UNIVERSIDAD CENTRAL "MARTA ABREU" DE LAS VILLAS

VERITATE SOLA NOBIS IMPONETUR VIRILIS TOGA. 1948

Facultad de Ingeniería Eléctrica.

**Centro de Estudios de Electrónica**

**y Tecnologías de la Información**

# TRABAJO DE DIPLOMA

## Clasificación de células de la prueba de Papanicolaou mediante el uso de Redes Neuronales Multicapa

**Yasel Batista Pelaez**

**Tutor:**

**Maykel Orozco Monteagudo**

**Santa Clara**

**2012**

# Universidad Central “Marta Abreu” de Las Villas

## Facultad de Ingeniería Eléctrica

### Centro de Estudios de Electrónica y Tecnologías de la Información



# TRABAJO DE DIPLOMA

## Título: Clasificación de células de la prueba de Papanicolaou mediante el uso de Redes Neuronales Multicapa


**Autor: Yasel Batista Pelaez**

**Tutor: Maykel Orozco Monteagudo**

email:morozco@uclv.edu.cu


**Santa Clara**

**2012**

Hago constar que el presente trabajo de diploma fue realizado en la Universidad Central "Marta Abreu" de Las Villas como parte de la culminación de estudios de la especialidad de Ingeniería Biomédica, autorizando a que el mismo sea utilizado por la Institución, para los fines que estime conveniente, tanto de forma parcial como total y que además no podrá ser presentado en eventos, ni publicados sin autorización de la Universidad.

_____

Firma del Autor

Los abajo firmantes certificamos que el presente trabajo ha sido realizado según acuerdo de la dirección de nuestro centro y el mismo cumple con los requisitos que debe tener un trabajo de esta envergadura referido a la temática señalada.

_____ _____

Firma del Autor

Firma del Jefe de Departamento donde se defiende el trabajo

_____

Firma del Responsable de

Información Científico-Técnica

# PENSAMIENTO

**Cuanto más talento tiene un hombre,**

**más se inclina a creer en el ajeno.**

**Blaise Pascal**

# DEDICATORIA

**A mis padres**

# AGRADECIMIENTOS

Agradezco a todas las personas que de una forma u otra contribuyeron al desarrollo de este trabajo, en especial a mis familiares.

A mis amigos más cercanos que siempre me apoyaron y a mi tutor por su dedicación y esfuerzo.

# TAREA TÉCNICA

- Aprender los fundamentos de:
    - Procesamiento de imágenes.
    - Aprendizaje automatizado.
- Realizar un estudio específico sobre las siguientes técnicas de aprendizaje automatizado:
    - Redes neuronales artificiales.
    - Algoritmos genéticos.
- Realizar una revisión bibliográfica relacionada con la segmentación de imágenes de la prueba de Papanicolaou y la clasificación de las regiones.
- Aprender a usar la toolbox de Procesamiento de Imágenes de Matlab y el software HGANN-Trainer v. 1.0.
- Realizar el entrenamiento de las Redes Neuronales Artificiales usando Algoritmos Genéticos.
- Medir la calidad de los clasificadores obtenidos.
- Comparar los resultados obtenidos con los reportados en la literatura.

# RESUMEN

El cáncer cervical es una de las enfermedades que más afecta a las mujeres en todo el mundo. La detección temprana de células pre-malignas, así como su clasificación correcta, aumenta la posibilidad de vida de las mujeres afectadas.

En el presente trabajo se plantea la clasificación de este tipo de células obtenidas a través de la prueba de Papanicolaou usando clasificadores basados en Redes Neuronales Multicapa entrenados mediante el uso de Algoritmos Genéticos. Para el proceso de evaluación se usó la base de casos anotados de imágenes de la prueba de Papanicolaou del Hospital Universitario de Herlev, Dinamarca. Los resultados obtenidos son aceptables de acuerdo al desempeño de otros clasificadores reportados en la literatura para este mismo problema.

Tabla de contenido

# Introducción

El cerebro es una de las cumbres de la evolución biológica debido a su gran capacidad como procesador de información. Una de sus características principales es su capacidad para procesar a gran velocidad grandes cantidades de información procedentes de los sentidos, combinarla o compararla con la información almacenada y dar respuestas adecuadas [1]. Además de esto, destaca por su capacidad para aprender a representar la información necesaria para desarrollar tales habilidades, sin instrucciones explícitas para ello. Los científicos llevan años estudiándolo y se han desarrollado algunos modelos matemáticos que tratan de simular su comportamiento. Estos modelos se han basado sobre los estudios de las características esenciales de las neuronas y sus conexiones [2]. Aunque estos modelos no son más que aproximaciones muy lejanas de las neuronas biológicas, son muy interesantes por su capacidad de aprender y asociar patrones parecidos, lo que nos permite afrontar problemas de difícil solución con la programación tradicional. Estos modelos se han implementado en computadoras y equipos especializados para ser simulados [2].

Una red neuronal, según Freman y Skapura [3], es un sistema de procesadores paralelos conectados entre sí en forma de grafo dirigido. Esquemáticamente, cada elemento de procesamiento (neuronas) de la red se representa como un nodo. Las conexiones establecen una estructura jerárquica, que tratando de emular la fisiología del cerebro, busca nuevos modelos de procesamiento para solucionar problemas concretos del mundo real. Para que las Redes Neuronales Artificiales (RNA) se utilicen como una herramienta útil para resolver un problema dado, requieren de un proceso de aprendizaje, o sea, un entrenamiento que viene dado por la selección correcta de los parámetros libres de la red. En este sentido, el algoritmo de retropropagación del error (*backpropagation*, en inglés) [4, 5] es el más difundido en la comunidad científica, aunque el mismo presenta limitaciones que dan lugar al uso de otros métodos de entrenamiento.

Los algoritmos evolutivos [6, 7] presentan grandes ventajas en la solución de problemas complicados y con restricciones. Una extensión novedosa de los mismos son los Algoritmos Genéticos (AG), debido a que proporcionan grandes ventajas en la búsqueda de soluciones a problemas, donde la determinación de la estructura (desconocida a

priori) es de vital importancia. A todo esto se suma, el aumento de la capacidad de cómputo de las máquinas modernas y sus bajos costos, así como el hecho de que soluciones necesarias para problemas específicos puedan ser obtenidas automáticamente, aún cuando estas sean alcanzadas a partir de un esquema de cálculo de alto costo computacional.

Por otro lado, la detección temprana de presencia de células cancerosas en personas sanas, es de vital importancia dentro de la medicina preventiva. El cáncer cervical o de cérvix uterino es una de las enfermedades oncológicas más frecuentes en mujeres, por lo que una determinación a tiempo de células pre-cancerosas, aumenta la posibilidad de vida a las pacientes. Aunque el problema de clasificación automática de células cancerosas, ha sido abordado en la literatura [8-11], la clasificación de este tipo de célula usando RNA entrenadas con AG es de gran importancia ya que otros métodos reportados dejan margen para mejoras.

Por lo antes expuesto se propusieron los siguientes objetivos:

**Objetivo General:**

Construir clasificadores basados en Redes Neuronales Multicapa (RNM) para la clasificación de células de la prueba de Papanicolaou usando algoritmos genéticos.

**Objetivos Específicos:**

- ✓ Seleccionar los parámetros de entrada del algoritmo genético para la construcción de clasificadores basados en RNM:
    - Probabilidades de cruzamiento y mutación.
    - Funciones de transición.
    - Tipo de remplazo.
- ✓ Comparar los resultados obtenidos con otros métodos reportados en la literatura para la clasificación de este tipo de células.

El presente trabajo está estructurado con introducción, tres capítulos, conclusiones y recomendaciones, referencias bibliográficas y anexos. En el capítulo 1 se describen los fundamentos de las redes neuronales artificiales y los algoritmos genéticos. En el capítulo 2 se describe el método de entrenamiento de las redes neuronales artificiales y la base de casos de imágenes de la prueba de Papanicolaou utilizada. En el capítulo 3 se muestran los resultados obtenidos así como una comparación con otros métodos

reportados en la literatura científica. Los anexos presentan algunos contenidos que no se explican con profundidad en el trabajo.

# Capítulo 1. Fundamentos de las Redes neuronales artificiales y los algoritmos genéticos

## 1.1 La neurona real

Del cerebro, visto a alto nivel y simplificando su estructura, se podría decir que es un conjunto de millones de células especiales, llamadas neuronas, interconectadas entre ellas por sinapsis [12] (Figura 1.1). Las sinapsis son las zonas donde dos neuronas se conectan. La parte de la célula que está especializada en estas conexiones son las dendritas y las ramificaciones del axón.

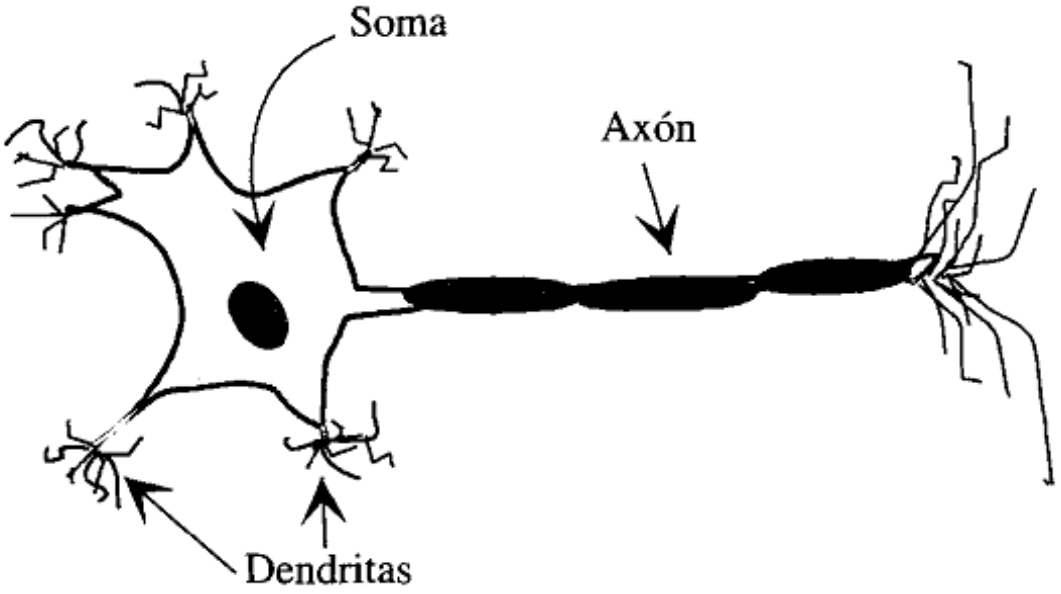



**Figura 1.1 Estructura de una neurona biológica.**

Cada neurona desarrolla impulsos eléctricos que se transmiten a lo largo de una parte de la célula muy alargada llamada axón, al final del axón este se ramifica en ramificaciones axonales que van a conectarse con otras neuronas por sus dendritas.

El conjunto de elementos que hay entre la ramificación axonal y la dendrita forman la sinapsis que regula la transmisión del impulso eléctrico mediante unos elementos bioquímicos, los neurotransmisores.

Los neurotransmisores liberados en las sinapsis al final de las ramificaciones axonales pueden tener un efecto negativo o positivo sobre la transmisión del impulso eléctrico de la neurona que los recibe en sus dendritas. La neurona receptora recibe varias

señales desde las distintas sinapsis y las combinan consiguiendo un cierto nivel de estimulación o potencial.

En función de este nivel de activación, la neurona emite señales eléctricas con una intensidad determinada mediante impulsos con una frecuencia llamada tasa de disparo que se transmiten por el axón.

Si se considera que la información en el cerebro esta codificada como impulsos eléctricos que se transmiten entre las neuronas y que los impulsos se ven modificados básicamente en las sinapsis cuando las señales son químicas, se puede intuir que la codificación del aprendizaje estará en las sinapsis y en la forma que dejan pasar o inhiben las señales segregando neurotransmisores.

## 1.2 Neurona artificial

Las neuronas artificiales [13] son modelos que tratan de simular el comportamiento de las neuronas biológicas. Cada neurona se representa como una unidad de proceso que forma parte de una entidad mayor, la red neuronal.

Dicha unidad de proceso consta de una serie de entradas $X_i$, que equivalen a las dendritas de donde reciben la estimulación, ponderadas por unos pesos $W_i$, que representan como los impulsos entrantes son evaluados y se combinan con la función de red que dará el nivel de potencial de la neurona.

La salida de la función de red es evaluada en la función de activación que da lugar a la salida de la unidad de proceso.

Por las entradas $X_i$ llegan unos valores que pueden ser enteros, reales o binarios. Estos valores equivalen a las señales que enviarían otras neuronas a la nuestra a través de las dendritas.

Los pesos que hay en las sinapsis *Wi*, equivaldrían en la neurona biológica a los mecanismos que existen en las sinapsis para transmitir la señal. De forma que la unión de estos valores (*Xi* y *Wi*) equivalen a las señales químicas inhibitorias y excitadoras que se dan en las sinapsis y que inducen a la neurona a cambiar su comportamiento.

Estos valores son la entrada de la función de ponderación o red que convierte estos valores en uno solo llamado típicamente el potencial que en la neurona biológica equivaldría al total de las señales que le llegan a la neurona por sus dendritas. La

función de ponderación suele ser la suma ponderada de las entradas y los pesos sinápticos.

La salida de la función de ponderación llega a la función de activación que transforma este valor en otro en el dominio que trabajen las salidas de las neuronas (Figura 1.2).

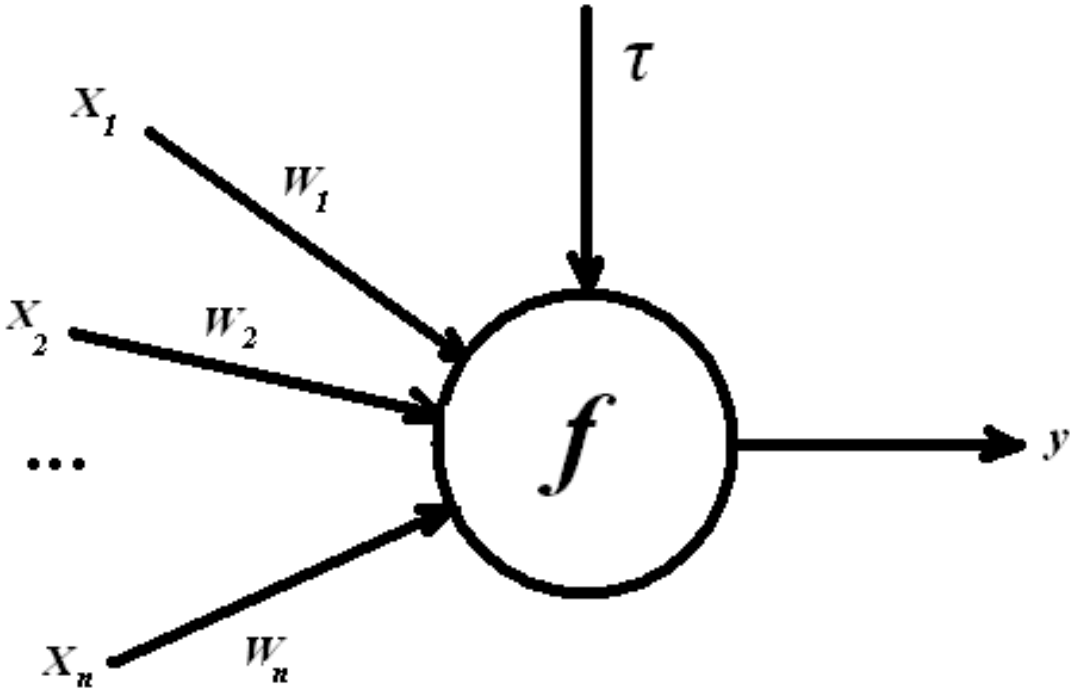



**Figura 1.2 Esquema de una neurona simple.**

La salida de una neurona es determinada como:

$$y = f\left(\sum_{i=1}^{n} W_i X_i + \tau\right) \tag{1.1}$$

donde $X_1$, $X_2$,..., $X_n$ son las señales de entrada, $W_1$, $W_2$,..., $W_n$ son los pesos de las conexiones de entrada, $\tau$ es el valor del sesgo de la neurona (*bias*, en inglés) y $f$ es una función de salida definida, que puede ser una tangente hiperbólica, paso unitario, etc. (Anexo 1).

## 1.2 Niveles de Neuronas

La distribución de neuronas dentro de la red se puede realizar formando niveles o capas de un número determinado de neuronas cada una. A este tipo de red neuronal se le conoce como Red Neuronal Multicapa o *Perceptron* Multicapa (MLP, *Multilayer*

*Perceptron* en inglés) (Figura 1.3). Las capas que componen a la red se pueden clasificar en tres tipos:

- De entrada: estas capas reciben la información desde el exterior.

- De Salida: estas envían la información hacia el exterior.

- Ocultas: son capas que solo sirven para procesar información y comunicar otras capas.

Partiendo de esta estructura, las MLP son ampliamente usadas en problemas de clasificación binaria, o sea, donde solo existen dos clases, generalmente catalogadas como positiva y negativa. Para este tipo de problemas, se tendrían tantas neuronas de entrada como rasgos, y una neurona de salida con una función de activación binaria (límite duro, por ejemplo) (Anexo 1).

## 1.3 Aprendizaje

En el contexto de las redes neuronales [13] puede definirse el aprendizaje como el proceso por el que se produce el ajuste de los parámetros libres de la red a partir de un proceso de estimulación por el entorno que rodea la red. El tipo de aprendizaje vendrá determinado por la forma en la que dichos parámetros son adaptados.

El aprendizaje adaptativo [13] es quizás la característica más importante de las redes neuronales, ya que pueden comportarse en función de un entrenamiento con una serie de ejemplos ilustrativos. De esta forma, no es necesario elaborar un modelo a priori, ni establecer funciones probabilísticas. Una red neuronal artificial es adaptativa porque puede modificarse constantemente con el fin de adaptarse a nuevas condiciones de trabajo.

La correcta funcionalidad de la topología de la red está determinada por un algoritmo de aprendizaje, capaz de modificar los parámetros de la red, específicamente sus conexiones y sesgos [13]. El algoritmo de retropropagación del error (*Backpropagation*, BP, en inglés) [4, 5, 13], que utiliza el método del gradiente en su proceso de actualización de los pesos, es el más utilizado por la comunidad científica. Sin embargo, el BP presenta algunas limitaciones y desventajas como las que se dan a continuación:

1. Necesita que la topología esté predeterminada, ignorando así la existencia de topologías menos complejas.

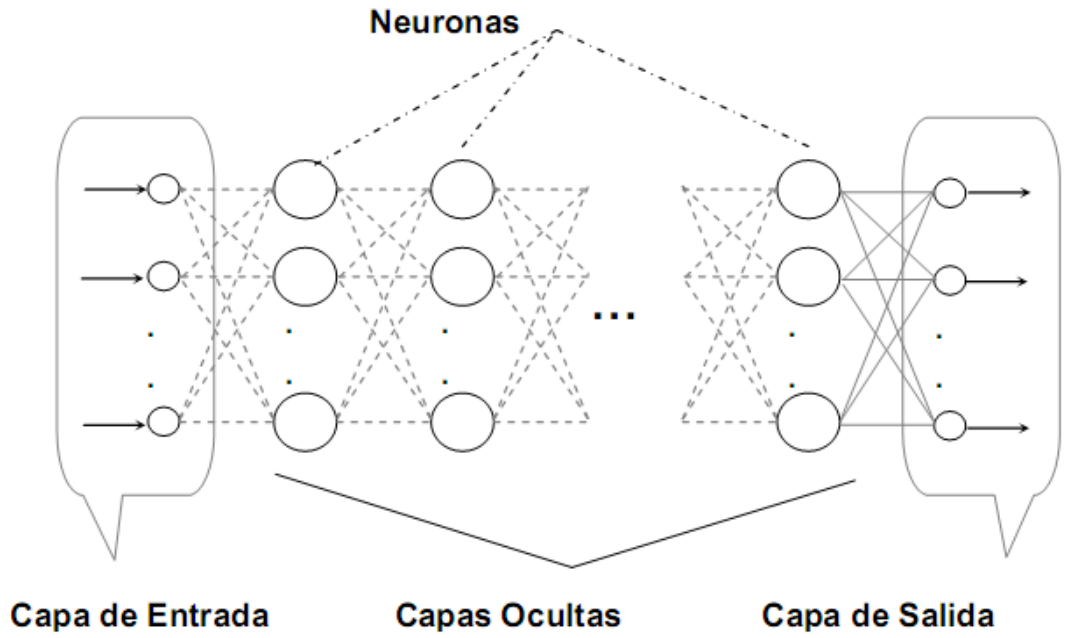



**Figura 1.3 Topología de una red multicapa *feedforward*.**

2. Mientras mayor sea la complejidad de la topología, el desempeño del BP es mucho menor.

3. Generalmente, la convergencia del método es lenta, pero una incorrecta selección de los pesos iniciales (solución inicial), puede hacerla aún más lenta.

4. El algoritmo puede estancarse en mínimos locales y nunca alcanzar el óptimo.

5. Una incorrecta selección de la tasa de aprendizaje, puede provocar que el algoritmo diverja en el proceso de búsqueda del mínimo.

Estas desventajas han permitido el desarrollado de otras heurísticas para el entrenamiento de las redes neuronales.

## 1.4 Algoritmos Genéticos

### *1.4.1 Algoritmos Genéticos Convencionales*

Los principios básicos de los Algoritmos Genéticos Convencionales (*AG*) fueron inicialmente propuestos por Holland [14], aunque muy pronto varios autores desarrollaron novedosos trabajos en esta temática [15-17].

Los *AG* están inspirados en el mecanismo de selección natural, en el cual los individuos con mejores características son los más propensos a ser seleccionados en un ambiente de competencia. En este último, los *AG* utilizan una analogía directa de la evolución natural. Mediante un método de evolución genético, podría ser alcanzada una solución óptima de un problema determinado y esta estaría representada por el individuo ganador de esta competencia genética.

Los *AG* codifican en un individuo o cromosoma, representado por un conjunto de parámetros, una solución potencial de un problema dado. Estos parámetros son los genes de un cromosoma y pueden ser estructurados en forma lineal, como una cadena de valores consecutivos (la codificación binaria es la más utilizada).

Un valor positivo, conocido como aptitud o idoneidad (*fitness value)* es asociado a cada cromosoma, y es utilizado como indicador del grado de aptitud o idoneidad del cromosoma para el problema en cuestión. La aptitud o idoneidad está estrechamente relacionada con la función objetico del problema.

Durante el proceso de evolución, los cromosomas con mayor grado de aptitud tienen la tendencia de producir descendientes de mejor calidad, lo que se traduce como individuos que codifican una mejor solución para el problema en cuestión.

En aplicaciones prácticas de los *AG*, una población inicial de individuos es construida con valores aleatorios. El tamaño de la misma varía de un problema a otro, aunque se han reportado algunas publicaciones referentes a la selección de este tamaño [18]. En cada ciclo genético del proceso evolutivo, una nueva generación es creada a partir de la actual. Esto se realiza cuando un grupo de individuos, usualmente llamados padres o progenitores (*parents)*, se aparean en el *matting pool*, seleccionados por un mecanismo de selección específico.

Los genes correspondientes a los padres son combinados para la producción de sus descendientes utilizando operadores genéticos específicos. Como resultado de este

proceso, se espera que el mejor cromosoma dé lugar a un mayor número de descendientes, emulando el mecanismo de la naturaleza de supervivencia del más fuerte.

El esquema de selección de ruleta (*Roulette Wheel Selection*) [16], es uno de los más utilizados. El ciclo de evolución se repite hasta que se cumpla un criterio de parada especificado. Este criterio de parada pudiera estar determinado por un número de ciclos predeterminado, la detección de una variación (entre individuos entre generaciones) menor que un umbral, o cuando se alcance o supere una idoneidad específica.

El mecanismo de un *AG* es como sigue:

1. El *AG* genera aleatoriamente una población de *n* estructuras (cadenas, cromosomas o individuos)

2. Sobre la población actúan los operadores transformando la población. Una vez completada la acción de los tres operadores se dice que ha transcurrido un ciclo generacional.

3. Luego se repite el paso anterior mientras no se garantice el criterio de parada del *AG*.

Las razones para utilizar *AG* y no otros métodos son varias.

Los métodos conocidos son buenos mientras el problema no es muy complejo. Los *AG* permiten la solución eficiente de funciones extremadamente complejas.

La mayoría de los especialistas en este tema coinciden en que los *AG* pueden resolver las dificultades representadas en los problemas de la vida real que a veces son insolubles por otros métodos.

Para Goldberg [15] el tema central de la investigación en *AG* consiste en la robustez: el balance entre la eficacia y la eficiencia necesaria para sobrevivir en muchos ambientes diferentes. Goldberg destaca las formas en que difieren los *AG* de los sistemas tradicionales [15, 19]:

1. Los *AG* trabajan con una codificación del conjunto de parámetros, no con los parámetros en sí.

2. Los *AG* realizan la búsqueda a partir de una población de puntos, no de un punto simple.

3. Los *AG* sólo utilizan la información de la función objetivo, sin derivadas u otro conocimiento auxiliar.

4. Los *AG* utilizan reglas de transición probabilísticas, no determinísticas.

Goldberg, además, expone algunos motivos por los que los *AG* pueden ser atractivos para el desarrollo de aplicaciones:

1. Pueden resolver problemas difíciles de forma rápida y confiable.

2. Son fáciles de enlazar a simulaciones y modelos existentes.

3. Son extensibles.

4. Son fáciles de hibridizar.

Los *AG* han sido utilizados tradicionalmente en problemas de búsqueda y optimización. Éste es el campo en que más aplicaciones se reportan, habiéndose realizado incluso trabajos con funciones muy complejas como en los trabajos de De Jong en 1975 que se reportan en [15]. También se ha trabajado en la obtención de soluciones a ecuaciones no lineales. Además de esto, los *AG* han sido ampliamente usados en aplicaciones de aprendizaje automatizado. Los sistemas genéticos de aprendizaje automático, según se refiere en [15] han tenido un desarrollo sostenido desde inicios de los años 60, reportándose variedad de trabajos, entre los más importantes se encuentran los relativos a los sistemas clasificadores [15], que son sistemas que aprenden reglas de inferencia simples que guían el comportamiento del sistema.

### *1.4.2 Elementos a tener en cuenta a la hora de aplicar un AG*

***Tamaño de la población****:* Muchos trabajos se han escrito relativos a la influencia del tamaño de la población en la convergencia del *AG* [18]. En principio, es lógico pensar que el trabajo con poblaciones pequeñas corren el riesgo de representar pobremente el espacio de soluciones. Por otro lado, las poblaciones de gran tamaño consumen mayor tiempo computacional. Sobre esta disyuntiva y como un trabajo teórico, se obtuvo que el tamaño óptimo de una población de cadenas binarias, crece exponencialmente con la longitud de la cadena [19].

***Generación de la población inicial***: Se describen fundamentalmente dos vías para obtener la población inicial con que el *AG* comienza su trabajo:

• Generación aleatoria de los individuos: Usualmente la población inicial es generada aleatoriamente.

• Sembrado de individuos: La cuestión de si es importante partir de un conjunto de soluciones suficientemente buenas como población inicial para el *AG*, también ha sido analizada. Se han desarrollado trabajos que han encontrado que el sembrado en una población con soluciones de alta calidad obtenidas de otra técnica heurística, ayuda al *AG* a encontrar mejores soluciones más rápidamente que con un comienzo aleatorio.

***Reemplazamiento****:* Cada iteración de un AG simple crea una subpoblación totalmente nueva de una población existente. Este es llamado *AG* generacional. El *AG* que reemplaza sólo una fracción pequeña de cromosomas a la vez, es llamado *AG* de estado fijo o *"steady-state*".

***Selección***: Existen varias técnicas de selección que han sido ampliamente abordadas en la literatura [20]. Algunos de estos son:

1. Muestreo estocástico con reemplazamiento (el mecanismo de ruleta se incluye dentro de esta técnica).

2. Muestreo estocástico con reemplazamiento parcial.

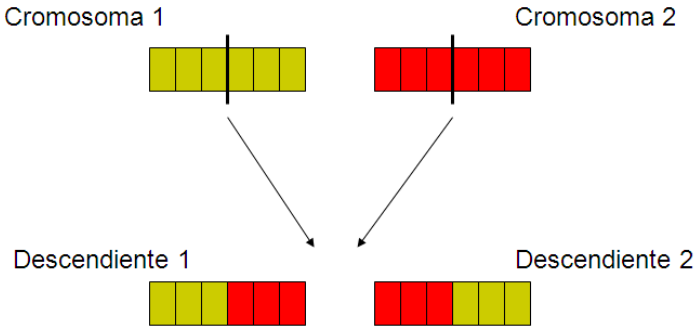
3. Muestreo estocástico universal.

***Cruzamiento***: Para el operador de cruzamiento en los *AG* diferentes técnicas han sido propuestas y analizadas (figura 1.4):

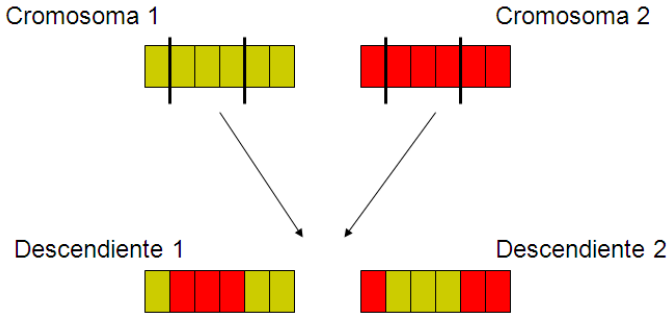
1. *Cruzamiento de uno, dos y múltiples puntos:* El cruzamiento de un punto de cruce es el que utiliza el *AG* simple. Ocurre seleccionando aleatoriamente un punto de cruzamiento y los segmentos que se forman a partir de este punto en los cromosomas padres son intercambiados. En el esquema de cruzamiento de dos puntos, dos puntos son seleccionados aleatoriamente y los segmentos de las cadenas entre ellos intercambiados. El cruzamiento de múltiples puntos trata cada cadena como un anillo de bits dividido por k puntos de cruce en k segmentos. Los segmentos alternados son intercambiados entre el par de cadenas a entrecruzar.

2. *Cruzamiento uniforme:* Es el intercambio de bits entre las cadenas, en vez de segmentos como los casos anteriores. En cada posición de la cadena los bits son probabilísticamente intercambiados con una probabilidad fija.

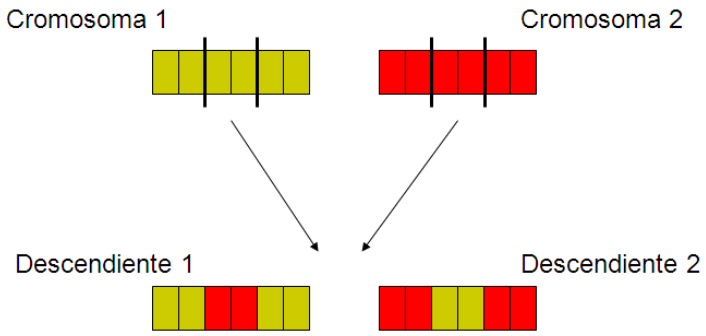
***Mutación****:* Este proceso es muy importante ya que puede ser que un individuo malo tuviera alguna característica muy buena. Cuando este individuo pasa por el proceso de reproducción existe una alta probabilidad de que sea eliminado y por lo tanto se pierda esa característica deseable. La recuperación de esta característica puede ser prácticamente imposible a través de los otros mecanismos genéticos. La mutación de bits se realiza simplemente cambiando el valor de bit que se desea mutar (figura 1.5). La mutación de otros valores depende de las características del problema.




(a)




(b)




(c)

**Figura 1.4 Tipos de cruzamiento. (a) En un punto. (b) En dos puntos. (c) Uniforme.**

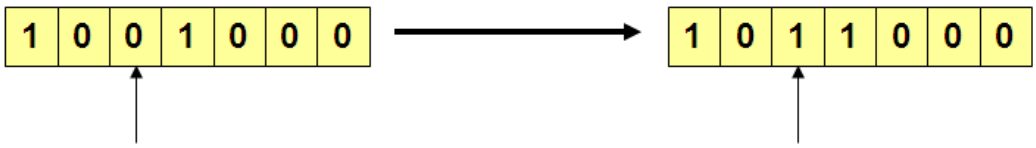



**Figura 1.5 Mutación de un bit.**

### *1.4.3 Algoritmos Genéticos con Codificación Jerárquica*

En [20], se presenta un nuevo método que imita la formulación de la estructura biológica del *ADN*, de manera que se forma una estructura genética jerárquica precisa para formular soluciones de problemas ingenieriles, en la cual los cómputos genéticos básicos permanecen invariables.

A partir de estas características, proponen una nueva formulación jerárquica de los cromosomas, estableciendo una analogía que consiste en:

1. Genes paramétricos (genes estructurales).

2. Genes de control (secuencias reguladoras).

La generalización de esta estructura, se realiza mediante la introducción de múltiples niveles de genes de control en una forma jerárquica, como se indica en la Figura 1.6.

Cada nivel asociado a los genes de control, está conformado por una cadena de bits, la activación de los cuales determina la influencia de estos sobre los niveles inferiores,

sucesivamente hasta llegar a los genes paramétricos. El uso de esta estructura jerárquica, permite que los cromosomas contengan mucha más información que los de los *AG* convencionales; además los genes inactivos siempre están presentes en la misma.

Las operaciones genéticas entre los cromosomas jerárquicos se realiza similarmente que en los *AG* convencionales. Las operaciones de cruzamiento y mutación pueden ser aplicadas similarmente que en los *AG* convencionales, solo que estas deben ser realizadas por nivel. Las probabilidades de cruzamiento y mutación en cada nivel no tienen que ser necesariamente iguales, inclusive en la mayoría de los problemas reales estas difieren debido a que el nivel de importancia de cada nivel es diferente.

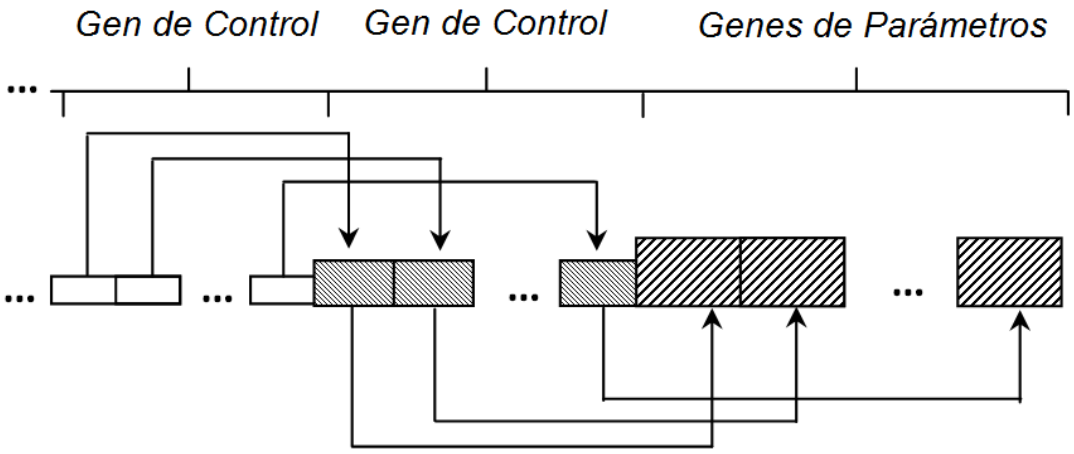



**Figura 1.6 Estructura Jerárquica del Cromosoma.**

El uso de *AG* con Codificación Jerárquica *(CJ)*, cobra particular importancia tanto en la optimización paramétrica como en la determinación de la estructura de la solución, en problemas donde la estructura de la solución es desconocida.

A partir del hecho de que la estructura de los cromosomas, en los *AG* con *CJ* es fija (genes de control y genes paramétricos), y que pueden aparecer parámetros de diferentes tamaños, las modificaciones a los operadores convencionales no son de gran complejidad.

En efecto, los métodos estándares pueden ser aplicados independientes a cada nivel de genes. Debe notarse que los cambios realizados a los niveles de orden superior, pueden determinar cambios en los de orden inferior. Esta última es la razón principal

por la que los *AG Jerárquicos* son no solamente buenos para obtener los parámetros de la solución, sino también su estructura.

# Capítulo 2. Materiales y Métodos

## 2.1 El cáncer cervicouterino

El cáncer cervicouterino es uno de los mayores factores de riesgo que afecta a miles de mujeres cada año. Es uno de los tipos cáncer más común en las mujeres a nivel mundial.

A través del uso de la prueba de Papanicolaou [21] se ha reducido considerablemente el número de mujeres que mueren a causa de esta enfermedad. Esta prueba se usa para descubrir cambios malignos y pre-malignos en la cérvix uterina. La prueba es un examen citológico en el que se toman muestras de células epiteliales en la zona de transición del cuello uterino, en busca de anormalidades celulares que orienten a (y no que diagnostiquen) la presencia de una posible neoplasia (proceso de proliferación anormal de células) de cuello uterino.

La cérvix es la porción más baja del útero, está cubierta por una capa delgada de células llamada epitelio, las células del epitelio tienen dos formas diferentes (escamosa y columnar) y están localizadas en áreas separadas: área escamosa y área columnar (figura 2.1). El área escamosa se localiza al fondo del canal de la cérvix (figura 2.1) y aquí las células se dividen en 4 capas: basal, parabasal, intermedia y superficial, cuando las células maduran se mueven a través de las capas y van cambiando de forma, color y otras características, el citoplasma aumenta su tamaño y el núcleo lo disminuye.

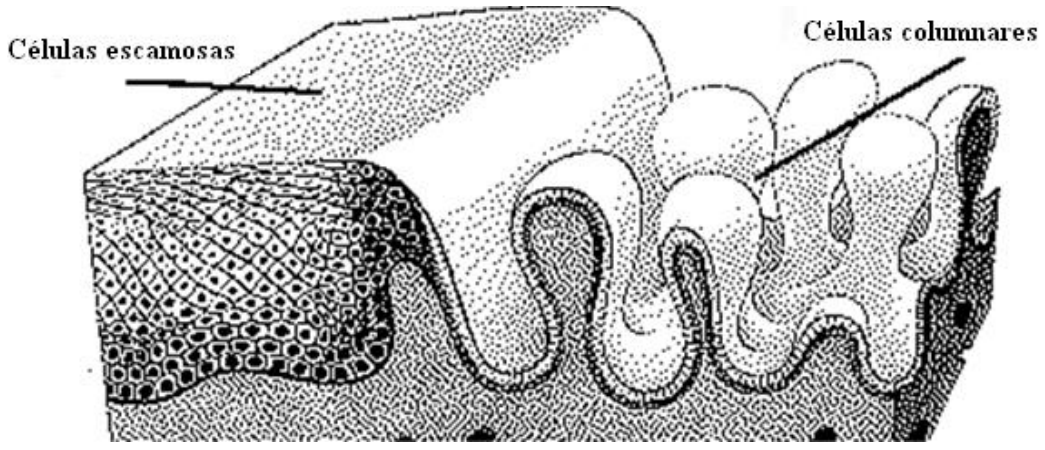



**Figura 2.1 Células escamosas y columnares.**

El área columnar [21, 22] se localiza más arriba, en el canal de la cérvix (figura 2.2). Las células columnares solo existen en la capa basal, tienen forma de columnas como su propio nombre lo indica con un citoplasma oblongo y un núcleo largo localizado en un extremo. La unión entre estas dos áreas se denomina unión escamocolumnar, y puede estar localizada dentro o fuera de la cérvix. La unión escamocolumnar es llamada también como zona de transición porque las células se transforman de una forma a otra.

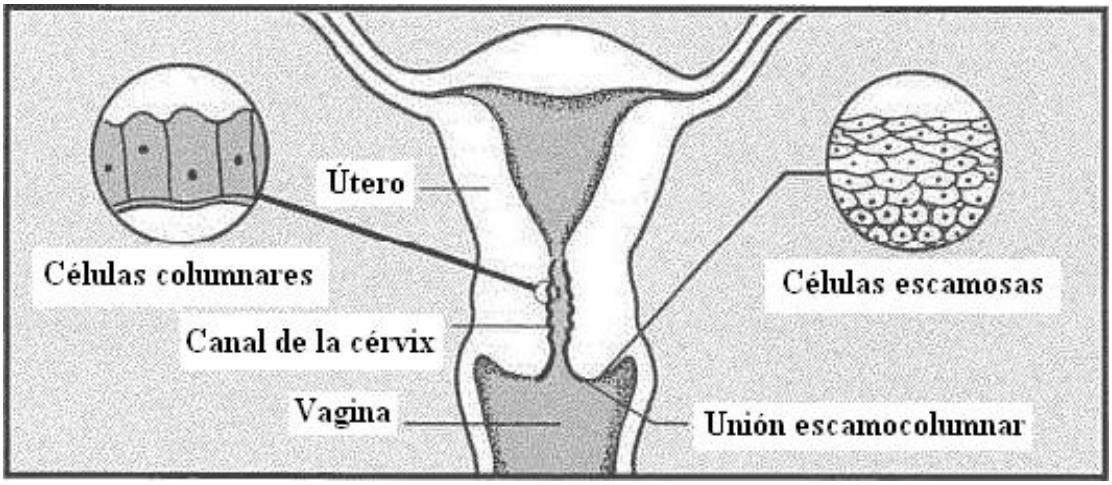



**Figura 2.2 Detalles del útero y localización de las células escamosas y columnares.**

Para la recolección de la muestra citológica se coloca a la paciente en posición ginecológica y se expone correctamente el cuello del útero utilizando un espéculo sin emplear lubricantes, se retira el exceso de secreción o mucus, si fuera necesario, sin tocar la superficie del cuello. El raspado generalmente se realiza en la línea de la unión escamocolumnar donde se unen los dos epitelios. Cuando esta línea coincide con el orificio externo del cuello, se introduce el extremo saliente de la espátula de Ayre en el mismo y se hace girar 360 grados en el sentido de las manecillas del reloj, con cierta presión.

Las muestras utilizadas para esta prueba se toman de tres sitios:

1. Endocérvix, que es el orificio que comunica con el útero;
2. Cérvix, que es la parte más externa del útero, y que comunica directamente con la vagina
3. Tercio superior de la vagina, que es la región que rodea el cuello del útero.

## 2.2 Clasificación de células de la prueba de Papanicolaou

La figura 2.3 muestra el modelo para la clasificación de células de la prueba de Papanicolaou usando redes neuronales multicapa.

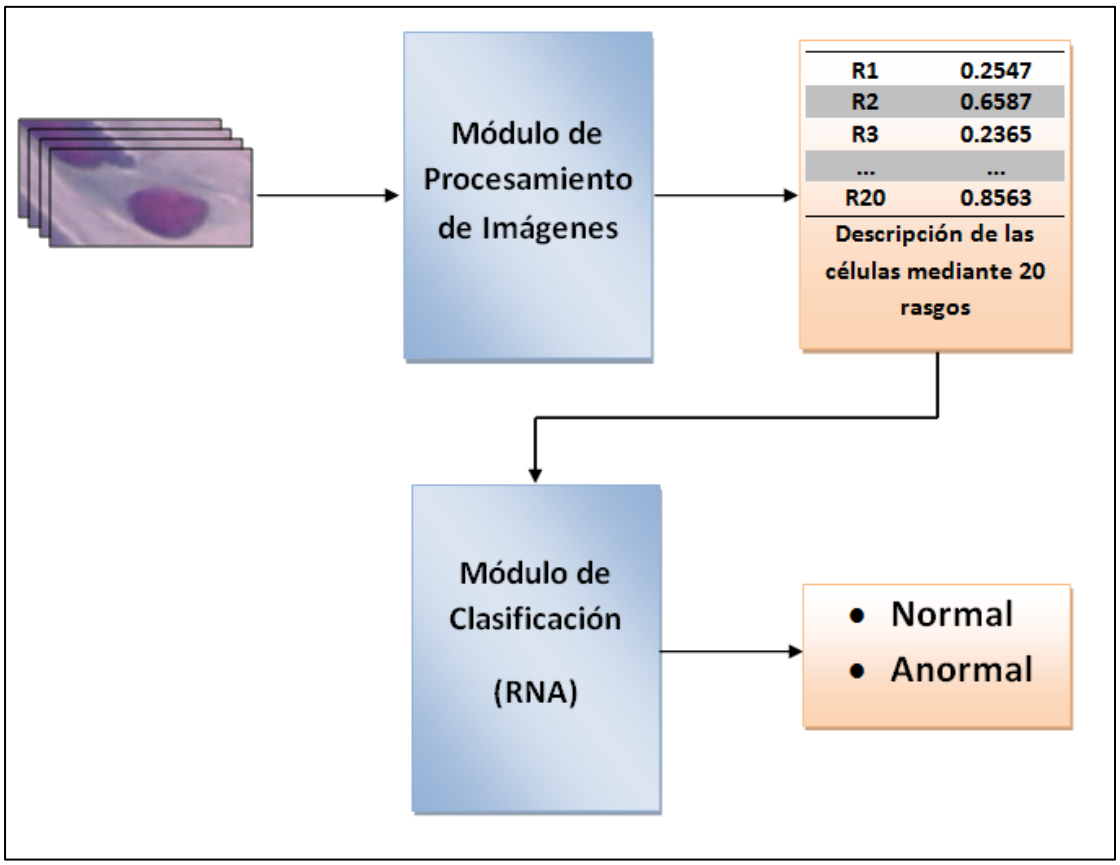



**Figura 2.3 Esquema de clasificación de células de la prueba de Papanicolaou usando redes neuronales multicapa.**

Primeramente se toma una imagen de la prueba de Papanicolaou y se extraen de las células presentes un conjunto de 20 rasgos con el objetivo de describirlas. Los 20 rasgos usados en el presente trabajo se describen en la sección 2.3. La construcción de este módulo será objetivo de otro trabajo. Por el momento, asumiremos que los rasgos en cuestión que describen a las células, están disponibles.

Una vez calculados los rasgos de las células, es necesario clasificarlas en normales o anormales. Para ello se utilizaron redes neuronales multicapa entrenadas mediante algoritmos genéticos. El método de clasificación, así como en entrenamiento de estos clasificadores se describen en la sección 2.4.

## 2.3 Base de casos de células cervicouterinas

La base de casos utilizada fue desarrollada por el Hospital Universitario Herlev, el departamento de Patología y el departamento de automatización en tecnología universitaria de Dinamarca. La misma contiene un total de 917 muestras, distribuidas en 7 clases diferentes, 3 normales y 4 anormales (tabla 2.1) [23-25]. En este trabajo solo se utilizaron dos clases o sea las normales como benignas y las anormales como malignas. La figura 2.4 muestra la forma general y las características de las células en cada una de las clases. El propósito de la prueba es diagnosticar los cambios de las células pre-malignas antes de que estas se conviertan en cáncer.

| **Tipo de células** | **No. de casos** | **Clasificación** |
|---|---|---|
| Epitelial escamoso (Superficial) | 74 | Normal |
| Epitelial escamoso (Intermedio) | 70 | Normal |
| Epitelial columnar | 98 | Normal |
| Displasia leve escamosa no queratinizante | 182 | Anormal |
| Displasia moderada escamosa no queratinizante | 146 | Anormal |
| Displasia severa escamosa no queratinizante | 197 | Anormal |
| Carcinoma escamoso *in situ* (Intermedio) | 150 | Anormal |
| Total | 917 | |

**Tabla 2.1. Células utilizadas en la construcción de la base de casos**

Cada célula de la base de casos es descrita por 20 rasgos que se muestran en la tabla 2.2. De los 20 rasgos, 6 son rasgos de intensidad/color y los 14 restantes son rasgos morfométricos o relacionados con la forma.

| No. | Rasgos | Tipo de Rasgo |
|---|---|---|
| R1 | Área del núcleo | Morfométrico |
| R2 | Área del citoplasma | Morfométrico |
| R3 | Proporción entre el núcleo y el citoplasma | Morfométrico |
| R4 | Brillo del núcleo | De Intensidad |
| R5 | Brillo del citoplasma | De Intensidad |
| R6 | Diámetro más corto del núcleo | Morfométrico |
| R7 | Diámetro más largo del núcleo | Morfométrico |
| R8 | Alargamiento del núcleo | Morfométrico |
| R9 | Redondez del núcleo | Morfométrico |
| R10 | Diámetro más corto del citoplasma | Morfométrico |
| R11 | Diámetro más largo del citoplasma | Morfométrico |
| R12 | Alargamiento del citoplasma | Morfométrico |
| R13 | Redondez del citoplasma | Morfométrico |
| R14 | Perímetro del núcleo | Morfométrico |
| R15 | Perímetro del citoplasma | Morfométrico |
| R16 | Posición relativa del núcleo | Morfométrico |
| R17 | Los máximos en el núcleo | De Intensidad |
| R18 | Los mínimos en el núcleo | De Intensidad |
| R19 | Los máximos en el citoplasma | De Intensidad |
| R20 | Los mínimos en el citoplasma | De Intensidad |

**Tabla 2.2. Rasgos extraídos de las células de la base de casos.**

A continuación se describen las formulaciones de los 20 rasgos usados.

**R1 y R2. Área del núcleo y del citoplasma**

Se calcula contando los píxeles correspondientes de la muestra segmentada. El área de un píxel es $(0.201\mu m)^2$.

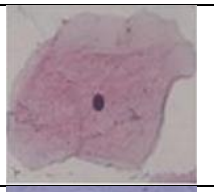
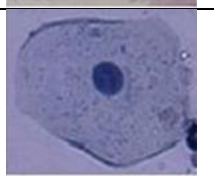
| CÉLULAS NORMALES | |
|---|---|
| **1. Escamoso Superficial**<br>• Forma: llana u ovalada.<br>• Núcleo muy pequeño.<br>• Relación núcleo-citoplasma muy pequeña. |  |
| **2. Escamoso Intermedio**<br>• Forma: redonda.<br>• Núcleo largo.<br>• Relación núcleo-citoplasma pequeña. |  |
| **3. Columnar**<br>• Forma: redonda.<br>• Núcleo largo.<br>• Relación núcleo-citoplasma media. |  |
| **CÉLULAS ANORMALES** | |
| **4. Displasia Leve**<br>• Núcleo claro o largo.<br>• Relación núcleo-citoplasma media. |  |
| **5. Displasia Moderada**<br>• Núcleo largo u oscuro.<br>• Citoplasma oscuro.<br>• Relación núcleo-citoplasma grande. |  |
| **6. Displasia Severa**<br>• Núcleo largo, oscuro o deforme.<br>• Citoplasma oscuro.<br>• Relación núcleo-citoplasma muy grande. |  |
| **7. Carcinoma in situ**<br>• Núcleo largo, oscuro o deforme.<br>• Relación núcleo-citoplasma muy grande. |  |

**Figura 2.4 Tipos de células y sus características obtenidas a través de la prueba de Papanicolaou.**

**R3. Proporción entre el núcleo y el citoplasma**

$$N/C = \frac{Área\ (Núcleo)}{Área\ (Núcleo)\ +\ Área\ (Citoplasma)} \tag{2.1}$$

**R4 y R5. Brillo del núcleo y del citoplasma**

Se calculada como el brillo promedio de los píxeles del núcleo y citoplasma respectivamente. El brillo de una región se calcula de la siguiente forma:

$$Y = 0.299 \cdot \mu_{ROJO} + 0.587 \cdot \mu_{VERDE} + 0.114 \cdot \mu_{AZUL} \quad (2.2)$$

donde $\mu_{ROJO}$, $\mu_{VERDE}$ y $\mu_{AZUL}$ son, respectivamente los colores rojo, verde y azul promedio de la región en cuestión. Las ponderaciones utilizadas vienen dadas por el brillo percibido por el ojo humano.

**R6 y R10. Diámetro más corto del núcleo y del citoplasma**

Es el diámetro mínimo de entre todas las circunferencias que se pueden inscribir en la región (Figura 2.5).

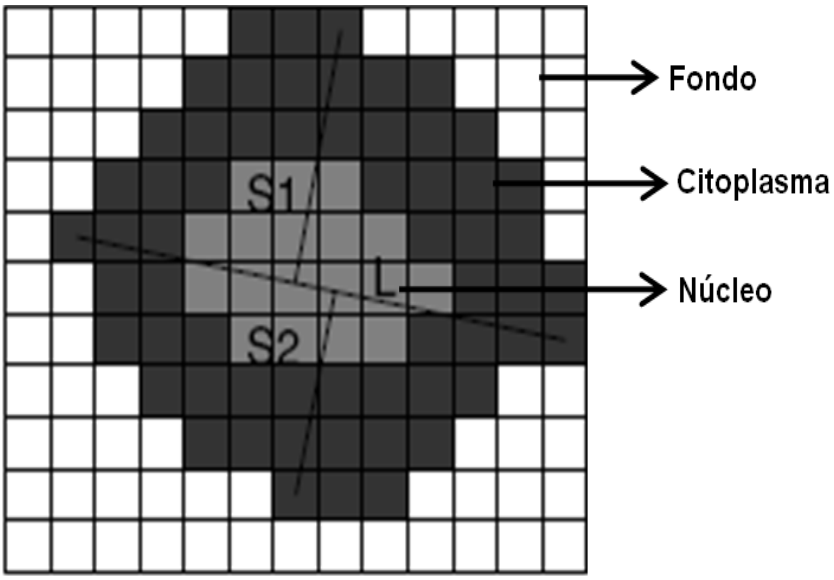



**Figura 2.5. Imagen binaria de una célula. Diámetro más largo (L), diámetro más corto (S1 y S2) para el citoplasma.**

**R7 y R11. Diámetro más largo del núcleo y del citoplasma**

Es el diámetro máximo de entre todas las circunferencias que pueden circunscribir la región (Figura 2.5).

**R8 y R12. Alargamiento del núcleo (N) y del citoplasma(C)**

$$Alargamiento(N) = \frac{Diámetro\ más\ corto\ (N)}{Diámetro\ más\ largo\ (N)} \quad (2.3)$$

$$Alargamiento(C) = \frac{Diámetro\ más\ corto\ (C)}{Diámetro\ más\ largo\ (C)} \quad (2.4)$$

**R9 y R13. Redondez del núcleo (N) y del citoplasma (C)**

$$A_c(N) = \frac{\pi}{4} \cdot [Diámetro\ más\ largo(N)]^2 \quad (2.5)$$

$$Redodez(N) = Área(N)/A_c(N) \quad (2.6)$$

$$A_c(C) = \frac{\pi}{4} \cdot [Diámetro\ más\ largo(C)]^2 \quad (2.7)$$

$$Redodez(C) = Área(C)/A_c(C) \quad (2.8)$$

**R14 y R15. Perímetro de núcleo y del citoplasma**

Perímetro (puntos de la frontera) del objeto.

**R16. Posición del núcleo (N)**

$$Posición(N) = \frac{2a\sqrt{(X_N + X_C)^2 + (Y_N + Y_C)^2}}{Diámetro\ más\ largo(N)} \quad (2.9)$$

aquí la posición está dada por el centro $(X_N; Y_N)$ y $(X_C; Y_C)$ para el núcleo y el citoplasma respectivamente.

**R17 y R19. Cantidad de máximos locales en el núcleo y en el citoplasma**

Los máximos locales se calculan como la cantidad de píxeles en la región en cuestión que son máximos globales en una vecindad de 8 pixeles.

**R18 y R20. Cantidad de mínimos locales en el núcleo y en el citoplasma**

Los mínimos locales se calculan como la cantidad de píxeles en la región en cuestión que son mínimos globales en una vecindad de 8 pixeles.

## 2.4 Entrenamiento una red neuronal multicapa usando algoritmos genéticos jerárquicos

El proceso de entrenamiento de RNM usando algoritmos genéticos jerárquicos [26, 27] se muestra en la figura 2.6.

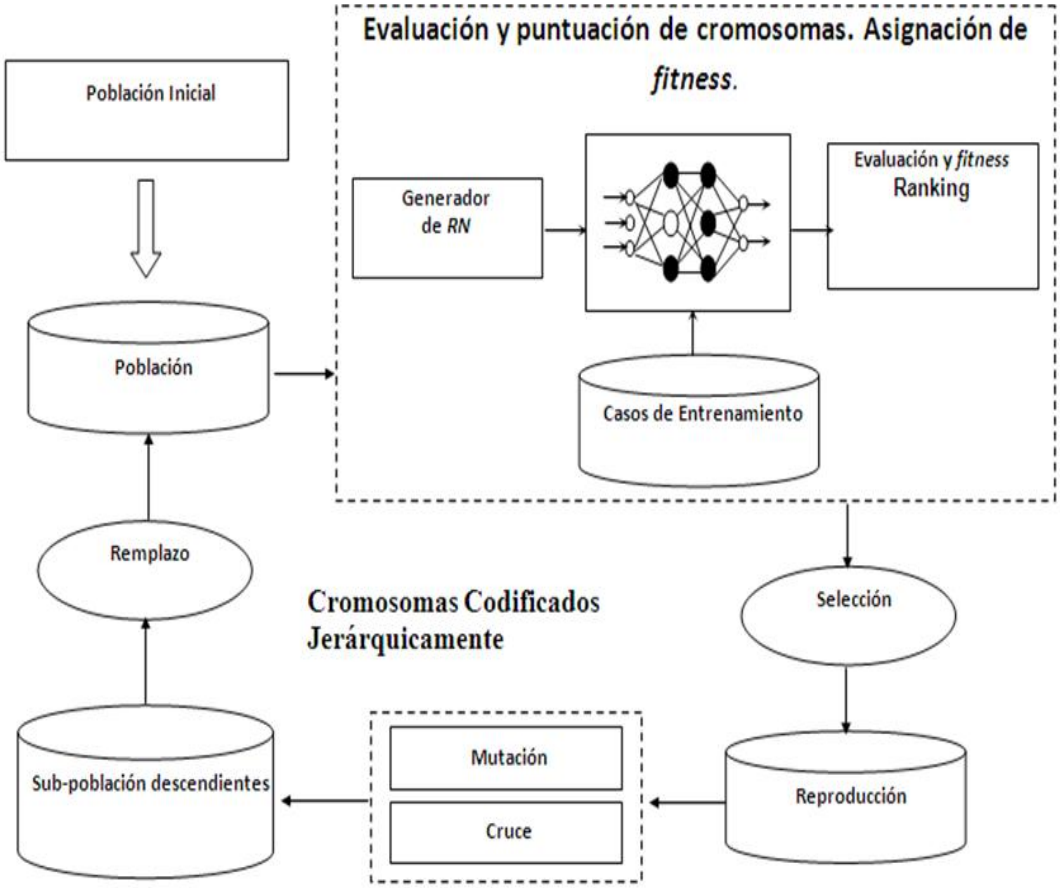



**Figura 2.6 Entrenamiento de la RNM usando AG**

Inicialmente, una población es generada aleatoriamente. Las redes son codificadas usando un cromosoma jerárquico como el que se muestra en la figura 2.7. El cromosoma está estructurado con dos capas de genes de control y una de genes paramétricos.

La primera capa de genes de control está conformada por los genes de control de capa. Estos controlan la existencia de cada capa de la red neuronal. La segunda capa de genes de control son los genes de control de neurona. Estos controlan la existencia o no de cada neurona por capa. Los genes de control, dentro de la estructura de datos del cromosoma, son de tipo booleano. Por último los genes paramétricos contienen los pesos de las conexiones de la red y sus sesgos y son de tipo real.

Para la evaluación de cada cromosoma (cálculo de la aptitud o idoneidad) a partir de cada cromosoma se genera la red correspondiente y se evalúa su desempeño de acuerdo a los casos de entrenamiento. A partir de este rating que se forma a partir de la aptitud, y siguiendo al algoritmo de la ruleta como mecanismo de selección, se

genera el *matting pool* o conjunto de individuos que formarán parte del proceso de reproducción.

La reproducción se realiza siguiendo los mecanismos clásicos de cruzamiento y mutación, controlado por sus respectivas probabilidades. En este caso cada tipo de gen (control de capas, control de neuronas y paramétricos) es controlado por una probabilidad que puede ser diferente.

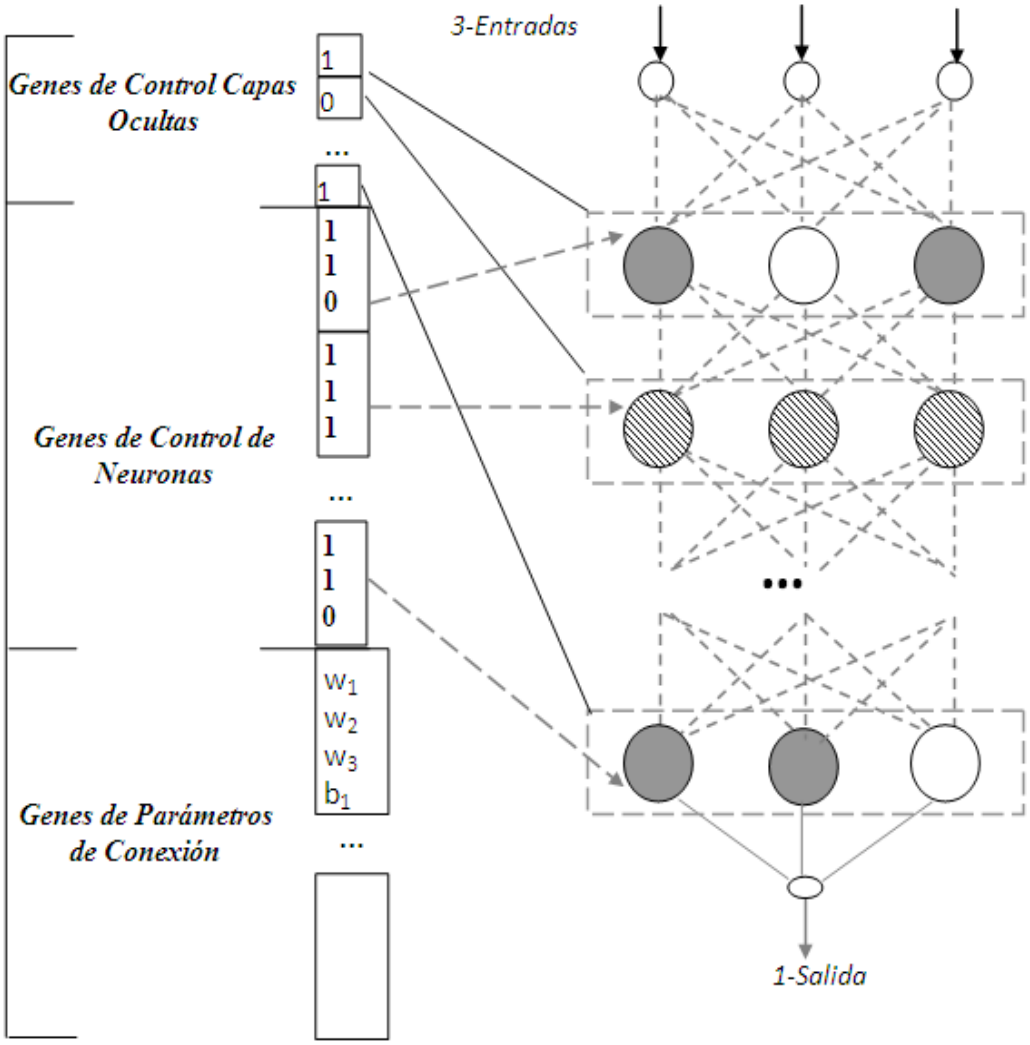



**Figura 2.7 Codificación usando un cromosoma jerárquico de una red neuronal multicapa.**

Por último, a partir de la subpoblación de descendientes, siguiendo un mecanismo de reemplazo, se conforma la población de la siguiente iteración. El proceso iterativo es controlado por el número de iteraciones. La complejidad de los cromosomas impide el uso de otro criterio de parada como la distancia entre los cromosomas de dos

iteraciones sucesivas. El uso de un umbral determinado para la idoneidad sería erróneo debido a que estaríamos acotando la calidad de los resultados que se pudieran obtener.

Para controlar fenómenos evolutivos no deseados como la deriva genética o la convergencia prematura se usaron los algoritmos propuestos en [27].

# Capítulo 3. Resultados y discusión

## 3.1 Evaluación de los clasificadores

Para realizar la evaluación de los clasificadores basados en RNM se usaron un grupo de medidas de calidad (tabla 3.1) que se describen a continuación [28, 29].

- TP (Verdaderos Positivos, *True Positives*): significa la clasificación correcta de un valor positivo (célula anormal).
- TN (Verdaderos Negativos, *True Negatives*): significa la clasificación correcta de un valor negativo (célula normal).
- FP (Falsos Positivos, *False Positives*): significa la cantidad de valores negativos dados como positivos (célula anormal).
- FN (Falsos Negativos, *False Negatives*): significa la cantidad de valores positivos dados como negativos (célula normal).

| | **Clasificada Normal (Negativo)** | **Clasificada Anormal (Positivo)** |
|---|---|---|
| **Células Normal (Negativo)** | TN | FP |
| **Células Anormal (Positivo)** | FN | TP |

**Tabla 3.1 Definición de algunas medidas de evaluación de los clasificadores.**

Los criterios de desempeño utilizados a partir de estos valores son:

- Tasa de clasificación (TC, *Accuracy*): es la razón de las clasificaciones correctas con relación a todas las clasificaciones.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.1}$$

- Sensibilidad (*Recall*): corresponde a la fracción de positivos detectados correctamente con relación a la cantidad total de positivos de la base de casos.

$$Recall = \frac{TP}{TP + FN} \tag{3.2}$$

- Predictividad Positiva (*Precision*): es la fracción de positivos detectados correctamente con relación a las clasificaciones dadas como positivas.

$$Precision = \frac{TP}{TP + FP} \tag{3.3}$$

- Medida-F (*F-Measure*)

$$F - Measure = \frac{1}{\lambda \frac{1}{Precision} + (1 - \lambda) \frac{1}{Recall}}, \qquad 0 \leq \lambda \leq 1 \tag{3.4}$$

Para este trabajo se utilizó $\lambda = 0.5$

Para la validación de los resultados se usó una validación cruzada con 5 grupos (*5-Fold Crossvalidation*) [30].

Esta técnica es bien simple pero garantiza conocer el grado de generalización adquirido por los clasificadores. En el caso de la validación cruzada con 5 grupos, la base de casos se divide en cinco partes iguales y se realizan 5 entrenamientos siempre dejando un grupo para la prueba y tomando para entrenar los 4 grupos restantes. La base de casos se divide de forma tal que los 5 grupos que se forman sean representativos de la base de casos completa, en nuestro caso es manteniendo la proporción de casos positivos y negativos.

## 3.2 Parámetros usados en el entrenamiento

Los parámetros que se usaron en el entrenamiento se describen en la tabla 3.2.

Los parámetros 1 y 2 no tienen margen de ser cambiados debido a la forma en que están codificados los casos en la base de casos.

Los parámetros 3 y 4 fueron seleccionados en esos valores debido a que está demostrado que una red neuronal no necesita más de dos capas para realizar una buena clasificación [26]. Con estos parámetros estamos acotando los resultados a redes de hasta tres capas con a lo sumo 15 neuronas cada una.

Debido a que las clases en la base de casos estaban codificadas como 1 y 0 para positivo y negativo entonces como función de transferencia para la capa de salida no había otra opción que usar el límite duro.

La cantidad de iteraciones se limitó a 500 debido a que no se estaban obteniendo variaciones en los resultados a partir de esta iteración.

| No. | Parámetros del AG | Valores usados | Mejores resultados |
|---|---|---|---|
| **1** | Cantidad de neuronas de entrada | 20 | - |
| **2** | Cantidad de neuronas de salida | 1 | - |
| **3** | Cantidad máxima de neuronas por capa oculta | 15 | - |
| **4** | Cantidad máxima de capas ocultas | 3 | - |
| **5** | Función de transferencia para las capas ocultas | Logaritmo Simoidal<br>Tangente Simoidal | Logaritmo Simoidal |
| **6** | Función de transferencia para la capa de salida | Límite Duro | - |
| **7** | Cantidad de iteraciones | 500 | - |
| **8** | Tamaño de la población | 120 | - |
| **9** | Cantidad de casos de la base de casos | 917 (validación cruzada de 5 grupos) | - |
| **10** | Porciento de remplazo | 30<br>50<br>70 | 50 |
| **11** | Probabilidad de cruzamiento en los genes de control de capas | 0.8<br>0.9<br>1.0 | 1 |
| **12** | Probabilidad de cruzamiento en los genes de control de neuronas | 0.8<br>0.9<br>1.0 | 1 |
| **13** | Probabilidad de cruzamiento en los genes paramétricos | 0.8<br>0.9<br>1.0 | 1 |
| **14** | Tipo de cruzamiento | En un punto<br>Uniforme | En un punto |
| **15** | Probabilidad de mutación en los genes de control de capas | 0.01<br>0.05<br>0.10 | 0.05 |
| **16** | Probabilidad de mutación en los genes de control de neuronas | 0.01<br>0.05<br>0.10 | 0.05 |
| **17** | Probabilidad de mutación en los genes paramétricos | 0.01<br>0.05<br>0.10 | 0.05 |

**Tabla 3.2 Parámetros usados en el entrenamiento.**

El tamaño de la población se limitó a 120 con el objetivo de realizar corridas viables en las computadoras que se tenían a disposición. Con este tamaño de población el

tiempo de cada corrida en la computadora con las especificaciones dadas no sobrepasó los 20 minutos.

Se probaron en total tres porcientos de remplazo, desde el conservador 30 % al arriesgado 70 %. Al final los mejores resultados se obtuvieron con un 50 % de remplazo, o sea, de una generación a la siguiente solo sobreviven el 50 % de los individuos.

Las probabilidades de cruzamiento se probaron de acuerdo a los resultados al respecto que se encuentran en la literatura [27]. Un valor menor que 0.8 es demasiado conservador y pudiera hacer la convergencia demasiado lenta o incluso hacer el AG divergente. La probabilidad seleccionada fue 1 al final, o sea, todas las parejas de individuos son aptas para el cruzamiento. El cruzamiento en un punto fue el utilizado en los experimentos por encima del cruzamiento uniforme.

Por otro lado las probabilidades de mutación nunca deben sobrepasar el 0.25 [27]. En nuestro trabajo se fue desde la conservadora 0.01 a 0.1. El valor que mejores prestaciones dio fue 0.05.

## 3.3 Resultados obtenidos

Las tablas 3.3 y 3.4 muestran los resultados de la clasificación para los 5 mejores clasificadores de los 5 grupos de la validación cruzada para los casos de entrenamiento y prueba, respectivamente.

| Grupo | TC | Sensibilidad | Especificidad | F-Measure |
|---|---|---|---|---|
| **1** | 0.9272 | 0.9737 | 0.9308 | 0.9517 |
| **2** | 0.9269 | 0.9685 | 0.9347 | 0.9512 |
| **3** | 0.9294 | 0.9756 | 0.9319 | 0.9531 |
| **4** | 0.9330 | 0.9652 | 0.9450 | 0.9549 |
| **5** | 0.9294 | 0.9693 | 0.9371 | 0.9528 |
| **Promedio** | 0.9292 | 0.9705 | 0.9359 | 0.9527 |

**Tabla 3.3 Resultados para el promedio de los cinco mejores clasificadores en los 5 grupos de la validación cruzada para los casos de entrenamiento.**

El promedio de la tasa de clasificación para estos 25 clasificadores fue del 91.71 % con una sensibilidad del 96.00 %. Estos valores altos de sensibilidad siempre son deseados en este tipo de clasificación debido a que los falsos negativos representan a

casos que fueron diagnosticados como normales y realmente no lo son. El hecho de que la sensibilidad sea mayor que la tasa de clasificación y la especificidad apoya aún más esta tesis pues reconoce que los clasificadores prefieren dar falsas alarmas ante los casos dudosos que declarar sana a una célula enferma.

| Grupo | TC | Sensibilidad | Especificidad | F-Measure |
|---|---|---|---|---|
| **1** | 0.9261 | 0.9719 | 0.9309 | 0.9507 |
| **2** | 0.9022 | 0.9393 | 0.9288 | 0.9337 |
| **3** | 0.9169 | 0.9748 | 0.9182 | 0.9454 |
| **4** | 0.9137 | 0.9422 | 0.9408 | 0.9415 |
| **5** | 0.9268 | 0.9719 | 0.9322 | 0.9514 |
| **Promedio** | 0.9171 | 0.9600 | 0.9302 | 0.9445 |

**Tabla 3.4 Resultados para los promedios de los cinco mejores clasificadores en los 5 grupos de la validación cruzada para los casos de prueba.**

El mismo análisis se puede realizar para el mejor clasificador obtenido para cada grupo de la validación cruzada (tablas 3.5 y 3.6).

| | TC | Sensibilidad | Especificidad | F-Measure |
|---|---|---|---|---|
| **1** | 0.9332 | 0.9593 | 0.9505 | 0.9548 |
| **2** | 0.9332 | 0.9833 | 0.9299 | 0.9559 |
| **3** | 0.9332 | 0.9630 | 0.9472 | 0.9550 |
| **4** | 0.9401 | 0.9741 | 0.9460 | 0.9599 |
| **5** | 0.9401 | 0.9944 | 0.9291 | 0.9606 |
| **Promedio** | 0.9360 | 0.9748 | 0.9405 | 0.9572 |

**Tabla 3.5 Resultados para el mejor clasificador en los 5 grupos de la validación cruzada para los casos de entrenamiento.**

En la figura 3.1 se muestran las tasas de clasificación para los 5 grupos de la validación cruzada. Aunque, como es normal, siempre la tasa de clasificación desmejora de los casos de entrenamiento a los de prueba, se puede observar que este decremento es a lo sumo de 2.69 puntos porcentuales. Este resultado refuerza la idea de la posibilidad del uso de los algoritmos genéticos jerárquicos en el entrenamiento de redes neuronales artificiales.

| | TC | Sensibilidad | Especificidad | F-Measure |
|---|---|---|---|---|
| **1** | 0.9293 | 0.9556 | 0.9485 | 0.9520 |
| **2** | 0.9130 | 0.9630 | 0.9220 | 0.9420 |
| **3** | 0.9290 | 0.9556 | 0.9485 | 0.9520 |
| **4** | 0.9071 | 0.9333 | 0.9403 | 0.9368 |
| **5** | 0.9071 | 0.9778 | 0.9041 | 0.9395 |
| **Promedio** | 0.9171 | 0.9571 | 0.9327 | 0.9445 |

**Tabla 3.6 Resultados para el mejor clasificador en los 5 grupos de la validación cruzada para los casos de prueba.**

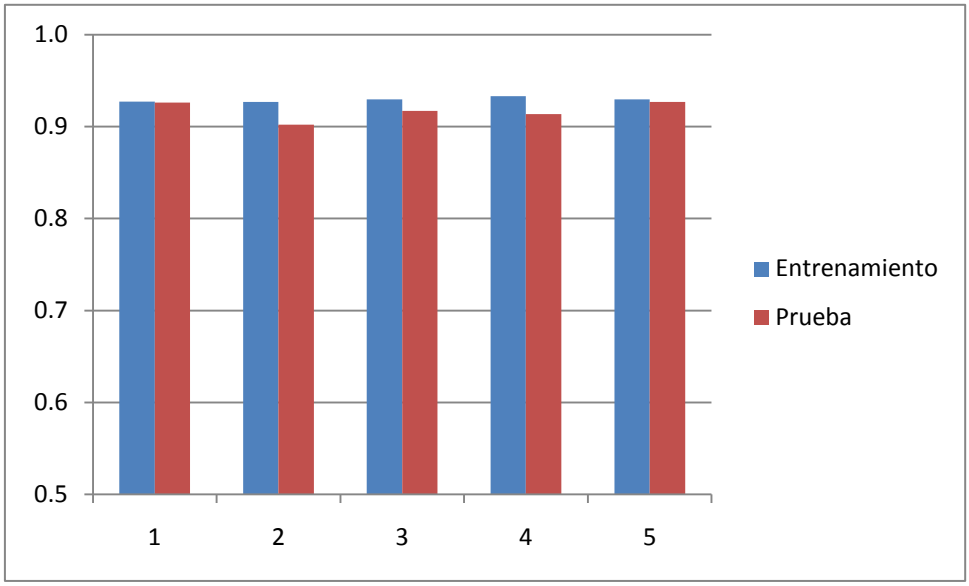



**Figura 3.1 Tasa de clasificación para los 5 mejores casos de los 5 grupos de la validación cruzada.**

## 3.4 Comparación con otros métodos

Los resultados descritos en la sección anterior se comparan con los resultados obtenidos usando Máquinas de Soporte Vectorial (SVM, *Support Vector Machine*) usando diferentes parámetros de estas.

### *3.4.1 Máquinas de soporte vectorial*

Las SVM son una herramienta poderosa y robusta para tareas de clasificación [31, 32]. Considere un conjunto de entrenamiento de $n$ observaciones para realizar una clasificación binaria, $(x_1, y_1) .. (x_n, y_n) \in X \times Y$, donde $x_i = (x_{i1}, .., x_{id})$, corresponde al vector de rasgos para la $i^{th}$muestra, en un espacio $d$-dimensional, y $y_i \in \{-1,1\}$ denota estas dos clases involucradas. Cualquier punto $x$ en el hiperplano $X$ debe

satisfacer la condición $\mathbf{w} \cdot \mathbf{x} + b = 0$, donde **w** es normal al hiperplano. En general, la SVM no lineal viene dada por el siguiente problema de optimización:

$$\min_{\mathbf{w},b,\xi_i \geq 0} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i$$
$$\text{S. A: } y_i(\mathbf{w} \cdot \Phi(x_i) + b) \geq 1 - \xi_i, \qquad \forall i \in 1..n \tag{3.5}$$

En la ecuación anterior, un caso $x_i$ es mapeado a un espacio con una dimensionalidad mayor por medio de la función núcleo (*kernel*) Φ, mientras que C es una penalización especificada al mecanismo de entrenamiento. Resolviendo el problema enunciado en la ecuación (3.5), obtenemos el margen máximo de separación entre el hiperplano de separación y el caso. Un caso de prueba **z** puede ser clasificado entonces de acuerdo a la siguiente regla de decisión:

$$D(\mathbf{z}) = \text{signo}(h(\mathbf{z}) = \mathbf{w} \cdot \mathbf{\Phi}(\mathbf{z}) + b) \tag{3.6}$$

### *3.4.2 Comparación de los resultados*

| | Clasificador | TC |
|---|---|---|
| **C1** | SVM con kernel Gaussiano. $\sigma = 0.5$ | 0.7394 |
| **C2** | SVM con kernel Gaussiano. $\sigma = 1$ | 0.8987 |
| **C3** | SVM con kernel Gaussiano. $\sigma = 2$ | 0.9510 |
| **C4** | SVM con kernel Gaussiano. $\sigma = 4$ | 0.9487 |
| **C5** | SVM con kernel Gaussiano. $\sigma = 8$ | 0.9367 |
| **C6** | SVM con kernel Lineal. | 0.9345 |
| **C7** | SVM con kernel Polinomial de grado 2 | 0.8986 |
| **RNA-HGA** | **RNA entrenadas con HGA** | **0.9171** |

**Tabla 3.7 Resultados de otros clasificadores para la base de casos de estudio.**

La tabla 3.7 y la figura 3.2 muestran la comparación de los resultados obtenidos con siete clasificadores SVM. Estos SVM fueron seleccionados con kernel Gaussiano con desviaciones estándar {0.5,1,2,4,8}, kernel lineal y kernel polinomial con grado 2. Los resultados obtenidos con el método propuesto son equiparables a los obtenidos utilizando SVM.

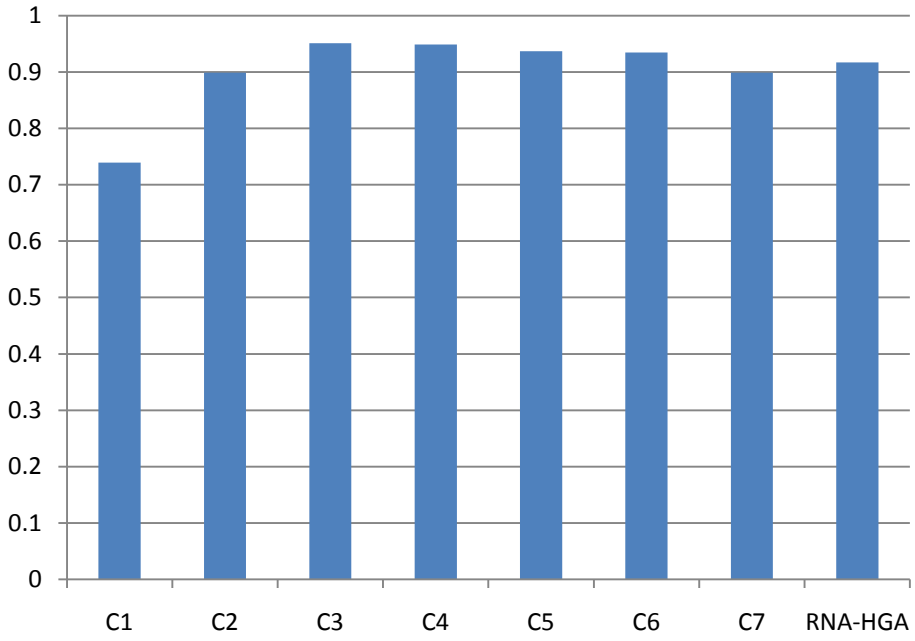



**Figura 3.2 Resultados obtenidos usando RNA-HGA con los siete SVM.**

# Conclusiones y Recomendaciones

Se construyeron clasificadores basados en Redes Neuronales Multicapa para la clasificación de células de la prueba de Papanicolaou usando algoritmos genéticos, con buenas prestaciones.

Los parámetros de entrada del AG que mejores resultados presentaron fueron:

- ✓ Probabilidades de cruzamiento y mutación: 1 y 0.05 respectivamente.
- ✓ Funciones de transición: logaritmo simoidal.
- ✓ Tipo de remplazo: en un punto.

Los resultados obtenidos superan o igualan los reportados en otras literaturas para este tipo de problema.

Al presente trabajo le realizamos las siguientes recomendaciones:

- ✓ Realizar el entrenamiento de las RNA usando la combinación de algoritmos genéticos y autómatas celulares (algoritmos genéticos celulares).
- ✓ Realizar un estudio del costo computacional del entrenamiento.

# Referencias Bibliográficas


[1] H. Damasio, *Human brain anatomy in computerized images*: Oxford University Press, USA, 2005.

[2] B. D. Ripley, *Pattern recognition and neural networks*: Cambridge Univ Pr, 2008.

[3] J. A. Freeman and D. Skapura, "Redes Neuronales. Algoritmos, aplicaciones y técnicas de propagación," *México: Addison-Wesley*, p. 306, 1993.

[4] B. Pradhan and S. Lee, "Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling," *Environmental Modelling & Software*, vol. 25, pp. 747-759, 2010.

[5] G. Puscasu, B. Codres, A. Stancu, and G. Murariu, "Nonlinear system identification based on internal recurrent neural networks," *International Journal of Neural Systems*, vol. 19, p. 115, 2009.

[6] M. Affenzeller and S. Winkler, *Genetic algorithms and genetic programming: modern concepts and practical applications* vol. 6: Chapman & Hall/CRC, 2009.

[7] V. Valls, F. Ballestin, and S. Quintanilla, "A hybrid genetic algorithm for the resource-constrained project scheduling problem," *European Journal of Operational Research*, vol. 185, pp. 495-508, 2008.

[8] O. L. Mangasarian, W. N. Street, and W. H. Wolberg, "Breast Cancer Diagnosis and Prognosis via Linear Programming," July, 2009.

[9] S. Yu and D. Dasgupta, "Conserved self pattern recognition algorithm with novel detection strategy applied to breast cancer diagnosis," *Journal of Artificial Evolution and Applications*, vol. 2009, p. 4, 2009.

[10] I. Maglogiannis, E. Zafiropoulos, and I. Anagnostopoulos, "An intelligent system for automated breast cancer diagnosis and prognosis using SVM based classifiers," *Applied Intelligence*, vol. 30, pp. 24-36, 2009.

[11] X. Yi, P. Wu, J. Li, and L. Liu, "Breast cancer diagnosis using WNN based on GA," *Life System Modeling and Intelligent Computing*, pp. 367-374.

[12] H. Peng, Z. Ruan, D. Atasoy, and S. Sternson, "Automatic reconstruction of 3D neuron structures using a graph-augmented deformable model," *Bioinformatics*, vol. 26, pp. i38-i46, 2010.

[13] S. S. Haykin, *Neural networks and learning machines* vol. 3: Prentice Hall, 2009.

[14] J. H. Holland, *Adaptation in natural and artificial systems*: University of Michigan press, 1975.

[15] D. E. Goldberg, *Genetic algorithms in search, optimization, and machine learning*: Addison-wesley, 1989.

[16] L. Davis and M. Mitchell, *Handbook of genetic algorithms*: Van Nostrand Reinhold, 1991.

[17] Z. Michalewicz, *Genetic algorithms+ data structures*: Springer, 1996.

[18] S. W. Mahfoud, "Population sizing for sharing methods," *Urbana,* vol. 51, p. 61801, 1994.

[19] D. E. Goldberg, "Sizing populations for serial and parallel genetic algorithms," 1989, pp. 70-79.

[20] K. F. Man, K. S. Tang, and S. Kwong, *Genetic algorithms: concepts and designs* vol. 1: Springer Verlag, 1999.

[21] A. Meisels, C. Morin, D. Giri, and R. S. Hoda, "Cytopathology of the uterus," *International Journal of Gynecologic Pathology,* vol. 17, p. 286, 1998.

[22] D. Landwehr, " Web based pap-smear classification," in *Master's thesis, Technical University of Denmark(DTU) Dept. of Automation, Bldg 326, 2800 Lyngby, Denmark.*, 2001.

[23] J. Jantzen, J. Norup, G. Dounias, and B. Bjerregaard, "Pap-smear benchmark data for pattern classification," *Proceedings of the NiSIS,* pp. 1-9, 2005.

[24] E. Martin and M. Exbrayat, "Pap-smear classification," *Technical University of Denmark - DTU,* 2003.

[25] J. Norup, "Classification of pap-smear data by transductive neuro-fuzzy methods," *Master Thesis,* 2005.

[26] M. Orozco-Monteagudo, A. Taboada-Crispi, and L. Gutierrez-Hernandez, "Optimal Parameter for the Training of Multilayer Perceptron Neural Networks by Using Hierarchical Genetic Algorithm," 2008, p. 3.

[27] M. Orozco-Monteagudo, A. Taboada-Crispi, and A. Del Toro-Almenares, "Training of multilayer perceptron neural networks by using cellular genetic algorithms," *Progress in Pattern Recognition, Image Analysis and Applications,* pp. 389-398, 2006.

[28] Y. Jiang, B. Cukic, and Y. Ma, "Techniques for evaluating fault prediction models," *Empirical Software Engineering,* vol. 13, pp. 561-595, 2008.

[29] M. V. Joshi, "On evaluating performance of classifiers for rare classes," 2002, pp. 641-644.

[30] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Statistics Surveys,* vol. 4, pp. 40-79, 2010.

[31] S. Abe, *Support vector machines for pattern classification*: Springer-Verlag New York Inc, 2010.

[32] N. Dehak, R. Dehak, P. Kenny, N. Brammer, P. Ouellet, and P. Dumouchel, "Support vector machines versus fast scoring in the low-dimensional total variability space for speaker verification," 2009.

# Anexo 1. Funciones de transferencia de las neuronas

La tabla A1.1 muestra las distintas funciones de transferencia que se usan dentro del campo de las redes neuronales artificiales. Para todas ellas, el dominio es el campo de los números reales, o sea:

$$f: \quad \begin{matrix} \mathbb{R} \longrightarrow \mathbb{R} \\ x \longrightarrow f(x) \end{matrix} \tag{A1.1}$$

| Función | Nombre (en Inglés) | Expresión |
|---|---|---|
| Límite Duro | Hard Limit | $f(x) = \begin{cases} 0 & si \ \ x \leq 0 \\ 1 & si \ \ x > 0 \end{cases}$ |
| Límite Duro Simétrico | Symmetric Hard Limit | $f(x) = \begin{cases} -1 & si \ \ x \leq 0 \\ 1 & si \ \ x > 0 \end{cases}$ |
| Lineal Pura | Pure Linear | $f(x) = x$ |
| Lineal Positiva | Positive Linear | $f(x) = \begin{cases} 0 & si \ \ x < 0 \\ x & si \ \ x \geq 0 \end{cases}$ |
| Lineal Saturada | Saturating Linear | $f(x) = \begin{cases} 0 & si \quad x < 0 \\ x & si \ \ 0 \leq x \leq 1 \\ 1 & si \quad x > 1 \end{cases}$ |
| Lineal Saturada Simétrica | Symmetric Saturating Linear | $f(x) = \begin{cases} 0 & si \quad x < 0 \\ x & si \ \ 0 \leq x \leq 1 \\ 1 & si \quad x > 1 \end{cases}$ |
| Logaritmo Simoidal | Sigmoig Logarithms | $f(x) = \dfrac{1}{1 + e^{-x}}$ |
| Tangente Simoidal o Tangente Hiperbólica | Sigmoig Tangent | $f(x) = \dfrac{e^{x} - e^{-x}}{e^{x} + e^{-x}} = \dfrac{e^{2x} - 1}{e^{2x} + 1}$ |

**Tabla A1.1 Funciones de Transferencia.**