diseño de una memoria asociativa, previamente a la fase de recuperación de patrones, se requiere la fase de aprendizaje, que es el proceso mediante el cual se crea la memoria asociativa, a través de la formación de asociaciones de patrones, las cuales son parejas de patrones, uno de entrada y uno de salida. Si en cada asociación sucede que el patrón de entrada es igual al de salida, la memoria es auto asociativa; en caso contrario, la memoria es hetero asociativa: esto significa que una memoria auto asociativa puede considerarse como un caso particular de una memoria hetero asociativa.

### 2.2 Memoria asociativa Lernmatrix, Correlograph, Linear Associator y Hopfield.

Las Memorias Asociativas han merecido la atención de numerosos investigadores internacionales desde hace más de cuatro décadas. Uno de los pioneros fue el científico alemán Karl Steinbuch quien, a principios de la década de los sesenta, ideó, desarrolló y aplicó la Lernmatrix. (Steinbuch, 1961; Steinbuch y Frank, 1963).

La Lernmatrix constituye un antecedente crucial en el desarrollo de los modelos actuales de memorias asociativas, y constituye uno de los primeros intentos exitosos de codificar información en arreglos cuadrículados conocidos como crossbar (Simpson, 1990).

No obstante que Steinbuch presentó la Lernmatrix hace más de cuatro décadas, ningún investigador, incluyendo al propio Steinbuch, se ha dado a la tarea de estudiar con rigor científico las condiciones necesarias y suficientes para recuperación perfecta del conjunto fundamental y de patrones que no pertenezcan a éste.[19]

Ocho años después de la *Lernmatrix*, en 1969, tres científicos escoceses (Willshaw, Buneman & Longuet-Higgins) crearon el *Correlograph*, el cual es un dispositivo óptico elemental capaz de funcionar como una memoria asociativa. En palabras de los autores "el sistema es tan simple, que podría ser construido en cualquier laboratorio escolar de física elemental".



El año de 1972 fue testigo de una febril actividad de los investigadores pioneros en memorias asociativas. Para corroborar esta aseercción, baste dar una vistazo a la siguiente información: apoyado por la UCLA, James A. Anderson prendió la mecha con su *Interactive Memory* ; en abril, Teuvo Kohonen, a la sazón profesor de la Helsinki University of Technology, presentó ante el mundo sus *Correlation Matrix Memories*; tres meses después, Kaoru Nakano de la University of Tokyo, dio a conocer su *Associatron* ; y en el ocaso del año, Shun-Ichi Amari, profesor de la University of Tokyo y uno de los más prolíficos escritores de artículos científicos hasta nuestros días, publicó un trabajo teórico donde continuaba con sus investigaciones sobre las *Self-Organizing Nets of Threshold Elements*, tópico cuyo estudio había iniciado un par de años atrás. Este trabajo de Amari constituye una valiosa muestra del estilo teórico característico de este personaje y representa, históricamente, un importante antecedente para la creación de lo que una década más tarde se convertiría en el modelo de memoria asociativa por antonomasia: la memoria Hopfield. Los trabajos de Anderson y Kohonen, y en cierta medida el de Nakano, dieron lugar al modelo que actualmente se conoce con el nombre genérico de *Linear Associator*.

Si 1972 fue el año de los pioneros en el área de las memorias asociativas, 1982 fue el año del científico estadounidense John J. Hopfield. Su artículo de ese año, publicado por la prestigiosa y respetada National Academy of Sciences (en sus Proceedings), impactó positivamente y trajo a la palestra internacional su memoria asociativa. Dos años después, publicaría su segundo artículo, donde presentaba una extensión de su modelo original.

Las autorizadas voces de los editores del compendio Neurocomputing aseguran que la era moderna de las memorias asociativas (y de las redes neuronales) nace a raíz de la publicación del

artículo de Hopfield; afirman que el éxito de este artículo se debe en gran parte a que, además de tener un estilo claro, coherente sofisticado, fue escrito por el distinguido y reconocido físico John J. Hopfield, en cuyas manos la teoría se convierte en algo legítimo y respetable.

El trabajo de Hopfield ha tenido un gran impacto en las áreas de memorias asociativas y redes neuronales. Invitamos al lector a que verifique esta afirmación mediante una búsqueda en Google con la palabra clave "Hopfield"; es sorprendente la cantidad de trabajos científicos, ingenieriles y de aplicación a que ha dado lugar el conjunto de ideas plasmadas por John Hopfield en sus trabajos científicos.

Los trabajos de Hopfield causaron excitación en el mundo de las memorias asociativas y las redes neuronales, de modo tal que muchos de los científicos que habían estado al margen, gracias a Hopfield se interesaron en estos temas y se generó una actividad importante. Así, en la segunda mitad de la década de los ochenta del siglo XX aparecieron investigadores que tomaron los modelos clásicos, los modificaron o extendieron y dieron lugar a nuevos tipos de memorias asociativas, cuya importancia



radica en que son consecuencia del trabajo de los grandes hombres involucrados en la concepción y desarrollo de los modelos clásicos.

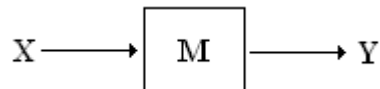
Entre la mirada de aportaciones e innovaciones en el campo de las memorias asociativas, después del espectacular éxito de la memoria Hopfield de 1982, no sucedió nada realmente trascendente hasta 1998, cuando aparecieron las memorias asociativas morfológicas.

La diferencia fundamental entre estas memorias y las memorias asociativas clásicas como el *Linear Associator* y la Memoria Hopfield, es que mientras éstas basan su operación en la suma y multiplicación usuales, las memorias morfológicas se basan en las operaciones morfológicas de dilatación y erosión. Estas memorias rompieron el esquema utilizado a través de los años en los modelos de memorias asociativas clásicas, que utilizan operaciones convencionales entre vectores y matrices para la fase de aprendizaje, y suma de productos para recuperar patrones.

Las memorias asociativas morfológicas usan máximos o mínimos de sumas para la fase de aprendizaje, y máximos o mínimos de sumas para la fase de recuperación. Con este nuevo esquema, superaron claramente a las memorias asociativas clásicas. [22]

### 2.3 Conceptos básicos

Una Memoria Asociativa puede formularse como un sistema de entrada y salida, idea que se esquematiza a continuación:



**Figura 2.1** Representación de una Memoria Asociativa

En la figura 2.1 se observa un vector columna de entrada 'X' y un vector columna de salida 'Y'. Cada uno de los patrones de entrada forma una asociación con el correspondiente patrón de salida, la cual es similar a la una pareja ordenada.

Los patrones de entrada y salida se denotarán con las letras negrillas, **X** y **Y**, agregándoles números naturales como superíndices para efectos de discriminación simbólica. Por ejemplo, a un patrón de entrada  $x^1$  le corresponderá el patrón de salida  $y^1$ , y ambos formarán la asociación  $(x^1, y^1)$ ; del mismo modo, para un número entero positivo  $k$  específico, la asociación correspondiente será  $(x^k, y^k)$ .

La memoria asociativa **M** se representa mediante una matriz, la cual se genera a partir de un conjunto finito de asociaciones conocidas de antemano: este es el conjunto fundamental de aprendizaje, o simplemente conjunto fundamental.



El conjunto fundamental se representa de la siguiente manera:

$$\{(\mathbf{x}^\mu, \mathbf{y}^\mu) \mid \mu = 1, 2, \dots, p\}$$

Donde  $p$  es un número entero positivo que representa la cardinalidad del conjunto fundamental.

A los patrones que conforman las asociaciones del conjunto fundamental se les llama patrones fundamentales. La naturaleza del conjunto fundamental proporciona un importante criterio para clasificar las memorias asociativas:

Una memoria es Auto asociativa si se cumple que  $\mathbf{x}^\mu = \mathbf{y}^\mu \forall \mu \in \{1, 2, \dots, p\}$ , por lo que uno de los requisitos que se debe de cumplir es que  $n = m$ .

Una memoria hetero asociativa es aquella en donde  $\exists \mu \in \{1, 2, \dots, p\}$  para que se cumple que:

$\mathbf{x}^\mu \neq \mathbf{y}^\mu$ . Nótese que puede haber memorias hetero asociativas con  $n = m$ .

En los problemas donde intervienen las memorias asociativas, se consideran dos fases importantes: La fase de aprendizaje, que es donde se genera la memoria asociativa a partir de las  $p$  asociaciones del conjunto fundamental, y la fase de recuperación que es donde la memoria asociativa opera sobre un patrón de entrada, a la manera del esquema que aparece al inicio de esta sección.

A fin de especificar las componentes de los patrones, se requiere la notación para dos conjuntos a los que llamaremos arbitrariamente **A** y **B**. Las componentes de los vectores columna que representan a los patrones, tanto de entrada como de salida, serán elementos del conjunto **A**, y las entradas de la matriz **M** serán elementos del conjunto **B**.

No hay requisitos previos ni limitaciones respecto de la elección de estos dos conjuntos, por lo que no necesariamente deben ser diferentes o poseer características especiales. Esto significa que el número de posibilidades para escoger **A** y **B** es infinito.

Por convención, cada vector columna que representa a un patrón de entrada tendrá  $n$  componentes cuyos valores pertenecen al conjunto **A**, y cada vector columna que representa a un patrón de salida tendrá  $m$  componentes cuyos valores pertenecen también al conjunto **A**. Es decir:

$$\mathbf{x}^\mu \in A^n \text{ y } \mathbf{y}^\mu \in A^m \forall \mu \in \{1, 2, \dots, p\}$$



La  $j$ -ésima componente de un vector columna se indicará con la misma letra del vector, pero sin negrilla, colocando a  $j$  como subíndice ( $j \in \{1, 2, \dots, n\}$  o  $j \in \{1, 2, \dots, m\}$  según corresponda). La  $j$ -ésima componente del vector columna  $\mathbf{x}^\mu$  se representa por:  $x_j^\mu$

Con los conceptos básicos ya descritos y con la notación anterior, es posible expresar las dos fases de una memoria asociativa:

1. **Fase de Aprendizaje** (Generación de la memoria asociativa). Encontrar los operadores adecuados y una manera de generar una matriz  $\mathbf{M}$  que almacene las  $p$  asociaciones del conjunto fundamental  $\{(\mathbf{x}^1, \mathbf{y}^1), (\mathbf{x}^2, \mathbf{y}^2), \dots, (\mathbf{x}^p, \mathbf{y}^p)\}$ , donde  $\mathbf{x}^\mu \in A^n$  y  $\mathbf{y}^\mu \in A^m \forall \mu \in \{1, 2, \dots, p\}$ .

Si  $\exists \mu \in \{1, 2, \dots, p\}$  tal que  $\mathbf{x}^\mu \neq \mathbf{y}^\mu$ , la memoria será hetero asociativa; si  $m = n$  y

$\mathbf{x}^\mu = \mathbf{y}^\mu \forall \mu \in \{1, 2, \dots, p\}$ , la memoria será auto asociativa.

2. **Fase de Recuperación** (Operación de la memoria asociativa). Hallar los operadores adecuados y las condiciones suficientes para obtener el patrón fundamental de salida  $\mathbf{y}^\mu$ , cuando se opera la memoria  $\mathbf{M}$  con el patrón fundamental de entrada  $\mathbf{x}^\mu$ ; lo anterior para todos los elementos del conjunto fundamental y para ambos modos: auto asociativo y hetero asociativo.

Se dice que una memoria asociativa  $\mathbf{M}$  exhibe recuperación correcta si al presentarle como entrada, en la fase de recuperación, un patrón  $\mathbf{x}^\omega$  con  $\omega \in \{1, 2, \dots, p\}$ , ésta responde con el correspondiente patrón fundamental de salida  $\mathbf{y}^\omega$ .

Una memoria asociativa bidireccional también es un sistema de entrada y salida, solamente que el proceso es bidireccional. La dirección hacia delante se describe de la misma forma que una memoria asociativa común: al presentarle una entrada  $\mathbf{X}$ , el sistema entrega una salida  $\mathbf{y}$ . La dirección hacia atrás se lleva a cabo presentándole al sistema una entrada  $\mathbf{y}$  y para recibir una salida  $\mathbf{X}$ . [23]

## 2.4 Memorias Asociativas Alfa-Beta

Las memorias Alfa-Beta utilizan máximos y mínimos, y dos operaciones binarias originales  $\alpha$  y  $\beta$  de las cuales heredan el nombre.

Para la definición de las operaciones binarias  $\alpha$  y  $\beta$  se deben especificar los conjuntos  $\mathbf{A}$  y  $\mathbf{B}$ , los cuales son:

$$A = \{0, 1\} \text{ y } B = \{0, 1, 2\}$$



La operación binaria  $\alpha: A \times A \rightarrow B$  se define como:

**Tabla 2.1**  $A \times A \rightarrow B$

x	y	$\alpha(x, y)$
0	0	1
0	1	0
1	0	2
1	1	1

La operación binaria  $\beta: B \times A \rightarrow A$  se define como

**Tabla 2.2**  $B \times A \rightarrow A$

x	y	$\beta(x, y)$
0	0	0
0	1	0
1	0	0
1	1	1
2	0	1
2	1	1

Los conjuntos  $A$  y  $B$ , las operaciones binarias  $\alpha$  y  $\beta$  junto con los operadores  $\wedge$ (mínimo) y  $\vee$ (máximo) usuales conforman el sistema algebraico  $(A, B, \alpha, \beta, \wedge, \vee)$  en el que están inmersas las memorias asociativas Alfa-Beta.



### 2.5. Operadores $\alpha$ y $\beta$

Las memorias asociativas Alfa-Beta son de dos tipos (máx. y min) y pueden operaren dos modos diferentes (auto asociativo y hereto asociativo). El operador  $\alpha$  es utilizado en la fase de aprendizaje, mientras que el operador  $\beta$  es útil durante la fase de recuperación. A continuación se incluyen las tablas que representan a los operadores  $\alpha$  y  $\beta$ , dados los conjuntos  $A=\{0,1\}$  y  $B=\{00,01,10\}$ :

**Tabla 2.3.** Definición del operador Alfa

$\alpha : A \times A \rightarrow B$		
$x$	$y$	$\alpha(x,y)$
0	0	01
0	1	00
1	0	10
1	1	01

**Tabla 2.4.** Definición del operador Beta

$\beta : B \times A \rightarrow A$		
$x$	$y$	$\beta(x,y)$
00	0	0
00	1	0
01	0	0
01	1	1
10	0	1
10	1	1

La operación binaria  $\alpha$  exhibe algunas propiedades algebraicas, expuestas en la Tabla 2.5.





**Tabla 2.5. Propiedades de la operación binaria  $\alpha$**

$\alpha$ 1.-isoargumentos en $\alpha$	$\alpha (x, x) = 1$
$\alpha$ 2.-intercambio de argumentos en $\alpha$	$(x \leq y) \leftrightarrow \alpha (x, y) \leq \alpha (y, x)$
$\alpha$ 3.- es creciente por la izquierda	$(x \leq y) \leftrightarrow [\alpha (x, z) \leq \alpha (y, z)]$
$\alpha$ 4.- $\alpha$ es decreciente por la derecha	$(x \leq y) \leftrightarrow [\alpha (z, x) \leq \alpha (z, y)]$
$\alpha$ 5.- $\alpha$ es distributiva por la derecha respecto al	$\alpha [x \bigvee y, z] = \alpha (x, z) \bigvee \alpha (y, z)$
$\alpha$ 6.- $\alpha$ es distributiva por la derecha respecto al	$\alpha [x \bigvee y, z] = \alpha (x, z) \bigwedge \alpha (y, z)$



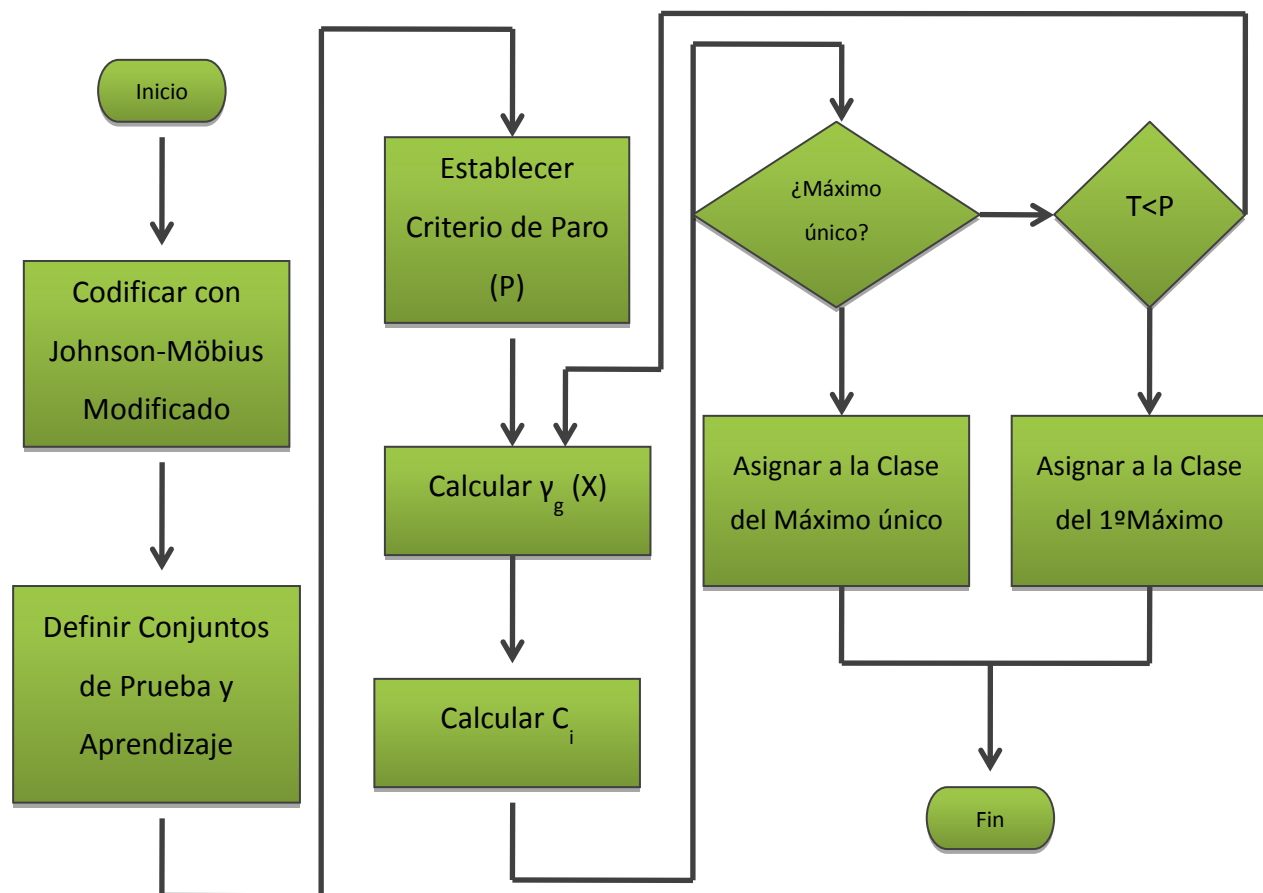
## CAPÍTULO 3 Métodos

En este Capítulo se dan a conocer los conceptos y métodos en los cuales se basa el desarrollo del Clasificador de Recurrencia Cáncer

En la primera sección se describe el Algoritmo que da lugar al Clasificador Gamma Generalizado, el cual fue desarrollado por el M. en C. Itzamá López Yáñez. En la segunda sección se presenta el operado  $u_{\beta}$ , el cual también es esencial para el desarrollo del Clasificador.

Dado que los vectores con que trabaja el algoritmo serán codificados con el código Johnson-Möbius modificado se presenta información de éste método así como algunos ejemplos numéricos.

### 3.1 Algoritmo





### 3.2. Operador $u_\beta$

“Se define el operador  $u_\beta(x)$  de la siguiente manera:  $u_\beta(x)$  tiene como argumento de entrada un vector binario n-dimensional  $x$  con la  $i$ -ésima componente representada por  $x_i$  y la salida es un número entero no negativo”[24] que se calcula así:

$$u_\beta(x) = \sum_{i=1}^n \beta(x_i, x_i)$$

**Ejemplo 3.1** Sea  $x = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}$ ; obtener  $u_\beta(x)$ .

Según la definición del operador  $u_\beta$  se tiene que:

$$u_\beta = \sum_{i=1}^5 \beta(x_i, x_i) \therefore u_\beta(x) = \beta(0,0) + \beta(1,1) + \beta(1,1) + \beta(0,0) + \beta(0,0) \\ = 0 + 1 + 1 + 0 + 0 = 2$$

Entonces,  $u_\beta=2$

**Ejemplo 3.2** Sean  $x = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, y = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ ; obtener  $u_\beta(x)$  y  $u_\beta(xy)$ .

Ahora  $u_\beta(x) = \beta(1,1) + \beta(1,1) + \beta(1,1) = 1 + 1 + 1 = 3$ , mientras que  $u_\beta(y) = \beta(0,0) + \beta(0,0) + \beta(0,0) + \beta(0,0) = 0 + 0 + 0 + 0 = 0$

### 3.3 Código Binario Johnson-Möbius Modificado

Existen tres tipos de ruido:

Ruido Aditivo: 0110  $\longrightarrow$  0111 “se añade un bit”

Ruido Sustractivo: 0111  $\longrightarrow$  0101 “se quita un bit”

Ruido Mezclado: 0111  $\longrightarrow$  1000 “se añade y se quita un bit”

Las Memorias Asociativas funcionan bien contra el ruido Aditivo y el ruido Sustractivo; sin embargo; son poco eficientes en el caso del Ruido Mezclado, es en estos casos cuando se usa el código Johnson-Möbius Modificado.

A continuación se presentan los pasos a seguir para utilizar el Método Johnson Möbius Modificado así como algunos ejemplos.



### 3.4 Algoritmo del Código Johnson-Möbius Modificado:

1. Sea un conjunto de números reales

$$\{r_1, r_2, \dots, r_i, \dots, r_n\}$$

Donde  $n$  es un número entero positivo fijo.

2. Si uno de los números del conjunto (por ejemplo  $r_i$ ) es negativo, se debe crear un nuevo conjunto transformado a través de la operación “restar  $r_i$  a cada uno de los  $n$  números”

$$\{t_1, t_2, \dots, t_i, \dots, t_n\}$$

Donde  $t_j = t_j - r_i \forall j \in \{1, 2, \dots, n\}$  y particularmente  $t_i = 0$ . Nota: si hay más de un negativo, se trabaja con el menor. Con esto se logra que el menor número sea el cero; es decir, el conjunto es no negativo.

3. Escoger un número fijo  $d$  de decimales y truncar cada uno de los números del conjunto transformado (los cuales son no negativos) precisamente a  $d$  decimales. Este número  $d$  depende de la exactitud requerida para el problema específico que se estén tratando.
4. Realizar un escalamiento de  $10d$  en el conjunto del paso 3, para obtener un conjunto de  $n$  **enteros no negativos**.

$$\{e_1, e_2, \dots, e_i, \dots, e_m, \dots, e_n\}$$

Donde  $e_m$  es el número mayor.

5. El código Johnson-Möbius modificado para  $j=1, 2, \dots, n$  se obtiene al generar  $(e_m - e_j)$  ceros concatenados por la derecha con  $e_j$  unos.

Para aclarar el anterior método a continuación se presentan algunos ejemplos:

**Ejemplo 3.3** Representar los números 3, 7, 11, 17 y 19 con 20 bits usando el código Johnson-Möbius modificado, se tiene lo siguiente:

1. Cada código tendrá 20 bits.
2. Para el número 3 se tendrán  $20 - 3 = 17$  ceros seguidos de 3 unos.
3. Para el número 7 se tendrán  $20 - 7 = 13$  ceros seguidos de 7 unos.



- 4. Para el número 11 se tendrán  $20 - 11 = 9$  ceros seguidos de 11 unos.
- 5. Para el número 17 se tendrán  $20 - 17 = 3$  ceros seguidos de 17 unos.
- 6. Para el número 19 se tendrán  $20 - 19 = 1$  cero seguidos de 19 unos.

Los códigos correspondientes se muestran en la **Tabla 3.1**.

**Tabla 3.1.** Código Johnson-Möbius Modificado para 3, 7, 11, 17 y 19

NUMERO	Código Johnson-Möbius Modificado
3	00000000000000000111
7	00000000000001111111
11	00000000111111111111
17	00011111111111111111
19	01111111111111111111

**Ejemplo 3.4** Representar los números 0.2, 1.25, -0.3 y 2.147 con 20 bits usando el código Johnson-Möbius modificado, se tiene lo siguiente:

- 1. Se tiene el conjunto  $r = \{0.2, 1.25, -0.3, 2.147\}$
- 2. Dado que existe un número negativo en el conjunto (-0.3) éste deberá restarse a todo el conjunto para obtener  $t = \{0.5, 1.55, 0, 2.447\}$
- 3. Se escoge un número fijo  $d = 1$ , con el cual el número de decimales se reducirá a uno para obtener  $t = \{0.5, 1.5, 0, 2.4\}$
- 4. Se realiza un escalamiento  $10d$  para obtener  $e = \{5, 15, 0, 24\}$  donde  $e_m = 24$  (el número mayor del conjunto resultante).
- 5. Para cada número  $e_i$  del conjunto e, se generan  $e_m - e_i$  ceros concatenados con  $e_i$  unos.

Los códigos correspondientes se muestran en la **Tabla 3.2**.

**Tabla 3.2.** Ejemplo Código Johnson-Möbius Modificado

NUMERO	Código Johnson-Möbius Modificado
5	000000000000000000011111
10	00000000000000011111111111
0	00000000000000000000000000
24	1111111111111111111111111111



### 3.5. Operadores de Cadena Binaria Mínima y k-Binario de Expansión

Sea  $r$  un número entero no negativo. Se define el **operador de cadena binaria mínima**  $k_r$  de la siguiente manera:  $k_r$  tiene como argumento de entrada  $r$ , y la salida es el menor de los logaritmos de base 2 de entre todos los números mayores que  $r$  de la forma  $2^k$ , donde  $k$  es un número entero positivo. [24]

**Ejemplo 3.5** Calcular  $k(6)$ .

De acuerdo con la notación de la Definición, se observa que  $r = 6$ , por lo que los primeros elementos del conjunto de números enteros mayores que 6 son 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, ... y así sucesivamente. Es claro que sólo algunos de estos números son de la forma  $2^k$ ; algunos ejemplos, cuando  $k$  toma los valores 3, 4, 5, 6, 7 y 8, son  $8=2^3$ ,  $16=2^4$ ,  $32=2^5$ ,  $64=2^6$ ,  $128=2^7$ ,  $256=2^8$ . Sin embargo, hay dos casos que resaltan: por un lado, existen algunos valores de  $k$  cuyos correspondientes números de la forma  $2^k$  no están en el conjunto mencionado, y por otro, hay números en ese conjunto para los que no existe un número entero positivo  $k$  que cumpla con la forma  $2^k$  y sea igual al número mencionado. Para el primer caso, se observa que el número que corresponde a  $k = 2$  no se encuentra en este conjunto, dado que  $2^2 = 4$  y 4 es menor que 6; y para el segundo caso, es posible notar que no existe un número  $k$  que sea entero positivo y que cumpla con  $6=2^k$ . Ahora, al tomar los logaritmos de base 2 de los números que sí cumplen con las dos condiciones para este ejemplo, se tienen los siguientes valores:

$$\begin{aligned} \log_2 8 &= \log_2 2^3 = 3 \\ \log_2 16 &= \log_2 2^4 = 4 \\ \log_2 32 &= \log_2 2^5 = 5 \\ \log_2 64 &= \log_2 2^6 = 6 \\ \log_2 128 &= \log_2 2^7 = 7 \\ \log_2 256 &= \log_2 2^8 = 8 \\ \log_2 512 &= \log_2 2^9 = 9 \end{aligned}$$

,... así sucesivamente, siendo 3 el menor de todos. Es decir,  $k(6) = 3$ .

“Sean:  $r$  un número entero no negativo y  $k$  un número entero positivo, los cuales cumplen con la expresión  $k \geq k(r)$ . Se define el **operador k-binario de expansión**  $\varepsilon(r, k)$  de la siguiente manera:  $\varepsilon(r, k)$  tiene como argumentos de entrada a  $r$  y  $k$ , y la salida es un vector columna binario  $k$ -dimensional cuyas componentes corresponden a la expansión binaria de  $k$  bits de  $r$ , con el bit menos significativo en la parte inferior”. [24]

**Ejemplo 3.6** Obtener  $\varepsilon(6,3)$

En este caso  $r = 6$  y  $k = 3$  y de acuerdo con el Ejemplo 3.5 se tiene que  $k(6) = 3$  por lo que se cumple la condición  $k \geq k(r)$ . Dado que la cadena binaria que representa al



número 6 es 110, siendo 0 el bit menos significativo, el operador 3-binario de expansión para el número 6 produce el vector columna siguiente:

$$\varepsilon(6,3) = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

**Ejemplo 3.7** Obtener  $\in(6,2)$

En este caso  $r=6$  y  $k=2$ , y ya se ha definido que  $k(6) = 3$ , entonces se tiene que con  $k = 2$  la condición de  $k \geq k(r)$  no se cumple, y esto significa que el vector que corresponde a  $\in(6,2)$  no existe.

**3.6. Operador Gama de Similitud**

El operador gama de similitud, está basado en la operaciones  $\alpha$  y  $\beta$  de las memorias asociativas Alfa-Beta. Este operador indica si dos vectores son parecidos o no, dado un grado de disimilitud  $\theta$ ; por ello, este operador es de similitud.

Es importante resaltar que, debido a que el operador Gamma se basa en las operaciones  $\alpha$  y  $\beta$ , los dos vectores que compara  $x$  y  $y$  son binarios  $n$ -dimensionales.

El grado de similitud  $\theta$  es un número entero no negativo.

“Se define el operador gama de similitud  $\gamma(x, y, \theta)$  de la siguiente manera:  $\gamma(x, y, \theta)$  tiene como argumentos de entrada dos vectores binarios  $n$ -dimensionales  $x$  y  $y$ , un número entero no negativo  $\theta$ , y la salida es un número binario”[24] que se calcula así:

$$\gamma(x, y, \theta) = \begin{cases} 1 & \text{si } n - u_{\beta}[\alpha(x, y) \text{ mod } 2] \leq \theta \\ 0 & \text{en otro caso} \end{cases}$$

**Ejemplo 3.8** Sean  $x = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}$ ,  $y = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$  y  $\theta = 3$ ; calcular  $\gamma(x, y, \theta)$ .



En este caso se pudo observar que  $n = 8$ . Al calcular  $\alpha(x, y)$  se obtiene  $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 2 \\ 0 \\ 2 \\ 2 \\ 1 \end{pmatrix}$ ;

Calculando el módulo 2 de cada componente, se obtiene el vector  $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$ ; ahora bien,

$u_\beta \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = 2$ , que al resultado de  $n = 8$  da como resultado  $8 - 2 = 6$ ; pero  $6 > 3$ , por lo que  $6 \not\leq 3$ , entonces  $\gamma(x, y, \theta) = 0$ .

**Ejemplo 3.9** Sean  $x = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$ ,  $y = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$  y  $\theta = 5$ ; calcular  $\gamma(x, y, \theta)$ . Se puede observar que para





este caso  $n = 10$ . Luego,  $\alpha(x, y) = \begin{pmatrix} 1 \\ 2 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 0 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 0 \\ 2 \\ 1 \end{pmatrix} \text{ mod } 2 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, u_\beta \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} = 5, 10 - 5 = 5, y$

como  $5 = 5$  entonces  $5 \leq 5$ , por lo que  $\gamma(x, y, \theta) = 1$ .

**Ejemplo 3.10** Sean  $x = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, y = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$  y  $\theta = 3$ ; calcular  $\gamma(x, y, \theta)$ .

Ahora se tiene que  $n = 5$ . Entonces  $\alpha(x, y) = \begin{pmatrix} 1 \\ 2 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 0 \\ 1 \\ 1 \end{pmatrix} \text{ mod } 2 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix},$

$u_\beta \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} = 3, 5 - 3 = 2, y$  como  $2 < 3$ , entonces  $2 \leq 3$ , y el resultado es  $\gamma(x, y, \theta) = 1$ .

### 3.7 Algoritmo Clasificador de Patrones

El Algoritmo Clasificador utiliza los operadores  $\alpha$  y  $\beta$ , así como sus propiedades, usa el operador de similitud, que, a su vez requiere del operador  $u_\beta$  y trabaja con datos codificados con el Código Johnson-Möbius Modificado, para poder clasificar patrones.

Sean los números  $k, m, n, p \in \mathbb{Z}^+, \{x^\mu \mid \mu = 1, 2, \dots, p\}$  el conjunto fundamental de patrones de cardinalidad  $p$ , donde  $\forall \mu x^\mu \in \mathbb{R}^n$  y  $y \in \mathbb{R}^n$  un vector real  $n$ -dimensional a ser clasificado. Se asume que el conjunto fundamental está particionado en  $m$  clases diferentes, donde cada clase tiene cardinalidad  $k_i, i = 1, 2, \dots, m$ , por lo que:

$$\sum_{i=1}^m k_i = p$$



Para clasificar el patrón  $y$ , se realiza lo siguiente:

1. Codificar todas las componentes de cada uno de los patrones que servirán como conjunto de aprendizaje con el código Johnson-Möbius modificado; para ello, se debe calcular la magnitud de  $e_m$ , equivalente al número de bits necesarios para la codificación (como se mostró en el tema 3.4). Para elegir la magnitud  $e_m$  se consideran las componentes de todos los patrones del conjunto fundamental que tienen el mismo índice, en cada uno de los pasos de la codificación usando el código Johnson-Möbius modificado; es decir, para cada  $j= 1,2, \dots, n$  se calcula

$$e_m(j) = \bigvee_{i=1}^p x_j^i$$

Una vez que se tiene la magnitud  $e_m$  las componentes  $x_j^i$  han de ser transformadas y escaladas, de modo que todos los valores sean enteros no negativos para después generar un vector binario con  $(e_m - e_j)$  ceros concatenados con  $e_j$  unos. Así, la componente  $x_j^i$  se transforma en un vector binario de dimensión  $e_m(j)$ .

De lo anterior se puede deducir que el vector de patrones se transformara en una matriz, donde cada fila será un patrón codificado.

2. Calcular el parámetro de paro, para ello se elije el más pequeño de los valores máximos ( $e_m$ ), este valor delimita las iteraciones que realizará el clasificador.

$$\rho = \bigwedge_{j=1}^n e_m(j).$$

3. Codificar cada componente del patrón a clasificar con el código *Johnson-Möbius* modificado, utilizando las mismas condiciones que se utilizaron para codificar las componentes de los patrones fundamentales; es decir, para cada índice de componente, se utilizan los mismos valores de  $r_i$ ,  $d$  y  $e_m$  que se describen en el Tema 3.4.
4. Realizar una transformación de índices en los patrones del conjunto fundamental, de manera que el índice único que tenía un patrón originalmente en el conjunto fundamental, por ejemplo  $x^\mu$ , se convierta en dos índices: uno para la clase (por ejemplo la clase  $\hat{i}$ ) y otro para el orden que le corresponde a ese patrón dentro de esa clase (por ejemplo  $\omega$ ). Bajo estas condiciones ejemplificadas, la notación para el patrón  $x^\mu$  será ahora, con la transformación  $x^{i\omega}$ . Lo anterior se realiza para todos los patrones del conjunto fundamental.



- Inicializar  $\theta$  a 0. Ya que  $\theta$  es el grado de similitud el igualarlo a cero significa que se exige que los patrones sean iguales.
- Realizar la operación  $\gamma(x_j^{i\omega}, y_j, \theta)$  para cada clase y para cada componente de cada uno de los patrones fundamentales que corresponden a esa clase, y del patrón a clasificar:

$$\gamma(x_j^{i\omega}, y_j, \theta) \text{ donde } i = 1, 2, \dots, m, \omega = 1, 2, \dots, k_i, j = 1, 2, \dots, n$$

- Calcular la suma ponderada  $c_i$  de los resultados obtenidos en el paso 6, para cada clase  $i=1, 2, \dots, m$ :

$$c_i = \frac{\sum_{\omega=1}^{k_i} \sum_j^n \gamma(x_j^{i\omega}, y_j, \theta)}{k_i}$$

- Si existe más de un máximo entre las sumas ponderadas por clase, incrementar  $\theta$  en 1 y repetir los pasos 6 y 7 hasta que exista un máximo único, o se cumpla con la condición de paro  $\theta \geq \rho$ .

- Si existe un máximo único, asignar al patrón a clasificar la clase correspondiente a ese máximo:

$$C_y = C_j \text{ tal que } \bigvee_{i=1}^m c_i = c_j$$

- En caso contrario: si  $\lambda$  es el índice más pequeño de clase que corresponde a uno de los máximos, asignar al patrón a clasificar la clase  $C_\lambda$ .

**Ejemplo 3.11** Sean los patrones fundamentales  $x^1 = \begin{pmatrix} 2 \\ 8 \end{pmatrix}, x^2 = \begin{pmatrix} 1 \\ 9 \end{pmatrix}, x^3 = \begin{pmatrix} 6 \\ 3 \end{pmatrix}, x^4 = \begin{pmatrix} 7 \\ 4 \end{pmatrix}$ , agrupados en dos clases  $C_1 = \{x^1, x^2\}, C_2 = \{x^3, x^4\}$ ; clasificar los patrones:

$$y^1 = \begin{pmatrix} 6 \\ 2 \end{pmatrix}, y^2 = \begin{pmatrix} 3 \\ 9 \end{pmatrix}$$

Para este caso, la dimensión de los patrones es  $n = 2$ , y se tienen  $m = 2$  clases, donde las clases agrupan  $k_1 = 2, k_2 = 2$  patrones, respectivamente.

Entonces, siguiendo los pasos del algoritmo:

- Se codifican las componentes de los patrones fundamentales, resultando que para la primera componente,  $e_m(1) = 7$  y para la segunda  $e_2(2) = 9$ . Los patrones fundamentales codificados quedan como sigue.

$$x^1 = \begin{pmatrix} 0000011 \\ 011111111 \end{pmatrix}, x^2 = \begin{pmatrix} 0000001 \\ 111111111 \end{pmatrix}, x^3 = \begin{pmatrix} 0111111 \\ 000000111 \end{pmatrix}, x^4 = \begin{pmatrix} 1111111 \\ 000001111 \end{pmatrix}$$



2. Puesto que  $e_m(1) = 7$  y  $e_m(2) = 9$ , el parámetro de paro es en este caso (el menor de las máximos):

$$\rho = \bigwedge_{j=1}^2 e_m(j) = 7$$

3. Usando las mismas  $e_m(j)$ , los patrones a clasificar quedan codificados como sigue:

$$y^1 = \begin{pmatrix} 0111111 \\ 000000011 \end{pmatrix}, y^2 = \begin{pmatrix} 0000111 \\ 111111111 \end{pmatrix}$$

4. Los patrones fundamentales se agrupan por sus clases,  $C_1 = \{x^1, x^2\}$ ,  $C_2 = \{x^3, x^4\}$  por lo que  $x^1 = x^{11}$ ,  $x^2 = x^{12}$ ,  $x^3 = x^{21}$ ,  $x^4 = x^{22}$ .

5.  $\theta = 0$

Pasos restantes del algoritmo para la clasificación de  $y^1$ :

6.  $\gamma(x_1^{11}, y_1^1, o) = 0, \gamma(x_2^{11}, y_2^1, o) = 0, \gamma(x_1^{12}, y_1^1, o) = 0, \gamma(x_2^{12}, y_2^1, o) = 0;$   
 $\gamma(x_1^{21}, y_1^1, o) = 1, \gamma(x_2^{21}, y_2^1, o) = 0, \gamma(x_1^{22}, y_1^1, o) = 0, \gamma(x_2^{22}, y_2^1, o) = 0;$

7. Se suman los resultados anteriores, obteniéndose:  $c_1 = \frac{0+0+0+0}{2} = \frac{0}{2}, c_2 = \frac{1+0+0+0}{2} = \frac{1}{2}$

8. Con lo anterior se obtiene un máximo único:

$$\bigvee_{i=1}^2 c_i = c_2 = \frac{1}{2}$$

9. Entonces, a  $y^1$  se le asigna la clase  $c_2$ .

Pasos restantes del algoritmo para la clasificación de  $y^2$ :

6.  $\gamma(x_1^{11}, y_1^2, o) = 0, \gamma(x_2^{11}, y_2^2, o) = 0, \gamma(x_1^{12}, y_1^2, o) = 0, \gamma(x_2^{12}, y_2^2, o) = 1;$   
 $\gamma(x_1^{21}, y_1^2, o) = 0, \gamma(x_2^{21}, y_2^2, o) = 0, \gamma(x_1^{22}, y_1^2, o) = 0, \gamma(x_2^{22}, y_2^2, o) = 0;$

7. Se suman los resultados anteriores, obteniéndose:  $c_1 = \frac{0+0+0+0}{2} = \frac{0}{2}, c_2 = \frac{1+0+0+0}{2} = \frac{1}{2}$

8. Con lo anterior se obtiene un máximo único:

$$\bigvee_{i=1}^2 c_i = c_1 = \frac{1}{2}$$

9. Entonces, a  $y^2$  se le asigna la clase  $c_1$ .



Es importante aclarar que para el caso específico del Clasificador Automático de Recurrencia de Cáncer, si se llega a un punto tal que  $\theta$  no puede seguir creciendo y aun no existe un máximo único, el patrón se clasifica como recurrente, esto es porque si una paciente tiene, según el Clasificador, las mismas posibilidades de tener un Cáncer de tipo Recurrente que uno de tipo No Recurrente es mejor que lo sepa y así no correrá riesgos innecesarios.



## CAPITULO 4 Pruebas

### 4.1 Pruebas de Eficiencia para clasificadores

En el presente capítulo se describen las pruebas realizadas para comprobar la efectividad del Clasificador Automático

Se realizaron tres tipos de pruebas:

- *Hold Out*
- *K-Fold Cross Validation*
- *Take One Out*

### 4.2 Descripción de la Base de Datos

La Base de Datos fue obtenida de la página de la Universidad de Irvine California:

<http://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin>

**Título:** *Wisconsin Prognostic Breast Cancer (WPBC)*

#### **Fuente de información:**

##### **a) Creadores:**

Dr. William H. Wolberg, Departamento de Cirugía General, Universidad de Wisconsin, Centro de Ciencias Clínicas, Madison, WI 53792

wolberg@eagle.surgery.wisc.edu

W. Nick Street, Departamento de Ciencias de la Computación, Universidad de Wisconsin, 1210 West Dayton St., Madison, WI 53706

street@cs.wisc.edu 608-262-6619

Olvi L. Mangasarian, Departamento de Ciencias de la Computación, Universidad de Wisconsin, 1210 West Dayton St., Madison, WI 53706

olvi@cs.wisc.edu

##### **b) Donador: Nick Street**

##### **c) Fecha de Donación: Diciembre 1995**



✓ **Información Relevante**

Cada registro representa un caso de cáncer de mama. Son pacientes consecutivos revisados por el Dr. Wolberg desde 1984, e incluye sólo aquellos casos que mostraron cáncer de mama invasivo y no mostraron evidencia alguna de metástasis distante al momento del diagnóstico.

Los primeros 30 elementos se calculan a partir de una imagen digitalizada de una punción-aspiración con aguja fina de una masa de una mama. Se describen las características de los núcleos de las células presentes en la imagen.

*Número de Registros:* 194

*Número de Atributos:* 34 (ID del paciente, Clase, 32 características con valores de entrada reales)

**Información de los Atributos**

1) Número de Identificación (ID)

2) Clase (R = recurrente, N = no recurrente)

3) Tiempo (tiempo de recurrencia si el campo 2 = R, el tiempo libre de enfermedad si el campo 2 = N)

4-33) Diez características con valores reales fueron computadas por cada uno de los núcleos celulares:

- a) radio (distancia del centro a los puntos en el perímetro)
- b) textura (Desviación estándar de los valores en escala de grises)
- c) perímetro
- d) área
- e) suavidad (variaciones locales en la medida del radio)
- f)  $\text{perímetro}^2 / \text{área} - 1.0$
- g) concavidad (gravedad de puntos cóncavos del contorno)
- h) Puntos Cóncavos (número de porciones cóncavas en el contorno)
- i) simetría
- j) dimensión fractal ("aproximación de la línea de la costa" - 1)



Los campos se distribuyen de la siguiente forma: el error estándar, la media, y "el peor" o más grande (media de los tres valores más grandes) estas características se calculan para cada imagen, resultando en 30 funciones. Por ejemplo, el campo 4 es radio medio, el campo Radio 14 es Error estándar, campo 24 es el "peor" Radio.

Los valores para las características de la 4 a la 33 contiene cuatro cifras decimales.

34) El tamaño del tumor - diámetro del tumor extirpado en centímetros.

35) Estado de los ganglios linfáticos - número de ganglios linfáticos axilares positivo observados al momento de la operación

Distribución de las Clases: 148 No recurrentes, 46 Recurrentes [25]

### 4.3 Programación del Algoritmo

El Clasificador Automático de Recurrencia de Cáncer se programó en Microsoft Visual C# 2008 Express Edition ®, este lenguaje fue elegido debido a que la documentación acerca de este lenguaje es muy amplia, además de que brinda la facilidad de realizar interfaces amigables para el usuario de manera rápida y eficaz.

El Clasificador se programó y probó en una PC con Sistema Operativo Microsoft Windows XP Professional®, procesador Intel Pentium 4®y con un 1GB de memoria RAM.

### 4.4 Prueba *Hold-Out*

En la prueba *Hold Out* el total de datos se divide en dos subconjuntos:

- Conjunto de Aprendizaje
- Conjunto de Pruebas

Cada uno de estos subconjuntos tiene distintos elementos.

La selección de los elementos que contendrá cada uno de los conjuntos se realiza al azar.

El tamaño de los conjuntos no es fijo, en la primera prueba se toma 99% para aprendizaje y 1% para pruebas, en la segunda prueba se toma 90% para aprendizaje y el 10% para pruebas, y así hasta llegar al 1% de aprendizaje y el 99% de prueba.





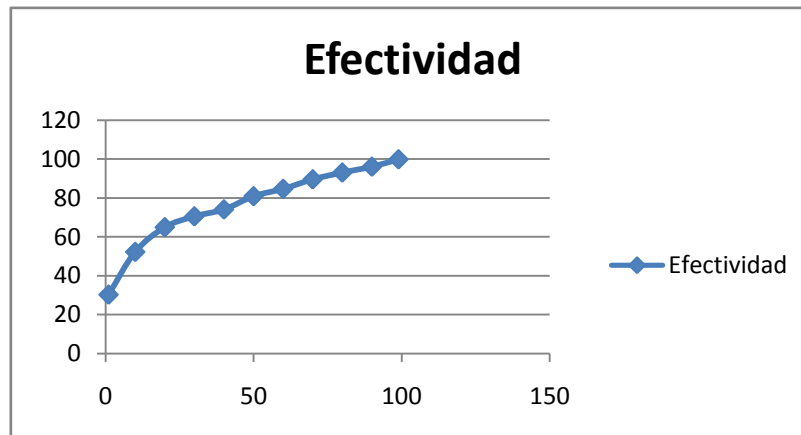
Para cada división distinta de porcentajes se realizaron veinte pruebas, para obtener un promedio de efectividad fiable.

A continuación se presentan los resultados:

**Tabla 4.1 Prueba Hold Out**

Porcentaje de Aprendizaje	Efectividad Obtenida
1	30.2
10	52.2
20	64.9
30	70.5
40	74.05
50	80.85
60	84.65
70	89.55
80	93.05
90	96.05
99	99.85

**Grafica 4.1 Prueba Hold Out**



Es una convención el tomar como porcentaje de aprendizaje valores entre 70% y 80%, es por eso que durante las pruebas Hold-Out comprobamos la efectividad con todos los valores intermedios (71, 72, 73,... 80).

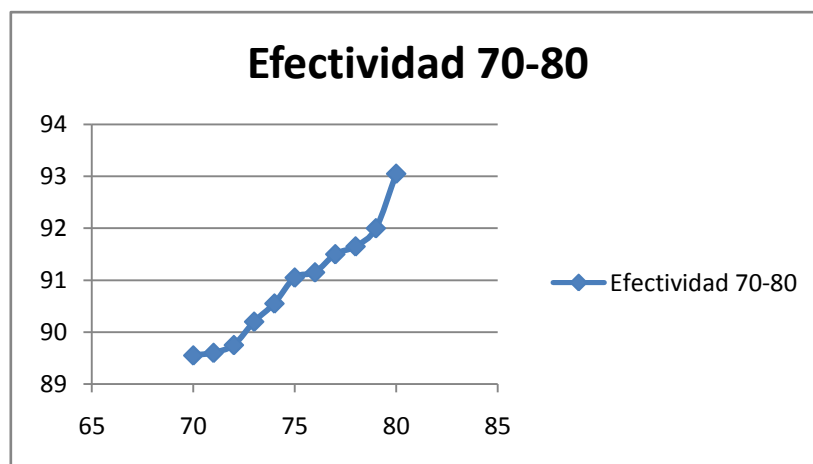


Los resultados obtenidos fueron:

**Tabla 4.2 Prueba Hold Out(2)**

Porcentaje de Aprendizaje	Efectividad Obtenida
70	89.55
71	89.6
72	89.75
73	90.2
74	90.55
75	91.05
76	91.15
77	91.5
78	91.65
79	92
80	93.05

**Grafica 4.2 Prueba Hold Out(2)**



La eficiencia del clasificador cuando se usan porcentajes de aprendizaje entre 70% y 80% se mantiene entre 89% y 93%, además es creciente, entre más datos se tengan en el conjunto de aprendizaje aumenta la eficiencia; sin embargo, un conjunto de



aprendizaje del 80% sería demasiado grande, es por eso que se ha optado por utilizar un conjunto de aprendizaje de 73% con el cual se tiene una efectividad de 90.2%

## 4.5 Prueba K-Fold Cross Validation

En ésta prueba el conjunto se dividió en 10 partes iguales, algunos de éstos conjuntos tenían en común dos o tres registros para poder completar los conjuntos de igual tamaño.

Cada conjunto tuvo 20 registros, 15 de la clase No recurrente y 5 de la clase de recurrente.

Se tomó el primer subconjunto  $K1$  para ser probado y, se tomó como conjunto de aprendizaje el resto de los subconjuntos, después se tomó el conjunto  $K2$  para prueba y el resto para aprendizaje así sucesivamente hasta probar con los 10 subconjuntos.

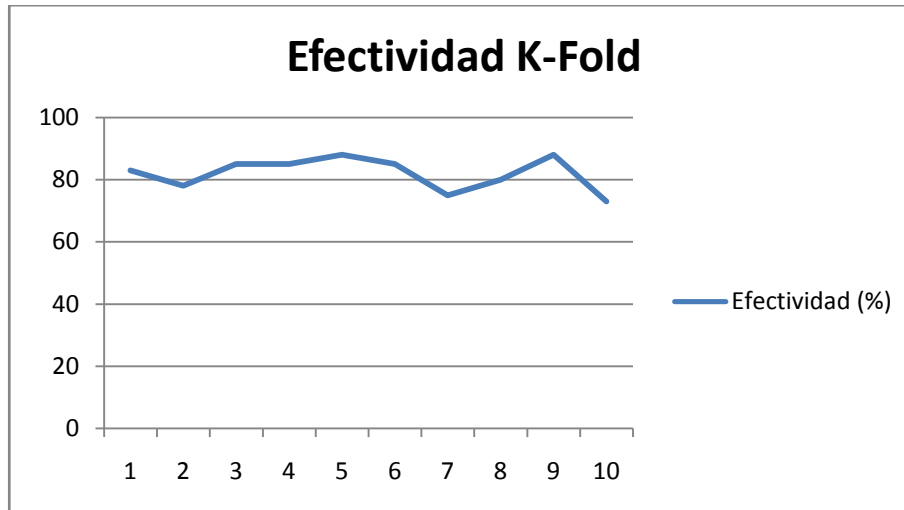
Los resultados fueron los siguientes:

**Tabla 4.3:** Resultados K-Fold Cross Validation

Promedio Total	
Prueba	Eficiencia
K1	83
K2	78
K3	85
K4	85
K5	88
K6	85
K7	75
K8	80
K9	88
K10	73
Promedio	82



**Gráfica 4.3:** Resultados K-Fold Cross Validation



En el caso de la prueba *K-Fold Cross Validation* se observa que se comporta de manera diferente con cada uno de los subconjuntos, esto se debe al hecho de que todos los conjuntos son diferentes. El conjunto de prueba es del mismo tamaño en cada uno de los casos pero podría ser posible que en un subconjunto queden registros con muy pocas coincidencias con los demás y de ésta manera ese subconjunto en especial mantendría resultados de efectividad bajos; en este caso el subconjunto K10 fue el que obtuvo la efectividad más baja con 73% y el mejor resultado se obtuvo con los subconjuntos K5 y K9 con una efectividad de 88%.

El promedio de efectividad en la prueba *K-Fold Cross Validation* fue de 82%, con una diferencia de 15% entre el mejor y el peor de los caso lo cual es un margen muy amplio.

#### 4.6 Take One Out

En esta prueba se tomó uno a uno cada registro y se usó el resto de la base de datos como prueba, los resultados obtenidos fueron los siguientes:

**Tabla 4.4:** Take One Out

TakeOneOut	
Clase	Errores
Recurrente	9
No-Recurrente	8
<b>Efectividad</b>	<b>92%</b>



En la tabla se observa que al probar con todos los registros, el clasificador cometió 9 errores en la clase Recurrente y 8 errores en la clase no recurrente y clasificó correctamente 177 registros, es decir; su eficiencia en ésta prueba fue del 92%.

Es importante mencionar que en ésta prueba si bien los registros que constituyen el conjunto de prueba y aprendizaje cambian, se mantiene la proporción de un registro de prueba por 193 de aprendizaje.

### 4.7 Resultados Generales

Como resultados generales se presentan los porcentajes de efectividad del mejor subconjunto  $K$  de la prueba *K-Fold Cross Validation*, el porcentaje de efectividad de la prueba *Hold Out* con 73% de conjunto de aprendizaje y el porcentaje de Efectividad de la prueba *Take One Out*

**Tabla 4.5: Resultados Generales**

Resultados Generales	
Prueba	Efectividad
Mejor K-Fold	88%
HoldOut	90.2%
TakeOneOut	92%

El mejor resultado podría parecer el 92% de *Take One Out*, sin embargo; por las características de la prueba el conjunto de aprendizaje fue excesivamente grande comparado con el conjunto de prueba.

El resultado que se reporta como final es la efectividad de 90.2% producto de la prueba *Hold Out*, en la cual se encuentran distribuidos de mejor manera los conjuntos de aprendizaje y prueba, siendo el conjunto de prueba del 27% y el conjunto de aprendizaje de 73%.



## CAPITULO 5 Conclusiones y Trabajo Futuro

### 5.1 Conclusiones

Como ya se mencionó anteriormente, en la prueba de *Hold Out* los porcentajes del conjunto de aprendizaje más utilizados para implementar un clasificador se encuentran entre el 70% y el 80%, en nuestro caso tomamos el **73%**, con lo cual se obtiene una efectividad de **90.2%**.

De la tabla 4.14 se toma el resultado de *Hold Out* debido que es la prueba más común y, con la efectividad de **90.2%** el clasificador del presente trabajo se ubica por encima de los clasificadores:

- 2007 Chih-LinChia, W. Nick Street b, William H. Wolberg, Application of Artificial Neural Network Based Survival Analysis on Two Breast Cancer Datasets, Universidad de Iowa Estados Unidos. Efectividad de: Clase No-Recurrente: 63.49%, Clase Recurrente: 47.92%.
- 2006 I. Maglogiannis, K. Karpouzis y M. Bramer, Automated Diagnosis of Breast Cancer Based on the WDBC Patient Data, New York. Efectividad: 90%.

Cabe señalar que se desconocen los métodos que se usaron para realizar las pruebas con las cuales se obtuvo la efectividad que reportan los anteriores trabajos.

Los trabajos que se mencionan (tanto el de Iowa como el de Nueva York), utilizan enfoques distintos al asociativo, en el caso del Clasificador de la Universidad Iowa se usaron redes Neuronales y en el caso del Clasificador desarrollado en Nueva York se usó un clasificador basado en un nuevo algoritmo llamado *Support Vector Machine* con expansión lineal de datos.

De acuerdo a lo anterior concluimos que el modelo asociativo Alfa-Beta actúa de una forma muy eficiente como clasificador, superando incluso a métodos que fueron creados expresamente para clasificar patrones como es el caso del proyecto diseñado en Nueva York, esto resulta importante ya que se demuestra la amplia gama de posibilidades que se tienen para aprovechar el modelo Asociativo Alfa-Beta.

Hablando específicamente del Clasificador Gamma Generalizado, se ha demostrado su alto rendimiento, ya que anteriormente (en la tesis de su autor) se probó su eficiencia con Bases de Datos distintas a la usada en este trabajo, obteniendo con todas resultados sobresalientes, este caso no fue la excepción y se ha logrado un muy buen resultado.



## 5.2 Dificultades Durante las Pruebas

La Base de Datos *Wisconsin Breast Cancer* no fue la primera opción para probar el algoritmo Gamma Generalizado como Clasificador de Recurrencia de Cáncer.

La primer Base de Datos de Cáncer con la que se probó el Algoritmo fue *Breast Cancer Data Set*. [26]

La *Breast Cancer Data Set* cuenta con 286 registros, 85 Recurrentes y 201 No Recurrentes, cada registro con nueve rasgos, estos rasgos eran en su mayoría categóricos (texto).

Para poder utilizar ésta base de datos se tuvieron que escalar los datos de texto a numéricos; sin embargo al probar la base de datos la efectividad fue demasiado baja, obteniendo por resultado 80.1%.

Para poder utilizar ésta base de datos resultaba necesario utilizar un método de *Feature Selection* lo cual cae fuera de los límites de éste trabajo, es por eso que se optó por buscar otra Base de Datos.

## 5.3 Trabajo Futuro

Buscar el método de *Feature Selection* adecuado que permita implementar la Base de Datos con la que se tuvo problemas y comparar la efectividad con los resultados obtenidos en el presente trabajo.

Comprobar si utilizando un método de *Feature Selection* se puede mejorar la efectividad del Clasificador desarrollado en éste Proyecto.

Implementar una interfaz para el programa que sea fácil de usar para usuarios no especializados.

Realizar una Base Datos con pacientes mexicanas.



## Anexo 1: Código Johnson-Möbius Modificado

```
//arreglo global que sirve para mantener momentaneamente la conversion Möbius
codificacion = new int[rasgos][em]; //numero de rasgos=33 em= rasgo mas grande

//funcion que realiza la conversion
public void Mobius(int k)
{
    for (int i = 0; i < rasgos; i++)
    {
        zerosrasgo[i] = em[,i] - datos[k,i];
        unosrasgo[i] = datos[k,i];
        for (int j = 0; j < zerosrasgo[i]; j++)
            codificacion[i][j] = 0;
        for (int j = zerosrasgo[i]; j < zerosrasgo[i] + unosrasgo[i]; j++)
            codificacion[i][j] = 1;
    }
}

//segmento del código en el que se guardan los registros convertidos
for (int i = 0; i < registros; i++)
    for (int j = 0; j < rasgos; j++)
        for (int k = 0; k < registros; k++)
        {
            Mobius(i);
            Codificados[i,j,k]=codificacion[j,k]; //arreglo de tres dimensiones
        }
```





## Anexo 2: Código Gamma Generalizado

En este anexo se presenta el fragmento de código donde se encuentran las coincidencias que hay entre los registros de prueba y las clases del conjunto de aprendizaje, antes de éste código se seleccionan los conjuntos de prueba y aprendizaje, y una vez establecida la clase de un registro se deben realizar otras operaciones a fin de presentar los resultados de una forma más clara.

El siguiente fragmento de código es el núcleo del Clasificador Gamma Generalizado.

//función gamma para la clase Recurrente

```
private int gammarec (int j)
{
  int aux=0;
  for(int i=0;i<em;i++)
  {
    //3 if que realizan las operaciones del operador alfa
    if(pr[i]==recurr[j],i)
      gamma[i]=1;
    if(pr[i]<recurr[j],i)
      gamma[i]=0;
    if(pr[i]>recurr[j],i)
      gamma[i]=2;
    gamma[i]=gamma[i]%2;//operacion modulo
  }
  for(int i=0;i<em;i++)
    aux+=gamma[i];
  if((em-aux)>tetha)
    return (0);
  if((em-aux)<=tetha)
    return (1);
}
```

//funcion gamma para la clase NO-Recurrente

```
private int gammanorec (int j)
{
  int aux=0;
  for(int i=0;i<em;i++)
  {
    //3 if que realizan las operaciones del operador alfa
```



```
    if(pr[i]==norecurr[j,i])
        gamma[i]=1;
    if(pr[i]<norecurr[j,i])
        gamma[i]=0;
    if(pr[i]>norecurr[j,i])
        gamma[i]=2;
        gamma[i]=gamma[i]%2;
    }
for(int i=0;i<em;i++)
    aux+=gamma[i];
if((em-aux)>tetha)
    return (0);
if((em-aux)<=tetha)
    return (1);
}

//fragmento del código donde se llama el código para clasificar

for(int k=0;k<lmpr;k++)
{
    tetha=0;
    bandera=0;

    for (int d=0;d<em;d++)
    {
        pr[d]=prueba[k,d];
    }
    while (bandera==0)//este ciclo se ejecuta hasta que se clasifica el registro
    {
        cl1=0;
        cl2=0;
        for(int j=0;j<lmrec;j++)
        {
            cl1+=gammarec(j);//se guardaran todas las coincidencias con la clase recurrente
        }

        for(int j=0;j<lmnorec;j++)
        {
            cl2+=gammanorec(j);//se guardaran todas las coincidencias con la clase NO-
recurrente
        }

        if(cl1<cl2)
        {
            clase[k]=2;//Clase no Recurrente
        }
    }
}
```



```
        bandera=1;
    }

    if(c1>c2)//Clase Recurrente
    {
        clase[k]=1;
        bandera=1;
    }

    if(c1==c2&&tetha<epm) //si no se ha clasificado pero tetha es menor al criterio
de paro, //tetha aumenta
        tetha++;
    else
        if(c1==c2&&tetha>=epm)//si se rebasa el criterio de paro y no se ha clasificado,
se //asigna recurrente
        {
            clase[k]=1;
            bandera=1;
        }
    }
}
```



## GLOSARIO

**BIT:** *Binary Digit* o Dígito Binario; es la unidad mínima de información, puede tener dos estados "0" o "1".

**Canales linfáticos:** Vinculan la linfa (a nivel intestinal), Se hallan agrupados en dos sectores:

### *Sector periférico absorbente*

Se inician en el canal linfático, formando parte de la unidad micro circulatoria, de allí forman los canales pre linfáticos, ambos poseen penetraciones intercelulares que permiten el paso de macromoléculas, y los capilares linfáticos.

### *Sector de conducción*

A partir del anterior punto el sistema linfático se convierte en recolectores, colectores y vasos linfáticos. Todos estos vasos poseen vaso motricidad. Se reúnen en dos sistemas colectores:

- Conducto linfático derecho
- Recoge parte de la linfa del tronco y de la cabeza
- Canal torácico o conducto torácico izquierdo
- Drena la mayor parte del cuerpo
- Ambos confluyen en la vena subclavia a nivel de la yugular.

**Carcinoma:** Cáncer de los tejidos epiteliales con tendencia a difundirse.

**Células Epiteliales:** Células que recubren las superficies interna y externa del cuerpo, formando masas o capas celulares (epitelio). Las células epiteliales que revisten la piel, boca, nariz y el canal anal derivan del ectodermo; las que revisten el sistema respiratorio y el sistema digestivo derivan del endodermo; las otras (sistema cardiovascular y sistema linfático) del mesodermo.

**Conjunto fundamental:** El conjunto fundamental o también llamado espacio muestral es el conjunto formado por todos los resultados posibles de un experimento o fenómeno.

**Ganglios linfáticos:** Parte del sistema linfático; órganos en forma de frijol que se encuentran debajo de la axila, en la ingle, el cuello y en el abdomen, y que actúan como filtro del líquido linfático a medida que éste los atraviesa.



**Gen:** Unidad hereditaria básica que controla cada carácter en los seres vivos. A nivel molecular corresponde a una sección de ADN, que contiene información para la síntesis de una cadena proteínica.

**Lobulillos:** Son las glándulas productoras de leche en los senos de la mujer.

**Metástasis:** Foco canceroso aparecido secundariamente a distancia del tumor origina.

**Micro calcificaciones:** Son depósitos de calcio en el espesor de la glándula mamaria. Se aprecian en la mamografía como pequeños puntitos blanquecinos.

**Neoplasia:** Formación nueva de tejido. Proceso o tumor maligno o benigno.

**Órgano:** una parte del cuerpo constituida por tejidos y células, que le permite realizar una función específica.

**Patrón:** Descriptor de un motivo. Un patrón es una representación abreviada de la secuencia con censo de un motivo. Los patrones, también conocidos como expresiones regulares, se han utilizado para la identificación de motivos en el análisis de secuencias.

**Proteínas:** Macromoléculas formadas por cientos o miles de aminoácidos, encargadas de diversas funciones en los seres vivos, como transportadores, catalizadores (enzimas), estructuras, etcétera.

**Tejido adiposo:** tejido biológico compuesto principalmente de células grasas.

**Tejido:** Agrupación de células en el cuerpo que tienen un origen común, son semejantes entre sí, y cumplen funciones específicas en conjunto.

**Tumor:** Masa celular anormal que resulta de la excesiva multiplicación celular, variable en tamaño, forma y color.

**Vasos sanguíneos:** son las venas, arterias y capilares a través de los cuales circula la sangre. Los vasos sanguíneos pueden ser donados y trasplantados.

**Estadio:** Momento, periodo o estado que forma parte de una serie o de un proceso, etapa, fase.



## Referencias

- [1] <http://www.lasnoticiasmexico.com/59482.html>
- [2] <http://mx.news.yahoo.com/s/26022009/7/mexico-crece-incidencia-cancer-mama-poblacion.html>
- [3] Torella M. & Villa S., 1998, *Bases Genéticas del Cáncer*, Fondo de Cultura Económica, México, D.F.
- [4] <http://www.salud.com/cancer-de-mama.html>
- [5] <http://www.infocancer.org.mx/que-es-el-cancer-de-mama>
- [6] <http://www.infocancer.org.mx/estructura-mama>
- [7] <http://www.infocancer.org.mx/tipos-de-cancer-de-mama>
- [8] <http://www.infocancer.org.mx/factores-de-riesgo-cancer-de-mama>
- [9] <http://www.infocancer.org.mx/sintomas-cancer-de-mama>
- [10] <http://www.infocancer.org.mx/deteccion-de-cancer-de-mama>
- [11] <http://www.infocancer.org.mx/tratamiento-cirugia-cancer-de-mama>
- [12] <http://www.infocancer.org.mx/tratamiento-terapia-hormonal-cancer-de-mama>
- [13] <http://www.infocancer.org.mx/tratamiento-quimioterapia-cancer-de-mama>
- [14] [http://www.esperanzayvida.com/conozca\\_cancer\\_3.asp](http://www.esperanzayvida.com/conozca_cancer_3.asp)
- [15] <http://dollar.biz.uiowa.edu/~street/research/AMIA2007.pdf>
- [16] I. Maglogiannis, K. Karpouzis y M. Bramer "Automated Diagnosis of Breast Cancer Based on the WDBC Patient Data" *Artificial Intelligence Applications and Innovations*, Editorial Springer 2006 New York USA, paginas 501-504.
- [17] [http://ieeexplore.ieee.org/xpl/freeabs\\_all.jsp?tp=&arnumber=1222259&isnumber=27452](http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?tp=&arnumber=1222259&isnumber=27452)
- [18] <http://www.springerlink.com/content/q71j44897k55062v/>
- [19] <http://ryanpotter.com/conference1.pdf>



[20] Burke, H.B., Goodman P.H., Rosen, D.B, (1997). Artificial Neural Networks Improve the Accuracy of Cancer Survival Prediction. *Cancer* Vol. 79, pp. 857–62.

[21] Street, W. N., Kim, Y. (2001). A streaming ensemble algorithm (SEA) for large-scale classification. In *Proceedings of the Seventh ACM SIGKDD international Conference on Knowledge Discovery and Data Mining (San Francisco, California, August 26 - 29, 2001)*. KDD '01. ACM, New York, NY, pp. 377-382.

[22] [http://148.204.63.175/MemAs\\_ABMorf.html](http://148.204.63.175/MemAs_ABMorf.html)

[23] Acevedo-Mosqueda, M.E.: *Memorias Asociativas Bidireccionales Alfa-Beta*. Tesis doctoral, Centro de Investigación en Computación, IPN, México (2006).

[24] López-Yáñez, I.: *Clasificador Automático de Alto Rendimiento*. Tesis Maestría, Centro de Investigación en Computación, IPN, México (2007).

[25] <http://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/wpbc.names>

[26] <http://archive.ics.uci.edu/ml/datasets/Breast+Cancer>