

Universidad de Granada

**DEPARTAMENTO DE ARQUITECTURA Y TECNOLOGÍA DE
COMPUTADORES**



**Desarrollo de Sistemas Inteligentes Para
Clasificación y Diagnóstico en Medicina**

TESIS DOCTORAL

Suhail Musa Issa Odeh

Granada, 2006

D. Eduardo Ros Vidal, Profesor Titular de Universidad, y D. Ignacio Rojas Ruiz, Profesor Titular de Universidad, ambos del Departamento de Arquitectura y Tecnología de Computadores de la Universidad de Granada,

CERTIFICAN:

Que la memoria titulada **“Desarrollo de Sistemas Inteligentes Para Clasificación y Diagnóstico en Medicina”**, ha sido realizada por D. Suhail Musa Issa Odeh bajo nuestra dirección en el Departamento de Arquitectura y Tecnología de Computadores de la Universidad de Granada para optar al grado de Doctor por la Universidad de Granada.

Granada, a 16 de Mayo de 2006

Fdo. Eduardo Ros Vidal
Director de la tesis

Fdo. Ignacio Rojas Ruiz
Director de la tesis

Universidad de Granada

**DEPARTAMENTO DE ARQUITECTURA Y TECNOLOGÍA DE
COMPUTADORES**



**Desarrollo de Sistemas Inteligentes Para
Clasificación y Diagnóstico en Medicina**

**Memoria presentada por
Suhail Musa Issa Odeh**

Para optar al grado de
DOCTOR POR LA UNIVERSIDAD DE GRANADA

Fdo.: Suhail Musa Issa Odeh

Agradecimientos

A mis directores de tesis, **Eduardo Ros Vidal**, el esfuerzo y el tiempo dedicados en la dirección de esta tesis, y por haber sabido transmitirme el ánimo necesario en todo momento. E **Ignacio Rojas Ruiz**, por su apoyo e incondicional colaboración en todo este trabajo, sus acertadas correcciones y comentarios han hecho que este trabajo sea más interesante y riguroso.

A mi madre, que aunque no lo haya vivido directamente, ha estado ahí para darme su apoyo y comprensión cuando a veces trabajaba noche y día sin cesar, a mis hermanos por el cariño y confianza que me brindan en todo momento, pero sobre todo por estar, cada uno a su manera, respaldándome para alcanzar mis objetivos.

Quiero expresar mi agradecimiento, por su apoyo económico, a las distintas instituciones que me han ayudado durante estos casi cuatro años de trabajo: fundamentalmente la AECI (Agencia Española de Cooperación Internacional), al plan propio de la Universidad de Granada y al grupo de investigación CASIP.

Además, me gustaría agradecer al grupo de investigación CASIP del departamento de Arquitectura y Tecnología de Computadores su apoyo en diversas facetas imprescindibles para un trabajo de investigación largo y continuado (espacios de investigación, medios, etc.).

Particularmente, quiero hacer explícito mi agradecimiento al Profesor Elmar Lang de la Universidad de Regensburg y su grupo de investigación la cesión a nuestro grupo de investigación de las imágenes de fluoroscopia que he utilizado en esta tesis doctoral.

Por fin, a todas aquellas personas que de verdad puedo llamarles amigos.

Cada persona tiene una sola madre, pero yo afortunadamente tengo dos, a quien dedico con mucho honor este trabajo: a mi primera madre “Palestina” quien me ha dado la identidad, y a mi segunda madre quien me ha dado la vida. y como no, a la memoria de mi padre

*A mi madre
Y
A la memoria de mi padre*

Índice General

Capítulo 1	1
<i>Introducción.....</i>	<i>1</i>
Capítulo 2	9
<i>Antecedentes en sistemas de diagnóstico para la clasificación de enfermedades de piel mediante técnicas de procesamiento de imágenes.....</i>	<i>9</i>
2.1 Introducción.....	10
2.2 Análisis de Componentes Independientes ICA para clasificar lesiones de piel	12
2.3 Imágenes espectroscópicas de reflectancia difusa en incidencia oblicua para clasificar lesiones de piel.....	14
2.4 Espectroscopia Raman para clasificar lesiones de piel	16
2.5 Sistema CAD (Computer Aided Diagnosis) para clasificar la lesiones de piel con pigmentación.....	17
2.6 Métodos de clasificación para diagnóstico de lesiones de piel	20
2.7 Contexto	21
Capítulo 3.....	23
Métodos de clasificación.....	23
3.1 Introducción	24
3.2 Descripción de un sistema de clasificación basado en KNN ..	29
3.2.1 El proceso de evaluación “dejar uno fuera” (leave one out)	35
3.2.2 Proceso de clasificación.....	36
3.3 Descripción de un sistema de clasificación basado en redes neuronales.....	38
3.3.1 Introducción.....	38
3.3.2 Arquitectura de la red	41
3.3.3 Entrenamiento de la red.....	41
3.3.4 Generalización	45
3.3.5 Validación cruzada	46
3.4 Descripción de un sistema de clasificación basado en la combinación de los sistemas de inferencia difusa y redes neuronales adaptables (ANFIS).....	47
3.4.1 ANFIS (Adaptive Neuro Fuzzy Inference System)	48
3.4.2 Arquitectura del modelo ANFIS.....	49
3.4.3 Aprendizaje en el modelo ANFIS	51

Capítulo 4	55
<i>Estrategias de optimización</i>	55
4.1 Métodos de exploración secuencial	57
4.1.1 Método hacia atrás	57
4.1.2 Método hacia adelante	58
4.1.3 El algoritmo “Greedy Feature Flip” (G-flip).....	60
4.2 Métodos de exploración multi-camino	62
4.2.1 Algoritmos Genéticos	62
4.2.1.1 Implementación del algoritmo genético	63
4.2.1.2 Representación	66
4.2.1.3 Población Inicial	67
4.2.1.4 Medida de Evaluación	67
4.2.1.5 Criterio de selección.....	68
4.2.1.6 Operador de cruce.....	69
4.2.1.7 Operador de Mutación	70
Capítulo 5	73
<i>Aplicación a diagnóstico de patologías de la piel</i>	73
5.1 Patologías de la piel	74
5.1.1 Introducción.....	74
5.1.2 El cáncer de piel y otras patologías	75
5.1.2.1 Keratosis Actínica	76
5.1.2.2 Carcinoma basal	77
5.1.2.3 Psoriasis	78
5.1.3 Objetivo	79
5.2 Descripción de técnica de adquisición de imágenes.....	80
5.3 Descripción de la base de datos	81
5.4 Extracción y análisis de parámetros	83
5.4.1 Tratamiento de imágenes a bajo nivel	85
5.4.1.1 Detección de bordes	86
5.4.1.1.1 Gradiente	88
5.4.1.1.2 El detector de borde de Sobel	90
5.4.1.1.3 El detector de borde de Prewitt.....	94
5.4.1.1.4 El detector de borde de Canny.....	97
5.4.1.1.5 El detector de bordes Laplaciano.....	99
5.4.1.2 Transforma de Fourier	101
5.4.2 Tratamiento de imágenes a nivel medio. Morfología matemática.....	103
5.4.2.1 Morfología binaria.....	105
(a) Binarización.....	105
(b) Dilatación	106
(c) Erosión.....	107
(d) Otras características morfológicas.....	111
(e) Componentes conexas conectadas.....	111
5.4.3 Parámetros obtenidos a partir de morfología matemática .	117
5.5 Resumen de parámetros	121

5.6 El coeficiente de correlación: Redundancia de parámetros 125

Capítulo 6	129
<i>Resultados de clasificación</i>	129
6.1 Los resultados del clasificador K-vecinas más cercanas (KNN).....	132
6.2 Optimización mediante algoritmos genéticos	140
6.3 Clasificación con redes neuronales artificiales (RNA)	144
6.4 Clasificación con ANFIS	147
6.5 Resumen de resultados	149
Capítulo 7	151
<i>Discusión y conclusiones</i>	151

Índice de siglas

a, b, c	<i>El conjunto de parámetros.</i>
A	<i>Una imagen binaria.</i>
AG	<i>Algoritmos Genéticos.</i>
AKB	<i>Keratosis actínica (Pre-cáncer).</i>
ANFIS	<i>Adaptive Neuro Fuzzy Inference System.</i>
B	<i>Un elemento estructurante.</i>
BAS	<i>Carcinoma basal (cáncer).</i>
BPN	<i>Backpropagation Net.</i>
CAD	<i>Computer Aided Diagnosis.</i>
c_k	<i>Las salidas objetivo.</i>
CR	<i>Tasa de Clasificación.</i>
D	<i>El diámetro de un círculo.</i>
DFT	<i>Transformada de Fourier Discreta.</i>
d_k	<i>Corresponden a los patrones de entrenamiento proporcionados (salidas deseadas).</i>
DRSI	<i>Diffuse Reflectance Spectroscopic Imaging.</i>
e(F)	<i>Función de la evaluación.</i>
E	<i>Número de los vecinos más cercanos de la clase ω_i.</i>
ELM	<i>Epiluminescence Microscopy.</i>
f(x)	<i>Función de densidad desconocida.</i>
F	<i>El conjunto de las características.</i>
FFT	<i>La transformada rápida de Fourier (Fast Fourier Transform).</i>
FN	<i>Falsos Negativos.</i>
FP	<i>Falsos Positivos.</i>
$F_y(x)$	<i>Función de distribución condicional a la clase.</i>
G-flip	<i>Greedy Feature Flip.</i>
g_k	<i>El gradiente actual.</i>
GTM	<i>Generativo topográfico.</i>
I(x, y)	<i>La magnitud del vector gradiente en la dirección de la máxima variación de la imagen en el punto (x, y).</i>
ICA	<i>Análisis de Componentes Independientes.</i>
KNN	<i>K vecinos más cercanos.</i>
$\mu(x)$	<i>Función de pertenencia.</i>
m	<i>Número del patrón del entrenamiento de la clase ω_i.</i>
M	<i>Número de las columnas de una imagen.</i>
MLP	<i>Perceptron multicapa.</i>
MM	<i>Melanoma Maligno.</i>
MRS	<i>La suma de errores cuadráticos.</i>
N	<i>Vectores del entrenamiento.</i>
N	<i>Número de las filas de una imagen.</i>
NOC	<i>Cerca del óptimo del clasificador (near óptima classifier).</i>
NOR	<i>Piel Normal.</i>
NV	<i>Pigmentado Nevé.</i>
P	<i>Medidas de valor real denominadas características.</i>

P	<i>Parámetro.</i>
P'	<i>Parámetro seleccionado.</i>
PCA	<i>Análisis de Componentes Principales.</i>
PSO	<i>Psoriasis.</i>
r	<i>Un coeficiente de una imagen.</i>
RNA	<i>Redes neuronales.</i>
S	<i>Conjunto del entrenamiento.</i>
SE	<i>Sensibilidad.</i>
SK	<i>Kerotasis Seborreica.</i>
SP	<i>Especificidad.</i>
Strel	<i>Structure element (elemento de estructura).</i>
SVM	<i>Support Vector Machines</i>
tansig	<i>Tangente sigmoide.</i>
TSK	<i>Mecanismo de inferencia Takagi-Sugeno.</i>
$V_k(x)$	<i>El volumen de un elipsoide centrado en x y de radio la distancia de x al k- vecino más cercano.</i>
VN	<i>Verdaderos Negativos.</i>
VP	<i>Verdaderos Positivos.</i>
w'_{kj}	<i>Los pesos de la capa de salida.</i>
w_k	<i>Un vector de pesos.</i>
w_{ij}	<i>Los pesos de la capa oculta.</i>
w	<i>Clases de objetos de interés.</i>
W	<i>Un vector de peso.</i>
x	<i>Vector de características.</i>
x_i	<i>Las n entradas de la red.</i>
y^{im}	<i>Implicado en el conflicto.</i>
y_j	<i>Las o salidas de la capa oculta.</i>
z_k	<i>Correspondiente salida de la red.</i>
z_k	<i>Las s salidas de la capa final.</i>

Otros símbolos

$\Delta\alpha$	<i>Una magnitud para actualizar parámetros de las funciones de pertenencia por el método descendente.</i>
η	<i>La tasa de aprendizaje.</i>
θ'_k	<i>Umbrales respectivos de la capa de salida.</i>
Θ	<i>La probabilidad de que un patrón x caiga dentro de volumen V centrado en un patrón.</i>
θ_j	<i>Umbrales correspondientes de la capa oculta.</i>
α_k	<i>La tasa de aprendizaje (o tasa de entrenamiento).</i>
$\theta_{s/x}^w$	<i>La función de margen.</i>
$\alpha(x, y)$	<i>La dirección del vector gradiente.</i>

Capítulo 1

Introducción

En este capítulo se da una breve introducción del trabajo realizado en la tesis. En él se describen brevemente algunos sistemas de clasificación general y particularmente los de diagnóstico automático así como sus aplicaciones en diversos campos dentro de la medicina (cáncer de piel y otras patologías). También se esboza un esquema de toda la memoria e incluye un resumen de los contenidos de cada capítulo de esta tesis.

Este trabajo se centra en el desarrollo de esquemas de diagnóstico automático y flexible. Para ello se exploran distintas alternativas capaces de utilizar eficientemente la información de un grupo de casos “etiquetados” para el diagnóstico de otros casos nuevos. Este tipo de herramientas de diagnóstico pueden servir para diversas patologías. En este trabajo hemos explorado diferentes alternativas y finalmente las hemos utilizado en una aplicación de diagnóstico concreta. Los últimos capítulos de este trabajo presentan el desarrollo de un sistema automático de diagnóstico de patologías de piel basado en procesamiento de imágenes. El trabajo utiliza imágenes de fluoroscopia (que contienen información de la dermis profunda) de pacientes con diversas patologías. En esta tesis se realiza un estudio de diferentes parámetros extraídos mediante procesamiento de imágenes y se evalúa la importancia de las distintas características para el problema de diagnóstico de cáncer de piel. Para ello, en vez de acometer un estudio estadístico convencional buscando correlaciones y grados de significancia entre las características concretas, se desarrolla un sistema de clasificación automático modular y se optimiza con esquemas de búsqueda combinatoria y algoritmos genéticos (búsqueda multi-camino). De esta forma avanzamos un paso más allá del simple estudio de significancia de las características extraídas diseñando un sistema de diagnóstico automático de cáncer de piel basado en imágenes de fluoroscopia. Además, el sistema de clasificación y su esquema de optimización son generales y constituyen una aplicación biomédica de diagnóstico que puede ser aplicada a otras patologías.

Para la implementación de un sistema automático en la ayuda a la clasificación y decisión en problemas relacionados con la medicina, se pretende desarrollar un sistema inteligente que hibride diferentes

paradigmas de lo que hoy en día se denomina *soft-computing*. En principio se utiliza un sistema clasificador de las diferentes patologías, y se optimiza con algoritmos evolutivos responsables de seleccionar de forma acertada las características más relevantes del sistema clasificador.

La clasificación es la atribución de una clase específica a un objeto. Esta atribución, necesita un cierto grado de abstracción para poder extraer generalidades a partir de los ejemplos disponibles. Para una máquina, la clasificación de rostros, de datos médicos o de formas son tareas bastante difíciles, mientras que para un humano son cuestiones cotidianas. Por ejemplo, en el caso de reconocimiento de caracteres manuscritos, es difícil enunciar una descripción general que tenga en cuenta todas las variaciones particulares de cada carácter. Una técnica que puede ser utilizada para resolver este problema es el aprendizaje. Así, el criterio para decidir si una imagen corresponde a una letra "A" consiste en comparar si esta imagen es lo suficientemente similar a otras "A"s vistas anteriormente; con este enfoque, uno no calcula la clasificación de letras, sino que se aprenden a partir de ejemplos etiquetados [TOR98].

El aprendizaje -desde un punto de vista pragmático- consiste en la adaptación de los parámetros de un sistema (sea artificial o natural) para obtener una respuesta deseada frente a un estímulo externo. Esta definición amplia del aprendizaje puede ser formalizada con el paradigma del aprendizaje supervisado. Este ajuste o aprendizaje se puede realizar por técnicas iterativas de minimización de un costo (cuantificación de los errores en las respuestas). De lo único que se dispone son datos en forma de pares de entrada-salida a los que llamaremos ejemplos. Si se posee un cierto número de estos ejemplos,

entonces se tiene un *Conjunto de Aprendizaje*. Un clasificador se construye a partir de un conjunto de aprendizaje.

La clasificación supervisada se caracteriza por tener unas clases ya determinadas y objetos caracterizados por atributos continuos o discretos que pertenecen a dichas clases. Al contrario, en la clasificación no supervisada, no se tiene las clases determinadas sino que se van creando de acuerdo a las características de los objetos, es decir, los objetos que más se parezcan pertenecerán a una misma clase. Se da una breve introducción de los diferentes algoritmos de clasificación supervisada tales como K vecinos más cercanos (KNN), redes neuronales y redes neuro-difusas en el capítulo tres, así como también los criterios de evaluación. Para evaluar la eficiencia de un algoritmo es importante considerar los siguientes aspectos: la tasa de error, la rapidez en clasificar, la interpretabilidad y la simplicidad del modelo.

Dentro de los métodos de evaluación en el marco de una aplicación de diagnóstico se consideran el acierto, la sensibilidad que en el área médica es importante para medir la proporción de verdaderos positivos y la especificidad que mide la proporción de verdaderos negativos. Para la evaluación se divide aleatoriamente la base de datos en dos grupos, uno para entrenar y el otro de prueba. La base de datos que hemos utilizado para la aplicación de diagnóstico consiste de 167 imágenes de fluoroscopia de tres tipos de lesiones de piel: 50 imágenes de keratosis actínica, 50 imágenes de carcinoma basal y 67 imágenes de psoriasis. Cada imagen ha sido procesada para extraer un total de 75 características (descritas en el Capítulo 5). Para evitar características altamente redundantes hemos extraído la matriz de coeficientes de correlación y hemos descartado características con un coeficiente de correlación mayor de 0.99 con respecto a otra

característica. De esta manera solamente 39 de las características iniciales se han seleccionado para su estudio posterior.

Por otra parte, la importancia del proceso de selección de características en cualquier problema de clasificación, se pone de manifiesto puesto que permite eliminar las características que puedan inducir a error (características ruidosas), las características que no aporten información (características irrelevantes) o aquellas que incluyen la misma información que otras (características redundantes) [BLU97]. Este proceso tiene varias ventajas: una disminución en los tiempos de procesamiento de los datos, un menor requerimiento de espacios para almacenar la información, un menor costo de obtención de los datos (la definición de características específicas permite desarrollar sensores adaptados para obtenerlas). Por lo tanto es muy importante la selección de un subconjunto de las características originales que aporten la mayor cantidad de información para un problema en particular. En este trabajo se presentan dos metodologías para seleccionar las características más importantes para fines de clasificación: la exploración secuencial (que utiliza las técnicas hacia delante y hacia atrás para seleccionar las mejores características) y la exploración multi-camino, (utiliza algoritmos genéticos para elegir las características más relevantes o significativas).

Se consideran cuatro modelos de la clasificación. 1) El método de K vecinos más cercanos (KNN) usando la técnica de la exploración secuencial (hacia delante “*forward method*” y hacia atrás “*backward method*”) para selección de los variables que conducen al mejor resultado de clasificación. 2) El segundo modelo utiliza KNN combinado con la técnica de la exploración multi-camino (algoritmos genéticos) para elegir las características que conducen al mejor

resultado de clasificación. 3) El tercero utiliza una red neuronal, concretamente el Perceptron Multicapa (MLP) y el proceso de selección de características de entrada se realiza por medio un algoritmo genético. 4) El cuarto modelo se basa en un modelo neuro-difuso ANFIS.

Los mejores resultados fueron obtenidos utilizando KNN optimizado con algoritmo genético y el ANFIS.

El trabajo tiene por objetivo mostrar cómo diferentes algoritmos de clasificación supervisada dan buenos resultados al aplicarlos en el dominio de la medicina. Estos esquemas también se pueden aplicar en diferentes dominios tales como: finanzas, reconocimiento de voz, verificación de firmas, clasificación de cromosomas humanos para la detección de anomalías, detección de fraudes, etc.

Pero más concretamente, el sistema desarrollado se ha validando utilizándolo en una aplicación real. Los diferentes algoritmos de inducción de modelos se han aplicado a la detección de las lesiones de la piel. Por lo tanto, este trabajo tiene también los siguientes objetivos concretos:

A- Aplicaciones biomédicas genéricas:

- Desarrollo de un sistema de clasificación modular para aplicaciones de diagnóstico.
- Estudio de esquemas de optimización de sistemas de diagnóstico modulares.
- Desarrollo de sistemas evolutivos para la selección de las características más relevantes en un sistema de diagnóstico.

B- Diagnostico de cáncer de piel:

- Evaluación de diversas características extraídas mediante procesamiento de imágenes de fluoroscopia para el diagnóstico de cáncer de piel.
- Implementación de un esquema de diagnóstico de cáncer de piel basado en procesamiento de imágenes de fluoroscopia.

A continuación se hace una breve descripción de la estructura de esta memoria haciendo referencia al contenido de los capítulos que la componen:

Capítulo 2: *Antecedentes en sistemas de diagnóstico para la clasificación de enfermedades de piel mediante técnicas de procesamiento de imágenes.* En este capítulo se presenta una breve discusión de trabajos similares y anteriores. En concreto este capítulo se centra en antecedentes de sistemas de clasificación de enfermedades de piel mediante técnicas de procesamiento de imágenes.

Capitulo 3: *Métodos de clasificación.* En este capítulo se describen los métodos de clasificación utilizados en este trabajo de investigación, estos métodos están basados en los algoritmos de K vecinos más cercanos (KNN), red neuronal perceptron multicapa (MLP) y el modelo red neuro-difuso (ANFIS).

Capitulo 4: *Estrategia de optimización.* En este capítulo se describen los métodos de selección de parámetros para mejorar los clasificadores. Se describen los métodos secuenciales (hacia delante *Forward* y hacia atrás *Backward*) y el método de búsqueda multi-camino (Algoritmos Genéticos).

Capítulo 5: *Aplicación a diagnóstico de patologías de la piel.* En este capítulo se presenta el problema de diagnóstico de la patología de la piel. Se extrae un vector de características de cada imagen, el material base del que partimos son imágenes por lo que esta etapa se centra en el procesamiento de imágenes digitales. Para caracterizar una imagen digital se buscan rasgos comunes entre las clases y utilizamos distintas técnicas de procesamiento. Una vez finalizada la fase de extracción de características se pasa a la reducción. Se observa que existen parámetros o características redundantes, el coeficiente de correlación entre características se utiliza para comprobar si hay dos o más características que tienen la misma información.

Capítulo 6: *Resultados de la clasificación.* En este capítulo se experimenta con diferentes configuraciones para llegar a nuestro objetivo de diagnóstico eficiente de cáncer. Con las pruebas verificamos lo planteado en la especificación del problema. Del análisis de los resultados realizados en este capítulo, se comprueba que es difícil discriminar entre Keratosis Actínica (pre-cáncer) y Carcinoma Basal (cancer), con este tipo de imágenes y características. Pero es sencillo discriminar entre psoriasis y las patologías cancerosas consideradas.

Capítulo 7: *Discusión y conclusiones.* Se enumeran y comentan las conclusiones y principales aportaciones que se derivan de la investigación realizada.

Capítulo 2

Antecedentes en sistemas de diagnóstico para la clasificación de enfermedades de piel mediante técnicas de procesamiento de imágenes.

Este capítulo representa una revisión de trabajos anteriores relacionados con esquemas de diagnóstico de lesiones de piel basados en técnicas de procesamiento de imágenes. Existen diversas metodologías que se referencian en la introducción de este capítulo. Por otro lado, los trabajos más relacionados (que utilizan técnicas muy similares a la de esta tesis o la misma base de datos) se comentan brevemente resaltando sus resultados y conclusiones finales.

2.1 Introducción

El diagnóstico preciso de lesiones de piel es difícil, incluso la discriminación entre lesiones malignas y benignas ya que son visualmente similares. El diagnóstico correcto de Carcinoma Basal es aproximadamente de un 65% cuando lo realizan dermatólogos practicantes [PRE87]. Médicos dermatólogos especialistas diagnostican correctamente melanomas malignos de un 63% [LIN94] a un 75% [KOH89] de los casos. De hecho, dermatólogos con menos de un año de experiencia sólo diagnostican lesiones de melanoma maligno correctamente en un 31% de los casos [LIN94].

Existen muchos trabajos relacionados con este tema, La mayoría de ellos adoptan el mismo esquema de clasificación: 1) extracción de características relevantes de las imágenes de la base de datos, 2) construcción de un sistema de clasificación automático basado en estas características (KNN, redes neuronales u otros sistemas de clasificación) y 3) análisis los resultados.

La mayoría de los trabajos relacionados con el diagnóstico basado en procesamiento de imágenes se centran en la discriminación del melanoma maligno [SCH03, SIG04, MEH02, KUS98] de otras lesiones de piel. La mayoría de estos sistemas se centran en distinguir lesiones de melanomas malignos de otras imágenes de piel sana, obteniendo muy buenos resultados. Por ejemplo en [GAN01] se definen tres grupos de lesiones: a) lesiones benignas que no requieren intervención quirúrgica (tres dermatólogos diferentes etiquetan estas lesiones como benignas), b) lesiones “displásticas” [ELD82, SOB83] que también son benignas pero pueden ser precursoras de melanomas

y c) melanomas malignos. Este trabajo obtiene como resultados relevantes una sensibilidad de un 87% y especificidad de 92% cuando sólo se trata de discriminar entre dos grupos: lesiones que requieren intervención quirúrgica (en la misma clase se incluyan melanomas malignos y lesiones displásicas, es decir los grupos b y c comentados anteriormente) y lesiones benignas que no requieren biopsia (grupo a definido anteriormente). Sin embargo cuando tratan de distinguir entre lesiones de tipo b y c (definidos anteriormente) los resultados son muy pobres (61% de acierto de clasificación). El sistema de clasificación que utiliza es el modelo KNN con selección secuencial de características [PUD94].

Las lesiones de tipo keratosis actínica y carcinoma de basal son muy difíciles de discriminar [MIE01, ABE04, SIG04, ODE05] como comentaremos en las siguientes secciones.

En cuanto al tipo de esquema de clasificación, muchos autores han utilizado distintos métodos de clasificadores como: KNN [MIE01, MEH02, ODE05, GAN01, SBO01, ROS95], análisis de discriminantes [GRE91, GRE94, POM99], redes neuronales [SIG04, COM99, HAS03, KUS98, SBO04, MAG04, BIN94, ERC94, ERC99, HIN96, DRE01, RUB02, PIC02, BIN98, KJO95, BIN94b, BOS93, DUR93, UMB91], regresión logística [DRE01, ALT91] y árboles de decisión [DRE01]. De hecho existen trabajos que explícitamente comparan distintas técnicas de clasificación para lesiones de la piel [DRE01], los resultados de este trabajo se resumen brevemente en la Sección 6 de este capítulo.

El proceso de extracción de características relevantes de las imágenes es crucial y puede involucrar distintas técnicas de procesamiento de imágenes, por ejemplo PCA e ICA [MIE01, KUS98, STO04], espectroscopia de diferentes tipos (como Raman) [SIG04, MEH02], CAD [SCH03, SBO04, COM99] (véase Sección 5 de este capítulo), y otros métodos de procesamiento de imágenes básicos (como extracción de bordes, contrastes, etc.).

El presente trabajo se centra en el procesamiento de imágenes de fluoroscopia [ACK01]. Existen otros trabajos que tratan de evaluar si esta técnica proporciona información válida para la detección de cáncer de piel, en la siguiente sección revisamos brevemente algunos trabajos relacionados con este tópico.

2.2 Análisis de Componentes Independientes ICA para clasificar lesiones de piel

Mies et al. [MIE01] y Stockmeier [STC04] utilizan la misma base de datos que en este trabajo de investigación (véase el Capítulo 5) lo cual facilita la comparación de sus resultados con los nuestros.

Mies et al. [MIE01] han utilizado en su trabajo técnicas como PCA “Análisis de Componentes Principales”, ICA “Análisis de Componentes Independientes”, e ICA local para extraer características relevantes de imágenes de fluoroscopia. Este trabajo se basa en la misma base de datos que utilizamos nosotros con tres tipos de lesiones: a) keratosis actínica, b) carcinoma basal y c) psoriasis.

En este trabajo se han utilizado tres métodos para clasificar las lesiones de piel:

- a. *PCA*: Utilizando PCA para extraer un “vector característico” de cada imagen de entrada y realizando la clasificación con el método de las KNN.
- b. *ICA*: Calcula las componentes independientes (ICA) de los “vectores característicos” anteriormente extraídos, y utiliza el clasificador sólo con esas componentes independientes.
- c. *ICA local*: Utiliza el modelo generativo topográfico (GTM) para extraer las características locales independientes (ICA-local) y utiliza el clasificador sólo con estas componentes.

Con estas técnicas se concluye que distinguir las lesiones de piel (keratosis actinica y carcinoma basal) de psoriasis es sencillo, pero por otro lado la discriminación entre keratosis actínica y carcinoma basal es muy difícil. De hecho en este trabajo se obtienen resultados muy pobres en esta segunda tarea como se muestra en la siguiente tabla [MIE01]:

Algoritmo	Psoriasis (%)	Carcinoma basal (%)	Kerotosis actinica (%)
PCA	82	68	54
ICA	87	72	55
ICA local	81	69	66

Tabla 2.1 *Acierto de clasificación resultados del trabajo Mies et al [MIE01].*

Por otra parte Kusumoputro et al. [KUS98] utiliza en su trabajo la técnica PCA para reducir el número de características extraídas de las imágenes seleccionando sólo características ortogonales que contienen la mayoría de la información extraída de las imágenes. Esto es importante para la red neuronal (perceptron multicapa, MLP) que utiliza como clasificador.

Este trabajo se centra en la discriminación entre lesiones de piel de tipo melanoma maligno de lesiones benignas. La base de datos que han utilizado está compuesta de 63 imágenes en color (29 casos de melanoma maligno y 34 de lesiones benignas). Los resultados de clasificación que se obtienen son $91,83 \pm 1,34\%$ de clasificación correcta sobre todos los casos (utilizando validación cruzada con tres divisiones de entrenamiento y test). Este trabajo estudia cómo se va degradando la capacidad de clasificación de la red neuronal cuando se reduce el número de características, y resalta que las características más significativas son de color.

2.3 Imágenes espectroscópicas de reflectancia difusa en incidencia oblicua para clasificar lesiones de piel

Mehrubeogul et al. [MEH02] han utilizado en su trabajo este tipo de imágenes para clasificar y distinguir las lesiones de la piel. Los autores han definido dos grupos: a) (grupo 1) tiene dos tipos de lesiones benignas (keratosis actínica y carcinoma basal) y b) (grupo 2) compuesto por dos tipos de lesiones que se consideran precursoras de potenciales cánceres malignos (nevé común y nevé displástico).

Este trabajo adopta un proceso complejo para la adquisición de imágenes, Se construye un sistema con “*Diffuse Reflectance Spectroscopic Imaging*” (DRSI) para controlar la longitud de onda y el ángulo de reflexión con que se iluminan las zonas para captar la imagen. El trabajo explora distintas características (que se pueden extraer con esta técnica de iluminación) y evalúa su capacidad de discriminación. El objetivo del trabajo es doble, por un lado validar este sistema de captación de imágenes (DRSI) y por otro lado indicar el tipo de características relevantes para la diagnosis de lesiones de piel.

Estas características son:

1. Características estadísticas sencillas (media, varianza, asimetría, energía, entropía, etc).
2. Características de textura (basados en un filtrado espacial de la imagen, para la descomposición en componentes específicas o “*wavelets*”).
3. Características introducidas basadas en técnicas fractales.
4. Componentes principales (PCA).

Utilizando un clasificador Bayesiano con técnicas de validación cruzada (“dejar uno fuera” como se explica en el capítulo 4 de esta memoria) obtienen buenos resultados: 95.7% de acierto de discriminación en el grupo 1 (especificidad) y 94.1% para el grupo 2 (sensibilidad).

2.4 Espectroscopia Raman para clasificar lesiones de piel

Sigurdsson et al. [SIG04] han utilizado un sistema de clasificación basado en redes neuronales y espectroscopia Raman. Las componentes espectrales de Raman se obtienen irradiando con láser la muestra. El haz láser excita las moléculas produciendo cierta diseminación de energía y un desplazamiento en las componentes espectrales reflejadas. Este desplazamiento depende del tipo de moléculas excitadas en la muestra, por lo que la técnica de espectroscopia Raman es útil para la clasificación de componentes químicos y como demuestra este trabajo puede ser de utilidad para el diagnóstico de lesiones de piel [SIG04, GNI97, GRI95]. Esta técnica se ha usado en otras aplicaciones biomédicas [SCH95, SHI97].

El trabajo incluye un estudio sobre distintas lesiones de piel: carcinoma basal (BCC) 48 muestras, melanoma maligno (MM) 21 muestras, piel normal (NOR) 89 muestras, pigmentado nevé (NV) 41 muestras y kerotasis seborreica (SK) 23 muestras. NV y SK son lesiones de menor importancia, BCC y MM son lesiones de mayor gravedad (aunque cabe insistir en que BCC es un cáncer benigno).

Para la extracción de características los autores segmentan la zona de la lesión y utilizan la técnica PCA. En este trabajo se ha usado una red neuronal para la clasificación obteniendo resultados de clasificación de MM con $80.5\% \pm 5.3\%$ de acierto, BCC con

95.8% \pm 2.7% y el acierto total sobre todas las lesiones de piel incluidas en el estudio de 94.8% \pm 3.0%.

2.5 Sistema CAD (Computer Aided Diagnosis) para clasificar la lesiones de piel con pigmentación

En este trabajo Schmid-Saugeon et al. [SCH03] se han utilizado las imágenes obtenidas por la técnica ELM (*Epiluminescence Microscopy*) [STO94]. La evaluación visual de imágenes de este tipo es compleja por la gran cantidad de información que contienen. Por ello, existen protocolos para la ayuda al diagnóstico con este tipo de imágenes [STO94] que indican características relevantes para la diagnosis correcta de melanoma maligno: a) asimetría en uno o dos ejes perpendiculares teniendo en cuenta el color, textura y forma de la lesión, b) bordes radiales para la evaluación del grado de contraste y su distribución, c) Color, se deben detectar 6 colores característicos de este tipo de lesiones (blanco, rojo, marrón claro, marrón oscuro, azul pizarra y negro) y d) estructuras diferenciales (patrones de formas concretas).

Los autores utilizan una base de datos reducida de 100 muestras: 50 imágenes de melanomas malignos y 50 imágenes de nevé benigno (de los cuales 20 son de nevé displástico). El nevé displástico es una lesión que muchos especialistas consideran de cierta gravedad ya que tiene riesgo de convertirse en melanoma maligno [STO94]. El estudio define dos tipos de problemas:

A) discriminación entre las 50 imágenes de melanoma maligno y las 50 de lesiones de tipo nevé.

B) discriminación entre las 50 imágenes de melanoma maligno y las 30 de nevé (excluyendo las de nevé displástico).

En este trabajo se han centrado en la evaluación de distintos tipos de características (textura, textura-color y textura-color-forma) para el diagnóstico de lesiones de piel. Concretamente, utiliza un sencillo clasificador lineal para evaluar el tipo de características que son más significativas.

Los resultados que han conseguido en este trabajo se dividen a tres grupos según de las características que han utilizado en el proceso de la clasificación:

1. Para tres combinaciones de las características: textura (2D), textura y color (4D), y textura, color, y forma (6D). El acierto de la clasificación entre las lesiones MM y nevé (incluyendo las lesiones de nevé displástico) son:

	Textura	Textura y color	Textura, color y forma.
Sensibilidad	66%	64%	78%
Especificad	90%	92%	90%
Acierto de la clasificación	60%	59.3%	72.2%

Tabla 2.2 *Los resultados de la clasificación para el problema A (definido anteriormente).*

2. Resultados que han conseguido para el problema B (segundo grupo). Se muestran en la siguiente tabla:

	Textura	Textura y color	Textura, color y forma.
Sensibilidad	76%	78%	82%
Especificad	93.3%	80%	80%
Acierto de la clasificación	69.1%	69.6%	73.2%

Tabla 2.3 *Los resultados del segundo grupo (problema B).*

3. Resultados que han conseguido; incluyendo las muestras del problema A (tercer grupo). Cuando utiliza análisis de componentes principales (PCA) en niveles de gris.

	Binario	Nivel de gris
Sensibilidad	60%	70%
Especificad	82%	92%
acierto de la clasificación	50.8%	64.8%

Tabla 2.4 *Los resultados del problema A cuando utiliza PCA.*

2.6 Métodos de clasificación para diagnóstico de lesiones de piel

Aunque existen diversas técnicas de clasificación que se han utilizado para esta tarea de diagnóstico (ver sección de introducción de este capítulo). El trabajo de Dreiseitl et al. [DRE01] se ha centrado en la comparación entre diferentes métodos de clasificación para distinguir tres tipos de las lesiones de piel pigmentada: nevé común (1290 imágenes), nevé displástico (224 imágenes) y melanoma maligno (105 imágenes). Los métodos que se han comparado en este estudio son los siguientes:

1. K vecinos más cercanos.
2. Regresión logística.
3. Redes neuronales.
4. Árboles de decisión,
5. Red neuronal de tipo SVM (*support vector machines*).

De los resultados que han conseguido de este trabajo, Dreiseitl et al. [DRE01] indica que los cinco métodos de clasificación se pueden utilizar para lesiones de piel. Entre estos métodos, los árboles de decisión son los que peores resultados generan. Las K vecinas más próximas obtienen resultados buenos, aunque los otros tres métodos la superan (regresión logística, redes neuronales y SVM).

En todo caso, es necesario indicar que en este trabajo los resultados obtenidos de redes neuronales son buenos porque el número de muestras es grande. En muchas aplicaciones biomédicas sólo se dispone de un número muy reducido de casos etiquetados por lo que

las K vecinas más próximas es un método de clasificación muy válido.

2.7 Contexto

Como se ha comentado, existen muchos tipos de esquemas de clasificación que se pueden aplicar a problemas de diagnóstico, redes neuronales, k -vecinos más cercanos, discriminadores lineales, árboles de reglas de decisión, etc. En particular en este proyecto nos centramos en clasificadores modulares (es decir, que puedan fácilmente incorporar/eliminar características) de forma que son susceptibles de ser optimizados fácilmente habilitando/deshabilitando las características en las que el sistema basa la decisión de diagnóstico. La utilización de esquemas de búsqueda secuencial combinatoria [ROS04, HAN81] y algoritmos genéticos [MOT03] para optimizar esquemas de diagnóstico es un campo de investigación interesante porque además sirve para resaltar la importancia (relevancia) relativa de características concretas en la tarea de diagnóstico. El capítulo 6 de esta memoria aplica este tipo de técnicas al problema de diagnóstico de lesiones de piel.

Capítulo 3

Métodos de clasificación

En este capítulo se describen los métodos de clasificación utilizados en esta memoria de investigación, estos métodos están basados en los algoritmos de K vecinos más cercanos (KNN), red neuronal perceptron multicapa (MLP) y el modelo red neuro-difuso (ANFIS).

3.1 Introducción

Tareas de clasificación tienen lugar en un extenso campo de acción de la actividad humana. En su concepto más amplio el término *clasificación* podría incluir cualquier contexto en el cual alguna deliberación o pronóstico es hecho sobre la base de una información disponible, y un *procedimiento de clasificación* es entonces algún método formal para repetidamente realizar tales valoraciones ante nuevas situaciones. El problema de clasificación de patrones es formulado en la siguiente manera. Hay J clases de objetos de interés, los cuales utilizarían el subíndice j con $j = 1, \dots, J$, para cada estado respectivamente. La información que poseemos sobre los objetos es resumida en un número, p , de medidas de valor real denominadas *características* (“features”). Todas juntas forman un *vector de características* $x \in R^p$ es decir $x = (x_1, x_2, \dots, x_p)$ [KUL98, DEV96]. Se puede asumir que el problema tiene que ver con la construcción de un procedimiento que será aplicado a una secuencia de *casos*, en el cual cada nuevo caso debe ser asignado a una clase de un conjunto predefinido de ellas sobre la base de un grupo de rasgos observados. Para modelar la relación entre el vector de características se asume que un objeto de la clase $y \in \{1, 2, \dots, J\}$ es una realización del vector de variables aleatorias con función de distribución condicional a la clase $F_y(x)$.

Muchos de los problemas surgidos en la ciencia, la industria y el comercio, que requieren el uso de datos complejos y a menudo de gran extensión, pueden ser considerados como problemas de clasificación o decisión. En su forma más simple, la clasificación se presenta como un proceso de reconocimiento de determinados

patrones. En este sentido contribuyen mucho el conjunto formado por disciplinas científicas tales como Medicina, la Bioinformática, las Ciencias Biológicas, las Geociencias, las Ciencias Sociales entre otras, debido a la cantidad de problemas relacionados con el Reconocimiento de Patrones que allí aparecen en toda su magnitud y complejidad. En estas disciplinas, denominadas ciencias poco formalizadas, se presentan problemas de clasificación, de diagnóstico, de pronóstico, de predicción, de determinación de factores de influencias y otros que pertenecen claramente a los que estamos aludiendo.

Los procesos de clasificación generalmente se presentan con dos enfoques diferentes. El primer enfoque trata de cómo a partir de un conjunto de observaciones establecer la existencia de clases o agrupaciones en éstas mientras que en un segundo enfoque se parte de un número conocido de clases y el propósito es establecer una regla con la cual podamos clasificar una nueva observación en una de las clases existentes [MIC94].

A partir de este análisis, se pueden identificar tres familias de problemas:

Clasificación con aprendizaje (o supervisada): se conoce como un universo de objetos que se agrupan en un número dado de clases. Cada una de las muestras de entrenamiento, se sabe a qué clase pertenece. Dada una nueva muestra, el dilema es establecer a qué clase pertenece con mayor probabilidad. Un ejemplo de esta variante de problemas lo constituye el diagnóstico diferencial de

enfermedades (al que se refieren los Capítulos 5 y 6 de esta memoria).

Clasificación con aprendizaje parcial: es análogo al supervisado, excepto en que hay al menos una clase de objetos de la que no se tiene ninguna muestra.

Clasificación sin aprendizaje (o no supervisada): no se conoce cómo se agrupan las entidades, es justamente el objetivo que se persigue.

El proceso de selección del conjunto de rasgos o características de las observaciones que nos permitan clasificar a estas en uno u otro grupo es un problema central en un proceso de clasificación. En general, en este punto dos directivas son fundamentales:

Reducir el número de rasgos en términos de los cuales se deben describir los objetos en modo eficiente.

Encontrar los rasgos que inciden de manera determinante en la clasificación.

Un ejemplo de encontrar rasgos que inciden en la clasificación de forma determinante es la identificación de los factores de riesgo de una enfermedad.

Otro problema esencial que se presenta en la clasificación es la identificación de las clases y la manera en que estas son definidas. Podemos distinguir tres casos frecuentes en este sentido:

Las clases corresponden a etiquetas para diferentes poblaciones, la pertenencia de las poblaciones no es cuestionada. Las clases son el resultado de un problema de predicción o agrupación.

Las clases son predefinidas por una partición del espacio de muestras.

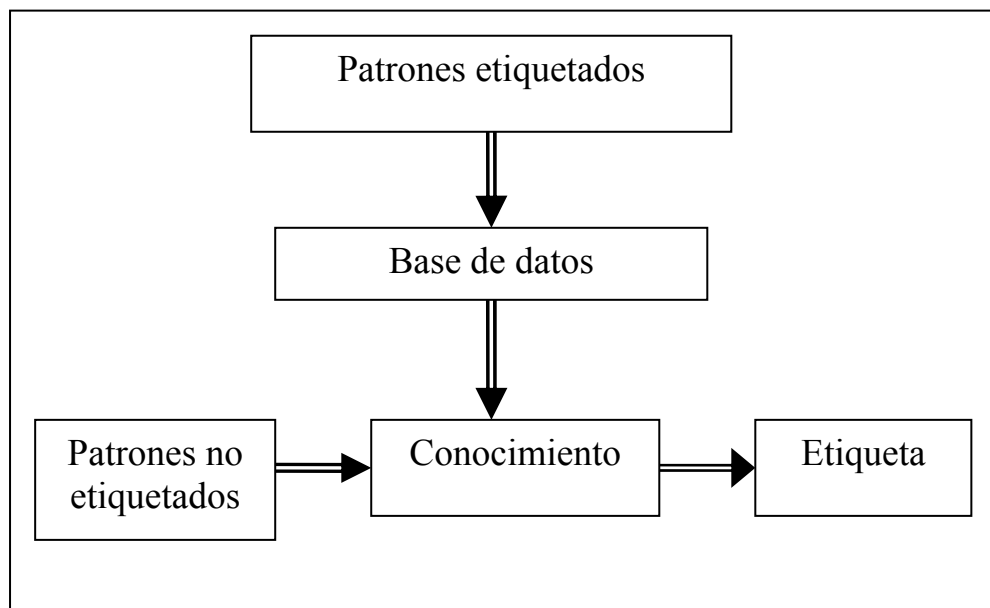


Figura 3.1 *Esquema de la clasificación.*

En este capítulo se describen algoritmos de clasificación supervisados que dan buenos resultados al aplicarlos en diferentes dominios tales como: medicina, finanzas, reconocimiento de voz, etc. Sin embargo el principal objetivo es comparar diferentes algoritmos de inducción de modelos aplicados a un dominio médico. Más concretamente, para la detección de cáncer de piel (Capítulos 5 y 6 de esta memoria).

El desarrollo de un sistema clasificador modular supervisado eficiente conlleva una serie de pasos. En principio señalar que se debe producir un entrenamiento previo del sistema, antes de que éste sea capaz de clasificar cualquier muestra. Es decir, el sistema debe observar una serie de muestras tomadas previamente, y de las que conocemos la clase a la que pertenecen cada una, para poder aprender a clasificar las muestras no etiquetadas (Fig. 3.1).

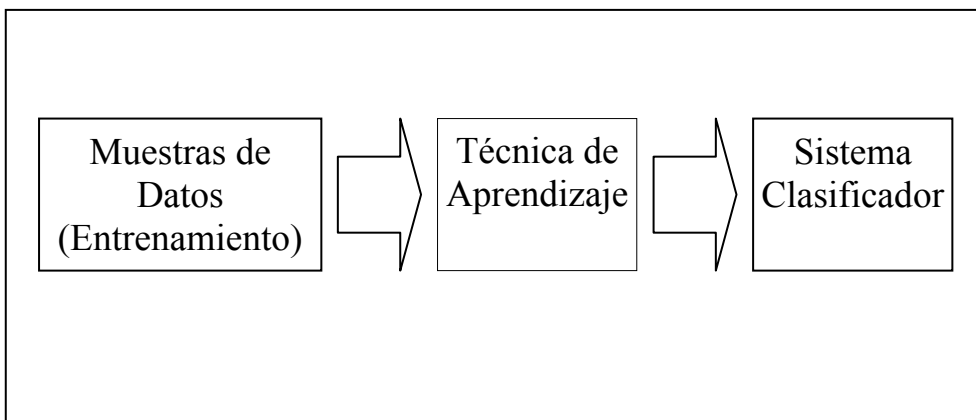


Figura 3.2 *Esquema básico de desarrollo de un clasificador con aprendizaje.*

El esquema básico que se sigue entonces hasta obtener el clasificador final es el mostrado en la Fig. 3.2. Es decir necesitamos disponer de una técnica de aprendizaje que partiendo de una serie de datos de entrenamiento sea capaz de generar un sistema clasificador de información. La técnica de aprendizaje lo que se busca es abstraer conocimiento representativo de entre todas las muestras de entrenamiento que se le pasan al sistema.

La clasificación supervisada se caracteriza por tener una clase ya determinada y objetos caracterizados por atributos que indican la pertenencia a dicha clase. Al contrario de la clasificación no

supervisada en la que no se tienen las clases determinadas sino que se van creando conforme a las características de los objetos. Es decir, los objetos que más se parezcan pertenecerán a una misma clase. En las siguientes secciones se describen diferentes algoritmos de clasificación supervisada tales como KNN (K vecinos más cercanos) y redes neuronales así como también los criterios de evaluación.

3.2 Descripción de un sistema de clasificación basado en KNN

El método de los K vecinos más cercanos (“*K- Nearest Neighbours*” o KNN) se ha utilizado como un instrumento importante del reconocimiento de patrones. El modelo de los vecinos más cercanos (*K- Nearest Neighbours*) ha sido descrito en muchas fuentes, por ejemplo por Theodoridis y Koutroumbas [THE99].

El método del vecino más cercano es de utilidad para reconocer los datos donde las distribuciones de la clase tienen poco solapamiento y las características usadas son discriminatorias. Sin embargo, las distribuciones de los datos para varias clases que están solapadas requieren de más de un vecino cercano y se utiliza la votación de la mayoría.

En el método de K vecinos más cercanos, se toman ciertas decisiones a priori para mejorar el rendimiento del proceso de reconocimiento. La primera establece que los vectores de las características individuales; para varias clases; sean discriminatorios. El segundo prerrequisito, que se presupone requiere que la

característica única de un patrón no sea dependiente en la interacción entre las varias características [SIN99].

El modelo KNN propuesto se puede describir en los siguientes pasos:

- Con N vectores del entrenamiento (etiquetados), se identifica a K vecinos más cercanos independientemente de la etiqueta de la clase. K se elige que sea impar.
- Con las etiquetas de las muestras de K , se identifica el número de los vectores k_i , que pertenecen a cada clase $\omega_i, i=1,2,\dots,M$. obviamente $\sum_i k_i = K$.
- Asigna x a la clase ω_i con el máximo número k_i de las muestras (con esa misma etiqueta).
- Si hay dos o más clases $\omega_i, i \in [1\dots M]$ tienen el mismo número igual E del máximo vecinos más cercanos, en este caso tenemos un conflicto.
 - Para cada clase implicada en el conflicto, se determina la distancia d_i entre el patrón del prueba $x = \{x_1, \dots, x_N\}$ y las clases ω_i basado a E vecinos más cercanos del clase ω_i . Si el patrón del entrenamiento m de la clase ω_i implicado en el conflicto se representa como $y^{im} = \{y_1^{im}, \dots, y_N^{im}\}$ entonces la distancia entre el patrón de prueba x y el clase ω_i es:

$$d_i = \frac{1}{E} \sum_{j=1}^E |x_j - y_j^{im}| \quad (3.1)$$

- Asigna x a la clase C si la distancia d_i es menor, es decir, $x \in \omega_c$, si $\omega_c < \omega_i$, para $i, C \in [1 \dots M]$ y $i \neq C$.

Es decir, los métodos de los vecinos más cercanos trabajan mejor con datos donde las características son estadísticamente independientes. Estos métodos se basan en una forma de medida de la distancia y la detección de las vecinas más cercanas de los datos de prueba no es dependiente de la interacción de las características.

El algoritmo KNN es un método de clasificación de vectores nuevos, es decir, es un método en el que el sistema computacional contiene una serie de datos repartidos en clases, que han sido previamente almacenados y clasificados correctamente por expertos humanos, para una tarea determinada y que constituye la base de conocimientos del sistema. En el método de KNN, sea x_1, x_2, \dots, x_n una muestra con función de densidad desconocida $f(x)$. Se estima $f(x)$ a partir de una celda de centro en x y que crece hasta capturar k elementos, donde k es definido arbitrariamente o como función de n . Estas muestras son las k vecinos más cercanos a x . Se tiene entonces:

$$f(x) = \frac{k/n}{V_k(x)} \quad (3.2)$$

Donde $V_k(x)$ es el volumen de un elipsoide centrado en x y de radio la distancia de x al k - vecino más cercano [FIX51].

La forma en que el algoritmo clasifica nuevos patrones (un nuevo objeto de estudio) es simple. Se encuentran, en la base de datos, los k -vecinos más cercanos al nuevo patrón a clasificar y después se somete a una *votación* para elegir la clase a la que

pertenece el nuevo patrón. Hay una alta probabilidad de encontrar un patrón en una región densamente poblada y baja en regiones poco pobladas en las que las observaciones están más dispersas. La probabilidad de que un patrón x caiga dentro de volumen V centrado en un patrón o (punto) x es dada por [WEB03]:

$$\theta = \int_{v(x)} p(x) dx \quad (3.3)$$

En el proceso de diseño del clasificador KNN, la fase de desarrollo incluye dos tareas: aprendizaje y validación. El conjunto de aprendizaje de este clasificador consiste en una tabla que almacena los patrones disponibles incluyendo la clase asociada a cada uno de estos patrones (es decir, la etiqueta de cada patrón). Para un nuevo patrón se calcula su distancia a los otros patrones de la tabla y se consideran los (K) casos más cercanos. Usualmente se utiliza la *distancia Eculídea* [BRE95]. Este nuevo patrón se clasifica con la clase mayoritaria de los K patrones etiquetados más cercanos.

El algoritmo KNN para clasificar patrones de dos clases distintas, se hará observando la distancia euclídea entre los patrones que representan las características de éstas clases. Así cuanto menor sea dicha distancia, más similares serán las muestras, y por tanto se puede considerar que representan a muestras de un mismo tipo. En la siguiente figura (Fig. 3.3) se muestra un ejemplo de dos clases.

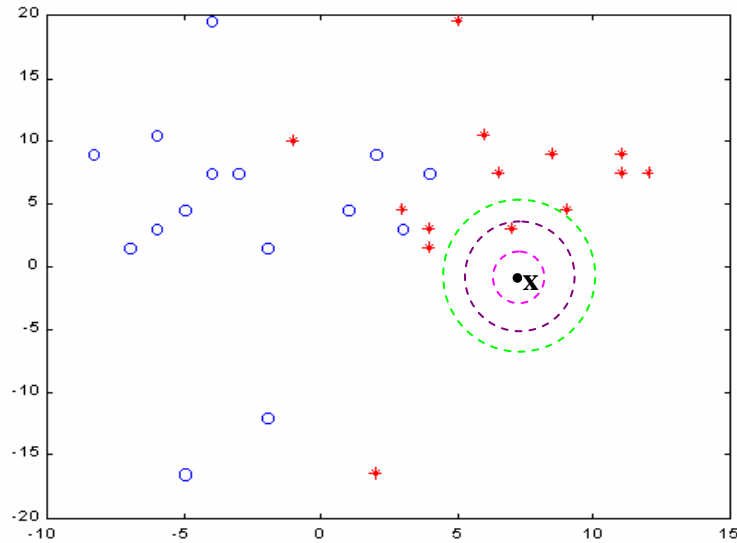


Figura 3.3 *Problema de dos clases.*

La métrica más comúnmente utilizada para la distancia de una muestra nueva de un prototipo es la distancia euclídea. Por consiguiente, puesto que todas las variables se tratan igualmente, las variables de la entrada se deben escalar para asegurarse de que la regla de KNN es independiente de unidades de la medida. Se utiliza la siguiente fórmula para calcular la distancia euclídea d entre los dos vectores x e y :

$$d = \sqrt{\sum (x - y)^2} \quad (3.4)$$

En la mayoría de estudios se utilizan valores distintos para K . En el capítulo 6 se utiliza este clasificador con $K = 1, 3, 5, 7, 9$.

Siempre se utiliza en número impar de vectores para la vecindad de forma que el esquema de votación siempre tiene una mayoría bien definida (si clasificamos en sólo dos clases).

Aquí, el clasificador 1NN clasifica el nuevo patrón como perteneciente a la clase del caso más cercano a él en la tabla de la base de datos etiquetados. Ese clasificador calcula la distancia euclídea con todos los patrones de la tabla de la base de datos etiquetados y asigna al patrón a clasificar (x) la clase del patrón más cercano encontrado como se puede ver en la siguiente figura.

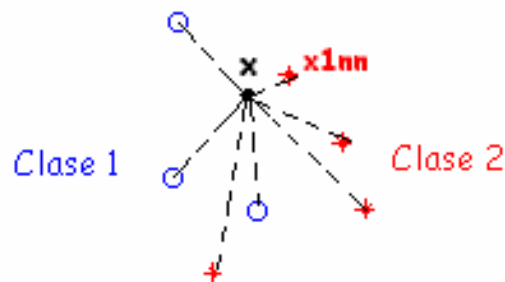


Figura 3.4 Clasificación 1NN. Clase 1 son los puntos (O) y la clase 2 son los puntos (+). El vector nuevo queda clasificado como de la clase 2.

Para los otros clasificadores (3, 5, 7, 9) NN se clasifica el nuevo patrón como perteneciente a la clase de la mayoría de los más cercanos a este patrón. Si la mayoría de esos nuevos patrones pertenecen a la clase 1 entonces el nuevo patrón considerado se etiqueta como de clase 1.

3.2.1 *El proceso de evaluación “dejar uno fuera” (leave one out)*

En el desarrollo de un clasificador se suelen utilizar dos conjuntos (aprendizaje y validación) que se explicarán en la siguiente sección. Si el número de muestras (patrones) disponibles es relativamente pequeño, se utiliza un procedimiento de validación cruzada denominado *dejar uno fuera*. En este procedimiento se deja cada vez una muestra fuera, utilizando el resto para el cálculo del modelo. Repitiendo el proceso un número de veces igual al número de muestras.

La decisión afecta negativamente al resultado de la clasificación cuando existen patrones erróneos cercanos al nuevo patrón. Con el algoritmo de las KNN, no es posible calcular el porcentaje de acierto sobre el conjunto de entrenamiento de un modo directo. En este caso se utiliza el procedimiento “*dejar uno fuera*” (*leave one out*) para calcular el error sobre el conjunto de vectores de entrenamiento. Este esquema consiste en tomar del conjunto de entrenamiento un patrón y quitarle la etiqueta “real”. Se aplica el algoritmo de las KNN con el resto de vectores de entrenamiento y se obtiene una etiqueta “calculada” que se compara con la etiqueta “real”. Esto se realiza tomando uno a uno todos los vectores del conjunto de entrenamiento para calcular el porcentaje de error de clasificación. Es decir, en cada iteración se extrae un único vector como patrón de test.

3.2.2 *Proceso de clasificación*

En el proceso de clasificación se divide el conjunto de ejemplos en dos subconjuntos disjuntos (Entrenamiento y Test) [ROS03]:

- 1- Entrenamiento: Este subconjunto tiene un mayor tamaño comparado con el de test y se emplea como referencia. Por ejemplo el 66 % del tamaño total del conjunto de ejemplos que se han elegido presentaría un buen tamaño para el conjunto de entrenamiento.

- 2- Test: Este subconjunto presenta un menor tamaño. Es el restante del conjunto de ejemplos que puede ser por ejemplo el 34% del tamaño total del conjunto de ejemplos. Normalmente se utiliza este subconjunto para evaluar el clasificador y para conocer su poder de generalización y de acierto. Por eso, se calcula el porcentaje de clasificación sobre los ejemplos de test (patrones desconocidos por el clasificador).

En la siguiente figura se puede ver las dos tareas del proceso del clasificador: entrenamiento y test.

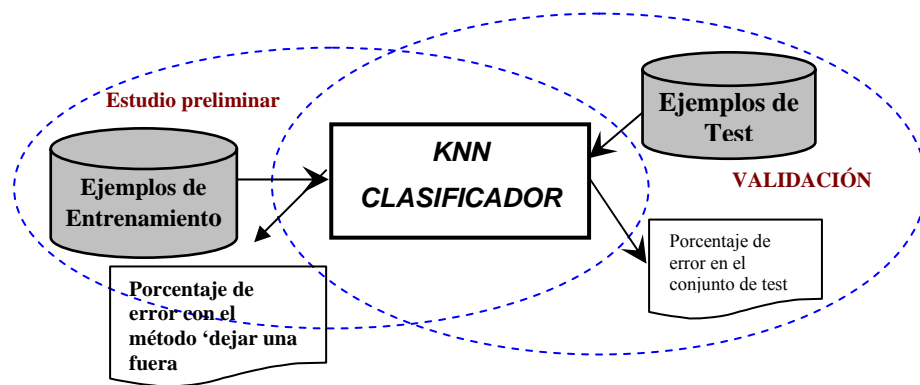


Figura 3.5 Esquema de aprendizaje y validación. Si se utiliza sólo el método de “dejar uno fuera”, los resultados son fiables, ya que se ha realizado una validación cruzada exhaustiva.

Se suelen realizar varias particiones del conjunto de ejemplos en entrenamiento y test, ya que se pretende evitar en la medida de lo posible dependencias de los cálculos de error con la partición realizada. Esto se denomina “validación cruzada”. En el caso del algoritmo KNN, se puede utilizar la estrategia de “dejar uno fuera” para evaluar el sistema de clasificación, como se ha explicado anteriormente.

3.3 Descripción de un sistema de clasificación basado en redes neuronales

3.3.1 Introducción

Una Red Neuronal Artificial (RNA) es un modelo computacional inspirado en redes neuronales biológicas que pueden ser consideradas como un sistema de procesamiento de información con características como aprendizaje a partir de ejemplos, adaptabilidad, robustez, capacidad de generalización y tolerancia a fallos.

Uno de los modelos de RNA más conocido y utilizado es el Perceptron Multicapa (*Multilayer Perceptron*, MLP), también conocido como Red de Retropropagación (Backpropagation Net, BPN). Con este modelo, se presentan repetidamente los datos de entrada a la red. Cada presentación de los datos de entrada a la red MLP se compara la salida de la red a la salida deseada y se computa el error. Este error después se propaga hacia atrás en la red MLP y se ajustan los pesos para que el error disminuya con cada iteración. De esta forma el modelo consigue cada vez una salida más cerca de la deseada. Este proceso se conoce como "entrenamiento".

La red neuronal artificial perceptron multicapa (MLP) usando el algoritmo de retropropagación ha sido aplicada a la solución de diversos problemas prácticos [HAY99]. El MLP consiste en al menos tres capas: una capa de entrada, una o más capas ocultas y una capa de salida. Aquí el número de neuronas de las capas de entrada y salida depende de cada aplicación en particular. Sin embargo, aunque el funcionamiento de la red depende en forma importante del número

de nodos en las capas ocultas, no existe aún un método fiable que permita determinar con precisión el número óptimo de estos [SAN04].

La estructura del MLP con una única capa oculta se muestra en las Figuras 3.6 y 3.7.

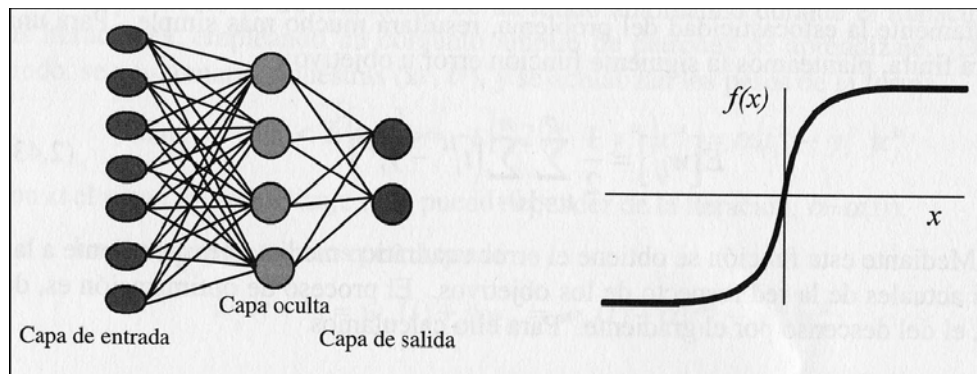


Figura 3.6 *Arquitectura (izquierda) y función de activación (derecha) para el MLP.*

Denotamos por x_i a las n entradas de la red, y_j a las o salidas de la capa oculta y z_k a las s salidas de la capa (final por tanto a las salidas de la red) las cuales deben de ser comparadas con las salidas objetivo c_k . Además, w_{ij} representan los pesos de la capa oculta, θ_j sus umbrales correspondientes, w'_{kj} los pesos de la capa de salida y θ'_k sus umbrales respectivos.

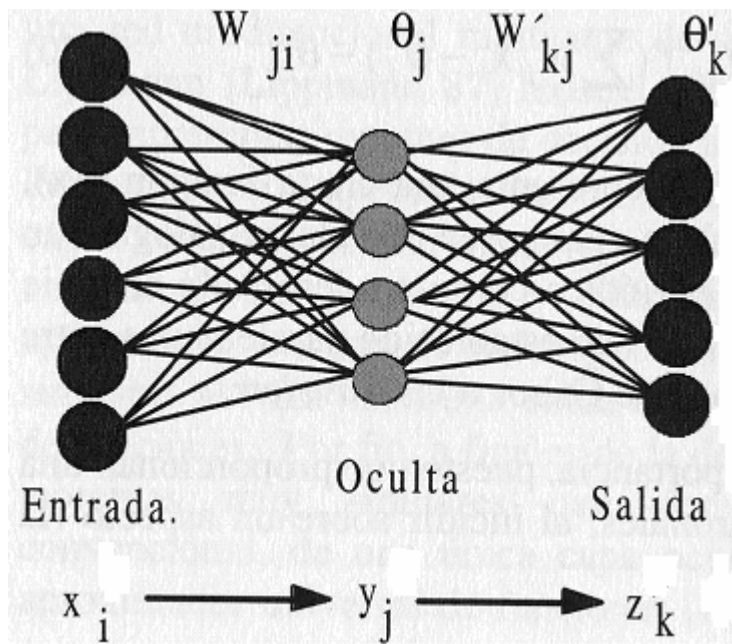


Figura 3.7 Arquitectura del MLP.

Las operaciones efectuadas por un MLP con una única capa oculta y con funciones de activación para la capa oculta y capa final de tipo *sigmoide* y lineal, son las siguientes:

$$z_k = \sum_{j=1}^0 w'_{kj} y_j - \theta'_k = \sum_{j=1}^0 w'_{kj} f \left(\sum_{i=1}^n w_{ji} x_i - \theta_j \right) - \theta'_k \quad (3.5)$$

Siendo $f(x)$ una función de tipo *sigmoide* o lineal (como se muestra en la Figura 3.6).

La popularidad de la arquitectura MLP se debe al hecho de que un MLP con una única capa oculta puede aproximar cualquier función continua en un intervalo hasta el nivel deseado, cuestión demostrada por Funahashi (1989) [FUN89] y que proporciona una base sólida en el campo de las redes neuronales, aunque el resultado no informa

sobre el número de nodos ocultos necesarios para llevar a cabo la aproximación de forma óptima.

3.3.2 *Arquitectura de la red*

Una red neuronal multicapa que hemos utilizado es:

- *La capa de entrada*: un conjunto de neuronas que reciben las señales del exterior.
- *La capa oculta*: las neuronas de esta capa reciben las señales de la capa de entrada y realizan una transformación local y no lineal sobre dichas señales. Esta parte tiene dos capas ocultas:
 1. *primera capa oculta*: constará con el doble de entradas de la capa de entrada.
 2. *segunda capa oculta*: constará con el mismo número de entradas que la capa de entrada.
- *La capa de salida*: mantiene la respuesta de la red.

3.3.3 *Entrenamiento de la red*

La arquitectura multicapa suele entrenarse con algoritmos de retropropagación. En esta red, se interconectan varias unidades de procesamiento en capas. Cada neurona de una capa proporciona una entrada a cada una de las neuronas de la siguiente capa.

El término retropropagación se refiere al método para calcular el gradiente de error en una red hacia delante, que es una aplicación

de la regla de la cadena de cálculo elemental. El entrenamiento consiste básicamente en lo siguiente:

1. Pasada hacia delante (*forward pass*): las salidas son calculadas y el error en las unidades de salida es calculado.
2. Pasada hacia atrás (*backward pass*): el error de las salidas es utilizado para alterar los pesos de las unidades de salida. Luego el error en las neuronas de las capas ocultas es calculado mediante propagación hacia atrás del error en las unidades de salida y los pesos en las capas ocultas son alterados usando esos valores.

Muchos modelos de redes neuronales emplean propagación hacia atrás para los diagnósticos médicos. Ya que estos modelos han logrado resultados buenos con esta técnica. La técnica de propagación hacia atrás se basa en la forma en que el gradiente es computado para redes multicapa no lineales. Hay muchas variantes de este algoritmo básico que se basan en otras técnicas de optimización estándar, como la del *gradiente conjugado* y *método de Newton* [PAI82, DEN83].

Es tarea difícil el decidir qué algoritmo de entrenamiento puede ser el más rápido de entre los que se basan en el gradiente conjugado, ya que esto depende de muchos factores.

En MATLAB se puede encontrar una serie de conjunto de evaluación (benchmarks), una lista de seis problemas, de varios algoritmos de entrenamiento. Unos de los cuales se corresponde con una aplicación de diagnóstico de cáncer. Es el *Quatre benchmark*, el objetivo de la red es clasificar un tumor como maligno o benigno. La

red usada por los benchmarks fue de 9-5-5-2 con la función de activación tangente sigmoide *tansig* en todas las capas. En la siguiente tabla se resumen los resultados de entrenamiento para nueve algoritmos distintos:

algoritmo	Media tiempo (s)	cociente	Min. tiempo (s)	Max. tiempo (s)	Std (s)
CGB	80.27	1.00	55.07	102.31	13.17
RP	83.41	1.04	59.51	109.39	13.44
SCG	86.58	1.08	41.21	112.19	18.25
CGP	87.70	1.09	56.35	116.37	18.03
CGF	110.05	1.37	63.33	171.53	30.13
LM	110.33	1.37	58.94	201.07	38.20
BFG	209.60	2.61	118.92	318.18	58.44
GDX	313.22	3.90	166.48	446.43	75.44
OSS	463.87	5.78	250.62	599.99	97.35

Tabla 3.1 *Resultados de entrenamiento para nueve algoritmos distintos.*

Donde los símbolos de los algoritmos en la tabla anterior se definen en la siguiente tabla:

símbolo	función	Algoritmo
LM	trainlm	Levenberg-Marquardt
BFG	trainbfg	BFGS Quasi-Newton
RP	trainrp	Resilient retropropagación. Solo se tiene en cuenta el signo de la derivada y no su valor
SCG	trainscg	Escales Conjúgate Gradiente
CGB	traincgb	Reanuda búsqueda de de dirección cada ciertas iteraciones
CGP	traincgp	Conjúgate Gradiente por Polak-Ribière
OSS	trainoss	<i>One-Step Secant</i>
GDX	traingdx	Momentum y Variable Learning Rate Backpropagation

Tabla 3.2 *Los símbolos de los algoritmos distintos. Los algoritmos se describen en [WOL90, MAN90].*

La implementación más simple del aprendizaje de retropropagación para que las actualizaciones de los pesos y las tendencias en la dirección en la cual la función disminuye lo más rápidamente posible - la negativa del gradiente. Una iteración de este algoritmo puede ser escrita

$$w_{k+1} = w_k - \alpha_k g_k \quad (3.6)$$

Donde w_k es un vector de pesos y de tendencias actuales, g_k es el gradiente actual y α_k es la tasa de aprendizaje (o tasa de entrenamiento).

Hay dos maneras diferentes en las que este algoritmo de gradiente se puede aplicar: el modo incremental y el modo serie o por lotes. En el modo incremental, el gradiente se computa y los pesos se actualizan después que cada entrada es aplicada a la red. En el modo de serie todas las entradas son aplicadas a la red antes de que los pesos se actualicen. En tratamiento por lotes (serie) los pesos y tendencias de la red se actualizan sólo después que el conjunto entero de entrenamiento ha sido aplicado a la red. Los gradientes calculados en cada ejemplo del entrenamiento se suman juntos para determinar el cambio en los pesos y tendencias.

3.3.4 Generalización

Uno de los aspectos fundamentales sobre las RNA's es la capacidad de la red neuronal de generalizar a partir de ejemplos, lo que se denomina en ocasiones "*problema de la memorización frente al aprendizaje*". Por generalización entendemos la capacidad de la red de proporcionar una respuesta correcta ante patrones que no han sido empleados en su entrenamiento. Una red neuronal correctamente entrenada debe poseer una buena capacidad de generalización, lo que significa que ha aprendido correctamente la representación (*mapping*) en cuestión, y que responderá adecuadamente ante nuevos patrones no vistos con anterioridad.

En un proceso de entrenamiento se puede considerar, por una parte, un error en aprendizaje, que se suele calcular como el error cuadrático medio de los resultados proporcionados por la red para el conjunto de patrones de aprendizaje. Con una red suficientemente grande, puede reducirse tanto como se quiera, sin más que llevar a

cabo más iteraciones. Por otra parte, existe un error en generalización, que se puede medir empleando un conjunto representativo de patrones diferentes a los utilizados en el entrenamiento. De esta manera, podemos entrenar una red neuronal haciendo uso de un conjunto de aprendizaje, y comprobar su eficiencia real, o error de generalización, mediante un conjunto de test.

3.3.5 Validación cruzada

El método de evaluación utilizando validación cruzada, fue introducido por Stone en 1974 [STO74]. La idea aquí es estimar el error de predicción dividiendo al azar el conjunto de datos en varias partes. En cada paso una de las partes se convierte en una muestra de prueba que sirve para evaluar el modelo y las restantes partes constituyen lo que es llamado una muestra de entrenamiento que sirve para construir el modelo.

Es ampliamente utilizado en la red de MLP. El proceso de entrenar y validar a la vez para detenerse en el punto óptimo se puede describir en los siguientes pasos [OLM93]: principio, la red se adapta progresivamente al conjunto de aprendizaje, acomodándose al problema y mejorando la generalización. Sin embargo, en un momento dado el sistema se ajusta demasiado a las particularidades de los patrones empleados en el entrenamiento, aprendiendo incluso el ruido presente en ellos, por lo que crece el error que comete ante patrones diferentes a los empleados en el entrenamiento (error de generalización). En este momento la red no ajusta correctamente el *representación (mapping)*, sino que está memorizando patrones del conjunto de aprendizaje, constituyendo lo que por el momento denominaremos sobre entrenamiento (*overtraining*). Idealmente, dada

una arquitectura de red, ésta debería entrenarse hasta alcanzar el punto óptimo en el que el error de generalización es mínimo (que idealmente debería ser igual que el error en aprendizaje).

3.4 Descripción de un sistema de clasificación basado en la combinación de los sistemas de inferencia difusa y redes neuronales adaptables (ANFIS)

Un sistema neuro-difuso que combina las técnicas de redes neuronales artificiales y las técnicas de inferencia difusa es un sistema híbrido. Los sistemas híbridos, que pueden consistir en diferentes combinaciones de lógica difusa y redes neuronales, proporcionan un método eficiente para resolver una gran variedad de problemas. Esta técnica tiene propiedades computacionales particulares (por ejemplo: habilidad de aprender) que la hace óptima para resolver ciertos problemas. Los modelos difusos son sistemas basados en reglas de inferencia difusos cuya complejidad tiene una relación directa con el número y tipo de reglas utilizadas, cantidad y tipo de partición, así como características de antecedentes y consecuentes seleccionados. Las reglas son expresadas como implicaciones lógicas (Si- Entonces).

En las redes neuronales, el conocimiento se adquiere automáticamente por el algoritmo de aprendizaje, pero este proceso es relativamente lento (gran cantidad de épocas de entrenamiento) y el análisis de la red entrenada es difícil (modelo de caja negra). No es posible extraer el conocimiento estructural (reglas) de la red neuronal ni puede éste integrarse con información especial sobre el problema en la red neuronal con el fin de simplificar el procedimiento de

aprendizaje. Los sistemas difusos son más favorables porque su comportamiento puede ser explicado por medio de reglas difusas y, de esta forma, su desempeño puede ser ajustado modificando estas reglas. Sin embargo, la adquisición del conocimiento es difícil, y, además, el universo de discurso de cada variable necesita ser dividido en intervalos, por lo que las aplicaciones de los sistemas difusos se restringen a problemas en los cuales el conocimiento está disponible en un número de variables de entrada pequeño. Para superar el problema de la adquisición del conocimiento, las redes neuronales son extendidas para extraer automáticamente las reglas difusas de los datos numéricos.

Los sistemas neuro-difusos combinan la capacidad de aprendizaje de las redes neuronales con el poder de interpretación lingüística de los sistemas de inferencia difusa. En la siguiente sección se presenta un modelo neuro- difuso.

3.4.1 ANFIS (Adaptive Neuro Fuzzy Inference System)

El modelo ANFIS fue desarrollado por Jang en 1993 [JAN93], y es funcionalmente equivalente a los sistemas de inferencia difusos. Es un método que permite sintonizar o crear la base de reglas de un sistema difuso, utilizando el algoritmo de entrenamiento de retropropagación a partir de la recopilación de datos de un proceso. Su arquitectura es funcionalmente equivalente al mecanismo de inferencia Takagi-Sugeno (TSK) [TAK85, JAN93].

3.4.2 Arquitectura del modelo ANFIS

En este modelo de un sistema híbrido neuro-difuso, la red neuronal artificial implementa un sistema difuso con su mecanismo de razonamiento, donde las reglas de inferencia son del tipo Takagi–Sugeno (TSK) [JAN93]. Para un sistema de inferencia TSK de primer orden, un conjunto de reglas difusas SI-ENTONCES es el siguiente [TAK85]:

$$\text{Si } x \text{ es } A \text{ y } y \text{ es } B \text{ entonces } z=f(x,y)$$

Donde (A y B) son conjuntos difusos, (x, y) son variables difusas, y f es una función de x e y que aproxima el valor de z . Dicha función es usualmente una combinación lineal de las variables de entrada, cuyos coeficientes son estimados usando mínimos cuadrados. Cada capa de la red neuronal realiza un proceso específico en la inferencia de la salida del sistema. La red neuronal híbrida que representa este tipo de inferencia es una red adaptable con 5 capas, donde cada capa representa una operación del mecanismo de inferencia difusa. Esta red se muestra en la Figura 3.8.

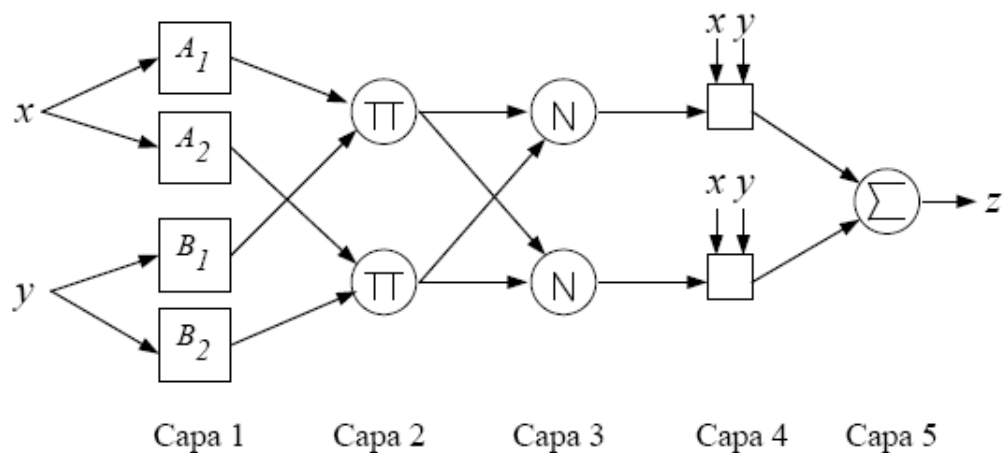


Figure 3.8 *Arquitectura típica de un modelo ANFIS.*

En esta arquitectura, todos los nodos de una misma capa tienen la misma función (los nodos representados con cuadros son nodos adaptables, es decir, sus parámetros son ajustables). La estructura de la red ANFIS consiste de cinco capas [JAN93]:

Capa 1: Los nodos de esta capa son adaptativos, y las entradas en esta capa corresponden a las entradas x e y . La salida del nodo es el grado de pertenencia para el cual la variable de entrada satisface el término lingüístico asociado a este nodo. Esta capa calcula el valor de la función de pertenencia, que puede definirse de varias maneras, teniendo en cuenta que debe ser una función de membresía diferenciable:

$$\mu(x) = \left(1 + \left(\frac{x-c}{a}\right)^{2b}\right)^{-1} \quad (3.7)$$

Donde a, b, c es el conjunto de parámetros.

$$\mu(x) = e^{-\left(\frac{x-c}{a}\right)^2} \quad (3.8)$$

Donde x es la entrada y $\{a, c\}$ es el conjunto de parámetros.

Capa 2: Cada nodo de esta capa es no adaptativo y cada uno calcula el grado de activación de la regla asociada a dicho nodo. Ambos nodos están representados con un Π en la Figura 3.8, por el hecho de que ellos pueden representar cualquier t-norma para modelar la operación lógica *AND*. Los nodos de esta capa son conocidos como *nodos de reglas*.

Capa 3: Los nodos de esta capa son no adaptativos. Y cada nodo está representado por un N en la Figura 3.8, para indicar la

normalización de los grados de activación. La salida del nodo es el grado de activación normalizado (con respecto a la suma de los grados de activación) de la regla i .

$$\bar{w}_i = \frac{w_i}{\sum_{j=1}^N w_j}, i = 1, 2, \dots, n \quad (3.9)$$

Capa 4: Cada nodo de esta capa es adaptativo y la salida de los nodos corresponde al producto entre el grado de activación normalizado por la salida individual de cada regla.

$$\bar{w}_i z_i = \bar{w}_i (p_i x + q_i y + r_i), i = 1, 2, \dots, n \quad (3.10)$$

Capa 5: El único nodo de esta capa es no adaptativo y calcula la salida total del sistema (agregación) como la suma de todas las entradas individuales de este nodo.

$$z = \sum_i \bar{w}_i f_i \quad (3.11)$$

Frecuentemente los parámetros de los nodos de la capa 1 se denominan parámetros de las premisas, y los parámetros de los nodos de la capa 4 se denominan parámetros de los consecuentes, ya que se corresponden con las premisas y los consecuentes de las reglas SI-ENTONCES del modelo de inferencia TSK [TAK85].

3.4.3 Aprendizaje en el modelo ANFIS

El modelo ANFIS tiene dos conjuntos de parámetros que deben ser entrenados: los parámetros del antecedente (constantes que caracterizan las funciones de pertenencia) y los parámetros del

consecuente (parámetros lineales de la salida del modelo de inferencia). El paradigma de aprendizaje del modelo ANFIS emplea algoritmos de gradiente descendiente para optimizar los parámetros del antecedente y el algoritmo de mínimos cuadrados para determinar los parámetros lineales del consecuente. Debido a esta combinación se lo conoce como *regla de aprendizaje híbrida* [JAN93].

Para aplicar el aprendizaje híbrido en grupo, en cada época de entrenamiento debe ejecutarse un paso *hacia adelante* y un paso *hacia atrás* [JAN92]. En el paso *hacia adelante*, los parámetros de las funciones de pertenencia son inicializados y se presenta un vector de entrada-salida, se calculan las salidas del nodo para cada capa de la red y entonces los parámetros del consecuente son calculados usando el método de mínimos cuadrados. Una vez identificados los parámetros del consecuente, el error se calcula como la diferencia entre la salida de la red y la salida deseada presentada en los pares de entrenamiento. Una de las medidas más usadas para el error de entrenamiento es la suma de errores cuadráticos (MRS), definida como:

$$MRS = \sum_{k=1}^N (d_k - z_k)^2 \quad (3.12)$$

Los d_k corresponden a los patrones de entrenamiento proporcionados (salidas deseadas) y z_k es la correspondiente salida de la red. En el paso *hacia atrás*, las señales de error son propagadas desde la salida, en dirección de las entradas; el vector gradiente es acumulado para cada dato de entrenamiento. Al final del paso *hacia atrás* para todos los datos de entrenamiento, los parámetros en la capa

1 (parámetros de las funciones de pertenencia) son actualizados por el método descendente en una magnitud $\Delta\alpha$ igual a:

$$\Delta\alpha = -\eta \frac{\partial E}{\partial \alpha} \quad (3.13)$$

Donde η es la tasa de aprendizaje, que puede ser expresada como:

$$\eta = \frac{\rho}{\sqrt{\sum_{\alpha} \left(\frac{\partial E}{\partial x}\right)^2}} \quad (3.14)$$

ρ es el tamaño del paso (la longitud de cada transición a lo largo de la dirección del gradiente en el espacio de parámetros). Esto significa que la tarea de aprendizaje es variable y se ajusta automáticamente [JAN96].

Capítulo 4

Estrategias de optimización

En este capítulo se explican las técnicas de exploración de parámetros o características. Hemos utilizado dos técnicas de exploración: exploración secuencial y exploración multi-camino, para seleccionar las características que conducen al mejor resultado de clasificación.

En el problema de reconocimiento de patrones se desea clasificar ejemplos de determinados objetos en una de varias clases preestablecidas. La tarea de clasificación de patrones consiste en construir un mapa de relaciones entre el espacio de características y el conjunto de las clases. De este modo se puede reconocer a qué clase corresponde cualquier patrón de entrada representado por un vector de características.

En muchos problemas de clasificación, como el que se describe en el Capítulo 5, no se conoce a priori, el número de características (rasgos) relevantes que serán necesarios para la posterior distinción de las diferentes imágenes.

El problema de selección de características o parámetros consiste en seleccionar un subconjunto de m parámetros de entre un conjunto original de n parámetros candidatos, bajo algún criterio de desempeño. El número de subconjuntos posibles crece exponencialmente, haciendo imposible la búsqueda exhaustiva, incluso para valores pequeños de n .

Entre los propósitos de la selección de características se cuentan:

- a. Reducir la complejidad del clasificador y su implementación en hardware/software.
- b. Compresión de información (eliminar características redundantes o de poca relevancia).
- c. Reducir el coste de medición al disminuir el número de características.
- d. Mejorar el nivel de clasificación.

En este trabajo se ha utilizado una técnica de exploración secuencial de características, y también una técnica de exploración multi-camino. Estas técnicas se pueden desarrollar como se explica en las siguientes secciones.

4.1 Métodos de exploración secuencial

Este método de exploración se puede desarrollar con dos estrategias como se describe a continuación.

4.1.1 Método hacia atrás

En este caso se comienza con el modelo completo y en cada paso se va eliminando una característica. Si resultara que todas las características predictoras son importantes, entonces no se hace nada y el mejor modelo será aquel que tenga todas las características predictoras disponibles [EDG04]. En caso contrario, en cada paso el parámetro que se elimina del modelo es aquel que satisface cualquiera de estos requisitos equivalentes entre sí:

- a) Aquel parámetro que tiene el valor más pequeño entre los parámetros incluidos aún en el modelo.
- b) Aquel parámetro que produce la menor disminución en la clasificación al ser eliminado del modelo. Es decir, aquel parámetro que produce el más pequeño incremento en la suma de cuadrados del error.

- c) Aquel parámetro que tiene la correlación parcial (en valor absoluto) más pequeña con el valor de respuesta, tomando en cuenta las variables que quedarían en el modelo. La correlación parcial de Y con la variable X_i se define como la correlación entre los residuales de la regresión de Y con todas las variables predictoras, excepto X_i y los residuales de la regresión de X_i con todas las otras restantes variables predictoras.

El método hacia atrás padece del efecto de anidamiento ya que toda característica o parámetro que es eliminado del modelo ya no vuelve a entrar a él. En este trabajo se utiliza este método para seleccionar los mejores parámetros que dan mejor resultado en la clasificación. En el primer paso se utilizan todos los parámetros en el clasificador y en el segundo paso, uno de los parámetros se quita, dando como resultado la mejor clasificación posible. Esto se realiza con todos los parámetros uno a uno y finalmente como nuevo conjunto de parámetros se toma aquel que no incluyendo uno de los parámetros conduce al mejor resultado. Esto se repite eliminando un nuevo parámetro en cada iteración [ROS03].

Finalmente se obtiene, el conjunto de parámetros con una máxima clasificación y un mínimo número de los parámetros.

4.1.2 Método hacia adelante

Este método se empieza con la regresión lineal simple que considera como parámetro predictor a aquel que está más altamente correlacionado con la respuesta deseada.

Si este primer parámetro no es significativo entonces se para el proceso, de lo contrario se sigue y en el siguiente paso se añade al modelo el parámetro que reúne cualquiera de estos requisitos equivalentes entre sí:

- a) Aquel parámetro que tiene el valor más grande entre los parámetros no incluidos aún en el modelo.
- b) Aquel parámetro que produce el mayor incremento en la clasificación al ser añadida al modelo. Es decir, aquel parámetro que produce la mayor reducción en la suma de cuadrados del error.
- c) Aquel parámetro que tiene la correlación parcial más alta (en valor absoluto) con respuesta deseada, tomando en cuenta los parámetros ya incluidos en el modelo.

Para usar este método en nuestro trabajo es necesario tener una función objetivo f que evalúe el clasificador usando m parámetros. Esta función puede ser el acierto del clasificador de K vecinos más cercano (KNN). El método empieza con un parámetro ($m = 1$), es decir se busca en todos los parámetros, aquel que maximiza el acierto de clasificación. Luego se busca el segundo parámetro que combinado con el anterior maximiza nuevamente el acierto con dos parámetros ($m = 2$). De esta manera, no son considerados parámetros que estén muy correlacionados con el parámetro ya seleccionado, o bien que no proporcionen información para la detección, ya que no maximizan acierto. El método se repite hasta conseguir los mejores m parámetros.

La curva obtenida de este método, se corresponde a los parámetros con la clasificación donde el conjunto de los parámetros

crece, aumentando con los primeros parámetros y disminuyendo cuando son considerados demasiados parámetros.

En el último paso, el conjunto de los parámetros seleccionados se toma como el mínimo número de parámetros que dan lugar a un máximo acierto de clasificación.

4.1.3 El algoritmo “Greedy Feature Flip” (G-flip)

Este algoritmo se introduce por [RAN04], es un algoritmo *codicioso* de búsqueda para maximizar de función de evaluación $e(F)$ donde F es el conjunto de las características. El algoritmo itera repetidamente sobre el conjunto de características y actualizando el conjunto de las características seleccionadas. En cada iteración se decide quitar o agregar la característica actual al conjunto de las características seleccionadas evaluando el margen según la ecuación (4.2) con o sin esta característica.

$$e(w) = \sum_{x \in S} \theta_{s/x}^w(x) \quad (4.1)$$

Esta ecuación se usa como función de la evaluación para un conjunto del entrenamiento S y un vector peso w .

Este algoritmo converge a un máximo local de la función de la evaluación, pues cada paso aumenta su valor y el número de las posibles características que forman el conjunto de características es finito. La complejidad computacional de cada iteración sobre todas las características de G-flip es $\Theta(N^2 m^2)$ donde N es el número de las

características y m es el número de las características ya seleccionadas. Empíricamente el algoritmo G-flip converge en pocas iteraciones. La mejor propiedad de este algoritmo es que no necesita “*parámetros de configuración*” (*parameter free*). No hay necesidad de optimización del número de las características o ningún tipo de umbral. En la siguiente figura se muestra el esquema del algoritmo G-flip.

Algoritmo Greedy Feature Flip (G-flip)

1. Inicialmente se pone el conjunto de las características seleccionadas vacío: $F = \phi$
 2. para $t = 1, 2, \dots$
 - (a) se pone aleatoriamente permutación s de $\{1 \dots N\}$
 - (b) para $i = 1$ hasta N ,
 - i. evalúate $e_1 = e(F \cup \{s(i)\})$
y $e_2 = e(F \setminus \{s(i)\})$
 - ii. si $e_1 > e_2$, $F = F \cup \{s(i)\}$
y si $e_2 > e_1$, $F = F \setminus \{s(i)\}$
 - (c) si no ha cambiado algo en (b) entonces finaliza
-

Figura 4.1 Algoritmo G-Flip.

4.2 Métodos de exploración multi-camino

Para conseguir un mecanismo que optimice el conjunto de parámetros seleccionados se pueden utilizar Algoritmos Genéticos, que son algoritmos que tienen gran éxito en problemas de búsqueda y clasificación.

4.2.1 Algoritmos Genéticos

Los Algoritmos Genéticos (AG) fueron introducidos por John Holland en 1970 [Hol70] inspirándose en el proceso observado en la evolución natural de los seres vivos. Los algoritmos genéticos son métodos de optimización, que tratan de resolver el mismo conjunto de problemas que se ha contemplado anteriormente, es decir, hallar (x_1, \dots, x_n) tales que $F(x_1, \dots, x_n)$ sea máximo. En un AG, tras parametrizar el problema en una serie de variables, (x_1, \dots, x_n) se codifican en un cromosoma. Todos los operadores utilizados por un algoritmo genético se aplicarán sobre estos cromosomas, o sobre poblaciones de ellos. En el algoritmo genético se establece una analogía entre el conjunto de soluciones de un problema y el conjunto de individuos de una población natural, codificando la información de cada solución en un vector cadena (usualmente de elementos binarios) a modo de cromosoma. Hay que tener en cuenta que un algoritmo genético es independiente del problema, lo cual lo hace un algoritmo *robusto*, por ser útil para cualquier problema, pero a la vez *débil*, pues no está especializado en ninguno.

A tal efecto se introduce una función de evaluación de los cromosomas, que llamaremos "fitness" y que está basada en la función

objetivo del problema. Igualmente se introduce un mecanismo de selección de manera que los cromosomas con mejor evaluación sean escogidos para "reproducirse" más a menudo que los que la tienen peor. Los AG generan un conjunto de soluciones mediante la mutación y recombinación de parte del conjunto de soluciones conocido. En cada paso el conjunto de soluciones conocido como "población actual" se renueva reemplazando una proporción de esta población por los sucesores de las soluciones más "adecuadas" (mediante el uso de una función de evaluación). Por otra parte, los AG pueden realizar búsquedas en espacios de soluciones que contienen complejas interacciones entre las distintas partes, donde el impacto de cada parte sobre la función de evaluación es difícil de especificar.

Los AG generalmente encuentran soluciones con un alto grado de acierto. Estos algoritmos son métodos robustos de búsqueda, que permiten tratar problemas de optimización donde el objetivo es encontrar un conjunto de parámetros que minimizan o maximizan una función. La mencionada anteriormente "función de *fitness*" que posteriormente se describirá en mayor profundidad.

En nuestro trabajo el problema será el encontrar el conjunto de parámetros que maximizan la función de fitness, ésta se basará en el porcentaje de aciertos conseguidos por un clasificador.

4.2.1.1 Implementación del algoritmo genético

El algoritmo genético (AG) procede de la forma siguiente:

- 1- Evaluar la puntuación (fitness) de cada uno de los genes.

- 2- Permitir a cada uno de los individuos reproducirse, de acuerdo con su puntuación.
- 3- Emparejar los individuos de la nueva población, haciendo que intercambien material genético, y que alguno de los bits de un gen se vea alterado debido a una *mutación* espontánea (con una probabilidad concreta).

Cada uno de los pasos consiste en una actuación sobre las cadenas de números, es decir, la aplicación de un *operador* a una cadena binaria. Se les denominan, por razones obvias, *operadores genéticos*, y hay tres principales: *selección*, *crossover* o *recombinación* y *mutación*; aparte de otros operadores genéticos no tan comunes, todos ellos se verán a continuación.

Para llevar a cabo la estructuración del algoritmo genético seguiremos el siguiente esquema:

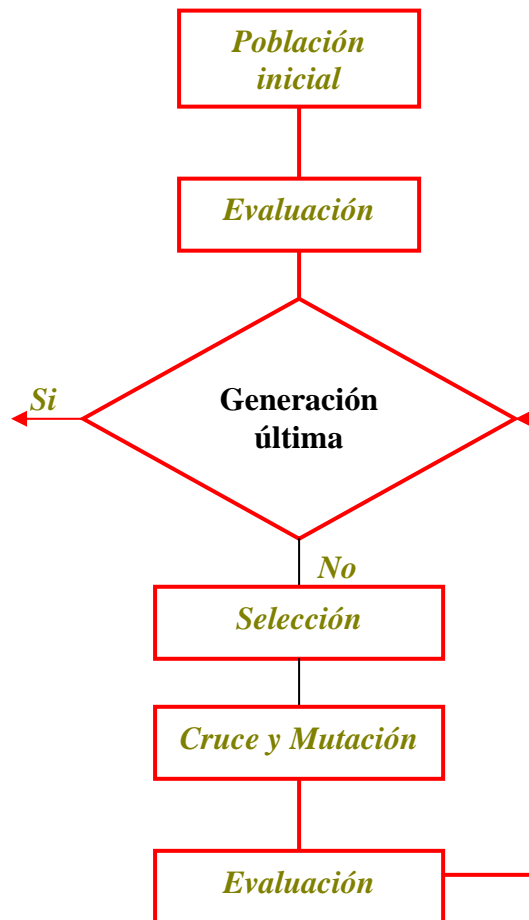


Figura 4.2 Esquema general de evolución de los algoritmos genéticos.

Un algoritmo genético tiene también una serie de parámetros que se tienen que fijar para cada ejecución, como los siguientes:

- **Tamaño de la población:** debe de ser suficiente para garantizar la diversidad de las soluciones, y, además, tiene que crecer más o menos con el número de bits del cromosoma, aunque nadie ha aclarado cómo tiene que hacerlo. Por supuesto, depende también del ordenador en el que se esté ejecutando.

- **Condición de finalización:** lo más habitual es que la condición de finalización sea la convergencia del algoritmo genético o un número prefijado de generaciones.

A continuación especificáramos los siguientes elementos para llevar a la práctica el esquema anterior (Figura 4.2) y concretarlo en un algoritmo, hay que especificar:

- Representación cromosómica
- Población inicial
- Medida de evaluación
- Criterio de selección (eliminación de cromosomas)
- Operaciones de recombinación
- Operaciones de mutación

4.2.1.2 Representación

La representación cromosómica utilizada es la que corresponde a un vector binario. En la literatura pueden encontrarse dos tendencias para representar las soluciones: a través de cadenas binarias o mediante otro tipo de configuraciones (vectores, árboles o grafos). Se ha elegido esta representación ya que es la que de un modo más natural y sencillo muestra nuestro objetivo, el cual no va a ser otro que el decidir si un parámetro se considera relevante o no en un modelo de clasificación.

Es decir, cada una de las posiciones del vector binario representa a uno de los parámetros, de tal modo que si en esa posición tenemos un “0” indicamos que no se considera ese parámetro y si por el contrario aparece un “1” sí se tendrá en cuenta. Por tanto la

configuración final que obtenga el algoritmo representa qué parámetros son válidos y cuáles pueden ser desechados.

4.2.1.3 Población Inicial

La población inicial suele ser generada aleatoriamente. Sin embargo, últimamente se están utilizando métodos heurísticos para generar soluciones iniciales de buena calidad. En este caso, es importante garantizar la diversidad estructural de estas soluciones para tener una "representación" de la mayor parte de población posible o al menos evitar la convergencia prematura.

El sembrado de soluciones, es decir la inclusión de soluciones predeterminadas en la población inicial, en general asegura una convergencia más rápida, pero se corre el riesgo de perder diversidad en la población.

Un aspecto importante a considerar en esta etapa es que la población tenga la suficiente diversidad, de modo de explorar zonas más amplias del espacio de búsqueda. Así se pueden obtener soluciones más cercanas al óptimo global y evitar la convergencia prematura.

4.2.1.4 Medida de Evaluación

Con el objeto de ser capaces de comparar dos o más soluciones, debe introducirse una forma de evaluar las soluciones, basándose en la función objetivo y en las restricciones del problema. En caso de no

tener la función objetivo definido en forma explícita, la evaluación puede llevarse a cabo con alguna subrutina de cálculo representativa. A la evaluación de cada solución se le suele denominar calidad (*fitness*).

Para controlar la factibilidad de las soluciones, al índice de calidad puede sumársele un factor de penalización proporcional a la violación de cada restricción. Este factor de penalización puede ser constante o ir ajustándose en la medida que transcurren las generaciones. Con el uso adecuado de factores de penalización adaptables se puede evitar la eliminación de buenas zonas de búsqueda potenciales en las primeras etapas de la optimización. Así, en la medida que las generaciones transcurren, las soluciones no factibles se van descartando y las poblaciones quedan formadas principalmente por soluciones factibles de alta calidad.

4.2.1.5 Criterio de selección

Para aplicar los operadores genéticos tendremos que seleccionar un subconjunto de la población. Algunas de las técnicas de que disponemos son:

- *Selección directa*: toma elementos de acuerdo a un criterio objetivo, como son «los x mejores», «los x peores»... los del tipo «el cuarto individuo a partir del último escogido» son empleados con mucha frecuencia cuando se quieren seleccionar dos individuos distintos, y se selecciona el primero por un método aleatorio o estocástico.

- *Selección aleatoria*: puede ser realizado por *selección equiprobable* o *selección estocástica*.
 - *Selección equiprobable*: todos tienen la misma probabilidad de ser escogidos.
 - *Selección estocástica*: la probabilidad de que un individuo sea escogido depende de una heurística. Hay muchos distintos procedimientos estocásticos, Uno de los procedimientos más utilizado es el denominado de la Ruleta de Baker [BAK87].

La experiencia indica que la estrategia óptima de selección, para este tipo de problemas, como el que nos ocupa, es escoger el método de *la Ruleta de Baker* [BAK87], en donde cada individuo tiene una sección circular de una ruleta que es directamente proporcional a su calidad. Para realizar una selección se realizaría una tirada de bola en la ruleta, tomando el individuo asociado a la casilla donde cayó la bola.

4.2.1.6 Operador de cruce

Se denomina *técnica de cruce* a la forma de calcular el cromosoma del nuevo individuo en función del cromosoma del padre y de la madre. El operador de cruce es fuertemente responsable de las propiedades del algoritmo genético, y determinará en gran medida la evolución de la población. Es el operador de búsqueda más importante dentro del algoritmo. Es un operador que intercambia el material genético de un par de progenitores produciendo dos cromosomas hijos que normalmente difieren de sus progenitores.

El operador de cruce opera con probabilidad p_c , lo cual permite que en algunos casos no haya recombinación y se mantengan los padres. Dados un par de padres, de largo k bits, se escoge aleatoriamente un punto $k \in \{1, \dots, k-1\}$ y se intercambian los bits a la derecha de esa posición entre ambos individuos, obteniéndose dos descendientes, tal y como se indica a continuación (Figura 4.3):

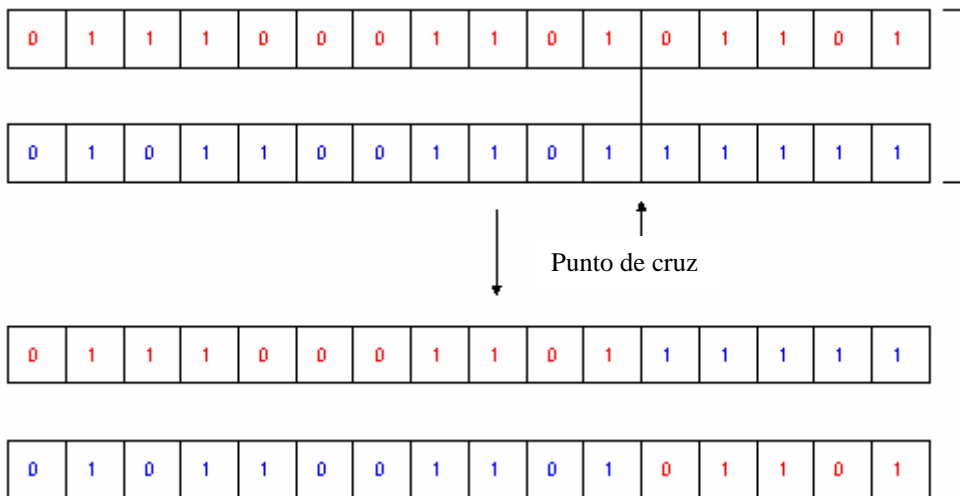


Figura 4.3 Operador de cruce.

4.2.1.7 Operador de Mutación

El operador de mutación realiza modificaciones aleatorias en bits o cadenas elegidas al azar. De esta forma, la mutación aporta diversidad en la población de soluciones, permitiendo una exploración más completa del espacio de búsqueda. Desde otro punto de vista, la motivación para usar mutación es prevenir la pérdida permanente de un bit determinado y así evitar la convergencia prematura permitiendo escapar de óptimos locales [WHI93].

El operador de mutación más sencillo corresponde a cambiar, con una determinada probabilidad, el valor de un bit (en caso de ser una cadena binaria). Durante las pruebas realizadas se observó que los operadores de mutación clásicos retrasaban considerablemente la convergencia del algoritmo genético, ya que lo obligaban a explorar regiones del espacio de búsqueda al azar, donde la mayor parte de las veces no se encontraban mejores soluciones. Al diseñar un algoritmo genético que pueda competir en términos de eficiencia con los algoritmos específicos del dominio de aplicación, debe sacrificarse en parte la amplitud de la exploración, restringiéndola a aquellas regiones donde es más probable que existan buenas soluciones. Esto puede lograrse diseñando operadores de mutación que no hagan cambios totalmente aleatorios, sino que, de la misma forma que el operador de cruce, tiendan a hacer cambios que aumenten la calidad de las soluciones, dejando algunas pocas decisiones libradas al azar.

La operación de mutación utilizada consiste en reemplazar con cierta probabilidad el valor de un bit (Fig.4.4).

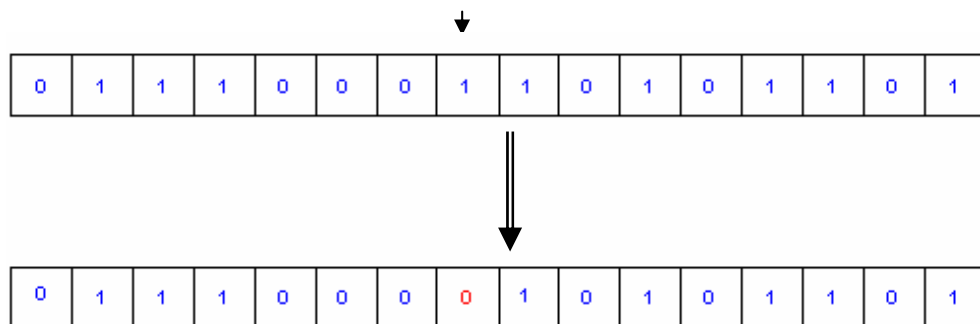


Figura 4.4 Operador de Mutación.

Esta operación juega un papel secundario. Tasas de mutación pequeñas garantizan que un individuo no difiera mucho de sus padres en el genotipo (cadena binaria). La mutación sirve para evitar la pérdida de diversidad producto de bits que han convergido a un cierto valor para toda la población y que por tanto no pueden ser recuperados por el operador de recombinación.

El operador de mutación implementado invierte cada bit de la cadena binaria con una determinada probabilidad (Fig. 4.4).

Capítulo 5

Aplicación a diagnóstico de patologías de la piel

Este capítulo se centra en la extracción de información relevante para distinguir patologías de la piel. Se describen las características que se han extraído de las imágenes fluoroscopia por el tratamiento de imágenes. Se evalúa de forma preliminar la relevancia de distintas características para la aplicación de diagnóstico. También se explica cómo se eliminan aquellas características altamente redundantes. La conclusión de este capítulo es una lista de características diversas que se utilizarán en el Capítulo siguiente en esquemas de clasificación.

5.1 Patologías de la piel

5.1.1 Introducción

¿Que es el cáncer? Las células, de órganos y tejidos del cuerpo, se pueden reproducir. Constantemente las células cumplen un ciclo de vida donde una vez llegadas a la vejez, mueren, y son reemplazadas por células nuevas.

Normalmente, la división y el crecimiento de células son ordenados y controlados; pero si este proceso se descontrola por alguna razón, las células continuarán dividiéndose y convirtiéndose en una masa a la cual se le denomina tumor (Fig.5.1). Los tumores pueden ser benignos o malignos. El cáncer es el nombre dado a un tumor maligno. En un tumor benigno las células no se separan y dispersan a otras partes del cuerpo, así que no son cancerosas (no dan lugar a metástasis). No obstante, si continúan creciendo en el sitio original, pueden causar un problema presionando en los órganos circundantes.

Las células cancerosas (cáncer maligno) a menudo se dispersan a otras partes del cuerpo donde comienzan a crecer y a reemplazar el tejido normal. Este proceso, llamado metástasis, ocurre a medida que las células cancerosas entran al torrente sanguíneo o vasos linfáticos del cuerpo

Es importante enfatizar que el cáncer no es una sola enfermedad con un solo tipo de tratamiento. Hay múltiples clases de cáncer, cada uno con su propio nombre y forma de tratamiento [CAN03].

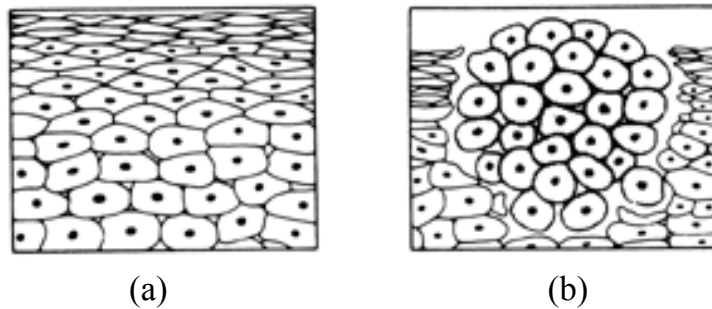


Figura 5.1 (a) *Células normales*, (b) *células que forman un tumor*

5.1.2 *El cáncer de piel y otras patologías*

El cáncer de piel es el más común de todos los cánceres y está en aumento [MAR95], probablemente debido a nuestra tendencia a exponer la piel a la luz del sol y a luces ultravioletas usadas para el bronceado. Una definición para el cáncer en general podría ser la siguiente: “crecimiento descontrolado de células anormales que han mutado en tejidos normales” [MAR95].

El cáncer puede causar la muerte cuando estas células impiden el funcionamiento normal de los órganos vitales afectados o se diseminan por todo el cuerpo y dañan otros sistemas esenciales. Personas de todas las edades puede padecer el cáncer de la piel, pero es raro en niños [NET05]. Existen distintos tipos de cáncer de piel, entre los que están la Keratosis Actínica (se considera más bien una lesión precancerosa benigna) y el carcinoma basal (cáncer de células basales, también benigno).

5.1.2.1 *Keratosi Actínica*

Se presenta en la superficie de la piel, posee un color que puede ser rosa, rojo o una combinación de éstos (de tonalidad clara u oscura). La Keratosi Actínica se considera una condición precancerosa de la piel debido a que es el primer paso en el desarrollo del cáncer de la piel. Por lo tanto la presencia de esta anomalía, que también se conoce como keratosi solar, indica que ha habido daños normalmente por exposición al sol y que se puede convertir en cualquier clase de cáncer de piel.

La keratosi actínica aparece con mayor frecuencia en la cara, los oídos, el cuero cabelludo de una zona calva, el cuello, las partes posteriores de manos y antebrazos, y los labios. Las lesiones no son generalmente peligrosas para la vida, mientras se detecten y se traten a tiempo. Sin embargo, si esto no se hace, pueden crecer e invadir los tejidos finos circundantes y, en raras ocasiones, propagarse a los órganos internos.

Por lo general, en los hombres suele aparecer en el tronco y en las mujeres, es más frecuente en la parte inferior de las piernas. Estas personas suelen tener la piel clara. Puede aparecer en personas de piel más oscura, en zonas de las manos, planta de los pies, y debajo de las uñas.

La exposición del sol es la causa de casi todos los keratosi actínicos. El daño del sol a la piel se acumula con el tiempo. Ciertos grupos de gente son de más riesgo que otros [UMG05]:

- La gente que tiene pelo corto de la piel, rubio o rojo. El color de los ojos azul, verde o grises son de alto riesgo. La gente morena solamente pueden desarrollar keratosis si se exponen al sol de forma continuada y sin protección.
- Los africano-americanos tienen raramente estas lesiones.
- Personas a las que se les ha aplicado quimioterapia, sufren de sida o se les ha trasplantado algún órgano, también son de alto riesgo.

Una de cada seis personas desarrollará un keratosis actínico en el curso de la vida, según algunas estimaciones. La gente mayor es más probable que sufra esta enfermedad, porque la exposición acumulativa del sol aumenta con edad [UMG05].

5.1.2.2 *Carcinoma basal*

Es el tipo más común de todos los cánceres, comienza en las células de la capa profunda de la epidermis (la capa externa de la piel). Es de seis a ocho veces más común que el melanoma maligno. Está situado a menudo en la cara y en el cuello [GNI97]. El carcinoma basal no es uno de los cánceres más peligrosos. Puede ser localmente destructivo y difícil de tratar si está cerca de una estructura anatómica importante tal como el ojo, la nariz, la boca o el oído.

La ' úlcera del roedor ' es un término alternativo aplicado a veces al carcinoma basal. La exposición crónica a la luz del sol es la causa del 90% de los casos, que ocurren con más frecuencia en las partes expuestas del cuerpo la cara, los oídos, el cuello, el cuero

cabelludo, los hombros, y las nalgas. Raramente, sin embargo, los tumores aparecen en áreas no expuestas. Hay otros factores que contribuyen a este problema, como por ejemplo el contacto con arsénico, la exposición a la radiación, y complicaciones de quemaduras, de las cicatrices, de las vacunaciones y los tatuajes.

Las primeras señales de peligro pueden ser masas pequeñas o cambios en la piel, como marcas parecidas a una cicatriz o formaciones parecidas a eczemas. La primera muestra puede también ser un dolor pequeño con los bordes levantados. Es un cáncer de crecimiento lento, nunca se propaga a otras partes del cuerpo. La localización, la evolución del tamaño y los colores son los síntomas indirectos principales que conducen a su diagnóstico [NET05].

5.1.2.3 Psoriasis

Es una patología común de la piel dando por resultado un calor rojizo y picazón. Se sabe que es una patología crónica de la piel que afecta entre 1-3% de la población. La psoriasis es una alteración inflamatoria de la piel, caracterizada normalmente por erupciones en forma de placas rojas escamosas, predominando en codos, rodillas, tronco, manos/uñas y cuero cabelludo. La psoriasis puede comenzar a cualquier edad. Áreas de la piel que crecen mucho más rápido, se enrojecen y se escaman. El cuero cabelludo, los codos, y las rodillas son las zonas más comunes para la psoriasis [PSO04].

La psoriasis a día de hoy no tiene cura, aunque los medicamentos y los tratamientos en ocasiones la alivian e incluso pueden hacer desaparecer los síntomas durante temporadas que

pueden variar de un individuo a otro. No existen dos casos iguales por decirlo de alguna forma. Al igual que no existen medicamentos que funcionen de igual forma en todos los enfermos.

Existen distintos tipos de psoriasis. Dependiendo de la severidad, duración, localización, y de la forma y patrón de las zonas afectadas. La forma más común comienza con los pequeños puntos rojos. La localización de las regiones afectadas en el cuerpo es un factor muy importante para su diagnóstico [PAM04]. Es una enfermedad molesta en algunos casos y en ocasiones, dependiendo del tipo de psoriasis que se padezca, puede ocasionar cierta invalidez. En la mayoría de los casos no afecta la salud en general a menos que se desarrolle artritis grave o descamación rebelde.

La psoriasis es de causa desconocida, aunque parece ser una enfermedad hereditaria, y estar relacionada con la respuesta inmune y la inflamación. En personas sanas, las células nuevas de la piel normal tardan aproximadamente 1 mes en emigrar desde las capas más profundas de la piel hasta la superficie; en la psoriasis, este proceso sólo lleva unos días, de forma que hay un recambio constante de células, las células muertas no pueden descamarse lentamente y persisten formando grandes escamas.

5.1.3 Objetivo

En este trabajo se propone evaluar la vialidad del diagnóstico diferencial de tres anomalías (psoriasis, keratoris actínica y carcinoma basal) basado en imágenes de fluoroscopia. Este tipo de imágenes son

monocromáticas pero contienen información de capas profundas de la piel.

5.2 Descripción de técnica de adquisición de imágenes

En este trabajo se utilizan imágenes de fluorescencia. Las imágenes de fluorescencia usadas en este trabajo han sido proporcionadas por el Instituto de Biofísica de la Universidad de Regensburg, Alemania. Se aplicarán diversas técnicas de tratamiento de imágenes para extraer parámetros que caractericen cada imagen. Estos parámetros (o características) se utilizarán para su clasificación (véase Capítulo 6).

Para conseguir las imágenes con lesiones de keratosis actínica, carcinoma basal y psoriasis, se grabaron 167 imágenes de fluoroscopia con un tamaño de 768x512. Se utilizó una cámara CCD convencional. El rango de niveles de gris va desde el 0 (blanco) hasta el 255 (negro).

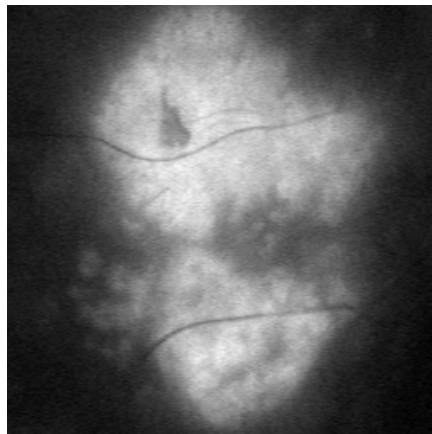
Estas imágenes habían sido pre-procesadas, fueron centradas y reducidas a un tamaño de 512x512. Seguidamente se les aplicó un factor de reducción que las dejó con el tamaño de 256x256.

El método de fluorescencia usado ha sido mediante la inducción de δ -aminolevulínico, más conocido como (ALA) [MIR01, ACK01].

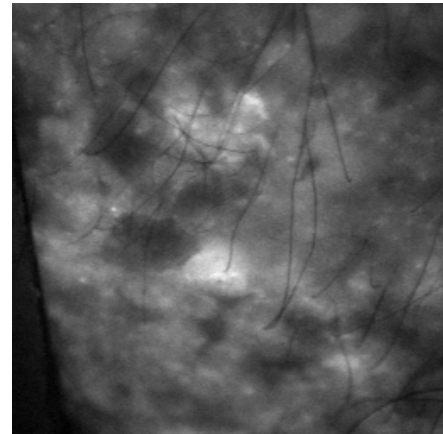
5.3 Descripción de la base de datos

La base de datos está compuesta por 167 imágenes de fluoroscopia que se reparten en tres grupos de imágenes del siguiente modo:

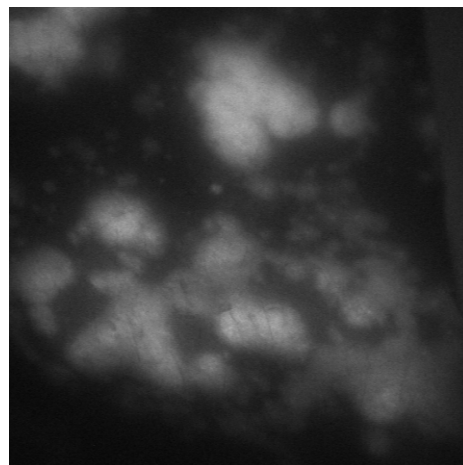
- 1- 50 imágenes de keratosis actínica (pre-cáncer).
- 2- 50 imágenes de carcinoma en células basales (lesión cancerosa benigna).
- 3- 67 imágenes de psoriasis (patología no cancerosa).



(a) Keratosis actínica.



(b) Carcinoma basal.



(c) Psoriasis

Figura 5.2 a) *Keratosis actínica*, b) *carcinoma basal* y c) *Psoriasis*.

Se ha reducido el tamaño de las imágenes para eliminar carga computacional del sistema y obtener mejores resultados durante el proceso de extracción de parámetros.

El objetivo de este estudio es diseñar un sistema clasificador de imágenes que distinga entre tres tipos de lesiones.

- ✧ Keratosis actínica (Pre-cáncer)
- ✧ Carcinoma basal (cáncer benigno)
- ✧ Psoriasis (lesión no cancerosa)

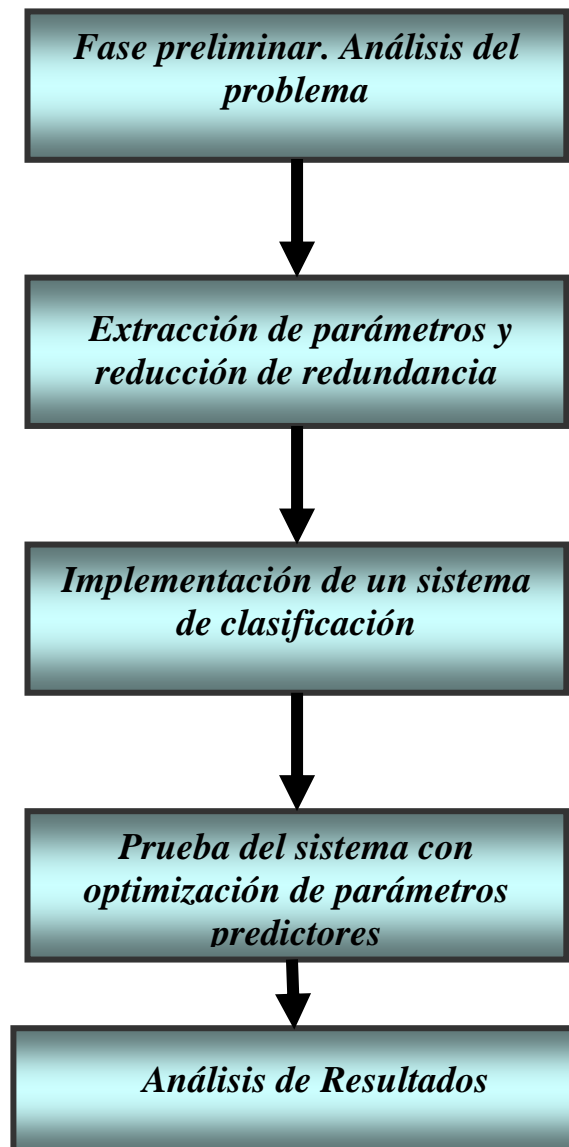


Figura 5.3 Esquema de desarrollo de la aplicación de diagnóstico.

En la Figura 5.3 se puede ver el esquema de trabajo seguido. En la primera fase se estudia el problema de detectar y clasificar las imágenes. La segunda fase se dedicó a la extracción de parámetros de las imágenes. En la tercera fase se implementan varios sistemas de clasificación. En la cuarta fase se prueban y optimizan los sistemas de clasificación y en la última fase se analizan los resultados obtenidos.

5.4 Extracción y análisis de parámetros

El objetivo de esta parte del estudio es procesar imágenes y realizar una extracción de características suficientes para su diferenciación. Se describen qué parámetros se han utilizado y el modo en que se han obtenido. Para ello se han seguido los siguientes pasos:

1. En un primer paso se extraen características diversas utilizando distintas técnicas de procesamiento de imágenes.
2. En un segundo paso se eliminan aquellas características más redundantes.
3. En un tercer paso se evalúa de forma preliminar el grado de relevancia de cada característica para la tarea de clasificación. Se ha realizado una representación gráfica sencilla para facilitar una evaluación preliminar basada en una simple inspección visual.

Al finalizar estos pasos se puede conocer qué parámetros son los que mayor información presentan para este estudio aunque es imposible evaluar si existen combinaciones de características que

llevan información cuando se presentan conjuntamente (esto requeriría un estudio estadístico incluyendo covarianzas). En este trabajo, tras la etapa de extracción de características, en vez de realizar un estudio estadístico sobre la significancia de cada una se representan gráficamente para facilitar su evaluación preliminar sencilla. Será en el Capítulo 6 donde herramientas de selección automática de características indiquen el conjunto más relevante para las tareas de diagnóstico que se definan.

Todo ello se acompaña de las representaciones gráficas. En todas las figuras se ha seguido el mismo esquema de colores, de tal modo que basta con comentar aquí qué colores representan a cada una de las imágenes:

- **Rojo (o)**: (AKB) Keratosis actínica (Pre-cáncer)
- **Azul (*)**: (BAS) Carcinoma basal (cáncer)
- **Verde (+)**: (PSO) Psoriasis

Es decir, hay 50 imágenes con lesiones de pre-cáncer, otras 50 con lesiones cancerosas y 67 de psoriasis.

Para la extracción de las características es necesario realizar un procesamiento de la imagen. El tratamiento de imágenes engloba una serie de técnicas que comprenden operaciones cuyo origen es una imagen de entrada a la cual se le aplica un procesamiento con el fin de obtener como resultado otra imagen diferente; o un número (que representa una característica o parámetro concreto).

El resultado de todo el tratamiento será usado para la interpretación humana (por ejemplo, un médico) o por un computador (aplicación de diagnóstico automático).

El objetivo del presente trabajo se centrará en el Capítulo 6 en la interpretación automática de características de imágenes para diagnóstico médico. El conjunto de técnicas que forman la etapa de procesamiento automático de imágenes se puede estructurar de la siguiente manera:

- 1- Tratamiento a bajo nivel.
- 2- Tratamiento a medio nivel.

5.4.1 Tratamiento de imágenes a bajo nivel

En este nivel se utilizan filtros de detección de bordes para resaltar los cambios de contraste con el fin de determinar y evaluar los contornos de la imagen. Posteriormente cuantificaremos de forma global o local la cantidad o forma de los bordes detectados.

Por la simplicidad del método, y su buen funcionamiento, se llevará a cabo esta tarea utilizando detectores de bordes basados en diferencias finitas. Éstos consisten en calcular en forma aproximada los gradientes en diferentes direcciones utilizando diferencias finitas de algún orden arbitrario [FLE92].

5.4.1.1 *Detección de bordes*

La detección de bordes es el método más común para detectar discontinuidades significativas en el nivel de gris. Un borde se define como la frontera entre dos regiones con nivel de gris relativamente diferente [CAN86]. Se supone que las regiones son suficientemente homogéneas para que la transición entre dos de ellas se pueda determinar mediante las discontinuidades del nivel de gris.

La idea básica detrás de cualquier detector de bordes es el cálculo de un operador local de derivación. La primera derivada en cualquier punto de la imagen vendrá dada por la magnitud del gradiente. En la Figura 5.4 se puede observar que la primera derivada es positiva para cambios a nivel de gris mas claro, negativa en caso contrario y cero en aquellas zonas con nivel de gris uniforme. Mientras que la segunda derivada vendrá dada por el operador Laplaciano [LIM90], se presenta el valor positivo en la zona oscura de cada borde, valor negativo en la zona clara de cada borde y valor cero en las zonas de valor de gris constante y justo en la posición de los bordes. El valor de la magnitud de la primera derivada nos sirve para detectar la presencia de bordes, mientras que el signo de la segunda derivada nos indica si el píxel pertenece a la zona clara o a la zona oscura. Además la segunda derivada presenta siempre un cruce por cero en el punto medio de la transición. Esto puede ser muy útil para localizar bordes en una imagen [FLE92]

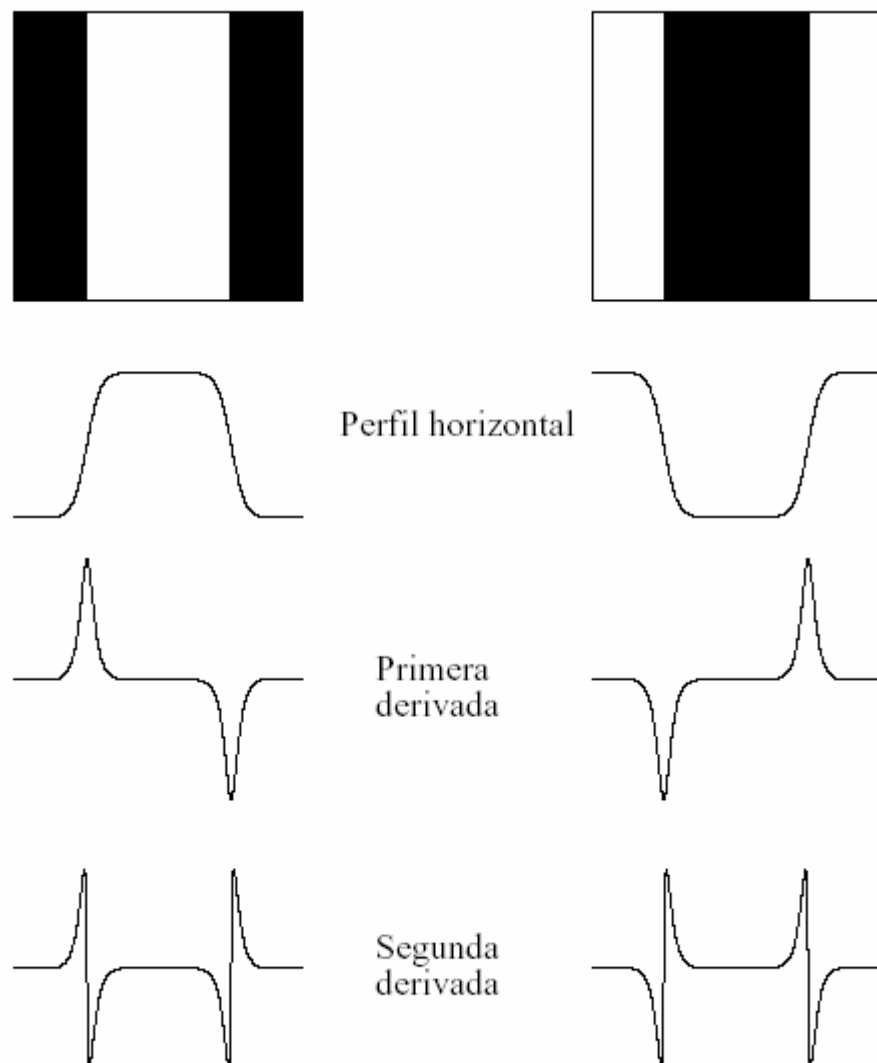


Figura 5.4 *Detección de bordes empleando operadores de derivación. La segunda derivada tiene un cruce por cero en la posición de cada borde.*

Uno de los objetivos principales de este tipo de técnicas es reducir de forma drástica el volumen de información de una imagen. Se trata de aislar los bordes de todos los objetos que aparecen en la imagen. De este modo posteriormente sólo se procesan un número limitado de líneas y contornos. Existen diferentes filtros para la extracción de bordes entre los cuales están:

- 1- El detector de bordes de Sobel.
- 2- El detector de bordes de Prewitt.
- 3- El detector de bordes de Canny.
- 4- El detector de bordes Laplaciano.

En este estudio extraemos bordes utilizando todos estos operadores para obtener diversas características relacionadas con la detección de bordes basada en distintos operadores. Dejaremos para etapas posteriores la evaluación de la redundancia entre estas técnicas en el marco de la aplicación en la que se centra la etapa final de esta memoria.

5.4.1.1.1 Gradiente

El borde de la imagen se define como variaciones locales de la intensidad de la imagen. Por lo tanto, las técnicas locales de diferenciación de la imagen pueden producir operadores detectores de borde. En la detección de bordes es muy importante la magnitud del vector gradiente en la dirección de la máxima variación de la imagen $I(x, y)$ en el punto (x, y) [PIT00].

$$\nabla I = \begin{bmatrix} I_x \\ I_y \end{bmatrix} = \begin{bmatrix} \frac{\partial I}{\partial x} \\ \frac{\partial I}{\partial y} \end{bmatrix} \quad (5.1)$$

EL módulo del gradiente se calcula como:

$$\|\nabla I\| = \sqrt{I_x^2 + I_y^2} \quad (5.2)$$

En general el gradiente se suele aproximar mediante la expresión:

$$\|\nabla I\| \approx |I_x| + |I_y| \quad (5.3)$$

Y la dirección del vector gradiente

$$\alpha(x, y) = \tan^{-1} \frac{I_y(x, y)}{I_x(x, y)} \quad (5.4)$$

El cálculo del gradiente se basa en obtener las derivadas parciales para cada píxel. La primera derivada en dos o más direcciones. Aplicando una primera derivada en la imagen, en la dirección de la variación más importante, demarcará los bordes y realizará la visibilidad de saltos pequeños y otros detalles.

Se puede producir una imagen derivada con apariencia suavizada con unos pocos pasos, aplicando una convolución con un núcleo (“Kernel”) asimétrico. Se recurre a una serie de máscaras donde las más comunes son las de Prewitt y Sobel [MUN01].

-1	-2	-1
0	0	0
1	2	1

-1	0	1
-2	0	2
-1	0	1

(a)

-1	-1	-1
0	0	0
1	1	1

-1	0	1
-1	0	1
-1	0	1

(b)

Figura 5.5 *Los operadores del Prewitt y Sobel (vertical y horizontal)*
 (a) *operador máscara de Prewitt (vertical a la izquierda y horizontal a la derecha)* (b) *operador máscara de Sobel (vertical a la izquierda y horizontal a la derecha).*

En la Figura 5.5 se muestran operadores de Prewitt y Sobel para determinar las derivadas parciales. El operador de Sobel tiene la ventaja de que proporciona un suavizado además del efecto de derivación.

5.4.1.1.2 *El detector de borde de Sobel*

El detector de bordes de Sobel utiliza un núcleo simple de convolución para crear una serie de magnitudes del gradiente. Podemos representar el núcleo (kernel) de la convolución matemáticamente como se muestra en la expresión (5.5).

$$N(x, y) = \sum_{k=-1}^1 \sum_{j=-1}^1 K(j, k) p(x-j, y-k) \quad (5.5)$$

Donde $N(x,y)$ es la imagen de salida, $p(x-j,y-k)$ es la imagen de entrada, y $K(j,k)$ es el núcleo de la convolución que se define como el siguiente matriz.

$$\begin{bmatrix} k_{-1,-1} & k_{-1,0} & k_{-1,1} \\ k_{0,-1} & k_{0,0} & k_{0,1} \\ k_{1,-1} & k_{1,0} & k_{1,1} \end{bmatrix} = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad (5.6)$$

El detector de bordes de Sobel se procesa en 2 D como medida espacial de gradiente en una imagen y resalta las regiones de alta frecuencia espacial que corresponden a los bordes. Se suelen usar dos mascarar: alisar la imagen obtenida suavizada para después filtrarla con las mascarar, una para la primera componente del vector y otra para la segunda componente [MUN01]. Estas máscarar son dos núcleos (kernel) de convolución, uno para detectar cambios en el contraste vertical (h_x) y el otro para detectar el contraste horizontal (h_y) como en la Figura (5.5).

Estas máscarar se diseñan para obtener una respuesta máxima a los bordes verticales y horizontales, una máscara para cada una de las dos orientaciones perpendiculares [FIS00, GON92]. Las máscarar se pueden aplicar por separado a la imagen de entrada, para producir las medidas separadas del componente de gradiente en cada orientación. Éstos se pueden entonces combinar para encontrar la magnitud absoluta del gradiente en cada punto y la orientación de ese gradiente.

En la base de datos a usar se ha aplicado el operador de Sobel en derivación horizontal y vertical con diferentes valores de umbral

como se puede ver en la tabla de parámetros (Tabla 5.1). Algunas características o parámetros obtenidos utilizando esta técnica son las siguientes:

P1: Contraste de la imagen (Sobel). Suma de todos los valores de los bordes de una imagen. Sin eliminar ninguno con el valor de umbral=0.

P2: Contraste de la imagen (Sobel) con umbralizado automático. Suma de todos los valores de los bordes de una imagen. Previamente se filtran los bordes usando un umbral automático (para que el porcentaje de píxeles representando bordes sea un 30%).

P3-P8: Contraste de la imagen (Sobel) umbralizado. Suma de todos los valores de los bordes de una imagen. Previamente se filtran los bordes usando diversos umbrales. (Valores de umbral: 1, 2, 3, 4, 5, 6. – se eliminan los bordes que son menores que estos valores-).

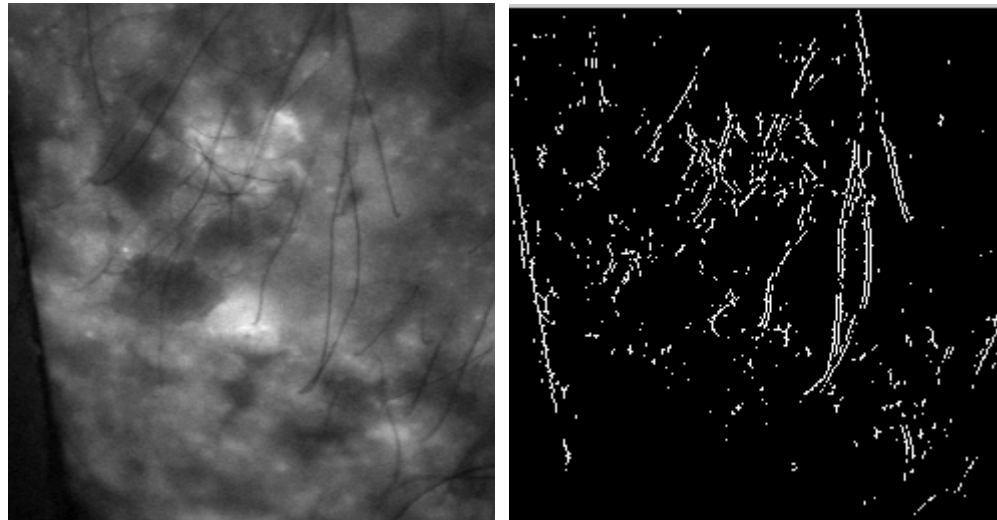
P9: Contraste vertical (Sobel). Detectamos solamente los bordes verticales y sumamos todos los valores de la imagen. Sin eliminar ninguno con el valor de umbral=0.

P10-P15: Contraste vertical (Sobel) umbralizado. Detectamos solamente los bordes verticales con diversos umbrales, y sumamos todos los valores de la imagen. (Valores de umbral: 1, 2, 3, 4, 5, 6).

P16: Contraste horizontal (Sobel). Detectamos solamente los bordes horizontales y sumamos todos los valores de la imagen. Sin eliminar ninguno con el valor de umbral=0.

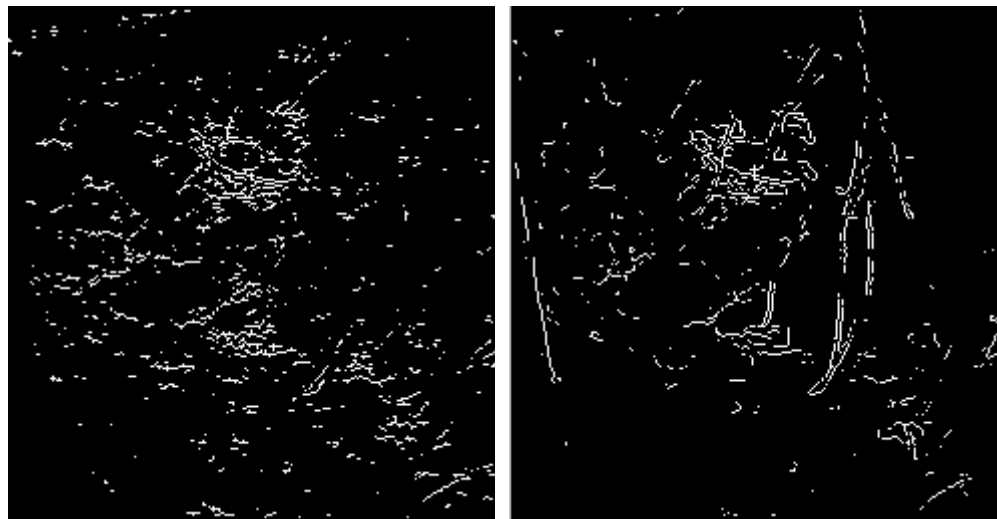
P17-P22: Contraste horizontal (Sobel) umbralizado. Detectamos solamente los bordes horizontales con diversos umbrales, y sumamos todos los valores de la imagen. (Valores de umbral: 1, 2, 3, 4, 5, 6).

En la siguiente figura se puede ver un ejemplo del resultado de aplicar las mascararas de Sobel a una imagen que tiene Keratosis actínica.



(a)

(b)



(c)

(d)

Figura 5.6 (a) Imagen original, (b) derivación vertical usando Sobel (umbral automático para conservar sólo el 30% de los píxeles) (c) derivación horizontal usando Sobel (umbral automático), (d) imagen gradiente usando Sobel con un umbral mayor (4).

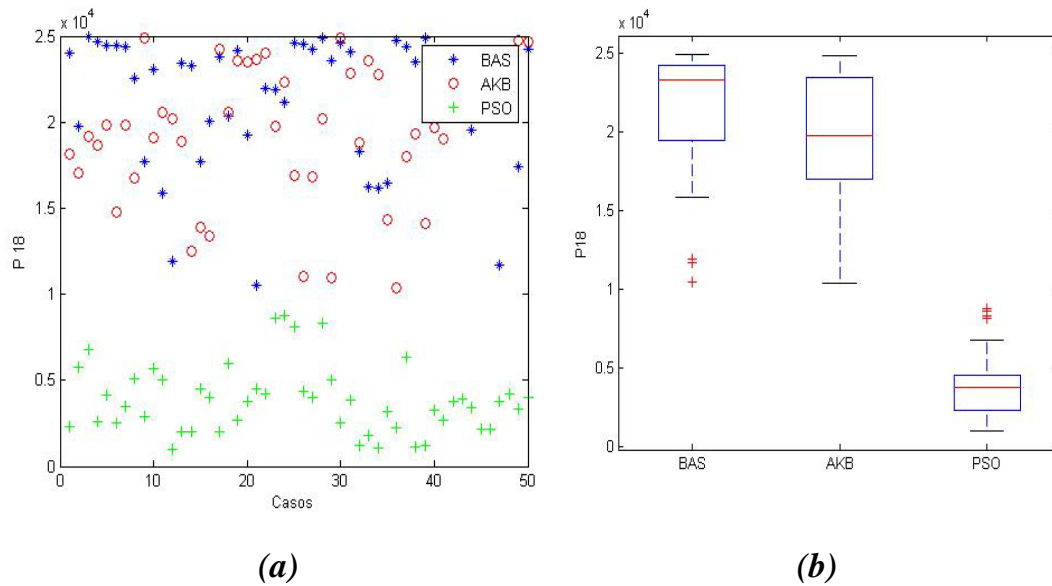


Figura 5.7 (a) Distribución de muestras de imágenes al utilizar el detector de borde Sobel, con un umbral= 3 (P18). (b) Media de los valores obtenidos.

En la Figura 5.7 se puede ver la distribución en el espacio del valor obtenido para cada una de las imágenes (siguen el esquema de colores comentado al comienzo de la sección 5.5). Tal y como se puede observar a simple vista, la distinción de zonas de agrupamiento según colores es difícil y no se observa que las dos clases (rojo y azul) sigan un patrón bien diferenciado de distribución. Por otro lado la clase que tiene el color verde (Psoriasis) se puede observar que está agrupada en una zona bien diferenciada de los demás.

5.4.1.1.3 El detector de borde de Prewitt

Este detector tiene las mismas características que el detector de Sobel con la diferencia de que utiliza otras máscaras que son usadas para

calcular el valor del píxel en la imagen. Este valor está limitado dentro de un rango valores mínimos y máximos de la correspondiente ubicación. La detección de bordes de Prewitt [JAM97] se realiza con operaciones convolutivas con las máscaras de la Figura 5.5, para calcular el valor del píxel en la imagen [GON92].

Se ha aplicado a la base de datos el operador de Prewitt en derivación horizontal y vertical con diferentes valores del umbral como se puede ver en la tabla de parámetros (Tabla 5.1). Algunas características obtenidas utilizando esta técnica son las siguientes:

P23: Contraste de la imagen (prewitt). Suma de todos los valores de los bordes de una imagen. Sin eliminar ninguno con el valor de umbral=0.

P24: Contraste de la imagen (prewitt) umbralizado automático. Suma de todos los valores de los bordes de una imagen. Previamente se filtran los bordes usando un umbral automático. (Para que el porcentaje de píxeles representando bordes sea un 30%).

P25-P32: Contraste de la imagen (prewitt) umbralizado. Suma de todos los valores de los bordes de una imagen. Previamente se filtran los bordes usando diversos umbrales. (Valores de umbral: 1, 2, 3, 4, 5, 6, 7, 8).

P33: Contraste vertical (prewitt). Detectamos solamente los bordes verticales y sumamos todos los valores de la imagen. Sin eliminar ninguno con el valor de umbral=0.

P34-P38: Contraste vertical (prewitt) umbralizado. Detectamos solamente los bordes verticales con diversos umbrales, y sumamos todos los valores de la imagen. (Valores de umbral: 1, 2, 3, 4, 5).

P39: Contraste horizontal (prewitt). Detectamos solamente los bordes horizontales y sumamos todos los valores de la imagen.

P40-P44: Contraste horizontal (prewitt) umbralizado. Detectamos solamente los bordes horizontales con diversos umbrales, y sumamos todos los valores de la imagen. (Valores de umbral: 1, 2, 3, 4, 5).

En la siguiente figura se puede ver un ejemplo del resultado de aplicar las mascararas de prewitt a una imagen que tiene keratosis actínica con un umbral automático.

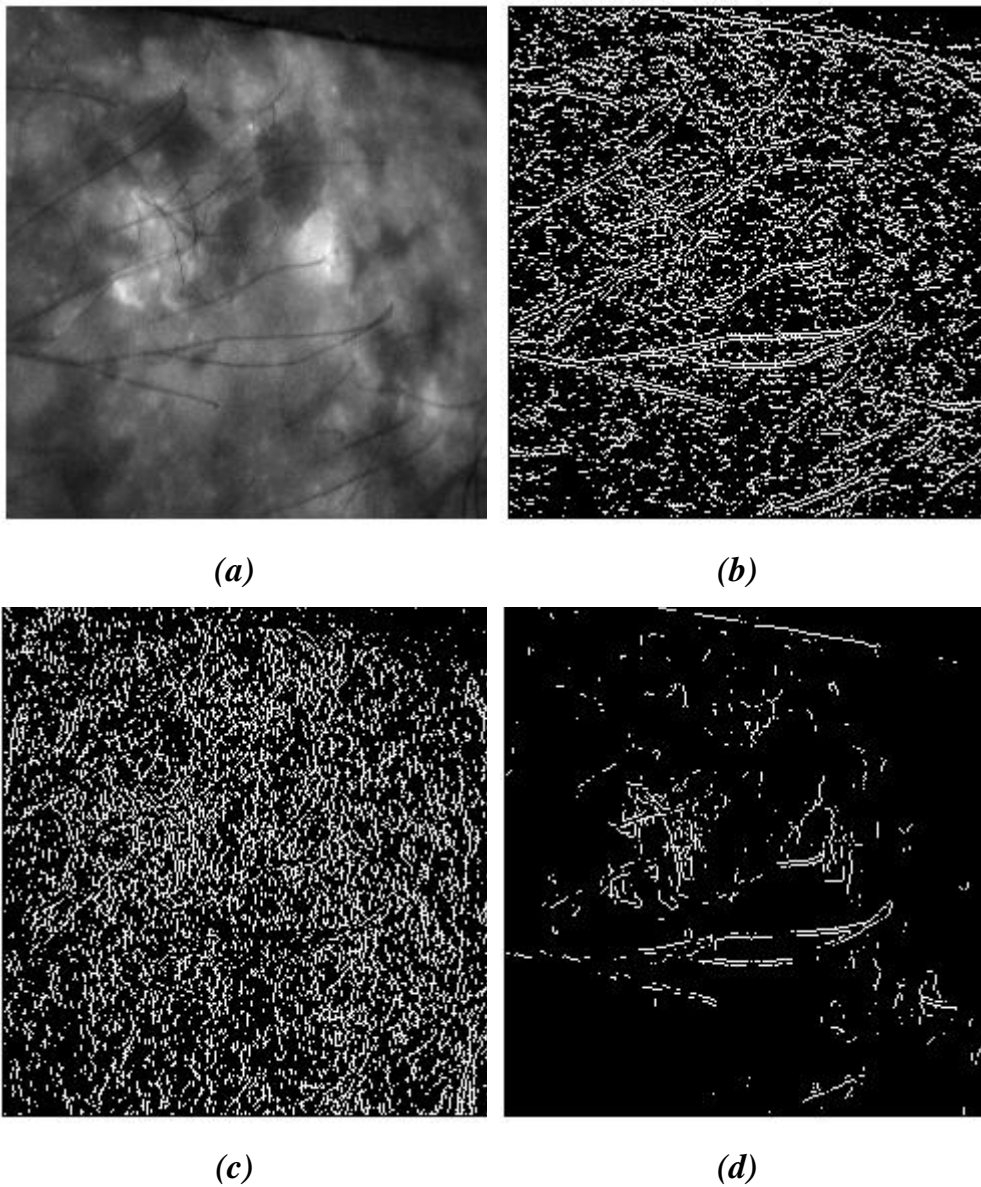


Figura 5.8 (a) Imagen original (b) Derivación horizontal usando prewitt. (Umbral automático). (c) Derivación vertical usando prewitt.

(Umbral automático). (d) Imagen gradiente usando prewitt con un umbral mayor (6).

Se puede ver en la siguiente figura de las medias de los valores obtenidos de este tipo de bordes. Se puede destacar la gran diferenciación que se consigue entre los casos de tumores con respecto a los de psoriasis.

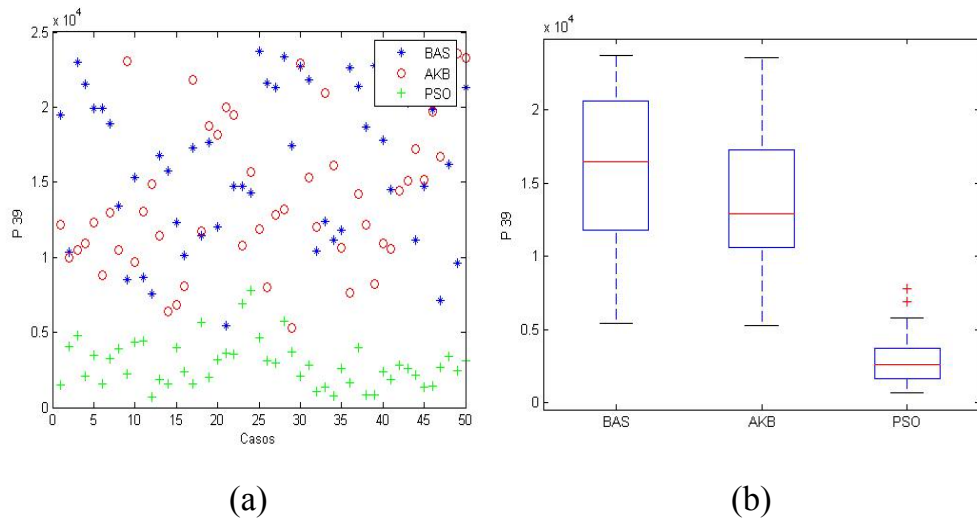


Figura 5.9 (a) Distribución de muestras al utilizar el detector de borde prewitt (P 36) (con un umbral=3). (b) Medida de la media y desviación estándar de los valores obtenidos.

5.4.1.1.4 El detector de borde de Canny

Es el detector de bordes más poderoso que existe actualmente. Elimina ruidos en la imagen (suavizada) con filtros gaussianos unidimensionales. Para determinar los píxeles donde se produce la máxima variación se calcula el gradiente de la imagen suavizada usando una aproximación del gradiente de la función gaussiana [CAN86].

La función gaussiana del gradiente tendrá valores grandes donde la variación local de la intensidad sea grande. Por lo tanto, se realiza la selección del umbral para eliminar los píxeles que no tienen magnitud alta. Posteriormente se realiza un proceso de eliminación de falsos bordes y realzado de bordes poco definidos. Este proceso se realiza eliminando aquellos píxeles que no son máximos locales.

El parámetro que se extrae con este filtro es:

P45: Contraste de la imagen (Canny). Suma de todos los valores de los bordes de una imagen. Previamente se filtran los bordes usando umbrales bajos y altos automáticamente (el alto se escoge para que sólo el 30% de los pixels representen bordes y el bajo para que sea un 40% del valor del umbral alto).

En la siguiente figura se puede ver un ejemplo del resultado de aplicar el detector de bordes de Canny a una imagen que tiene Keratosis actínica.

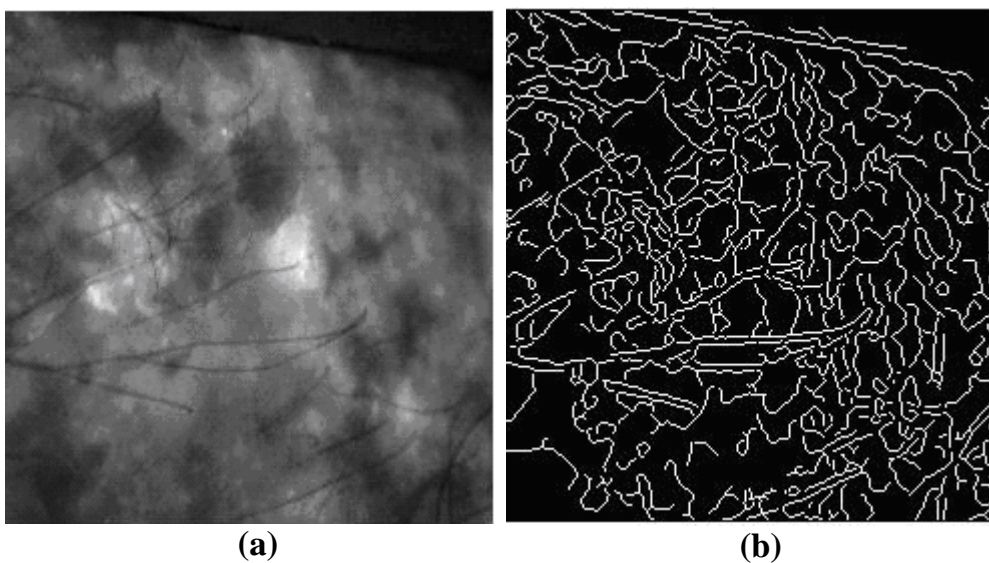


Figura 5.10 (a) Imagen original (b) Imagen de bordes usando Canny.

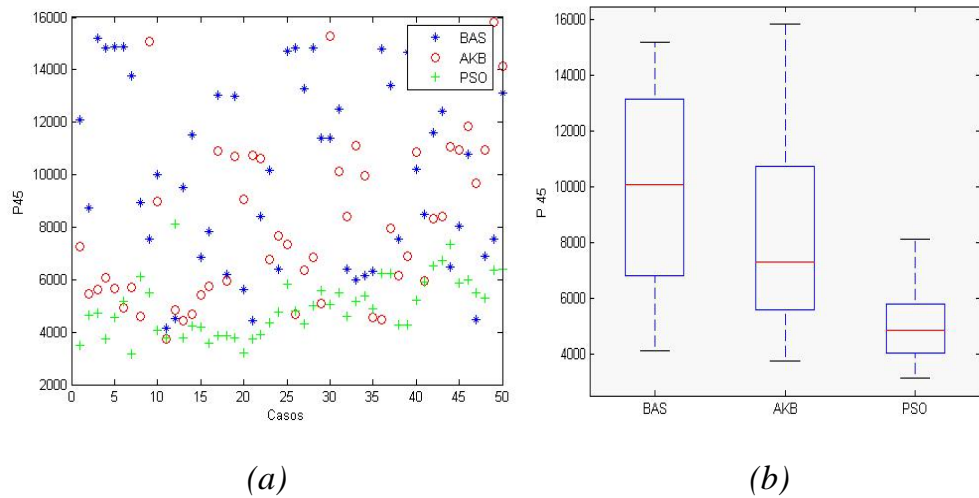


Figura 5.11 (a) Distribución de muestras al utilizar el detector de borde canny (P45). (b) Medida de la media y desviación estándar de los valores obtenidos.

Como se puede apreciar en la Figura 5.11, al igual que en casos anteriores es de destacar la diferencia entre las imágenes con psoriasis y las imágenes con keratosis actínica o células basales. Las distribuciones de muestras de lesiones precancerosas (AKB) y cancerosas (BAS) se superponen sin dividir zonas de forma clara y no se distingue entre el keratosis actínica y la carcinoma basal.

5.4.1.1.5 El detector de bordes Laplaciano

El operador Laplaciano [FIS00] marca la presencia de un borde cuando la salida del operador realiza una transición por cero. En principio, las localizaciones de los cruces por cero pueden ser estimadas a nivel de subpíxeles usando interpolación lineal; pero los resultados pueden ser inexactos. Además para las características que se describen en este apartado, la localización exacta de los bordes es irrelevante.

El Laplaciano de una función bidimensional $I(x, y)$ es una derivada de segundo orden y está definida por la ecuación siguiente:

$$\nabla^2 I = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \quad (5.7)$$

Normalmente se utiliza una máscara para calcular el operador de laplaciano que se muestra en la siguiente matriz:

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} \quad (5.8)$$

Al aplicar estas máscaras a las imágenes se pueden extraer algunas características como se puede ver en la siguiente Figura 5.12.

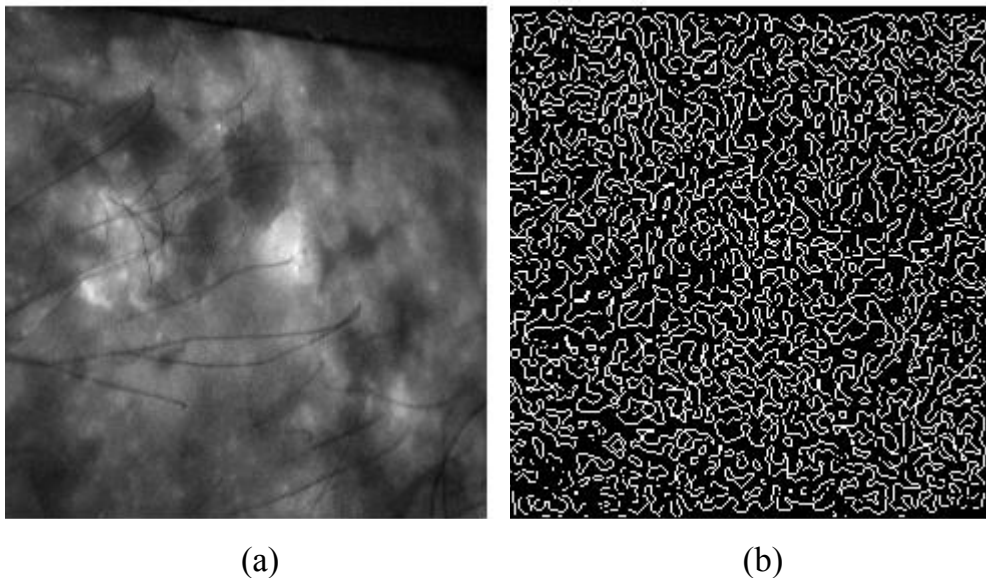


Figura 5.12 (a) Imagen original (b) Imagen gradiente usando el operador Laplaciano (con umbral= 0.5).

Las características que se han extraído utilizando esta técnica con diversos umbrales son las siguientes:

P46-P47: Contraste de la imagen (Laplaciano) umbralizado. Suma de todos los valores de los bordes de una imagen. Los bordes se filtran previamente usando diversos umbrales. (Umbrales=0.5, -0.5)

En la Figura 5.13 volvemos a encontrar ahora una situación similar a las anteriores, el único grupo que se diferencia claramente es el de psoriasis (PSO), pero esta característica no contribuye en nada a diferenciar entre las lesiones cancerosas que se consideran en este estudio.

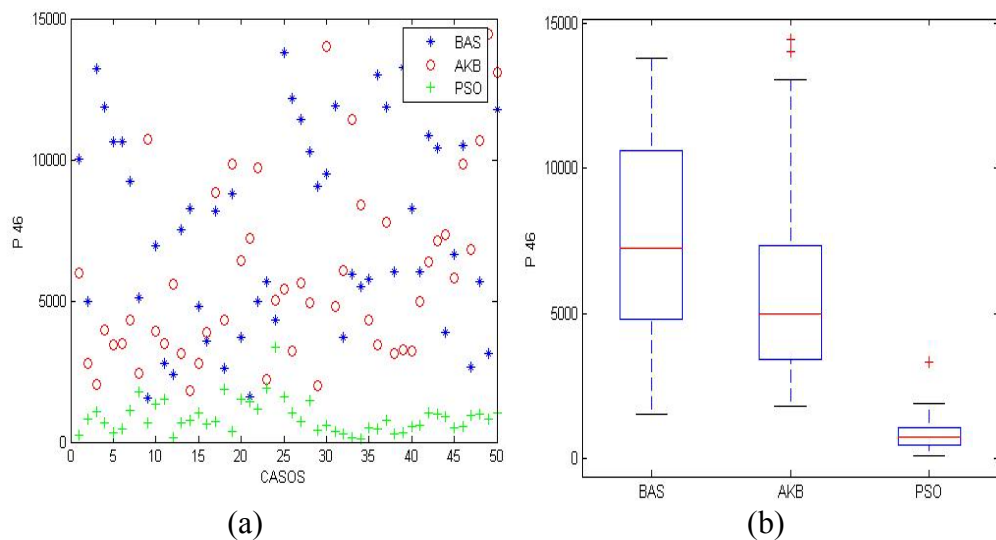


Figura 5.13 (a) Distribución de muestras al utilizar el detector de borde Laplaciano (P46) con umbral=0.5 (b) Medida de la media y desviación estándar de los valores obtenidos.

5.4.1.2 Transforma de Fourier

La transformada de Fourier es una herramienta muy importante en el procesamiento de imágenes digitales. Se define la Transformada de

Fourier Discreta (DFT) de una imagen $f(x, y)$ de tamaño N filas y M columnas como:

$$F(p, q) = \frac{1}{NM} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} f(x, y) e^{-j(2\pi/N)p_x} e^{-j(2\pi/M)q_y} \quad (5.9)$$

La transformada de Fourier da como resultado un número complejo, por lo tanto, queda definido por su módulo y su argumento. En este trabajo se ha utilizado el algoritmo de *la transformada rápida de Fourier (Fast Fourier Transform) (FFT)* para calcular la DFT, donde se han calculado dos dimensiones de la transformada Fourier para cada columna y fila de la imagen. En primer lugar se calculan para la matriz de imagen los valores máximos de cada una de las filas, para después comparar estos resultados obtenidos y sacar de todos ellos el valor máximo. En un segundo paso se obtienen los valores medios de cada fila que componen la matriz. Después se calcula el valor máximo de estas medias para cada imagen. También se calcula la desviación estándar del máximo de cada fila de esta forma, para cada imagen se obtienen las siguientes características:

P48: Valor máximo de frecuencia espacial (Fourier). 1) Sumamos todos los componentes en cada fila de la imagen, 2) nos centramos en la fila con un valor acumulado máximo, 3) P48 es el valor máximo dentro de esta fila.

P49: Componente medio de la frecuencia espacial (Fourier). Calculamos el promedio de los componentes a lo largo de cada fila y elegimos el valor máximo entre las diversas filas.

P50: Desviación de estándar de la frecuencia espacial (Fourier). Calculamos la desviación de estándar de todos los componentes de la frecuencia.

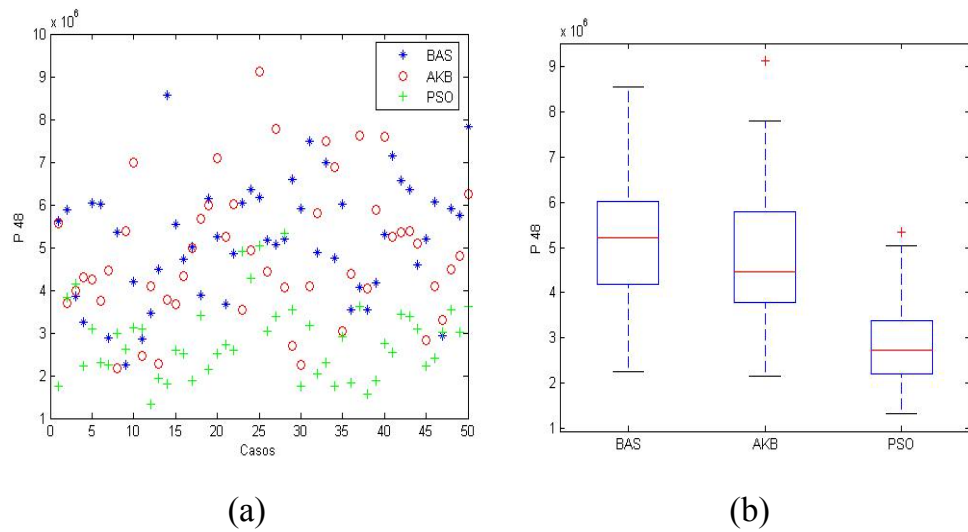


Figura 5.14 (a) Distribución de muestras al utilizar la transformada de Fourier (P_{48}) (b) Medida de la media y desviación estándar de los valores obtenidos.

5.4.2 Tratamiento de imágenes a nivel medio. Morfología matemática

La morfología matemática es una técnica de procesamiento y análisis de imágenes que se basa en el estudio de la forma que se consigue mediante una comparación del objeto de estudio con un elemento de prueba (llamado elemento estructurante, segmento, disco, línea, cuadrado, ..., etc.). Dependiendo de si se ajusta o no a la forma objeto de estudio, nos permite extraer propiedades e información de la forma del objeto de estudio. La morfología matemática ha demostrado gran capacidad para solventar una amplia gama de problemas sobre imágenes binarias (en blanco y negro) o continuas (en escala de grises o a color).

La morfología matemática se basa en geometría y forma, las operaciones morfológicas simplifican imágenes y conservan las principales características de formas de los objetos.

Un sistema de operadores como los de la morfología matemática es útil porque pueden formarse composiciones de sus operadores, que cuando actúan sobre formas complejas, son capaces de descomponerlas en partes con sentido y además separarlas en partes que se pueden marcar. Un sistema de operadores de este tipo y su composición permite que las formas subyacentes sean identificadas y reconstruidas de forma óptima a partir de sus formas distorsionadas y ruidosas. Además permite que cada forma se entienda en función de una descomposición, siendo cada entidad de esa descomposición una forma simple apropiada.

Las transformaciones morfológicas, si se usan, constituyen usualmente una parte intermedia de la secuencia del procesamiento de imágenes. En una primera fase, la imagen es digitalizada y preprocesada usando operadores de convolución locales y luego es segmentada para obtener una imagen binaria en la que se separan los objetos del fondo. Las operaciones morfológicas pueden formar una segunda fase que opera sobre la forma de esos objetos. El último paso del procesamiento evalúa los resultados de la morfología usando descriptores numéricos o sintácticos.

La morfología matemática se puede usar, entre otros, con los siguientes objetivos:

- Pre-procesamiento de imágenes (supresión de ruido, simplificación de formas),

- Destacar la estructura de los objetos (extraer el esqueleto, marcado de objetos, envolvente convexa, ampliación, reducción),
- Descripción cualitativa de objetos (área, perímetro, etc).

5.4.2.1 Morfología binaria

(a) Binarización

Se intentarán extraer algunas características a partir de las imágenes con el objetivo de obtener nuevos parámetros significativos para este trabajo. En muchas de las aplicaciones de tratamiento de imágenes es suficiente con trabajar con imágenes binarias para resolver el problema propuesto.

Una imagen binaria como su nombre indica no es más que una imagen en blanco (0) y negro (1) donde se diferencia claramente el fondo (=0) del objeto (=1).

Las imágenes binarias se obtienen como resultado de aplicar algunas técnicas de segmentación (ejemplo: umbralización) a una imagen de niveles de gris de modo que los objetos queden claramente diferenciados en el fondo.

Aplicando esta técnica a nuestro problema lo que se intentará será diferenciar ciertas zonas de la imagen sobre las que después se podrá calcular distintos parámetros como el área, si el objeto representa alguna forma concreta (por ejemplo los melanomas se

caracterizan por no tener una simetría respecto a un origen: este caso concreto se deja como posible ampliación futura si se estudian melanomas).

En ocasiones, llegar a la segmentación que representa la imagen binaria, es suficiente para obtener características discriminantes que nos permitan describir (extracción de características) correctamente los objetos presentes en la imagen.

El procesado morfológico se refiere al estudio de la topología de las estructuras de los objetos en imágenes. Este procesado tiene dos operaciones básicas: *dilatación* y *erosión*.

La dilatación y la erosión parten de la definición de suma y resta de Minkowski [GON92, SER88]. Se definen de la siguiente manera:

(b) Dilatación

La dilatación es la transformación morfológica que se aplica a dos elementos utilizando la suma. La dilatación de un conjunto A por un elemento estructurante B simétrico, es el lugar de las posiciones del centro del elemento estructurante B cuando B toca a A .

Más formalmente la dilatación de A por B se nota $A \oplus B$ y se define mediante la siguiente expresión:

$$A \oplus B = \{c \in X / c = a + b \text{ para algún } a \in A \text{ y } b \in B\}. \quad (5.10)$$

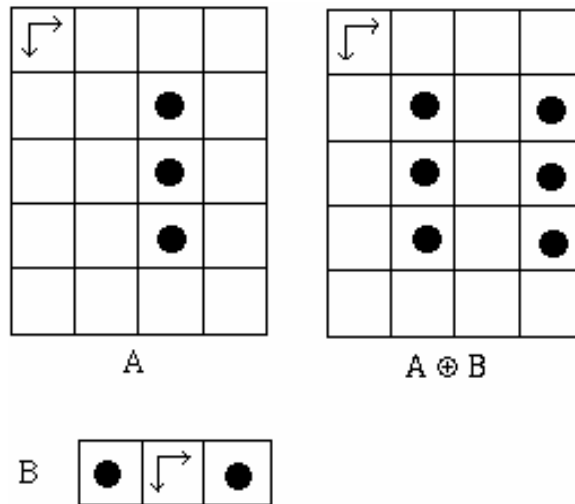


Figura 5.15 Dilatación $A \oplus B$. Imagen A . Elemento estructurante B .

(c) Erosión

La erosión es la operación morfológica dual de la dilatación, la erosión de un conjunto A por un elemento estructurante B simétrico, es el lugar de las posiciones del centro del elemento estructurante B cuando B está incluido en A [SER88].

La erosión de A por B se nota $A \ominus B$ y su definición es:

$$A \ominus B = \{x \in X / x + b \in A \text{ para cualquier } b \in B\}. \tag{5.11}$$

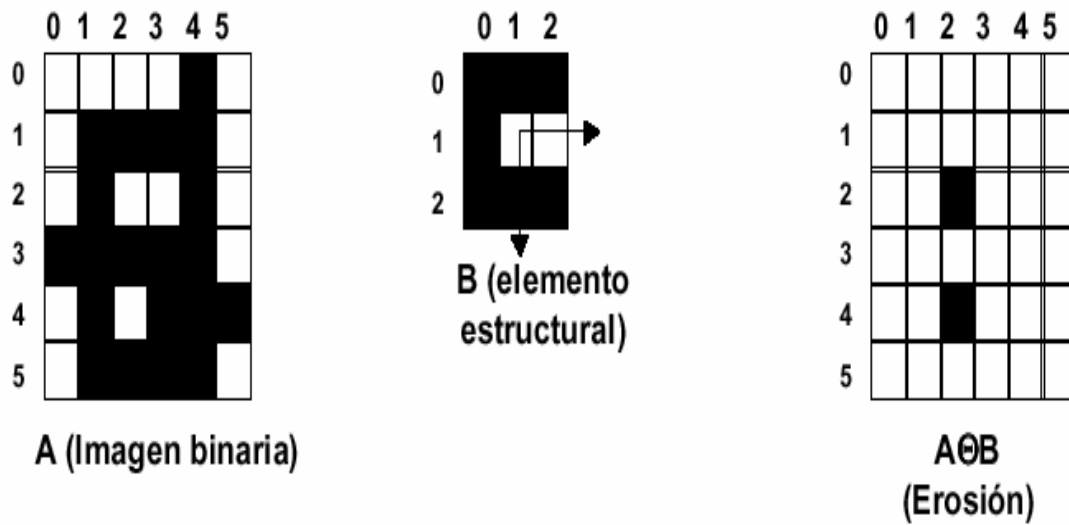


Figura 5.16 *Erosión.*

La utilidad de la erosión puede apreciarse mejor cuando ésta se expresa de forma diferente. La erosión de una imagen, A , por un elemento estructurante, B , es el conjunto de todos los elementos $x \in X$ para los cuales B trasladado por x está contenido en A .

Las transformaciones de dilatación y erosión tienen una marcada similitud, en la cual uno actúa sobre el primer plano, mientras que el otro lo hace sobre el fondo.

La dilatación en general causa que los objetos se dilaten o crezcan de tamaño, mientras que la erosión causa que los objetos se encojan. La cantidad que crecen o que se encogen los objetos depende de la elección del elemento estructurante.

Las dos estructuras de elementos más comunes son la conectividad a 4 y conectividad a 8 como se puede ver en la siguiente figura.

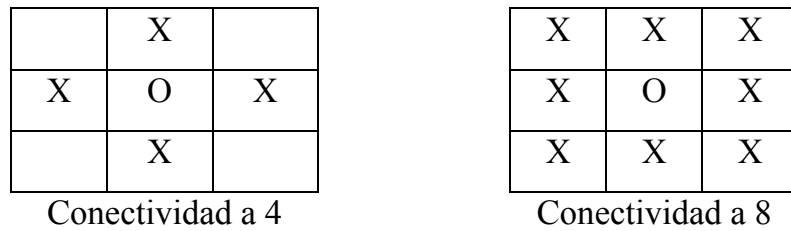


Figura 5.17 Conectividad a 4 y conectividad a 8.

En el trabajo realizado con el programa de MATLAB se recurre al comando *strel*. $SE = STREL('disk', R)$ Este comando crea elementos en forma de disco, donde el radio R , está a priori ya definido y debe de ser un número entero no negativo. Se pretende usar para la dilatación un *strel* del disco con diferentes radios y para la erosión se utilizarán dos tipos de operadores *strel* (con el disco y el octágono).

Los operadores morfológicos en general (como el de dilatación y erosión) se suelen utilizar sobre imágenes binarias. Así que todas las imágenes se han binarizado previamente utilizando un umbral automático (definido por el método de Otsu [OTS79] para quedarnos con un 30% de los píxeles activos. Tras esta binarización se aplican los operadores morfológicos.

Finalmente sólo se han seleccionado los siguientes parámetros:

P51: Dilatación de la imagen usando un elemento de la estructura (disco de la radio 1). Número de píxeles activos.

P52: Dilatación de la imagen usando un elemento de la estructura (disco de la radio 2). Número de píxeles activos.

P53: Erosión de la imagen usando un elemento de la estructura (disco de la radio 1). Número de píxeles activos.

P54: Erosión de la imagen usando un octágono (de tamaño 3 píxeles) como elemento de la estructura. Número de píxeles activos.

Se puede definir parámetros de dilatación y erosiones con radios mayores pero no modificaban las imágenes originales de forma significativa. De hecho, estas características (parámetros) se han extraído con los elementos de las estructuras de diversos tamaños, pero resultaron ser redundantes.

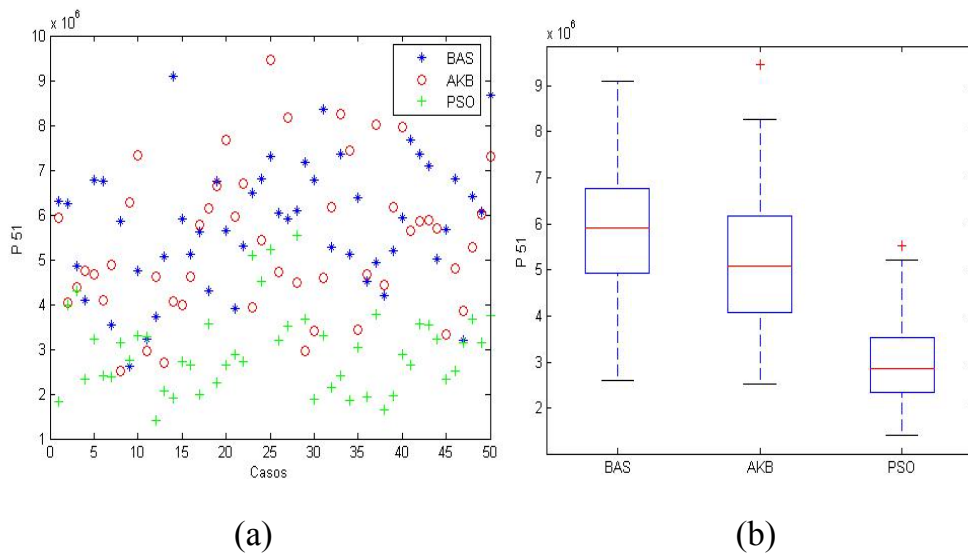


Figura 5.18 (a) *Distribución de muestras al utilizar la dilatación (P51)* (b) *Medida de la media y desviación estándar de los valores obtenidos.*

Volvemos a encontrar una situación similar a las anteriores, la única diferencia apreciable es en relación con las imágenes de psoriasis, pero no entre los distintos tipos de tumores.

(d) Otras características morfológicas

También se han utilizado algunas características morfológicas muy simples para extraer algunos parámetros como la *media de la imagen*, *desviación típica*, etc.

P55: Se calcula el valor medio de los niveles de gris que aparecen en la imagen.

P56: Se calcula la desviación típica de la imagen, es decir, la desviación típica de los niveles de gris de la imagen.

P57: Se calcula el máximo nivel de gris medio para cada componente y se toma el máximo de ellos.

(e) Componentes conexas conectadas

Las técnicas de agrupamiento de píxeles como su nombre indican, en esencia lo que hacen es agrupar píxeles pertenecientes a una misma región o segmento y se les asigna la misma etiqueta.

Una técnica de agrupamiento es el etiquetado de componentes conectadas, que consiste en intentar conectar todos los píxeles etiquetados con valor 1 a otros píxeles para formar un camino de píxeles todos con valor binario 1 que serán identificados con una única etiqueta (una región).

La conectividad entre píxeles es un concepto importante para establecer los límites de un objeto. Se dice que dos píxeles p y q están conectados, si son vecinos y tienen el mismo valor de píxel (0 ó 1). Denominaremos componentes conexas conectadas al conjunto de píxeles en los cuales cada píxel está conectada a los otros píxeles.

Existen diferentes variantes de conectividad: conectividad a 4, conectividad a 8,..., conectividad a m dependiendo del número de píxeles vecinos considerados. Una de las operaciones más usadas en visión es encontrar las componentes conectadas dentro de una imagen, es decir el etiquetado, y contar el número de objetos. Veamos un ejemplo con conectividad a 4.

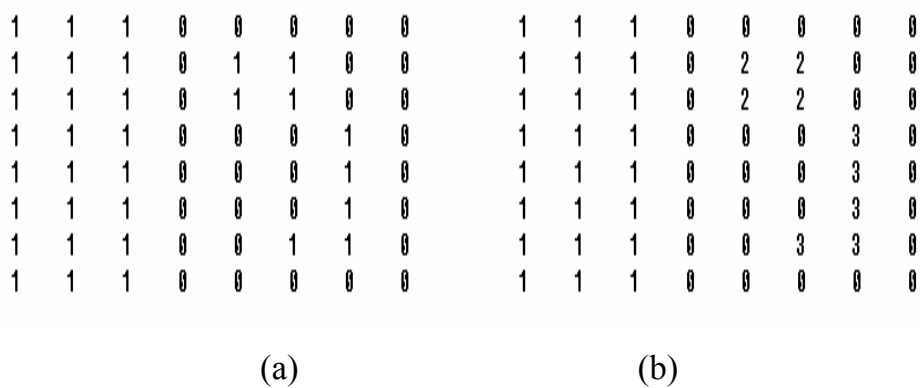


Figura 5.18 a) Imagen binaria. b) Identificación de las diferentes zonas. Cada una de las zonas es etiqueta con un entero.

Los parámetros obtenidos a partir de esta imagen son (tras umbralizar la imagen con *el método de Otsu [OTS79]*).

P58: Número de componentes blancas.

P59: Área de las zonas blancas.

P60: El máximo área de las zonas detectadas.

P61: Número de Euler.

Estas características se describen brevemente a continuación.

(e.1) Número de componentes blancas (parámetro 58)

Este parámetro contabiliza el número de componentes conexas conectadas en la imagen binaria. Mediante el etiquetado con enteros de cada una de las componentes distinguimos las zonas en la imagen binaria.

Una componente conexa es una región sin ningún agujero y una componente conectada es el conjunto de píxeles vecinos distintos de cero.

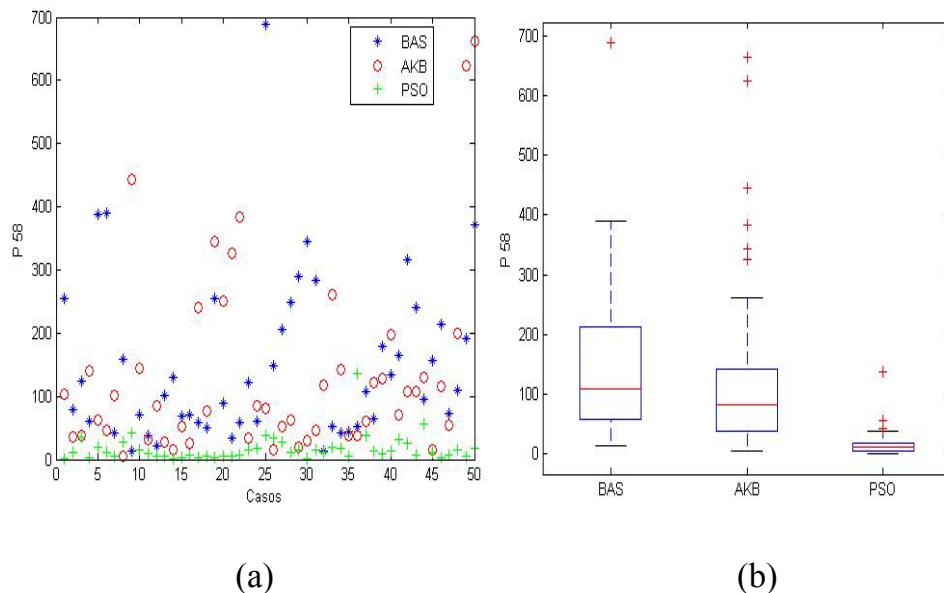


Figura 5.19 *Número de componentes blancas (P 58) (a) Distribución de muestras (b) Medida de la media y desviación estándar de los valores obtenidos.*

Encontramos que tal y como se esperaba, no se trata de una característica o parámetro que vaya a decidir de un modo excesivamente claro entre lesiones cancerosas. Ahora bien, vemos cómo nuevamente se vuelve a confirmar la diferencia tan evidente entre las características de los tumores con respecto a las lesiones de psoriasis.

(e.2) *Área de las zonas blancas (parámetro 59)*

Tal y como ya ha sido documentado, en nuestras imágenes desconocemos la escala con la que éstas han sido obtenidas. Esto es un problema importante, ya que en la documentación acerca de características de los tumores, se hace mucha referencia a aspectos relacionados con los tamaños de éstos.

A pesar de ello decidimos probar qué tal funcionaría un parámetro como éste, el cual aparentemente no tiene por qué presentar una distribución de valores significativa para la tarea de diagnóstico planteada.

Intentamos por tanto obtener el área dada en número de píxeles, y ver qué tal se comporta como característica o parámetro de las muestras.

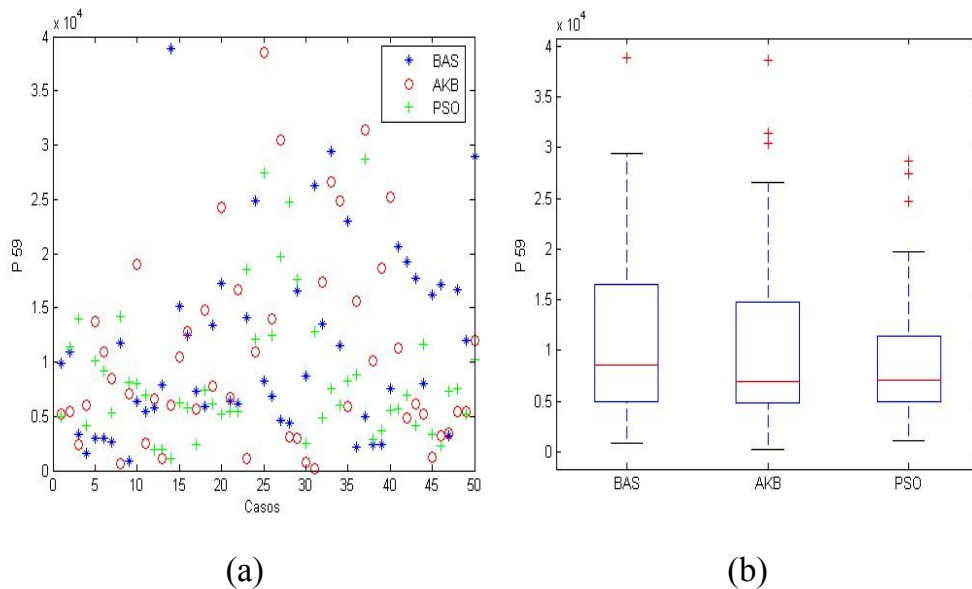


Figura 5.20 *Área de las zonas blancas (P 59) (a) Distribución de muestras (b) Medida de la media y desviación estándar de los valores obtenidos.*

La distribución de puntos que se sigue es similar para el pre-cáncer y el cáncer. No nos distingue nada entre ellos, la media no es una medida representativa para estas clases. La distribución de la psoriasis tampoco se demarca de las otras clases.

(e.3) Máximo área de las zonas detectadas (parámetro 60)

Se calcula el máximo área de las zonas detectadas en la imagen dada en número de píxeles. Es decir el área de la mayor componente conexa conectada.

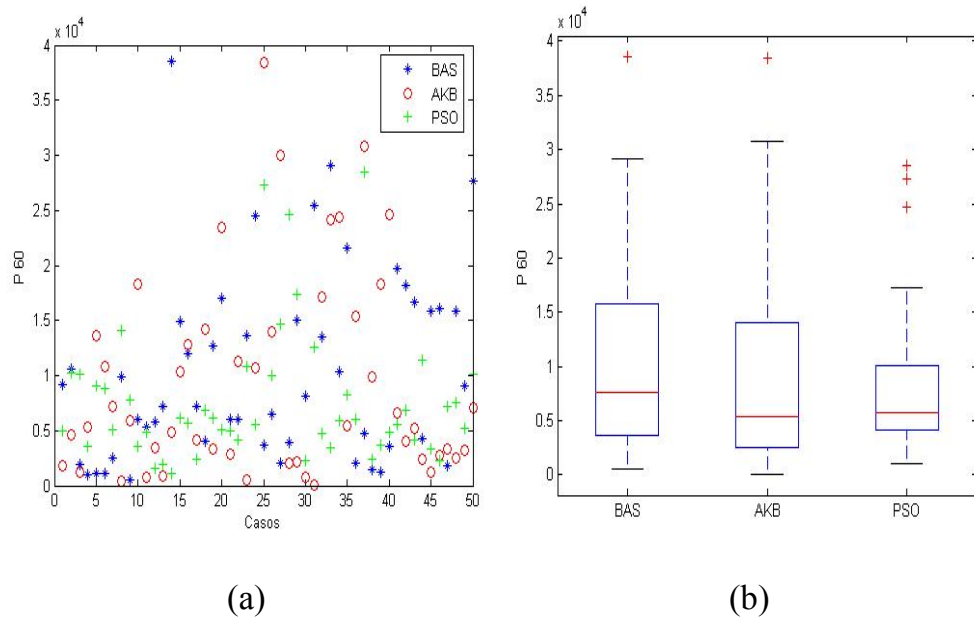


Figura 5.21 *El máxima área (P 60) (a) Distribución de muestras (b) Medida de la media y desviación estándar de los valores obtenidos.*

Como en los casos los anteriores la distribución de puntos es similar para el pre-cáncer y el cáncer, también para la psoriasis (no se puede distinguir entre ellos).

(e.4) Número de Euler (parámetro 61)

En imágenes binarias el número de Euler es definido como el número de objetos menos el número de agujeros en el interior de esos objetos:

$$\text{Numero_Euler} = \#\text{Objetos} - \#\text{Hoyos} \quad (5.12)$$

En nuestro caso los objetos vendrán representados por aquellos contornos bien definidos que se puedan apreciar en la imagen (posible tumor), siendo los hoyos los huecos apreciados dentro que representen un brusco cambio en el nivel de gris.

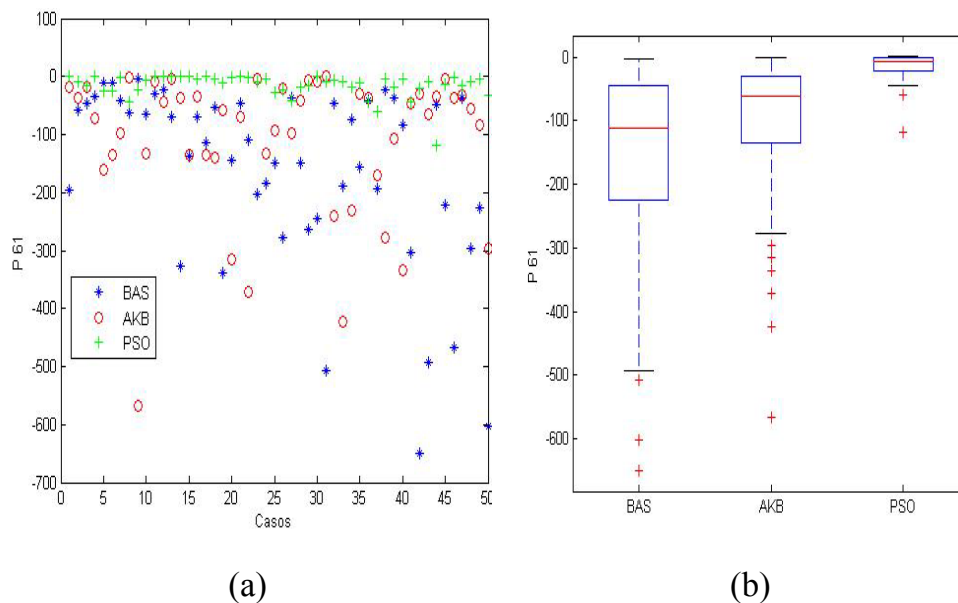


Figura 5.22 Número de Euler (P 61) (a) Distribución de muestras (b) Medida de la media y desviación estándar de los valores obtenidos.

En la distribución de muestras que los valores sean negativos indica tan sólo que es normal encontrar en los tumores un gran número de hoyos dentro del contorno del tumor.

En este caso las imágenes de la clase psoriasis, se destacan con respecto a las otras clases. La clase psoriasis sigue una distribución no muy negativa salvo en algunos casos.

Por un lado destacar la diferencia, como viene siendo habitual, entre psoriasis y los dos tipos de tumores.

5.4.3 Parámetros obtenidos a partir de morfología matemática

En el *MATLAB* utilizamos una función llamada *regionprops* que nos permite conseguir diversos valores suma de los componentes etiquetados.

En la siguiente figura se ilustran los ejes y la orientación de la elipse. La parte izquierda de la figura muestra una región de la imagen y su elipse correspondiente. La parte derecha demuestra la misma elipse, con las características indicadas gráficamente; las líneas azules sólidas son los ejes, los puntos rojos son los focos, y la orientación es el ángulo entre la línea punteada horizontal y el eje principal.



Figura 5.23 *Elipse.*

Con esta función hemos definido los siguientes parámetros:

P62: La longitud (en píxeles) del eje mayor de la elipse.

P63: La longitud (en píxeles) del eje menor de la elipse.

P64: La excentricidad de la elipse, donde la excentricidad es el cociente entre la distancia focal y la longitud de eje mayor.

P65: La orientación el ángulo (grados) entre el x- ordenada y el eje principal de la elipse.

Por otro lado se utiliza también la técnica de “*bounding box*” (caja de limitación) donde dicha caja es el rectángulo más pequeño que incluye completamente los objetos que hay en la imagen binaria. Con esta técnica se pueden extraer algunas características nuevas. En la siguiente figura se ilustra la caja de limitación que sería la región que incluye los píxeles blancos (recuadro verde). El punto rojo es el centro de gravedad de la figura.

Se definen cajas de limitación que incluyan componentes conectadas convexas.

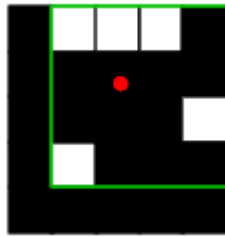


Figura 5.24 Caja de limitación “bounding box”.

Se pueden obtener los siguientes parámetros:

P66: El número de píxeles del objeto en la caja de limitación.(sin incluir huecos).

P67: El número de píxeles en la caja de limitación incluidos los píxeles de los huecos. Se considerará como tamaño de la imagen el tamaño de la caja de limitación de la región. Como se puede ver en la Figura 5.25.

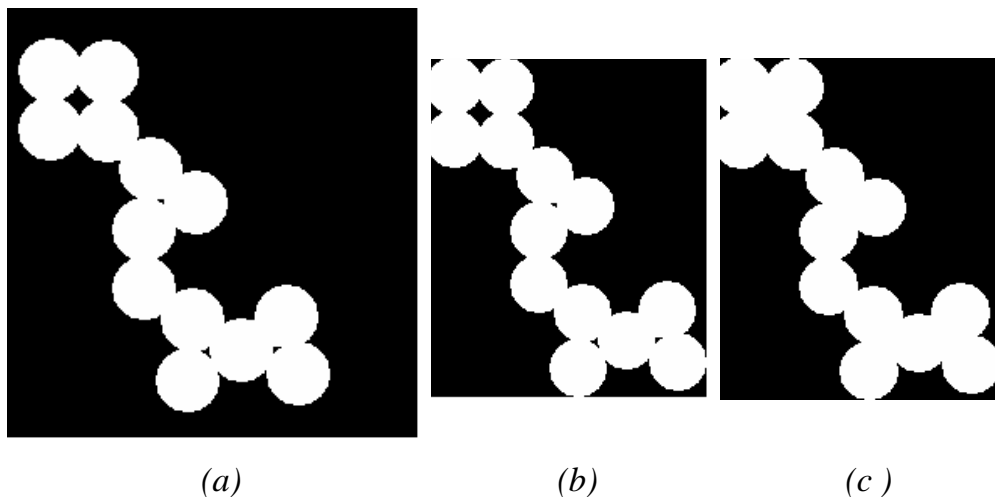


Figura 5.25 (a) Imagen original (b) parámetro 66 (c) parametro67

P68: El diámetro de un círculo con el mismo área que la región. Se calcula con la siguiente fórmula:

$$D = \sqrt{\frac{4 \cdot AREA}{\pi}} \quad (5.13)$$

Notamos casco convexo como el mínimo polígono convexo que contiene una determinada zona, su área la notamos como *ConvexArea*. Aunque los siguientes parámetros se pueden definir en zonas concretas nosotros los hemos extraído de la imagen completa (por lo tanto, en este caso la región es la imagen completa y su área es $Área = 256 \cdot 256 = 65536$ píxeles).

Definimos objeto como aquellos elementos formados por píxeles conectados.

P69: 'Solidez' la proporción de los píxeles en el casco convexo que está también en la región. Computado como $Area/ConvexArea$.

P70: 'Extensión' la proporción de los píxeles en la caja de limitación que están también en la región. Computado como el área de la región dividido por el área de la caja de limitación.

P71: Diámetro del máximo objeto redondo cuando se quitan previamente todos los objetos que contienen menos de 30 píxeles.

Descripción: cuando eliminamos los objetos pequeños (menores de 30 píxeles de área) y recalculamos de nuevo objetos redondos obtenemos objetos finales menores. Tras este procesado el parámetro P71 es el diámetro del objeto redondo máximo que se obtiene).

P72: Diámetro del máximo objeto redondo cuando se quitan previamente todos los objetos que contienen menos de 500 píxeles.

Descripción: preprocesado igual que el parámetro P71.

P73: Para este parámetro calculamos la media en cada columna de la imagen. Después tomamos el máximo entre estos valores de la

media y finalmente reprocesamos la imagen eliminando todos los píxeles de ella que sean menores que este valor. Finalmente el parámetro P73 es el umbral de esta nueva imagen utilizando el método de Otsu [OTS79].

P74: Este parámetro calcula la diferencia entre los valores de umbral (utilizando el método de Otsu [OTS79]) antes y después de quitar píxeles de acuerdo con el preprocesamiento descrito en P73.

P75: Se eliminan los objetos con menos de 500 píxeles de área y recalculamos los nuevos objetos. Este parámetro P75 es la suma de las áreas de los nuevos objetos (que quedan en la imagen).

5.5 Resumen de parámetros

Una vez procesada y segmentada la imagen se realiza la extracción de características. A continuación se muestra la tabla global de características consideradas:

Característica	Técnica usada	Descripción de la característica
1	Detector de bordes de Sobel	<i>Suma de todos los valores de los bordes de una imagen.</i>
2		<i>Suma de todos los valores de los bordes de una imagen. Previamente se filtran los bordes usando umbral automático para que queden sólo el 30% de los píxeles [OTS79].</i>
3		<i>Con un umbral de 1.</i>
4		<i>Con un umbral de 2.</i>
5		<i>Con un umbral de 3.</i>
6		<i>Con un umbral de 4.</i>
7		<i>Con un umbral de 5.</i>
8		<i>Con un umbral de 6.</i>
9		<i>Detectamos solamente los bordes verticales, y sumamos todos los valores de la imagen.</i>

10		<i>Con un umbral de 1.</i>
11		<i>Con un umbral de 2.</i>
12		<i>Con un umbral de 3.</i>
13		<i>Con un umbral de 4.</i>
14		<i>Con un umbral de 5.</i>
15		<i>Con un umbral de 6.</i>
16		<i>Detectamos solamente los bordes horizontales, y sumamos todos los valores de la imagen.</i>
17		<i>Con un umbral de 1.</i>
18		<i>Con un umbral de 2.</i>
19		<i>Con un umbral de 3.</i>
20		<i>Con un umbral de 4.</i>
21		<i>Con un umbral de 5.</i>
22		<i>Con un umbral de 6.</i>
23	Detector de bordes de Prewitt	<i>Suma de todos los valores de los bordes de una imagen.</i>
24		<i>Suma de todos los valores de los bordes de una imagen. Previamente se filtran los bordes usando un umbral automático [OTS79] para mantener sólo el 30% de los píxeles activos.</i>
25		<i>Con un umbral de 1.</i>
26		<i>Con un umbral de 2.</i>
27		<i>Con un umbral de 3.</i>
28		<i>Con un umbral de 4.</i>
29		<i>Con un umbral de 5.</i>
30		<i>Con un umbral de 6.</i>
31		<i>Con un umbral de 7.</i>
32		<i>Con un umbral de 8.</i>
33		<i>Detectamos solamente los bordes verticales, y sumamos todos los valores de la imagen.</i>
34		<i>Con un umbral de 1.</i>
35		<i>Con un umbral de 2.</i>
36		<i>Con un umbral de 3.</i>
37		<i>Con un umbral de 4.</i>
38		<i>Con un umbral de 5.</i>
39		<i>Detectamos solamente los bordes horizontales, y sumamos todos los valores de la imagen.</i>
40		<i>Con un umbral de 1.</i>
41		<i>Con un umbral de 2.</i>

42		<i>Con un umbral de 3.</i>
43		<i>Con un umbral de 4.</i>
44		<i>Con un umbral de 5.</i>
45	Detector de bordes de Canny	<i>Suma de todos los valores de los bordes de una imagen. Previamente se filtran los bordes usando diversos valores de umbrales bajos y altos automáticamente.</i>
46	Detector de bordes Laplaciana de Gaussianas	<i>Suma de todos los valores de los bordes de una imagen. Los bordes se filtran previamente usando un umbral de 0.5.</i>
47		<i>Suma de todos los valores de los bordes de una imagen. Los bordes se filtran previamente usando un umbral de -0.5.</i>
48	Discreta de Fourier transformada en dos dimensión	<i>Valor máximo de frecuencia espacial.</i>
49		<i>Componente medio de la frecuencia espacial.</i>
50		<i>Desviación de estándar de la frecuencia espacial.</i>
51	Morfología	<i>Dilatación de la Imagen usando un elemento de la estructura (disco de la radio 1).</i>
52		<i>Dilatación de la Imagen usando un elemento de la estructura (disco de la radio 2).</i>
53		<i>Erosión de la imagen usando un elemento de la estructura (disco de la radio 1).</i>
54		<i>Erosión de la imagen usando un octágono como elemento de la estructura.</i>
55		<i>Se calcula el valor medio de los valores de gris que aparecen en la imagen.</i>
56		<i>Se calcula la desviación típica de la imagen, es decir, la desviación típica de los valores de gris de la imagen.</i>
57		<i>Se calcula el máximo valor del medio componente de los píxeles</i>

		<i>de la imagen.</i>
58		<i>Número de componentes blancas.</i>
59		<i>Área de las zonas blancas.</i>
60		<i>La máxima área de las zonas detectadas.</i>
61		<i>Número de Euler.</i>
62		<i>Longitud (en píxeles) del eje mayor de la elipse.</i>
63		<i>La longitud (en píxeles) del eje menor importancia de la elipse.</i>
64		<i>La excentricidad de la elipse, donde la excentricidad es el cociente entre la distancia focal y la longitud de eje mayor.</i>
65		<i>La orientación el ángulo (grados) entre el eje x y el eje principal de la elipse.</i>
66		<i>El número de píxeles en la caja de limitación.</i>
67		<i>El número del píxel en el casco convexo.</i>
68		<i>El diámetro de un círculo con el mismo área que la región.</i>
69		<i>'Solidez' la proporción de píxeles en el casco convexo que están también en la región.</i>
70		<i>'Extensión' la proporción de los píxeles en la caja de limitación que están también en la región.</i>
71		<i>Diámetro del máximo objeto redondo cuando se quitan todos los objetos que contienen menos de 30 píxeles.</i>
72		<i>Diámetro del máximo objeto redondo cuando se quitan previamente todos los objetos que contienen menos de 500 píxeles.</i>
73		<i>El umbral utilizando el método de Otsu [OTS79] para que se mantengan sólo el 30% de los píxeles activos.</i>
74		<i>La diferencia entre los valores de umbral (utilizando el método de</i>

		<i>Otsu [OTS79]) antes y después de quitar píxeles.</i>
75		<i>Este parámetro calcula el área de los objetos cuando se quitan los objetos con menos de 500 píxeles.</i>

Tabla 5.1 *Las características y breve descripción.*

5.6 El coeficiente de correlación: Redundancia de parámetros

Cuando dos variables crecen o decrecen de forma simultánea y proporcional debido a factores externos, se dice que los fenómenos están positivamente correlacionados linealmente. Si uno crece en la misma proporción que el otro decrece, los dos fenómenos están negativamente correlacionados linealmente.

El coeficiente de correlación va desde -1 hasta +1. Una correlación positiva perfecta tiene un coeficiente +1, y para una correlación negativa perfecta tendremos un coeficiente de -1.

Las características de los coeficientes de correlación son:

1. Puede tener signo positivo o negativo, dependiendo del signo de covarianza que muestran de 2 variables, por tanto está entre -1 y +1.
2. Es simétrico: $r_{xy} = r_{yx}$
3. Un coeficiente $r = 0$ no necesariamente implica independencia estadística; pero si son independientes el coeficiente r es 0 (la $r = 0$ es necesaria pero no suficiente para la independencia estadística).

4. Es medida sólo de asociación lineal, por lo que pierde sentido en relaciones no lineales, además no implica relación de causalidad.

Al aplicar el coeficiente de correlación se eliminaron aquellos parámetros cuyo valor absoluto del coeficiente de correlación lineal con respecto a otro parámetro fuera ≥ 0.98 .

14	4	24	18	40	33
15	5	31	18	35	34
16	6	38	18	40	34
30	6	44	18	41	34
17	7	25	19	41	35
23	7	32	19	43	37
37	7	44	19	44	38
43	7	45	19	47	46
18	8	21	20	54	51
24	8	26	20	55	54
31	8	39	20	63	58
38	8	45	20	71	70
44	8	46	20		
19	9	26	21		
20	9	27	21		
25	9	46	21		
32	9	47	21		
45	9	42	22		
11	10	37	23		
20	10	43	23		
21	10	31	24		
26	10	38	24		
46	10	44	24		
12	11	45	25		
21	11	27	26		
27	11	45	26		
28	11	46	26		
46	11	28	27		
47	11	46	27		
13	12	47	27		
28	12	47	28		
47	12	37	30		
23	17	38	31		

30	17	44	31	
37	17	39	32	
43	17	39	33	

Tabla 5.2 Características redundantes con un valor absoluto de coeficiente de correlación $|\text{coeficiente de correlación}| \geq 0.98$

Por lo tanto, tras eliminar los parámetros redundantes, nos quedamos con los presentados en la Tabla 5.3.

P	P'	Técnica usada
1	1	Morfología
2	2	Morfología
3	3	Morfología
4	4	Detector de bordes de <i>Prewitt</i>
5	5	Detector de bordes de <i>Prewitt</i>
6	6	Detector de bordes de <i>Prewitt</i>
7	7	Detector de bordes de <i>Prewitt</i>
8	8	Detector de bordes de <i>Prewitt</i>
9	9	Detector de bordes de <i>Prewitt</i>
10	10	Detector de bordes de <i>Prewitt</i>
11	18	Detector de bordes de <i>Prewitt</i>
12	22	Detector de bordes de <i>Prewitt</i>
13	24	Detector de bordes de <i>Prewitt</i>
14	26	Detector de bordes de <i>Sobel</i>
15	29	Detector de bordes de <i>Sobel</i>
16	33	Detector de bordes de <i>Sobel</i>
17	34	Detector de bordes de <i>Sobel</i>
18	35	Detector de bordes de <i>Sobel</i>
19	36	Detector de bordes de <i>Sobel</i>

20	48	Detector de bordes de <i>Laplaciana de Gaussianas</i>
21	49	Detector de bordes de <i>Laplaciana de Gaussianas</i>
22	50	Detector de bordes de <i>Canny</i>
23	51	Discreta de Fourier transformada en dos dimensión
24	52	Discreta de Fourier transformada en dos dimensión
25	53	Discreta de Fourier transformada en dos dimensión
26	56	Morfología
27	57	Morfología
28	58	Morfología
29	59	Morfología
30	60	Morfología
31	61	Morfología
32	62	Morfología
33	64	Morfología
34	65	Morfología
35	66	Morfología
36	67	Morfología
37	68	Morfología
38	69	Morfología
39	70	Morfología

Tabla 5.3 *Parámetros no redundantes. Donde P es la nueva numeración de los parámetros y P' es la antigua numeración de los parámetros.*

En total son 39 características o parámetros que se van a utilizar en el sistema de clasificación en el Capítulo 6.

Capítulo 6

Resultados de clasificación

En este capítulo se presentan los resultados de los cuatro sistemas diagnósticos: KNN con exploración secuencial de características, KNN con AG, RNA con AG y ANFIS.

Además se indican las características más relevantes para las tareas de diagnóstico atacadas en este estudio. Esto es de gran importancia incluso para profesionales que quieran definir protocolos de diagnóstico basados en la técnica de imágenes de fluoroscopia.

El objetivo de esta investigación es el diagnóstico final de una enfermedad. Así, se estudian dos grupos de personas:

- 1- Un grupo de personas que sufren enfermedades de cierta patología.
- 2- Un grupo de personas sanas que no presentan esa patología.

El resultado de la clasificación puede dar los siguientes casos:

- A. Verdaderos Positivos VP: se clasifica al paciente enfermo como enfermo.
- B. Falsos Positivos FP: se clasifica al paciente sin la patología como enfermo.
- C. Verdaderos Negativos VN: se clasifica al sujeto sin la patología como sano.
- D. Falsos Negativos FN: se clasifica al paciente con la patología como sano.

Con estos casos se puede calcular ciertas medidas que ayudan a la evaluar el clasificador.

$$CR = \frac{VP + VN}{VP + VN + FP + FN}; SE = \frac{VP}{VP + FN}; SP = \frac{VN}{VN + FP} \quad (6.1)$$

Donde:

$CR = Tasa\ de\ Clasificación$: es la proporción de personas enfermas y sanas clasificadas correctamente del total de individuos.

$SE = Sensibilidad$: es la proporción de personas *enfermas* detectadas acertadamente entre el total de sujetos enfermos.

$SP = \textit{Especificidad}$: es la proporción de personas *sanas* detectadas acertadamente entre el total de sujetos sanos.

Se ha realizado la extracción de parámetros como se ha descrito en el Capítulo 5 obteniendo un “vector característico” para cada imagen. Seguidamente tomando como base estos vectores característicos hemos realizado una clasificación de las imágenes de forma que nos permitan identificar la imagen como de una lesión cancerosa o no.

Una vez extraídos los parámetros se seleccionan algunos que conforman el vector característico tipificado que determina el formato final de los vectores de entrada del clasificador. El vector característico tipificado indica qué parámetros se utilizarán en el proceso de clasificación. Se aplican técnicas avanzadas de selección de características. En nuestra investigación nos hemos centrado en algoritmos genéticos y técnicas de *búsqueda combinatorial (hacia delante y hacia atrás)* de componentes con vistas a automatizar el proceso de búsqueda y selección de características con el fin de mejorar la eficiencia del clasificador [ROS03].

Los esquemas de clasificación que son utilizados en este trabajo están basados en el algoritmo de los k-vecinos más próximos y redes neuronales. Los resultados de las clasificaciones se obtienen de dos subconjuntos de datos como se ha dicho antes:

- 1- conjunto de entrenamiento: compuesto por el 70% de los datos para cada una de las clases.

- 2- conjunto de test: compuesto por el 30% de los datos para cada una de las clases.

Aunque en el caso del clasificador de las K-vecinas más próximas también se utiliza una evaluación cruzada exhaustiva (dejar-uno-fuera) como se indica en la Sección 6.1.

Aleatoriamente se realizan tres particiones en tres conjuntos de entrenamiento y tres conjuntos de test, que son particiones diferentes de la base de datos para calcular las medidas de los diferentes resultados.

Es importante resaltar que en distintas secciones destacamos algunas características más relevantes, se debe tener en cuenta la Tabla 5.3 que relaciona los parámetros antes de prescindir de los redundantes con los que se han utilizado en este Capítulo 6.

6.1 Los resultados del clasificador K-vecinas más cercanas (KNN)

El método “*dejar-uno-fuera*” (*leave-one-out*) se ha utilizado para obtener los resultados de las clasificaciones del conjunto de entrenamiento (en el clasificador KNN).

Este método consiste en tomar un solo vector como conjunto de test. Se utiliza el clasificador y se evalúa el resultado como VP, VN,

FP o FN. En la siguiente iteración se utiliza otro vector único como grupo de test. Esto se repite tantas veces como vectores haya en la base de datos. De esta forma se utiliza en cada ciclo un solo vector de test y el resto se incluyen en el clasificador. Esto representa una validación cruzada con tantas iteraciones como vectores tenga la base de datos.

En las siguientes tablas se muestra los resultados obtenidos de las clasificaciones en la tarea fácil y la tarea difícil, utilizando el método hacia adelante y hacia atrás:

1- Tarea fácil

Método hacia adelante			
Número de vecinas	Resultados de entrenamiento		
	<i>Acierto</i>	<i>Sensibilidad</i>	<i>Especificidad</i>
1	100±0	100±0	100±0
3	100±0	100±0	100±0
5	100±0	100±0	100±0
7	100±0	100±0	100±0
9	100±0	100±0	100±0

(a)

Método hacia adelante			
Número de vecinas	Resultados de test		
	<i>Acierto</i>	<i>Sensibilidad</i>	<i>Especificidad</i>
1	100±0	100±0	100±0
3	100±0	100±0	100±0
5	100±0	100±0	100±0
7	98.52±1.28	97.69±0	100±0
9	98.52±1.28	97.69±0	100±0

(b)

Método hacia atrás			
Números de vecinas	Resultados de entrenamiento		
	<i>Acierto</i>	<i>Sensibilidad</i>	<i>Especificidad</i>
1	100±0	100±0	100±0
3	100±0	100±0	100±0
5	100±0	100±0	100±0
7	100±0	100±0	100±0
9	100±0	100±0	100±0

(c)

Método hacia atrás			
Número de vecinas	Resultados de test		
	<i>Acierto</i>	<i>Sensibilidad</i>	<i>Especificidad</i>
1	100±0	100±0	100±0
3	99.26±1.28	98.77±0	100±0
5	99.26±1.28	98.77±0	100±0
7	100±0	100±0	100±0
9	100±0	100±0	100±0

(d)

Tabla 6.1. Los resultados de clasificación con KNN en la tarea fácil. (a) Los resultados del entrenamiento con el método de selección de características hacia adelante. (b) Los resultados de test con el método adelante. (c) Los resultados de entrenamiento con el método hacia atrás. (d) Los resultados de test con el método hacia atrás.

En la Tabla 6.1 se muestra el resultado obtenido al trabajar sobre 3 particiones diferentes de la base de datos en la que incluimos las imágenes de tumores como un único grupo (keratosis actínica y células basales) y los imágenes de psoriasis. Podemos comprobar cómo los resultados obtenidos son muy buenos, ya que siempre rondan el 100 % de acierto.

2- Tarea difícil:

Método hacia adelante			
Número de vecinas	Resultados de entrenamiento		
	<i>Acierto</i>	<i>Sensibilidad</i>	<i>Especificidad</i>
1	73.70 ± 4.49	67.96 ± 9.50	78.35 ± 9.50
3	71.85 ± 0.64	71.10 ± 4.32	72.30 ± 4.32
5	74.07 ± 1.70	69.47 ± 3.40	77.83 ± 3.40
7	74.81 ± 3.21	72.90 ± 4.96	76.46 ± 4.96
9	72.22 ± 1.11	66.35 ± 3.06	77.15 ± 3.06

(a)

Método hacia atrás			
Número de vecinas	Resultados de entrenamiento		
	<i>Acierto</i>	<i>Sensibilidad</i>	<i>Especificidad</i>
1	65.83 ± 4.02	66.67 ± 8.66	65.00 ± 8.66
3	66.67 ± 4.73	62.50 ± 6.29	70.83 ± 6.29
5	61.25 ± 5.73	65.00 ± 10.00	57.50 ± 10.00
7	65.83 ± 3.15	65.83 ± 7.64	65.83 ± 7.64
9	62.50 ± 3.31	64.17 ± 7.64	60.83 ± 7.64

(b)

Método hacia atrás			
Número de vecinas	Resultados de test		
	<i>Acierto</i>	<i>Sensibilidad</i>	<i>Especificidad</i>
1	66.67 ± 17.56	66.67 ± 20.82	66.67 ± 20.82
3	65.00 ± 13.23	56.67 ± 11.55	73.33 ± 11.55
5	55.00 ± 18.03	46.67 ± 20.82	63.33 ± 20.82
7	60.00 ± 10.00	53.33 ± 11.55	66.67 ± 11.55
9	61.67 ± 12.58	50.00 ± 25.17	73.33 ± 25.17

(c)

Método hacia adelante			
Número de vecinas	Resultados de test		
	<i>Acierto</i>	<i>Sensibilidad</i>	<i>Especificidad</i>
1	56.67 ± 7.64	56.67 ± 15.28	56.67 ± 15.28
3	56.67 ± 7.64	56.67 ± 11.55	56.67 ± 11.55
5	65.00 ± 22.91	63.33 ± 40.41	66.67 ± 40.41
7	61.67 ± 12.58	66.67 ± 25.17	56.67 ± 25.17
9	56.67 ± 11.55	60.00 ± 20.82	53.33 ± 20.82

(d)

Tabla 6.2 Los resultados de clasificación en la tarea difícil. (a) Los resultados de entrenamiento con el método hacia adelante. (b) Los

resultados de test con el método adelante. (c) Los resultados de entrenamiento con el método hacia atrás. (d) Los resultados de test con el método hacia atrás.

En la tarea difícil se obtienen bajas tasas de clasificación. Esto indica pues la dificultad para distinguir entre keratosis actínica y carcinoma basal de un modo automático y con las características que hemos extraído. Por lo tanto con este esquema no podemos discriminar de forma fiable entre cáncer y pre-cáncer.

En la Tabla 6.3 se muestran los resultados obtenidos cuando utilizamos el método *dejar-uno-fuera* con todas las muestras de la base de datos (conjunto de entrenamiento y el conjunto de test).

Método hacia adelante (tarea fácil)			
Número de vecinas	<i>Acierto</i>	<i>Sensibilidad</i>	<i>Especificidad</i>
1	100±0	100±0	100±0
3	100±0	100±0	100±0
5	100±0	100±0	100±0
7	100±0	100±0	100±0
9	100±0	100±0	100±0

(a)

Método hacia atrás (tarea fácil)			
Número de vecinas	<i>Acierto</i>	<i>Sensibilidad</i>	<i>Especificidad</i>
1	100±0	100±0	100±0
3	100±0	100±0	100±0
5	100±0	100±0	100±0
7	100±0	100±0	100±0
9	100±0	100±0	100±0

(b)

Método hacia adelante (tarea difícil)			
Número de vecinas	<i>Acierto</i>	<i>Sensibilidad</i>	<i>Especificidad</i>
1	76.0±0	74.0±0	78.0±0
3	67.0±0	60.0±0	74.0±0
5	70.0±0	80.0±0	60.0±0
7	66.0±0	60.0±0	72.0±0
9	73.0±0	70.0±0	76.0±0

(c)

Método hacia atrás (tarea difícil)			
Número de vecinas	<i>Acierto</i>	<i>Sensibilidad</i>	<i>Especificidad</i>
1	68.0±0	72.0±0	64.0±0
3	61.0±0	56.0±0	66.0±0
5	58.0±0	66.0±0	50.0±0
7	63.0±0	68.0±0	58.0±0
9	60.0±0	58.0±0	62.0±0

(d)

Tabla 6.3 Los resultados de clasificación con la técnica de evaluación cruzada “dejar-uno-fuera”. (a) Los resultados con el método hacia adelante en la tarea fácil. (b) Los resultados con el método hacia atrás en la tarea fácil. (c) Los resultados con el método hacia adelante en la tarea difícil. (d) Los resultados con el método hacia atrás en la tarea difícil.

Vemos que en las mediciones considerando tan sólo dos clases los porcentajes de acierto son bajos en la tarea difícil. Recordemos que la mayoría de las características discriminaban fácilmente entre psoriasis y lesiones cancerosas, en cambio muchos de ellos encontraban dificultades para distinguir entre las imágenes de keratosis actínica y carcinoma basal. En la Tabla 6.4 se muestran las características que conducen a una mejor clasificación de las tareas fácil y difícil.

El tiempo de ejecución de este clasificador en un equipo con un procesador Pentium 4 a 2.80 GHz ha sido entre 90 a 100 minutos (para la extracción de todos los resultados, excluyendo la fase de procesamiento de imágenes).

Estas características que aparecen activas más frecuentemente en los clasificadores optimizados, son las que se muestran en la Figura 6.1. Se muestran las que más se repiten juntando los resultados extraídos con distinto número de vecinas (1, 3, 5, 7 y 9) y con los métodos de búsqueda hacia delante y hacia atrás. Las más repetidas son P10 y P26.

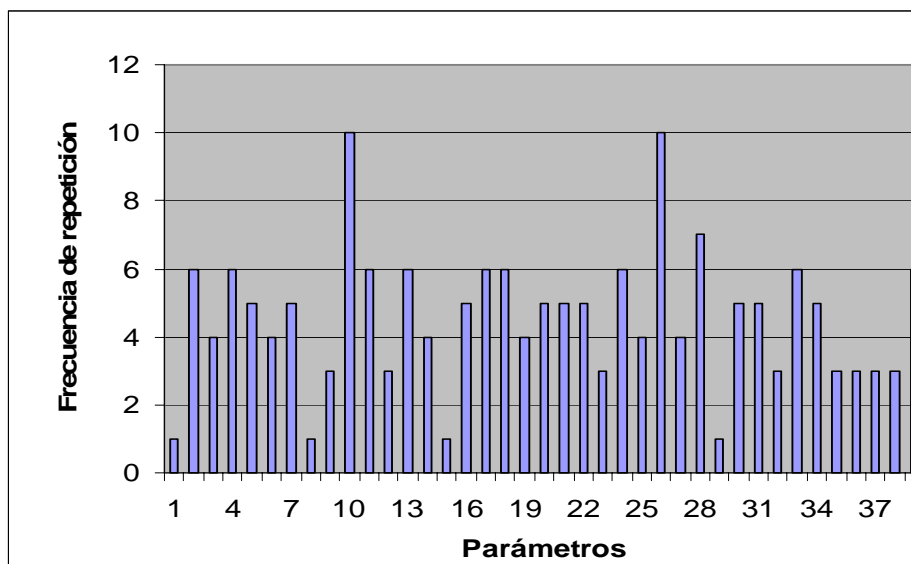


Figura 6.1 La frecuencia de las características seleccionadas utilizando exploración secuencial (hacia delante y hacia atrás) con distinto número de vecinas (1, 3, 5, 7, 9) para la tarea difícil.

Características		Tarea fácil	Tarea difícil
Detector de bordes	Prewitt	P1-P13	P4, P7, P10, P11, P13.
	Sobel	P14-P18	P17, P18.
	Laplacian	P19-P20	Nada
	Canny	P21	Nada
Transformada de Fourier		P22, P23, P24	P24.
Morfología		P26, P29, P30, P32, P33, P34, P36, P37, P39	P26, P28, P30, P33, P34, P36, P39.

Tabla 6.4 Características más discriminantes para las tareas fácil y difícil.

Tal vez debido a la dificultad que presenta el discriminar entre keratosis actínica y carcinoma basal, por el hecho de la poca diferencia entre las imágenes, encontramos poco eficientes los parámetros en este tipo de clasificador. Es decir, quizás el problema se encuentra en las características que se están seleccionados para formar nuestra base de datos y en que hemos utilizado las técnicas de *búsqueda combinatoria secuencial* (método *hacia adelante y hacia atrás*).

Esto plantea la necesidad de optimizar el proceso de selección de las características de que disponemos, intentando que el clasificador funcione del modo más certero posible. Por eso se ha utilizado la implementación de un algoritmo genético que nos ayude a optimizar estos resultados. Con ello se pretende conocer cuáles son las características más representativas de nuestra base de datos, para

comprobar si de este modo se consigue un clasificador con un porcentaje de acierto mayor. La ventaja del algoritmo genético es que permite buscar en el espacio de soluciones de forma más amplia. Es un mecanismo de búsqueda multi-camino. Además se puede repetir el experimento varias veces obteniendo soluciones optimizadas distintas, lo cual facilita el proceso de selección final de características más repetidas.

6.2 Optimización mediante algoritmos genéticos

El clasificador KNN se ha embebido dentro del algoritmo genético, el algoritmo va a buscar los mejores resultados posibles utilizando el mejor conjunto de características. Todos los resultados combinan las distintas características minimizando el error de clasificación final. El algoritmo genético se utiliza en nuestro estudio para desarrollar la selección de cromosomas, cada cromosoma tiene 39 genes, y cada gen consta de un bit, donde el “1” indica que la característica participa en la clasificación y un “0” indica que se omite. La función objetivo o *fitness* del cromosoma se usa para medir su capacidad de clasificar las muestras. Para este propósito el algoritmo genético busca un cromosoma con el mínimo números de 1`s. La selección del conjunto de características es lo que realiza el algoritmo genético con la siguiente configuración:

1. Inicialización: 100 generaciones de 50 individuos y 10 iteraciones.
2. Calcula el *fitness*, que es el cociente de error que se puede calcular usando el clasificador KNN que está basado en la

distancia euclidiana. Y la solución se considera cerca del óptimo del clasificador (*near optima classifier*) (NOC) y se almacena temporalmente para usarlo luego si produce un gran acierto.

3. Desarrolla nuevas cromosomas utilizando los operadores de mutación (probabilidad de 5%) y la cruce (probabilidad de 60%) si son necesarias más generaciones.
4. Termina el proceso después de 100 generaciones de poblaciones de 50 individuos. Por lo tanto 5000 cromosomas se utilizan en cada iteración completa.
5. Se repiten 10 iteraciones (experimentos completos). Las características (genes) seleccionadas con mayor frecuencia en las mejores 50 NOCs de cada iteración se utilizan al final para configurar el vector característico final.

La función de *fitness* se define como $1/\text{error}$ (sobre el conjunto de entrenamiento).

La Tabla 6.5 muestra los resultados obtenidos al aplicar el algoritmo genético en el clasificador KNN para seleccionar las características más representativas; que aparecen en la Tabla 6.6 (con las que hemos conseguido un porcentaje de acierto mayor).

Tarea fácil			
Número de vecinas	<i>Acierto</i>	<i>Sensibilidad</i>	<i>Especificidad</i>
1	100 ± 0	100 ± 0	100 ± 0
3	100 ± 0	100 ± 0	100 ± 0
5	100 ± 0	100 ± 0	100 ± 0
7	99.0 ± 1.0	100 ± 0	97.0 ± 3.0
9	99.3 ± 1.2	100 ± 0	98.0 ± 3.5

(a)

Tarea difícil			
Número de vecinas	<i>Acierto</i>	<i>Sensibilidad</i>	<i>Especificidad</i>
1	94.0±1.0	96.7±1.2	91.3±1.2
3	91.0±5.2	92.7±2.3	89.3±8.1
5	90.3±4.7	92.7±1.2	88.0±9.2
7	90.7±3.5	90.7±1.2	90.7±6.1
9	89.7±4.0	91.3±3.1	88.0±5.3

(b)

Tabla 6.5 Los resultados de la tarea fácil (a) y tarea difícil (b) al aplicar un algoritmo genético en el clasificador KNN.

Características		Tarea fácil	Tarea difícil
Detector de bordes	Prewitt	P4(10), P5(7), P11(6)	P4(10), P5(6)
	Sobel	P12(6), P15(8), P18(8)	P15(6), P17(7), P18(9)
	Laplacian	P20-P21(6)	P21(6)
	Canny	P22(6)	Nada
transformada de Fourier		P23(6)–P24(7)	P24(6)
Morfología		P26(6), P29-P34(8), P36(9), P38(8), P39(9)	P29(6), P32(6), P35(6), P36(8), P38(6)

Tabla 6.6 Las características seleccionadas en el clasificador KNN con optimización basada en algoritmo genético.

La Tabla 6.6 indica las características que con mayor frecuencia selecciona el algoritmo genético en el proceso de optimización. Entre paréntesis se indica el número de veces que aparece en la mejor solución tras repetir el experimento completo 10 veces (iteraciones).

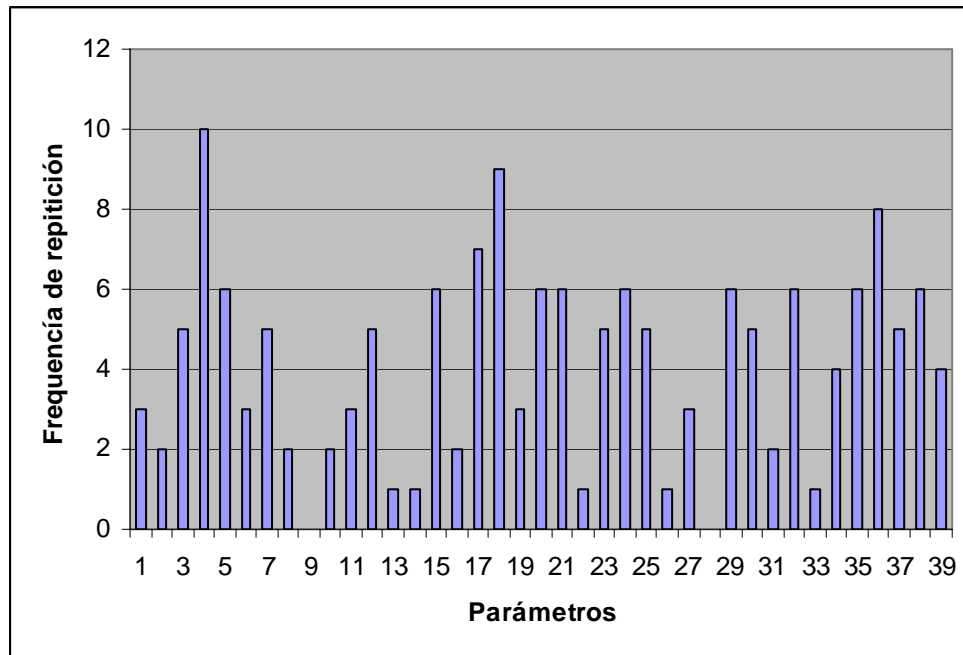


Figura 6.2 La frecuencia de los parámetros durante 10 iteraciones utilizando el algoritmo genético AG.

En la Tabla 6.5 se pueden ver los resultados obtenidos del clasificador KNN con optimización del algoritmo genético. Si lo comparamos con los resultados de las tablas anteriores, vemos que ahora los resultados son mucho mejores, llegando a alcanzarse prácticamente un 92% de acierto en media en la tarea difícil, con lo que ahora sí que se dispone de un clasificador que podemos considerar suficientemente eficiente como para poder ser usado en un diagnóstico eficaz de estos tres tipos de patologías.

En la Tabla 6.6 se indican las características seleccionadas para optimizar el clasificador KNN. En esta tabla se incluyen todos los parámetros que se seleccionan en este clasificador más de 5 veces (es decir, en más de 5 soluciones de las mejores de cada iteración). También se puede notar que algunos de ellos se han seleccionado 10 veces (parámetro 4(P4)) en las dos tareas, y hay otros que han sido

seleccionados en una tarea y no en la otra. (Hay algunos parámetros han salido 9 veces en una tarea y no en la otra tarea). Por ejemplo el parámetro 39 (P39) ha salido 9 veces en la tarea fácil mientras que ha salido menos de 5 veces en la tarea difícil por eso no ha ido seleccionado en el proceso de la clasificación en la tarea difícil. También el parámetro 17 (P17) ha sido seleccionado 7 veces en la tarea difícil mientras ha salido menos de 5 veces en la tarea fácil. Eso significa que algunos parámetros que son buenos para clasificar en la tarea fácil no son tan buenos para clasificar en la tarea difícil y viceversa.

El tiempo de ejecución de este clasificador en un equipo con un procesador Pentium 4 a 2.80 GHz ha sido 180 a 200 minutos (para la extracción de todos los resultados excluyendo el procesamiento de imágenes).

6.3 Clasificación con redes neuronales artificiales (RNA)

En esta clasificación se utiliza también el algoritmo genético para optimización de variables o características de entrada a la red neuronal.

En la clasificación con redes neuronales hemos utilizado distintos números de neuronas, donde el número de neuronas en la capa oculta, será igual al número de las entradas multiplicado por dos. Cada iteración el número de las entradas se cambia. En las siguientes tablas se puede ver el acierto de clasificación para tres versiones. Cada versión se refiere a una separación aleatoria en dos conjuntos

(entrenamiento y test). En la Tabla 6.8 las mejores características que se han seleccionado con este sistema y su número de repeticiones durante las 10 iteraciones del algoritmo genético.

	<i>Resultados (Aciertos)(%)</i>	
	Entrenamiento	Test
<i>Versión 1</i>	64.38	77.27
<i>Versión 2</i>	61.25	81.82
<i>Versión 3</i>	65.63	77.27
<i>Medida del acierto</i>	63.75 ± 2.25	78.79 ± 2.62

Tabla 6.7 *Los resultados de clarificación con RNAs.*

Se puede ver en la Tabla 6.7 los resultados de clasificación no son tan buenos como los resultados del clasificador del KNN con la optimización del algoritmo genético. Las características que se han seleccionado en este proceso se muestran en la siguiente Tabla 6.8.

Características	Parámetros
-----------------	------------

Detector de bordes	Prewitt	P2(7), P4(6), P5(6), P6(6), P7(7), P8(6), P10(6), P11(6)
	Sobel	P12(8), P14(5), P16(5)
	Laplacian	P20 (6), P21(5)
	Canny	P22(8)
transformada de Fourier		P24(6)
Morfología		P27(5), P28(5), P29(5), P30(8), P31(5), P32(6), P37(5), P39(6)

Tabla 6.8 Las características seleccionadas en el clasificador RNA con optimización del algoritmo genético.

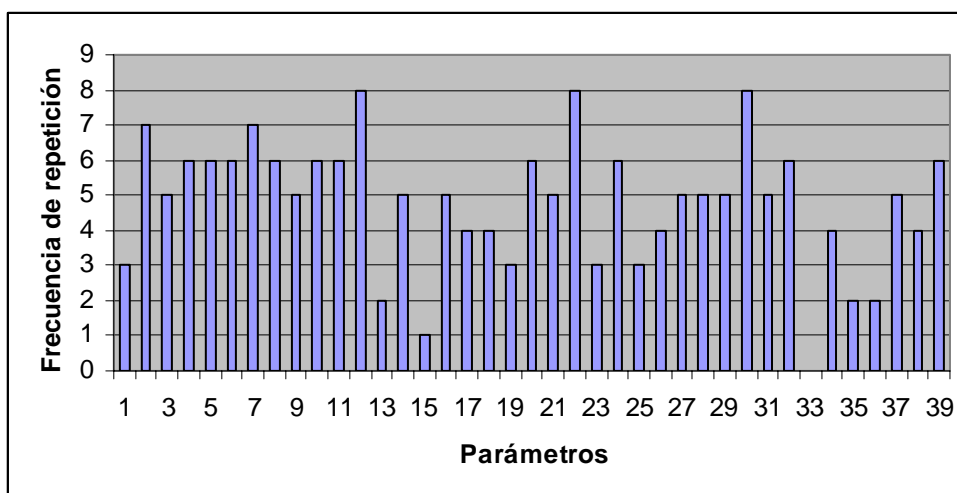


Figura 6.3 La frecuencia de los parámetros durante 10 iteraciones utilizando el algoritmo genético.

El tiempo de ejecución de este clasificador en un equipo con un procesador Pentium 4 a 2.80 GHz ha sido entre 6 horas a 7 horas (para la extracción de todos los resultados excluyendo la etapa de procesamiento de imágenes).

6.4 Clasificación con ANFIS

En este modelo de clasificador hemos utilizado un algoritmo para optimizar las mejores variables, que ordena los parámetros según su importancia. Hemos elegido los 6 primeros parámetros para optimizar la clasificación. Hemos hecho esto para evitar la “*maldición de la dimensionalidad*” [BEL61]. El número de reglas difusas relacionadas con el proceso se incrementa de forma exponencial con el número de variables relacionadas al sistema; por eso hemos elegido un número pequeño de variables que se consideran como una entrada a nuestro sistema de clasificación.

Como anteriormente se ha comentado, la idea principal de construcción en los sistemas difusos ANFIS para evitar la maldición de la dimensionalidad, es el la de colocar las variables de entrada del sistema en una colección de sistemas difusos de baja dimensión. En la siguiente tabla se muestran las variables en un orden de importancia como resultado del algoritmo (G-FLIP):

Parámetros	Características
P32	<i>Longitud (en píxeles) del eje mayor de la elipse</i>
P34	<i>Orientación el ángulo (grados) entre el x-ordenada y el eje principal de la elipse</i>
P35	<i>Número de píxeles en la caja de limitación</i>
P36	<i>Número de píxeles en el casco convexo</i>
P37	<i>Diámetro de un círculo con la misma área que la región</i>
P39	<i>Proporción de los píxeles en la caja de limitación que están también en la región</i>

Tabla 6.9 *Las características seleccionadas en el clasificador ANFIS con optimización del algoritmo G-flip.*

Las variables o parámetros que se van a utilizar en el clasificador son los 6 primeros parámetros que tienen el máximo acierto de clasificación utilizando el método “dejar-uno-fuera”.

En la siguiente tabla se muestran los resultados obtenidos por el clasificador ANFIS, donde hemos introducido los datos tres veces de forma aleatoria en dos conjuntos (un conjunto de aprendizaje con el 80% de los datos y un conjunto de test con el 20%) que son particiones diferentes de la base de datos para calcular las medidas de los diferentes resultados en cada una de las tres versiones.

	<i>Tarea fácil</i>	<i>Tarea difícil</i>
<i>Versión 1</i>	99.77	94.18
<i>Versión 2</i>	99.39	90.83
<i>Versión 3</i>	99.49	92.03
<i>Medida del acierto</i>	99.55 ± 0.2	92.35 ± 1.7

Tabla 6.10 *Los resultados del clasificador ANFIS.*

La información de la configuración del sistema ANFIS se muestra en la siguiente tabla:

El número de nodos	1503
El número de parámetros lineales	5103
El número de parámetros no lineales	54
El número total de parámetros	5157
El número de los datos (vectores etiquetados) de conjunto de aprendizaje	80
El número de los datos (vectores etiquetados) de conjunto de test	20
El número de reglas difusas	729

Tabla 6.11 *Configuración del sistema ANFIS.*

El tiempo de ejecución de este clasificador en un equipo con un procesador Pentium 4 a 2.80 GHz ha sido de 180 a 200 minutos.

6.5 *Resumen de resultados*

En esta sección se hace un resumen comparativo de los resultados obtenidos con distintos métodos de la clasificación por medio de un estudio estadístico de los vectores o parámetros que producen los mejores resultados.

En primer lugar como se puede observar de los resultados anteriores algunos de ellos se pueden considerar muy buenos resultados (KNN-GA, ANFIS). Por otro lado los resultados de los otros métodos de clasificación como la red neuronal y KNN con el método de búsqueda de parámetros combinatoria (hacia delante y

hacia atrás) no son tan buenos (aunque son muy buenos para distinguir la psoriasis de otras lesiones de piel).

Método de la clasificación	Método de la optimización		Acierto de la clasificación (Tarea difícil)
KNN	Búsqueda secuencial	Hacia adelante	65.00 ± 22.91
		Hacia atrás	66.67 ± 17.56
	Dejar uno fuera	Hacia adelante	76.0 ± 0
		Hacia atrás	68.0 ± 0
KNN	Algoritmo genético (GA)		94.0 ± 1.0
RNA	Algoritmo genético (GA)		78.79 ± 2.62
ANFIS			92.35 ± 1.7

Tabla 6.12 Resumen de los resultados de los clasificadores con los mejores resultados obtenidos para la tarea difícil.

Así, todos los resultados que aparecen en la Tabla 6.12 son los mejores resultados obtenidos. Se puede observar que los clasificadores ANFIS y KNN-AG son los que llevan a mejores resultados. Es decir, los sistemas automáticos de diagnóstico con estos clasificadores tienen mayor eficiencia para la aplicación de diagnóstico en la que se centra este trabajo y teniendo en cuenta el número reducido de muestras con el que se cuenta.

Por otro lado se puede notar que el tiempo de ejecución en estos clasificadores (ANFIS y KNN-AG) es similar y mucho menor que el tiempo necesario para ejecución la RNA-AG.

Capítulo 7

Discusión y conclusiones

Este trabajo representa en una importante contribución en dos campos diferentes. Por un lado el amplio estudio sobre diversas características de imágenes de fluoroscopia y su utilidad para el diagnóstico de cáncer ayudará a validar o descartar definitivamente la técnica de fluoroscopia para esta patología. Además el desarrollo de un sistema automático modular de diagnóstico y su estrategia de optimización constituye una aportación más general cuyos resultados se pueden aplicar a la detección de otras patologías.

En este trabajo se han obtenido características discriminatorias de las imágenes entre clases de cáncer de piel. Las características se han definido como parámetros que pueden abstraer de un modo razonable las características de las imágenes de que se dispone en este trabajo. Así, los parámetros que se utilizan en el presente trabajo dependen de técnicas de detección de bordes, morfología matemática y transformada de Fourier.

Como hemos definido diversos parámetros sin reparar en la posible redundancia a priori, se ha utilizado el coeficiente de correlación sobre los parámetros, encontrando que muchos de ellos eran equivalentes, con lo que se ha decidido no reducir la redundancia seleccionando sólo algunos de ellos para su estudio en el marco de la aplicación diagnóstico.

Se ha implementado un algoritmo de clasificación basado en las K-vecinas más próximas (KNN), donde el acierto de este clasificador depende de la representatividad de las características utilizadas. Dicho clasificador se ha aplicado sobre distintas particiones de la base de datos con dos métodos de optimización (búsqueda y selección) de las mejores características, uno de ellos

basado en mecanismo de búsqueda secuencial combinatoria (hacia delante “*forward*” y hacia atrás “*backward*”) y el otro basado en algoritmos genéticos (búsqueda multi-camino). Se puede notar que los resultados obtenidos de los dos métodos, el clasificador KNN da muy buenos resultados cuando se utilizan algoritmos genéticos para optimizar las mejores características. En este caso el acierto llega al 94% para la tarea difícil en el caso de $K=1$ vecinos como se puede ver en la Tabla 6.5 mientras que el mejor resultado obtenido en la técnica de búsqueda secuencial en método hacia atrás no llega al 67% (utilizando validación cruzada con tres particiones entrenamiento-test). Por otro lado el mejor resultado obtenido cuando se utiliza el método de evaluación cruzada dejar-uno-fuera (“*leave one out*”) es 76%.

Se ha desarrollado un sistema automático de diagnóstico utilizando redes neuronales con optimización de las características seleccionadas por medio de algoritmos genéticos. En este sistema se obtienen los resultados de 78.8% en el caso de entrenamiento y 63.8% en el test. Las redes neuronales no alcanzan mejores resultados debido al reducido número de muestras en el que se basa este trabajo.

Por otro lado se ha desarrollado también un sistema diagnóstico automático que está basado en un sistema neuro-difuso. Este sistema utiliza ANFIS como un algoritmo de clasificación y el algoritmo de G-flip para seleccionar las mejores características o parámetros. Este sistema da casi el mejor resultado con un acierto de la clasificación 92.4 %.

Así, los cuatro sistemas de diagnóstico automático que se han desarrollado en este trabajo representan la capacidad de las

características extraídas de las imágenes de la fluoroscopia para el diagnóstico de cáncer de piel y ayudarán para distinguir los tipos de lesiones cancerosas (keratosis actínica y carcinoma basal). También es importante el proceso de selección de las mejores características para la clasificación eficiente, en algunos sistemas el número de entradas es crítico que sea pequeño. Sobre todo en el sistema neuro – difuso, para evitar la “maldición de la dimensionalidad”.

Se puede observar la distribución de los resultados en dos conjuntos, ya que algunos de los parámetros encontraban dificultades para diferenciar bien las clases de las imágenes. Este tipo de dificultades se ha encontrado sólo para distinguir entre las imágenes de los tumores (keratosis actínica y carcinoma basal), ya que la psoriasis era fácilmente distinguible.

Así, se puede ver en las tablas del capítulo 6 los buenos resultados de la clasificación en la tarea fácil, (obteniendo un porcentaje de acierto casi del 100% en todos los casos). Por otro lado, los resultados en la tarea difícil no se consideran buenos en todos los casos como se aprecia en las tablas de resultados de las clasificaciones KNN con la técnica de búsqueda secuencial y las redes neuronales optimizadas con algoritmos genéticos. Finalmente los sistemas KNN optimizados con algoritmos genéticos y el ANFIS son capaces de distinguir eficientemente entre las dos lesiones cancerosas.

Por todo ello, se puede concluir que con el conjunto de características exploradas es posible distinguir entre la psoriasis y patologías cancerosas y distinguir entre distintas lesiones cancerosas (keratosis actínica y carcinoma basal). Con algunos sistemas (ANFIS y KNN-GA) como mejores esquemas de clasificación.

Conclusiones

Tras la discusión realizada en el apartado anterior podemos resumir los siguientes puntos como principales aportaciones del presente trabajo de investigación:

- Desarrollo, optimización y evaluación de esquemas modulares de diagnóstico automático. Diversos campos biomédicos requieren de estrategias de diagnóstico distintas, en esta parte del trabajo nos hemos centrado en desarrollar herramientas genéricas aplicables a diversos campos biomédicos:
 - Concretamente se han diseñado clasificadores basados en el algoritmo de las k-vecinas más próximas, redes neuronales y sistemas neuro-difusos.
 - Se han implementado y evaluado también esquemas de selección automática de características relevantes y optimización basados en búsqueda combinatorial y algoritmos genéticos.
- Aplicación al diagnóstico de patologías de la piel basado en procesamiento de imágenes de fluoroscopia. En este campo la hipótesis de trabajo consiste en que este tipo de imágenes son monocromo (por lo tanto perdemos detalles relacionados con la pigmentación de la piel) pero dan información sobre capas profundas de la piel. Para evaluar la utilidad de este tipo de imágenes se han utilizado diversas técnicas de procesamiento de imágenes y se ha evaluado su importancia en dos tareas de diagnóstico concretas:
 - Discriminación entre lesiones cancerosas y patologías sistémicas.

- Se ha demostrado que las imágenes de fluoroscopia son útiles para la discriminación entre patologías cancerosas (keratosis actínica y carcinoma basal) y patologías sistémicas (como por ejemplo psoriasis). La discriminación entre estos dos tipos de patologías basadas en imágenes de fluoroscopia es fácil y requiere sólo un procesamiento de imágenes sencillo.
- Diversas técnicas de procesamiento de imágenes se muestran de gran utilidad para este propósito (morfología matemática, transformada de Fourier, extracción de bordes, evaluación del contraste de las imágenes y otros procesamientos más complejos).
- Discriminación entre distintas lesiones cancerosas:
 - Se ha demostrado que las imágenes de fluoroscopia son también útiles para discriminar entre distintas patologías cancerosas (keratosis actínica y carcinoma basal).
 - Se han evaluado distintas técnicas de procesamiento de imágenes y ninguna de ellas ha resultado concluyente por sí sola.

Se han utilizado esquemas de diagnóstico que seleccionen automáticamente características basadas en distintas técnicas de procesamiento de imágenes y se han obtenido resultados alrededor del 90 % de clasificación correcta (considerados como muy buenos ya que superan en un 30 % los mejores resultados publicados hasta el momento). Las técnicas que se han revelado más eficientes para esta tarea de diagnóstico son las siguientes: KNN-GA, ANFIS.

BIBLIOGRAFÍA Y SITIOS DE WEBS

- [ACK01] G. Ackermann, “Photophysikalische Grundlagen zur Luoreszenzdiagnostik von Tumoren der Haut”, PhD. Dissertation, University of Regensburg, 2001.
- [AAD04] <http://www.aad.org/pamphlets/Psoriasis.html>, 2004.
- [ABE04] P. Aberg, I. Nicander, J. Hansson, P. Geladi, U. Holmgren, and S. Ollmar, “*Skin cancer identification using multifrequency electrical impedance—a potential screening tool*”, IEEE Transactions on Biomedical Engineering, Vol. 51, No. 12, pp. 2097 – 2102, 2004.
- [ALT91] D. Altman, “*Practical statistics for medical research*”, London/New York: Chapman & Hall, 1991.
- [ANC04] S. G. Anchez, H. Perez y M. Nakano, “*Red Neuronal Creciente Usando Perturbación Simultánea*”, Información Tecnológica, Vol. 15, No. 5, pp. 45-52, 2004.
- [AST96] R. K. Astleman, “*Digital Image Processing*”, Prentice-Hall, Inc., USA, pp. 470-473, 1996.
- [BAK87] J. Baker, “*Reducing bias and inefficiency in the selection algorithm*”, proceedings of the 2nd International Conference on Genetic Algorithms, pp. 14-21, 1987.
- [BEL61] R. Bellman, “*Adaptive Control Processes: A Guided Tour*”, Princeton University Press, 1961.
- [BIN94] M. Binder, A. Steiner, M. Schwarz, S. Knollmayer, K.

- Wolff, and H. Pehamberger, “*Application of an artificial neural network in epiluminescence microscopy pattern analysis of pigmented skin lesions: a pilot study*”, *J. Dermatol*, Vol. 130, pp. 460–465, 1994.
- [BIN98] M. Binder, H. Kittler, A. Seeber, and A. Steiner, “*Epiluminescence microscopy-based classification of pigmented skin lesions using computerized image analysis and an artificial neural network*”, *Melanoma Res.*, Vol. 8, pp.261–266, 1998.
- [BLU97] A. L. Blum, and P. Langley, “*Selection of relevant features and examples in machine learning*”, *Artificial intelligence*, Vol. 97, Nos. 1-2, pp. 245-271, 1997.
- [BOS93] R. Bostock, E. Claridge, and A. J. Harget, “*Towards a neural network based system for skin cancer diagnosis*”, *IEE International Conference on Neural Networks*, pp. 215–219, 1993.
- [BRE95] H. Breu, J. Gil, D. Kirkpatrick, and M. Werman, “*Linear Time Euclidean Distance Transform Algorithms*”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 17, No. 5, pp. 529-533, 1995.
- [CAN03] <http://www.cancerbacup.org.uk/Cancertype/Skin>, 2003
- [CAN86] J. Canny, “*A Computational Approach to Edge Detection*”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 8, No. 6, pp. 679-698, 1986.
- [COM99] D. Comaniciu, P. Meer, and D. J. Foran, “*Image-guided decision support system for pathology*”, *Machine Vision and Applications*, Vol. 11, pp. 213–224, 1999.
- [CUT92] D. R. Cutting, D. R. Karger, J. O. Pedersen, and J. W. Tukey, “*Scatter/Gather: A cluster-based approach to*

- browsing large document collections*”, Proceedings of the 15th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 318-329, 1992.
- [DEJ86] G. F. Dejong, and R. J. Mooney, “*Explanation-Based Learning: An Alternative View*”, Machine Learning, Vol. 1, No 2, pp. 145-176, 1986.
- [DEN83] J. Dennis, and R. Schnabel, “*Numerical Methods for Unconstrained Optimization and Nonlinear Equations*”, Prentice-Hall, Englewood Cliff, 1983.
- [DEV96] L. Devroye, L. Györfi, A. Krzyżak, and G. Lugosi, “*A Probabilistic Theory of Pattern Recognition*”, Springer-Verlag, New York, Chapter 5, 1996.
- [DOA92] J. Doak, “*An evaluation of feature selection methods and their application to computer security*”, Technical Report CSE-92-18, University of California at Davis, 1992.
- [DRE01] S. Dreiseitl, L. O.Machado, H. Kittler, S. Vinterbo, H. Billhardt, and Michael Binder, “*A Comparison of Machine Learning Methods for the Diagnosis of Pigmented Skin Lesions*”, Journal of Biomedical Informatics, Vol. 34, pp. 28–36, 2001.
- [DUR93] A. Durg, W. V. Stoecker, J. Cookson, S. E. Umbaugh, and R. H. Moss, “*Identification of variegated coloring in skin tumors neural networks vs. rule-based induction methods*”, IEEE Engineering Medicine and Biology, Vol.12, No.3, pp. 71–74, 1993.
- [ELD82] D. E. Elder, M. H. Green, D. P. Guerry IV, K. H. Kraemer, and W. H. Clark, “*The dysplastic nevus syndrome—Our definition*”, J. Dermatol., Vol. 4, pp.

- 455–460, 1982.
- [ERC94] F. Ercal, A. Chawla, W. V. Stoecker, H. C. Lee, and R. H. Moss, “*Neural network diagnosis of malignant melanoma from color images*”, IEEE Transactions on Biomedical Engineering, Vol. 41, pp. 837–845, 1994.
- [ERC99] F. Ercal, H. C. Lee, W. V. Stoecker, and R. H. Moss, “*Skin cancer classification using hierarchical neural networks and fuzzy systems*”, International J. Smart Engineering System Design, Vol. 1, pp. 273–289, 1999.
- [FIS96] R. B. Fisher, S. Perkins, A. Walker, and E. Wolfart, “*HIPR: Hypermedia Image Processing Reference*”, CDrom published by John Wiley and Sons, Chichester, 1996. HIPR2 published over the www at <http://homepages.inf.ed.ac.uk/rbf/HIPR2>.
- [FIX51] E. Fix, and J. L. Hodges, “*Discriminatory Analysis: Nonparametric Discrimination: Consistency Properties*”, Technical report number 4, Project number 21-49-004 USAF School of Aviation Medicine, Randolph Field, Texas, 1951.
- [FLE92] M. Fleco, “*Some Defects in Finite-Difference Edge Finders*”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14 No.3, pp. 337-345, 1992.
- [FUN89] K. I. Funahaski, “*On the approximate realization of continuous mappings by neural networks*”, Neural networks, Vol. 2, pp. 183-192, 1989.
- [GAN01] H. Ganster, A. Pinz, R. Roehrer, E. Wildling, M. Binder, and H. Kittler, “*Automated Melanoma Recognition*”, IEEE Transactions on Medical Imaging, Vol. 20, No. 3, pp. 233-239, 2001.
- [GAN95] H. Ganster, M. Gelautz, A. Pinz, M. Binder, H.

- Pehamberger, M. Bammer, and J. Krocza, "*Initial Results of Automated Melanoma Recognition*", In Theory and Applications of Image analysis II, Selected papers from the 9th Scandinavian Conference on Image Analysis G. Borgefors (Eds.), World Scientific, pp. 343-354, 1995.
- [GAR87] R. García Martínez, W. Fritz, y J. Blanqué, "*Un algoritmo de aprendizaje de conceptos para sistemas inteligentes*", Anales del V Congreso Nacional de Informática y Teleinformática, Buenos Aires, Argentina, pp. 91-96, 1987.
- [GIL04] R. Gilad-Bachrach, A. Navot, and N. Tishby, "*Margin based feature selection - theory and algorithms*", in proceedings of the 21'st international conference on Machine Learning ICML, Banff, Canada, 2004. Available at <http://www.cs.huji.ac.il/~ranb/bib/paper/giladbachrachnavottishby04b.pdf>
- [GNI97] M. Gniadecka, H. Wulf, N. Mortensen, O. Nielsen, and D. Christensen, "*Diagnosis of basal cell carcinoma by Raman spectra*", J. Raman Spectroscopy, Vol. 28, pp. 125-129, 1997.
- [GON92] R. Gonzalez and R. Woods, "*Digital Image Processing*", Addison Wesley, pp. 414 - 428, 1992.
- [GRE91] A. Green, N. Martin, G. McKenzie, J. Pfitzner, F. Quintarelli, B. W. Thomas, M. O'Rourke, and N. Knight, "*Computer image analysis of pigmented skin lesions*", Melanoma Res., Vol. 1, pp. 231-236, 1991.
- [GRE94] A. Green, N. Martin, J. Pfitzner, M. O'Rourke, and N. Knight, "*Computer image analysis in the diagnosis of melanoma*", J. Dermatol., Vol. 31, No. 6, pp. 958-964,

- 1994.
- [GRI95] P. Griffiths, I. Lewis, N. Chaffin, and J. Jegla, “*Remote characterization of materials by vibrational spectrometry through optical fibers*”, J. Molecular Structure, Vol. 347, pp. 169–185, 1995.
- [HAN81] D. Hand, “*Discrimination and Classification*”, Wiley & Sons, New York, 1981.
- [HAS03] R. R. Hashemi, M. Bahar, Alexander Tyler, Azita Bahrami, Nan Tang, and William Hinson, “*Development of Group's Signature for Evaluation of Skin Cancer in Mice Caused by Ultraviolet Radiation*”, International Conference on Information Technology: Computers and Communications ITCC, pp. 617-620, 2003.
- [HAY99] S. Haykin, “*Neural Networks*”, A Comprehensive Foundation, Second Edition, Prentice Hall, pp. 138-350, 1999.
- [HIL95] J. Hilera y V. Martínez, “*Redes Neuronales Artificiales. Fundamentos, modelos y aplicaciones*”, Madrid: Editorial RA-MA, pp. 45-54, 1995.
- [HIN01] M. Hintz-Madsen, L. Hansen, J. Larsen, and K. Drzewiecki, “*A probabilistic neural network framework for detection of malignant melanoma*”, in Artificial Neural Networks in Cancer Diagnosis, Prognosis and Patient Management, R. Naguib and G. Sherbet, CRC Press, pp. 141–183, 2001.
- [HIN96] M. Hintz-Madsen, L. Kai Hansen, J. Larsen, E. Olesen, and K. T. Drzewiecki, “*Detection of malignant melanoma using neural classifiers*”, in Solving Engineering Problems with Neural Networks, Proceedings International Conference Engineering

- Applications on Neural Networks EANN'96, pp. 395–398, 1996.
- [HOL70] J. H. Holland, *Genetic Algorithms*, Scientific American, 1970.
- [HOU95] C. R. Houck, J. A. Joines, and M. G. Kay, “A *Genetic Algorithm for function Optimization: A Matlab implementation*”, Technical Report NCSU-IE TR 95-09, North Carolina State University, 1995.
Also available as <http://www.fmmcenter.ncsu.edu/facstaff/joines/papers/gaot.ps>.
- [JAM97] R. James Parker, “*Algorithms for Image Processing and Computer Vision*”, John Wiley & Sons, Inc., New York, pp. 23-29, 1997.
- [JAN92] J. S. Jang, “*Neurofuzzy Modeling: Architecture, Analyses and Applications*”, PHD thesis, University of California, Berkeley, CA, USA, 1992.
- [JAN93] J. S. Jang. “*ANFIS: Adaptive-network-based fuzzy inference systems*”, IEEE Transactions on Systems, Man and Cybernetics, Vol. 23, No. 3, pp.665-685, 1993.
- [JAN96] J. S. Jang, “*Input selection for ANFIS learning*”, Proceedings of 5th IEEE International Conference on Fuzzy System, New York, USA, Vol. 2, pp. 1493-1499, 1996.
- [KAU90] L. Kaufmann, and P. J. Rousseeuw, “*Finding Groups in data: An introduction to Cluster Analysis*”, John Wiley & Sons, Inc., New York, USA, pp. 111-116, 1990.
- [KJO95] A. Kjoelen, M. J. Thomson, S. E. Umbaugh, R.H. Moss, and W. V. Stoecker, “*Performance of AI methods in detecting melanoma*”, IEEE Engineering Medicine and Biology, Vol. 14, No. 4, pp.411–416, 1995.

- [KOH89] H. Koh, R. Lew, and M. Prout, “*Screening for melanoma/skin cancer*”, J. Dermatol., Vol. 20, No. 2, pp. 159–172, 1989.
- [KUL98] S. Kulkarni, G. Lugosi, and S. Venkatesh, “*Learning pattern classification a survey*”, IEEE Transaction on Information Theory, Vol. 44, No 6, pp. 2178–2206, 1998.
- [KUS98] B. Kusumoputro and A. Ariyanto, “*Neural network diagnosis of malignant skin cancers using principal component analysis as a preprocessor*”, IEEE International Joint Conference on Neural Networks Proceedings, Vol.1, pp. 310-313, 1998.
- [LEO96] A. V. Leousky, and W. B. Croft, “*An evaluation of techniques for clustering search results*”, Technical Report IR-76, Department of Computer Science, University of Massachusetts, Amherst. 1996.
- [LIM90] J. S. Lim, “*Two-Dimensional Signal and Image Processing*”, Englewood Cliffs, NJ, Prentice Hall, pp. 478-488, 1990.
- [LIN94] B. Lindelöf, and M. Hedblad, “*Accuracy in the clinical diagnosis and pattern of malignant melanoma at a dermatological clinic*”, J. Dermatol., Vol. 21, No. 7, pp. 461–464, 1994.
- [MAG04] I. G. Maglogiannis, and E. P. Zafiroopoulos, “*Characterization of digital medical images utilizing support vector machines*”, BMC Medical Informatics and Decision Making, This article is available at: <http://www.biomedcentral.com/1472-6947/4/4>, 2004.
- [MAR95] R. Marks, “*An overview of skin cancers. Incidence and causation*”, U.S. National Library of Medicine, PubMed,

- Vol. 75, pp. 607–612, 1995.
- [MEH02] M. Mehrubeoglu, D. Nasser Kehtarnavaz, G. Marquez, M. Duvic, and V. W. Lihong, “*Skin lesion classification using oblique-incidence diffuse reflectance spectroscopic imaging Applied optics*”, Vol. 41, No.1 pp.182-192, 2002.
- [MIC82] R. S. Michalski, A. B. Baskin, and K. A. Spackman, “*A Logic-Based Approach to Conceptual Database Analysis*”, Sixth Annual Symposium on Computer Applications on Medical Care, George Washington University, Medical Center, Washington, DC, USA, pp. 792-796, 1982.
- [MIC83] R. S. Michalski, “*A Theory and Methodology of Inductive Learning*”, Artificial Intelligence, Vol. 20, pp. 111-161, 1983.
- [MIC94] D. Michie, D. J. Spiegelhalter, and C. C. Taylor, “*Machine Learning, Neural and Statistical Classification*”, Ellis Horwood, New York, USA, pp. 153-156, 1994.
- [MIC98] R. B. Michalski, and M. I. Kubat, (eds.), “*Machine Learning and Data Mining, Methods and Applications*”, John Wiley & Sons Ltd, West Sussex, England, 1998.
- [MIC99] Z. Michalewickz, “*Genetic Algorithms + Data Structures = Evolution Programs*”, Springer-Verlag, 1999.
- [MIE01] Ch. Mies, Ch. Bauer, G. Ackermann, W. Bäumler, C. Abels, C. G. Puntonet, M. R. Alvarez, and E. W. Lang, “*Can ICA help classify skin cancer and benign lesions?*”, Lecture Notes in Computer Science, Vol. 2085, pp. 328-335, 2001.

- [MIR01] J. Mira and A. Prieto (Eds), “*Bio-Inspired Applications of Connectionism*”, Lecture Notes in Computer Science, Vol. 2085, pp. 328-335, 2001.
- [MOT03] S. Mota, E. Ros, F. De Toro, and J. Ortega, “*Paroxysmal Atrial Fibrillation Prediction Application Using Genetic Algorithms*”, Lecture Notes in Computer Science, Vol. 2660, pp. 1011-1019, 2003.
- [MUÑ01] A. Muñoz, G. Osorio, F. Prieto, y F. Angulo, “*Detección de Patologías Celulares Empleando Visión Artificial*”, En DYNA, Revista de la Facultad de Minas, Universidad Nacional de Colombia Sede Medellín, Medellín, Colombia, Vol. 68, No. 133, pp. 73-82, 2001.
- [NET05] <http://www.netdoctor.co.uk/diseases/facts/skincancer/melanoma.htm>, 2005.
- [ODE05] S. M. Odeh, E. R. Vidal, I. R. Ruiz, “*Exploración de características para el diagnóstico de lesiones de piel basado en imágenes de la fluorescencia*”, I symposium Computational Intelligence, SICO'2005, Granada, pp. 431- 438 , 2005.
- [OLM93] I. Olmeda y B-R. Sergio, “*Redes neuronales artificiales. Fundamentos y aplicaciones*”. Servicio de Publicaciones de la Universidad da Alcalá, informe técnico, 1993.
- [OTS79] N. Otsu, “*A Threshold Selection Method from Gray-Level Histograms*”, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 9, No. 1, pp. 62-66, 1979.
- [PAI82] C. Paige, and M. Saunders, “*LSQR: an algorithm for sparse linear equations and sparse least squares*”, ACM Transactions Math. Software, Vol.8, No.1, pp. 43-71, 1982.
- [PIC02] D. Piccolo, A. Ferrari, K. Peris, R. Daidone, B. Ruggeri,

- and S. Chimenti, “*Dermoscopic diagnosis by a trained clinician vs. a clinician with minimal dermoscopy training vs. computer-aided diagnosis of 341 pigmented skin lesions: A comparative study*”, *J. Dermatol.*, Vol. 147, pp. 481–486, 2002.
- [PIT00] I. Pitas, “*digital image processing algorithms and applications*”, Hon Wiley & sons, pp. 242-243, 2000.
- [POM99] R. Pompl, W. Bunk, A. Horsch, W. Abmayr, G. Morfill, W. Brauer, and W. Stolz, “*Computer vision of melanocytic lesions using MELDOQ*,” in *Proceedings 6th Congress International Soc. Skin Imaging, London; Skin Research and Technology*, Vol. 5, pp. 150, 1999.
- [PRE87] S. Presser and J. Taylor, “*Clinical diagnostic accuracy of basal cell carcinoma*”, *J. Dermatol.*, Vol. 16, pp. 988–990, 1987.
- [PUD94] P. Pudil, J. Novovicová, and J. Kittler, “*Floating search methods in feature selection*”, *Pattern Recognition Lett.*, Vol. 15, pp. 1119–1125, 1994.
- [PUN93] W. F. Punch, E. D. Goodman, M. Pei, L. Chia-Shun, P. Hovland, and R. Enbody, “*Further Research on Feature Selection and Classification Using Genetic Algorithms*”, *International Conference on Genetic Algorithms 93*, Morgan Kaufmann, San Mateo, CA, USA, pp. 557-564, 1993.
- [QUI99] J. R. Quinlan, “*Learning Logic Definitions from Relations, Machine Learning*”, Kluwer Academic Publishers, Vol. 5, pp. 239-266, 1999.
- [QUN86] J. R. Quinlan, “*Induction of Decision Trees, Machine Learning*”, Kluwer Academic Publishers, Vol. 1, pp 81-106, 1986.

- [ROB03] S. L. Robert, and K. W. Peter, “*Bayesian Classification and Feature Reduction Using Uniform Dirichlet Priors*”, IEEE Transactions On Systems, Man And Cybernetics—Part B: Cybernetics, Vol. 33, No. 3, pp. 448- 464, 2003.
- [ROS03] E. Ros, M. M. Rodríguez, S. Mota, J. L. Bernier, I. Rojas, C.G. Puntonet and E.W. Lang, “*Skin Lesions Diagnosis Based on Fluorescence Image Processing: Simple Parameters Scanning*”, IbPRIA 2003, Vol. 2562, pp. 858-867, 2003.
- [ROS04] E. Ros, S. Mota, F. J. Fernández, F. J. Toro, and J. L. Bernier, “*ECG Characterization of Paroxysmal Atrial Fibrillation: Parameter Extraction and Automatic Diagnosis Algorithm*”, Computers in Biology and Medicine, Vol. 34, No. 8, pp. 679-696, 2004.
- [ROS95] T. Ross, H. Handels, J. Kreuzsch, H. Busche, H. H.Wolf, and S. J. Pöppel, “*Automatic classification of skin tumours with high resolution surface profiles*”, Lecture Notes in Computer Science, Vol. 970, pp. 368–375, 1995.
- [RUB02] P. Rubegni, M. Burrioni, G. Cevenini, R. Perotti, G. Dell’Eva, P. Barbini, M. Fimiani, and L. Andreassi, “*Digital dermoscopy analysis and artificial neural network for the differentiation of clinically atypical pigmented skin lesions: A retrospective study*”, J. Dermatol, Vol. 119, pp. 471–474, 2002.
- [SAN04] G. Sanchez, H. Perez y M. Nakano, “*Red Neuronal Creciente Usando Perturbación Simultánea*”, Información tecnológica, Vol.15, No.5, pp. 45-52, 2004.
- [SBO01] A. Sboner, E. Blanzieri, C. Eccher, P. Bauer, M.

- Cristofolini, G. Zumiani, and Stefano Forti, “*A Knowledge Based System for Early Melanoma Diagnosis Support*”, Proceedings of the 6th IDAMAP Workshop - Intelligent Data Analysis in Medicine and Pharmacology, London, UK, 2001
- [SBO02] A. Sboner, C. Eccher, G. Zumiani, P. Bauer and S. Forti, “*Exploiting clinical knowledge for supporting early diagnosis of melanoma*”, Proceedings of the IDAMAP2002, pp.83-87, 2002.
- [SBO04] A. Sboner, P. Bauer, G. Zumiani, C. Eccher, E. Blanzieri, S. Forti, and M. Cristofolini, “*Clinical Validation of an automated system for supporting the early diagnosis of melanoma*”, Skin Research and Technology, Vol. 10, No. 3, pp.184-192, 2004.
- [SCH03] P. Schmid-Saugeon, J. Guillod, and J. P. Thiran, “*Towards a Computer-Aided Diagnosis System for Pigmented Skin Lesions*”, Computerized Medical Imaging and Graphics Vol. 27, pp. 65–78, 2003.
- [SCH95] B. Schrader, S. Keller, T. Lochte, S. Fendel, D. Moore, A. Simon, and J. Sawatzki, “*NIR FT Raman spectroscopy in medical diagnosis*”, J. Molecular Structure, Vol. 346, pp. 293–296, 1995.
- [SER88] J. Serra, “*Image analysis and Mathematical Morphology*”, Vol. II, Theoretical Advances, Academic Press, London, 1988.
- [SHI03] YU Shixin, “*Feature Selection and Classifier Ensembles: A Study on Hyperspectral Remote Sensing Data*”, PhD thesis, University of Antwerp, 2003.
- [SHI97] M. Shim and B. Wilson, “*Development of an in vivo*

- Raman spectroscopy system for diagnostic applications*", J. Raman Spectroscopy, Vol. 28, pp. 131–142, 1997.
- [SIE98] B. Sierra, and P. Larrañaga, "*Predicting the survival in malignant skin melanoma using Bayesian networks automatically induced by genetic algorithms. An empirical comparison between different approaches*", Artificial Intelligence in Medicine. Vol. 14, Nos. 1, 2, pp. 215-230, 1998.
- [SIG04] S. Sigurdsson, P. A. Philipsen, L. K. Hansen, J. Larsen, M. Gniadecka, and H. C. Wolf. "*Detection of skin cancer by classification of Raman spectra*", IEEE Transactions on Biomedical Engineering, Vol. 51, No. 10, pp. 1784–1793, 2004.
- [SIN99] S. Singh, J.F. Haddon and M. Markou, "*Nearest Neighbour Strategies for Image Understanding*", Proceedings Workshop on Advanced Concepts for Intelligent Vision Systems (ACIVS'99), Baden-Baden, pp. 74-79, 1999.
- [SKI04] http://www.skinsite.com/info_psooriasis.htm, 2004.
- [SOB83] A. S. Sober, A. R. Rhodes, C. L. Day Jr., T. B. Fitzpatrick, and M. C. Mihm Jr., "*Primary melanoma of the skin. Recognition of precursor lesions and estimation of prognosis in stage I*", in Update: Dermatology in General Medicine, T. B. Fitzpatrick, A. Z. Eisen, K. Wolff, I. M. Freedberg, and K. F. Austen, Eds. New York: McGraw Hill, pp. 98–112, 1983.
- [STO04] H. G. Stockmeier, W. Bäumlner, H. Szeimies, F. J. Theis, E. W. Lang, and C. G. Puntonet, "*Classification of Skin Lesions by Fluorescence Diagnosis and Independent*

- Component Analysis*”, Proceedings 2nd International Conference on Biomedical Engineering (BIOMED 2004), pp. 204 – 208, 2004.
- [STO74] M. Stone, “*Cross-validation choice and assessment of statistical predictions*”, J. Royal. Statzs. Soc, Vol. 36, pp. 111-147, 1974.
- [STO94] W. Stolz, O. Braun-Falco, P. Bilek, M. Landthaler, and A. B. Cognetta, “*Color atlas of dermatoscopy*”. Oxford, Blackwell, 1994.
- [TAK85] T. Takagi, and M. Sugeno, “*Fuzzy identification of systems and its applications to modeling and control*”, IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-15, No. 1, pp. 116-132, 1985.
- [THE99] S. Theodoridis and K. Koutroumbas, “*Pattern Recognition*”, Academic Press, 1999.
- [TOR06] F. de Toro, E. Ros, S. Mota, and J. Ortega, “*Evolutionary Algorithms for Multiobjective and Multimodal Optimization of Diagnostic Schemes*”, IEEE Transactions on Biomedical Engineering, VOL. 53, NO. 2, pp. 178-189, 2006.
- [TOR98] J. M. Torres Moreno, “*Redes de Neuronas y Clasificación: De paradigmas conexionistas clásicos a redes que aprenden mientras crecen*”, LANIA, Vol. 23-24, artículo 2, 1998.
- [UMB91] S. E. Umbaugh, R. H. Moss and W. V. Stoecker, “*Applying artificial intelligence to the identification of variegated coloring in skin tumors*”, IEEE Engineering on Medicine and Biology, Vol. 10, No. 4, pp. 57–62, 1991.
- [UMG05] <http://www.umgcc.org/overview/nonmel.html>, 2005.

- [VAL99] P. E. Valencia, "*Selección de Características Para Clasificadores Neuronales*", Anales de instituto de Ingenieros de Chile. Diciembre, 1999.
- [WEB03] A. Webb, "*statistical pattern recognition*", second edition, Wiley press, England, pp. 93, 2003.
- [WHI93] D. Whitley, "*A Genetic Algorithm Tutorial*", Reporte Técnico CS-93- 103, Colorado State University, Marzo 1993.
- [MAN90] O. L. Mangasarian, R. Setiono, and W.H. Wolberg, "*Pattern recognition via linear programming: Theory and application to medical diagnosis*", in: "Large-scale numerical optimization", Thomas F. Coleman and Yuying Li, editors, SIAM Publications, Philadelphia, pp. 22-30, 1990.