

UNIVERSIDAD POLITÉCNICA DE VALENCIA
DEPARTAMENTO DE SISTEMAS INFORMÁTICOS Y COMPUTACIÓN



Tesis Doctoral

Aportaciones al Diagnóstico de Cáncer Asistido por Ordenador

Presentada por:
Rafael Llobet Azpitarte

Dirigida por los doctores:
D. Juan Carlos Pérez Cortés
D. Roberto Paredes Palacios

Julio de 2006

Aportaciones al Diagnóstico de Cáncer Asistido por Ordenador

Rafael Llobet Azpitarte

Trabajo realizado bajo la dirección de los Doctores
D. Juan Carlos Pérez Cortés
D. Roberto Paredes Palacios
y presentado en la Universidad Politécnica de Valencia
para optar al grado de Doctor en Informática

Valencia, 2006

*A Gloria, Mario y Pau,
por el tiempo robado.*

Agradecimientos

La realización de este trabajo no hubiera sido posible sin la colaboración de muchas personas que, de un modo u otro, me han aportado sus conocimientos y dado su apoyo.

En primer lugar quiero dar mi agradecimiento a mis directores de tesis, Juan Carlos Pérez y Roberto Paredes. Ellos me han proporcionado gran cantidad de los conocimientos que he empleado durante el desarrollo de esta tesis. Debo agradecer, además, su paciente revisión del texto presentado.

Alejandro Toselli y Alfons Juan participaron de forma activa en la primera parte de este trabajo, referente a la diagnosis del cáncer de próstata. A Alfons le debo, entre otras cosas, la idea de aplicar los Modelos de Markov en esta tarea, y a Alejandro la práctica totalidad de los experimentos realizados con este clasificador.

El equipo de urología del hospital La Fe de Valencia y, de forma muy especial, el doctor Álvaro Gómez-Ferrer, han participado también activamente en la tarea de diagnosis de cáncer de próstata. A ellos he de agradecer la aportación de los conocimientos médicos necesarios para el desarrollo de este trabajo y su participación en la evaluación del sistema de CAD implementado.

Agradecer igualmente a las instituciones que han subvencionado parcialmente este trabajo: Comisión Interministerial de Ciencia y Tecnología (CICYT) a través de sus proyectos TIC2000-1703-CO3-01 y TIC2003-08496-CO4 y Agencia Valenciana de Ciencia y Tecnología (AVCiT) a través de sus proyectos CTIDIA/2002/80 y GRUPOS03/031.

Quiero agradecer también a Filiberto Pla, Fransesc Ferri y Nicolás Pérez de la Blanca, por su participación en la revisión de este trabajo y a todo el grupo de PRHLT y RIVA por sus afortunados comentarios durante las sesiones de seminario. De manera particular a Ximo Arlandis y Javi Cano, con quienes he compartido muchas horas de trabajo (y de no trabajo).

Al meu amic Albert del Toro he d'agrair-li la traducció del text en valencià, inclòs aquest.

Finalmente, aunque no tengan el más mínimo interés en nada de lo que aquí se cuenta, quiero agradecer a Pau, Guillem, Sergio, Diego, Marina C., Mario, Joan, Helena, Marina G., Julián, Aïna, Elisa y Laura, que, con su grata compañía durante un buen número de excursiones y aventuras, me han proporcionado esos momentos de ocio y de “descanso” tan necesarios para finalizar con éxito una maratón como ésta.

Resumen

Para diagnosticar un cáncer se realizan, entre otras pruebas, tests de imagen, como pueden ser una radiografías, ecografías o resonancias magnéticas. Mediante estos tests pueden detectarse zonas con alta sospecha tumoral, cuyo diagnóstico debe confirmarse finalmente mediante la realización de una biopsia. Este tipo de imágenes, sin embargo, no son fáciles de interpretar, lo que provoca que el profesional encargado de analizarlas, a pesar de su experiencia, no sea capaz de detectar en ellas un porcentaje importante de tumores (falsos negativos).

Una posibilidad para mejorar el diagnóstico y disminuir el número de falsos negativos consiste en utilizar sistemas de diagnóstico asistido por ordenador o *computer-aided diagnosis* (CAD). Un sistema de CAD analiza la imagen médica y trata de detectar zonas sospechosas de contener alguna anomalía. Estas zonas son marcadas sobre la propia imagen con un doble objetivo: llamar la atención del profesional hacia la zona sospechosa y aportar una hipótesis plausible respecto al diagnóstico.

En esta tesis se presentan y evalúan diversas técnicas de visión por computador y reconocimiento de formas orientadas a la detección de tumores en imágenes médicas, con el objetivo de diseñar sistemas de CAD que permitan un mejor diagnóstico. El trabajo se ha centrado en el diagnóstico de cáncer de próstata a partir de imágenes de ecografía, y en el diagnóstico de cáncer de mama a partir de imágenes de radiografía. Se han evaluado diversos métodos de extracción de características basados en valores de intensidad, texturas o gradiente. En la etapa de clasificación se ha utilizado un clasificador no paramétrico basado en distancias (k -vecinos más cercanos) y otro paramétrico basado en modelos de Markov. Además se han evaluado las Máquinas de Vectores Soporte (SVM) y un clasificador de *Boosting*. A lo largo del trabajo se evidencian las distintas problemáticas que surgen en este tipo de tareas y se proponen soluciones a cada una de ellas.

El diagnóstico de cáncer de próstata asistido por ordenador es una tarea extremadamente complicada y que ha sido muy poco estudiada. A pesar de las dificultades, se han obtenido resultados que, aunque modestos, demuestran que es posible encontrar características en la imagen que discriminen entre tejido sano y canceroso. El caso del CAD de cáncer de mama ha sido investigado con mayor profundidad, y así lo demuestra el importante número de trabajos que han sido publicados al respecto. Los resultados obtenidos en este trabajo han sido prometedores, e igualan o superan a otros resultados publicados con anterioridad.

Resum

Per tal de diagnosticar un càncer es realitzen, entre altres proves, tests d'imatge com ara radiografies, ecografies o ressonàncies magnètiques. Per mitjà d'aquests tests poden detectar-se zones amb alta sospita tumoral, diagnòstic que ha de ser confirmat finalment mitjançant la realització d'una biòpsia. Aquest tipus d'imatges, però, no són fàcils d'interpretar, la qual cosa provoca que el professional encarregat d'analitzar-les, malgrat la seua experiència, no siga capaç de detectar-hi un percentatge important de tumors (falsos negatius).

Una possibilitat per a millorar el diagnòstic i disminuir el nombre de falsos negatius consisteix a utilitzar sistemes de diagnosi assistida per ordinador o *computer-aided diagnosis* (CAD). Un sistema de CAD analitza la imatge mèdica i tracta de detectar zones sospitoses de contenir alguna anomalia. Aquestes zones són marca-des sobre la imatge mateix amb un doble objectiu: cridar l'atenció del professional cap a la zona sospitosa i aportar una hipòtesi plausible quant al diagnòstic.

En aquesta tesi es presenten i avaluen diverses tècniques de visió per computador i de reconeixement de formes orientades a la detecció de tumors en imatges mèdiques, amb l'objectiu de dissenyar sistemes de CAD que permeten un diagnòstic més ajustat. El treball s'ha centrat en el diagnòstic de càncer de pròstata a partir d'imatges d'ecografia, i en el diagnòstic de càncer de mama a partir d'imatges de radiografia. S'han avaluat diversos mètodes d'extracció de característiques basats en valors d'intensitat, de textures o de gradient. En l'etapa de classificació s'ha utilitzat un classificador no paramètric basat en distàncies (k -veïns més pròxims) i un altre de paramètric basat en models de Markov. A més a més, s'han avaluat les Màquines de Vectors Suport (SVM) i un classificador de Boosting. Al llarg del treball s'evidencien les distintes problemàtiques que sorgeixen en aquest tipus de tasques i es proposen solucions per a cadascuna d'elles.

El diagnòstic de càncer de pròstata assistit per ordinador és una tasca extremadament complicada i molt poc estudiada. A pesar de les dificultats, s'han obtingut resultats que, tot i ser modestos, demostren que és possible trobar característiques en la imatge que discriminen entre teixit sa i cancerós. El cas del CAD de càncer de mama ha sigut investigat amb major profunditat, i així ho demostra l'important nombre de treballs que han sigut publicats en aquest camp. Els resultats obtinguts en aquest treball han sigut prometedors, i igualen o superen d'altres resultats publicats abans.

Abstract

In order to diagnose a cancer, medical imaging techniques like radiography, ultrasound or magnetic resonance are commonly used. By means of these techniques is possible to detect regions with high risk of containing cancer cells, whose diagnostic must be confirmed by getting a cell sample (biopsy). However, these kind of images are not easy to interpret, which causes that, sometimes, the physician is not able to detect the presence of a cancer (false negative), even when he or her has a large experience.

One possible way of improving the diagnosis and reducing the number of false negatives is to use computer-aided diagnosis (CAD). A CAD system analyses the medical image and tries to detect suspicious regions. These regions are highlighted in order to draw the attention of a human expert and to provide a second read of the image.

In this thesis, different methods of computer vision and pattern recognition aimed at cancer detection from medical images are presented and tested. The goal is to implement a CAD system that can help to improve the diagnosis. The study is focused on prostate cancer diagnosis from ultrasound images and on breast cancer diagnosis from mammographies. Different extraction features methods based on intensity values, textures and gradient have been tested. A non-parametric classifier based on distances (k -nearest neighbors) and a parametric classifier based on Markov Models have been used. Also, Support Vector Machines (SVM) and a *Boosting* classifier have been tested. The different difficulties that arise in computer-aided diagnosis are explained and solutions for each one are proposed.

Prostate computer-aided diagnosis is an extremely difficult task and with very few studies have been published in this area. In spite of the difficulties, promising results indicating that valid information for the diagnosis is present in the image have been obtained. On the other hand, breast computer-aided diagnosis has been researched in depth and many studies related to this topic have been published. The results obtained in this thesis have been encouraging, and are similar, or better in some cases, than other previously published.

Índice general

1. Introducción	1
1.1. Motivación	1
1.2. Objetivos generales	2
1.3. Estructura de la tesis	5
2. Conceptos de reconocimiento de formas	7
2.1. Introducción al reconocimiento de formas	7
2.1.1. Aproximación estadística	8
2.1.2. Aproximación estructural	9
2.1.3. Esquema general de un sistema de RF inductivo supervisado	9
2.2. Aproximación estadística al reconocimiento de formas	12
2.2.1. Clasificación probabilística: la regla de Bayes	13
2.2.2. Métodos de estimación paramétrica y no paramétrica . . .	15
2.3. Clasificadores basados en distancias	16
2.3.1. Ventanas de Parzen	16
2.3.2. El clasificador de los k -vecinos más cercanos	18
2.4. Modelos ocultos de Markov	21
2.4.1. Definición de HMM continuo	22
2.4.2. Probabilidad de emisión	23
2.4.3. Nomenclatura utilizada	23
2.4.4. Algoritmo “Forward”	24
2.4.5. Algoritmo “Backward”	24
2.4.6. Algoritmo “Viterbi”	25
2.4.7. Algoritmo “Backward-Forward” para entrenamiento de un HMM	25
2.5. Redes neuronales	27
2.6. Máquinas de Vectores Soporte	28
2.7. <i>Boosting</i>	31

3. Detección de lesiones en imágenes médicas	35
3.1. Introducción	35
3.2. Aproximaciones a la detección de tumores en imágenes médicas .	37
3.2.1. Preprocesamiento	38
3.2.2. Segmentación	38
3.2.3. Extracción de características y selección	40
3.2.4. Clasificación	41
3.2.5. Evaluación	42
3.3. Detección de cáncer de próstata	42
3.3.1. Métodos actuales en el diagnóstico del cáncer de próstata .	43
3.3.2. Limitaciones actuales al diagnóstico precoz del cáncer de próstata	44
3.3.3. Diagnóstico del cáncer de próstata asistido por ordenador .	46
3.4. Detección de cáncer de mama	50
3.4.1. Métodos actuales en el diagnóstico del cáncer de mama . .	51
3.4.2. Limitaciones actuales al diagnóstico precoz del cáncer de mama	53
3.4.3. Diagnóstico del cáncer de mama asistido por ordenador . .	54
4. Técnicas empleadas en la clasificación y detección de patrones en imágenes médicas	59
4.1. Introducción	59
4.2. Extracción de características	60
4.2.1. Niveles de gris	62
4.2.2. Filtro de Sobel	62
4.2.3. Matrices de Coocurrencia	63
4.2.4. AFUM	65
4.2.5. SFUM	66
4.3. Reducción de la dimensionalidad	66
4.3.1. La maldición de la dimensionalidad	67
4.3.2. Métodos de selección de características	67
4.3.3. Análisis mediante componentes principales	68
4.4. Tamaño de la ventana de características	71
4.5. Esquema multiescala	72
4.6. Búsqueda rápida de los k -NN mediante <i>kd-trees</i>	75
4.7. Modelos de Markov aplicados a la detección de cáncer en imágenes de ecografía transrectal	80
4.8. Análisis de resultados mediante curvas ROC y FROC	84
4.8.1. Conceptos previos	85
4.8.2. La curva ROC	89
4.8.3. Comparación de curvas ROC. El área bajo la curva ROC. .	90

4.8.4.	Curvas FROC	90
4.8.5.	Comparación de curvas FROC	92
5.	Aportaciones al diagnóstico de cáncer de próstata	93
5.1.	Introducción	93
5.2.	Adquisición y puesta a punto de una base de datos de imágenes ecográficas de próstata	96
5.2.1.	Población del estudio	97
5.2.2.	Equipo de adquisición	98
5.2.3.	Grabación de la biopsia	98
5.2.4.	Segmentación del tejido biopsiado	100
5.2.5.	El problema de la supervisión imperfecta	101
5.2.6.	Características del corpus	102
5.3.	Experimentos con clasificación automática de imágenes	103
5.3.1.	Conjuntos de entrenamiento y test	104
5.3.2.	Detalles de la experimentación	106
5.3.3.	Resultados	109
5.3.4.	Conclusiones	114
5.4.	Sistema de ayuda al diagnóstico	114
5.4.1.	Experimentos	114
5.4.2.	Resultados	118
5.5.	Conclusiones	126
6.	Aportaciones al diagnóstico de cáncer de mama	129
6.1.	Introducción	129
6.2.	Corpus	132
6.3.	Entrenamiento del modelo	134
6.3.1.	Invariantes a rotación y escalado	134
6.3.2.	<i>Bootstrapping</i>	135
6.4.	Experimentos de clasificación con regiones de interés	137
6.4.1.	Definición de las regiones de interés	138
6.4.2.	Preproceso	140
6.4.3.	<i>Bootstrapping</i>	141
6.4.4.	PCA	143
6.4.5.	Multiescalado	145
6.4.6.	Tamaño de la ventana de características y área de muestreo.	150
6.4.7.	Aproximaciones de extracción de características	152
6.4.8.	Comparación con otros clasificadores	155
6.4.9.	Conclusiones	160
6.5.	Experimentos de detección con imágenes completas	161
6.5.1.	Detección de regiones sospechosas	162

6.5.2. Trabajos previos	163
6.5.3. Conjuntos de evaluación	166
6.5.4. Resultados	167
6.5.5. Conclusiones y desarrollos futuros	169
7. Conclusiones finales y desarrollos futuros	171
7.1. Conclusiones finales	171
7.1.1. Publicaciones relacionadas con esta Tesis	174
7.2. Desarrollos futuros	175
I Apéndices	177
A. Cálculo de estadísticos en matrices de coocurrencia	179
B. Imágenes de los corpora utilizados	183
Bibliografía	192

Capítulo 1

Introducción

1.1. Motivación

El trabajo que se presenta en esta memoria es el resultado de varios años de investigación. La línea de trabajo que se ha seguido surge cuando, tras haber estado trabajando en algunos proyectos de visión por computador y reconocimiento de formas, se nos plantea el reto de abordar un problema de detección de cáncer de próstata a partir de imágenes de ecografía. El objetivo último era implementar un sistema de diagnóstico asistido por ordenador que ayudara al urólogo en la localización de tumores a partir de la imagen de ecografía. La tarea se adivinaba ciertamente difícil debido, principalmente, a que debíamos implementar un sistema que pudiera detectar tejido canceroso cuando, en la mayoría de los casos, ni siquiera un urólogo con varios años de experiencia, era capaz de encontrar tales patrones en la imagen ecográfica. El interés, sin embargo, era alto, debido a que una pequeña mejora en la precisión diagnóstica podía suponer un enorme beneficio para los pacientes que presentan este tipo de tumores. Pronto nos dimos cuenta de los escasos trabajos de investigación que se habían realizado en este campo y de lo poco que se había conseguido avanzar en el diagnóstico de cáncer de próstata asistido por ordenador, a pesar del indudable interés que presenta la tarea.

Para abordar el proyecto se creó un equipo de trabajo interdisciplinar, en el que, junto con investigadores de los grupos de Visión por Computador y de Reconocimiento de Formas y Tecnologías de la Percepción, de la Universidad Politécnica de Valencia, participaron activamente doctores del servicio de urología del hospital La Fe de Valencia. Puesto que no existía ninguna base de datos pública de imágenes de ecografía de próstata con la que trabajar, la primera tarea consistió en la adquisición de la misma. Durante un año se grabaron las ecografías de próstata realizadas a los pacientes que acudían al hospital La Fe, con lo que se lograron registrar más de 1.500 ecografías. Esta tarea supuso un arduo trabajo de puesta a punto de los equipos de adquisición y grabación, etiquetado, revisión y organización de las imá-

genes.

A la dificultad intrínseca de la tarea, se añadieron nuevas complicaciones en el etiquetado del corpus, causadas por un alto grado de incertidumbre respecto a la presencia o no de células cancerosas en algunas partes del tejido analizado. A pesar de las complicaciones, se obtuvieron resultados modestos pero prometedores. Las tasas de clasificación conseguidas, aunque lejos de alcanzar porcentajes que permitieran asegurar un diagnóstico preciso, demostraron que era posible obtener características de la imagen que mantuvieran cierta correlación con la presencia o no de células cancerosas en el tejido analizado, lo cual era algo que un experto humano podía hacer con dificultad.

Lejos de desalentarnos por los resultados obtenidos y convencidos de que las técnicas empleadas habían sido las adecuadas, decidimos abordar otra tarea similar aplicando esas mismas técnicas. Es así como se abrió la línea de investigación del diagnóstico de cáncer de mama a partir de imágenes de radiografía. Uno de los motivos de querer abordar esta tarea era el de tratar de confirmar la hipótesis de que, lo limitado de los resultados obtenidos en la tarea anterior, se debían principalmente a la dificultad de la misma y no a la metodología empleada. Para poder confirmar este extremo se tuvo cuidado de trabajar con imágenes obtenidas de bases de datos públicas con las que se hubiera publicado algún trabajo, y así poder contrastar nuestros resultados.

El problema del diagnóstico de cáncer de mama no entraña las grandes dificultades que tiene el diagnóstico de cáncer de próstata. Por un lado las imágenes de mamografía obtenidas mediante rayos X ofrecen una nitidez mucho mayor de lo que se consigue mediante ultrasonidos. Por otro lado, los patrones de un tumor de mama que aparecen en una mamografía son más conocidos que los de un tumor de próstata observados en una ecografía. A pesar de ello, entre un 10 % y un 25 % de los tumores incipientes pasan inadvertidos a los ojos del radiólogo responsable de analizar la imagen. Ello confirma que el análisis de este tipo de imágenes también entraña ciertas dificultades y pone de manifiesto la necesidad de investigar en nuevas técnicas que ayuden a realizar un diagnóstico más preciso.

1.2. Objetivos generales

El objetivo último de un sistema de CAD es servir de ayuda al experto humano para mejorar su rendimiento diagnóstico. En la Figura 1.1 se muestra de manera muy simplificada esta idea. Habitualmente, el experto humano trata de elaborar un diagnóstico a partir, entre otros tests, de una imagen médica que, dependiendo de las circunstancias, puede ser una radiografía, ecografía, resonancia magnética nuclear (RMN) o tomografía axial computerizada (TAC), por citar algunos ejemplos. Este tipo de imágenes, sin embargo, no son fáciles de interpretar. Aspectos como la escasa calidad de la imagen, ocultación de objetos, lesiones de reducido tamaño,

la falta de concentración o la subjetividad en la interpretación, provoca que el experto humano encargado de analizar la imagen no sea capaz de detectar en ellas un porcentaje importante de anomalías.

Un modo habitual de evaluar la precisión diagnóstica consiste en medir los valores de *sensibilidad* y *especificidad* (Sección 4.8). La sensibilidad indica la capacidad que se posee para detectar la enfermedad, mientras que la especificidad mide la capacidad de detectar correctamente la ausencia de enfermedad. Normalmente la mejora de uno de estos dos parámetros lleva parejo el empeoramiento del otro (es fácil conseguir un aumento en la sensibilidad diagnosticando la enfermedad ante el más mínimo indicio de que esta pueda existir, sin embargo ello provoca que baje la especificidad).

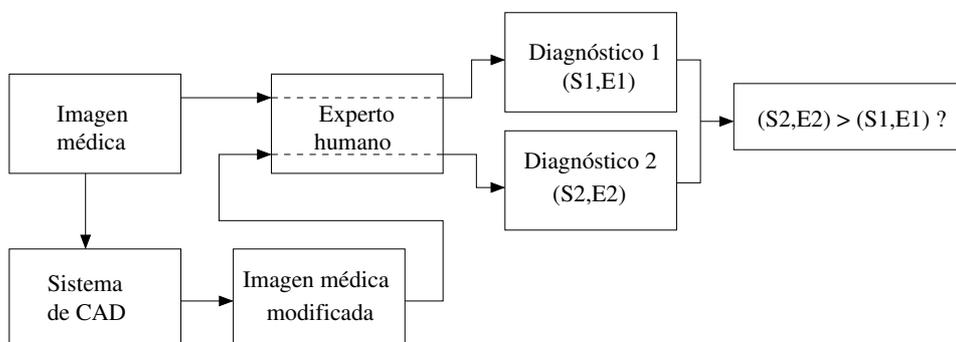


Figura 1.1: Diagnóstico con y sin la ayuda de un sistema de CAD. El objetivo final de un sistema de CAD es mejorar el rendimiento diagnóstico del experto humano encargado de interpretar la imagen médica, lo cual puede medirse mediante los valores de sensibilidad (S) y especificidad (E) conseguidos.

Cuando se cuenta con la ayuda de un sistema de diagnóstico asistido por ordenador (CAD), la imagen médica original es transformada en una segunda imagen en la que, mediante algún tipo de marcado, se resaltan las zonas sospechosas de contener alguna anomalía. Si el experto humano trabaja con esta imagen modificada, posiblemente elabore un diagnóstico distinto del que haría con la imagen original, y por tanto los valores de sensibilidad y especificidad en este caso (S_2, E_2), serán distintos a los que se lograrían sin el uso del sistema de CAD (S_1, E_1). Lógicamente, el sistema de CAD será de utilidad cuando los valores (S_2, E_2) sean mayores que (S_1, E_1) o, más concretamente, cuando para valores iguales de uno de los parámetros (S o E) el otro aumenta al utilizar el sistema de CAD.

El diagnóstico de una enfermedad no deja de ser un proceso de clasificación. Haciendo un abuso de las comparaciones, podría decirse que el uso de un sistema de CAD es equivalente al concepto ampliamente utilizado en la teoría de la decisión de *clasificadores combinados*. En este caso, el conocimiento aportado por el sistema de CAD se combina de algún modo con el propio conocimiento del experto humano para elaborar un diagnóstico final. La utilidad del sistema de CAD depen-

derá, en gran medida, de la capacidad que tenga éste de encontrar características en la imagen relacionadas con la enfermedad (o con la ausencia de la misma) que el experto humano no sea capaz de detectar con cierta facilidad.

Actualmente, atendiendo a los trabajos más recientes publicados en materia de CAD, el objetivo final de conseguir aumentar la precisión diagnóstica puede resultar ambicioso en algunas tareas. En el caso del CAD de cáncer de próstata no existe hoy en día ningún sistema en explotación de ayuda al diagnóstico, y ni siquiera se ha podido llegar a elaborar un prototipo con el que se demuestre claramente una mejora del rendimiento diagnóstico. En el caso del CAD de mama son muy pocos los sistemas actualmente en explotación y, a fecha de hoy, sólo con de uno de ellos se ha podido demostrar una mejora en el diagnóstico [Bird 05]. A pesar de ello, los esfuerzos de investigación en esta línea siguen siendo intensos, bajo el convencimiento de que el desarrollo de nuevos algoritmos y los avances en el campo de la computación, conducirán al desarrollo de sistemas con mayor capacidad diagnóstica. Por otro lado, los sistemas de CAD también son útiles como sistemas de autoevaluación y reciclaje o como herramienta de aprendizaje para médicos residentes en periodo de formación.

Para poder alcanzar el objetivo final de mejora de la precisión diagnóstica, deben marcarse unos hitos parciales que conduzcan a este fin. Hasta ahora se ha presentado el sistema de CAD como una “caja negra” que procesa cierta información de entrada, generalmente una imagen médica, y la transforma de algún modo para llamar la atención del experto humano hacia zonas sospechosas. En la Figura 1.2 se puede ver con más detalle las etapas principales que involucra el análisis y transformación de la imagen y que coinciden, en gran medida, con las etapas de cualquier sistema de reconocimiento de formas. Cada una de estas etapas juega un papel importante en el resultado final. Especialmente, la estrategia utilizada para la obtención de características influye de manera notable en el resultado final. Para acercarse al objetivo final de todo sistema de CAD es, por tanto, objetivo intermedio mejorar en la medida de lo posible cada una de las etapas involucradas en el proceso.

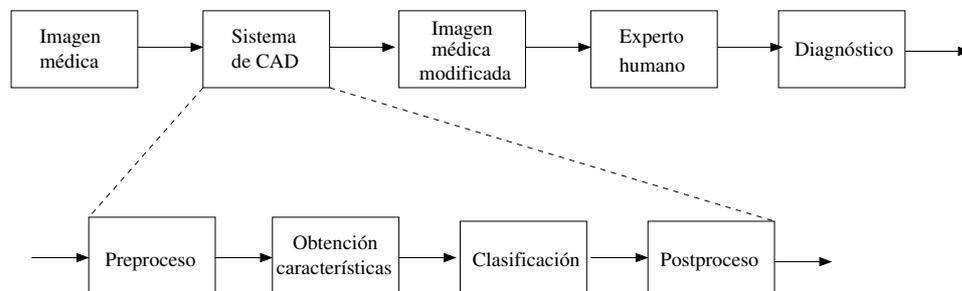
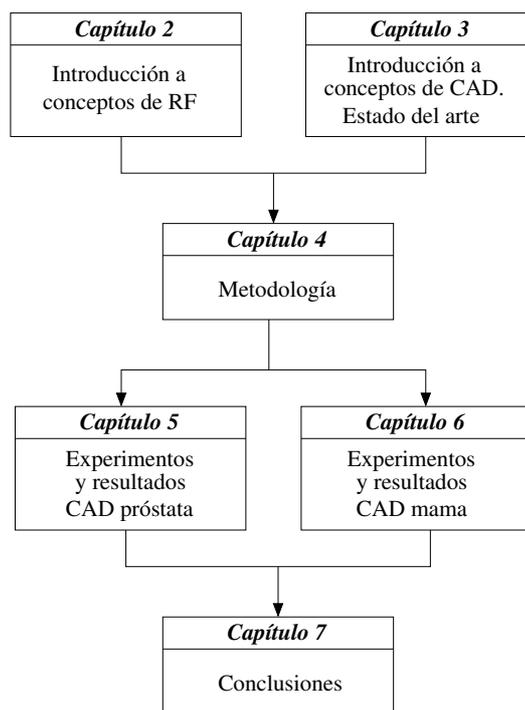


Figura 1.2: Etapas principales de un sistema de CAD.

1.3. Estructura de la tesis

El trabajo presentado en esta memoria se ha estructurado en 7 capítulos, incluido éste. El contenido de los 6 capítulos restantes se resume en el siguiente esquema:



- En el **Capítulo 2** se da una introducción general a conceptos importantes de reconocimiento de formas, los cuales sirven de base para presentar, en el Capítulo 4, las técnicas empleadas en este trabajo.
- El **Capítulo 3** comienza con una introducción al concepto de diagnóstico asistido por ordenador (CAD), para más adelante centrarse en las dos líneas de CAD estudiadas en esta tesis: ayuda al diagnóstico de cáncer de próstata y ayuda al diagnóstico de cáncer de mama. En ambos casos se presentan los métodos utilizados actualmente en el diagnóstico de este tipo de tumores y las limitaciones de los mismos. Finalmente se explica cómo un sistema de CAD puede mejorar el diagnóstico en estos casos y se presenta el estado del arte en CAD, tanto de próstata como de mama, mediante un resumen de los trabajos publicados más relevantes.
- En el **Capítulo 4** se exponen las técnicas empleadas para implementar los sistemas de ayuda al diagnóstico de próstata y de mama presentados en esta tesis. Concretamente se explican las distintas aproximaciones empleadas

para la extracción de características, el análisis de componentes principales (PCA), el problema de multiescalado, la clasificación mediante k -vecinos más cercanos (k -NN) con búsqueda rápida aproximada y mediante modelos ocultos de Markov (HMM), así como el análisis de resultados mediante curvas ROC y *free*-ROC (FROC).

- En el **Capítulo 5** se presenta el sistema de ayuda al diagnóstico de cáncer de próstata propuesto. Se explica el proceso de adquisición y etiquetado del corpus, se detallan los experimentos realizados y se exponen los resultados obtenidos. Finalmente se ofrecen las conclusiones.
- En el **Capítulo 6**, con una estructura muy similar al capítulo anterior, se presenta el sistema de ayuda al diagnóstico de cáncer de mama propuesto. Se explican las diversas baterías de experimentos realizados, se exponen los resultados obtenidos y se comparan con otros resultados previamente publicados. Finalmente se exponen las conclusiones.
- Finalmente, en el **Capítulo 7** se recogen las conclusiones finales de todo el trabajo realizado y se exponen las líneas de trabajo que quedan abiertas a futuras investigaciones.
- Adicionalmente, se han incluido dos apéndices en los que, respectivamente, se detallan las ecuaciones empleadas en el cálculo de descriptores de textura a partir de matrices de coocurrencia y se muestran imágenes de ejemplo de los corpora utilizados en este trabajo.

Capítulo 2

Conceptos de reconocimiento de formas

En este capítulo se hace un repaso de conceptos básicos de reconocimiento de formas (RF), haciendo mayor hincapié en los aspectos más directamente relacionadas con el trabajo realizado. Se presentan las dos aproximaciones generales en las que se pueden agrupar las distintas técnicas empleadas en reconocimiento de formas: la aproximación estadística o geométrica, y la aproximación sintáctica o estructural. Se aborda la regla de Bayes y se introducen dos clasificadores de distinta naturaleza que serán utilizados en este trabajo. Uno de ellos no paramétrico y basado en distancias, y otro paramétrico basado en modelos de Markov.

2.1. Introducción al reconocimiento de formas

Las aproximaciones generales al reconocimiento de formas se pueden clasificar desde dos enfoques distintos: aproximación deductiva *versus* aproximación inductiva y aproximación estructural o sintáctica *versus* aproximación geométrica o estadística.

Aproximación deductiva. La aproximación deductiva intenta abordar racionalmente el problema planteado, esto es, comprender su naturaleza y buscar la forma de resolverlo a partir de unas ideas lógicas que lo justifiquen, o bien intentar comprender la sistemática usada por los humanos en el reconocimiento de formas (imagen y sonido principalmente) para poder aplicarla al reconocimiento automático. Puesto que se basa en mecanismos de deducción bien establecidos, el conocimiento deductivo puede codificarse mediante la lógica matemática.

Aproximación inductiva. La aproximación inductiva se aplica en aquellos problemas en los que no se encuentra una explicación satisfactoria sobre los pasos

o mecanismos que conducen a la resolución de los mismos. Típicamente se trata de problemas en los que el ser humano es capaz de reconocer ciertas formas, aunque desconoce cómo lo hace. Esta metodología requiere del uso de un conjunto de aprendizaje, compuesto de una muestra de observaciones de objetos o *prototipos*, que son utilizados para crear un modelo. El clasificador usará este modelo en una etapa posterior para identificar (clasificar) objetos no vistos durante la etapa de aprendizaje. La aproximación inductiva lleva asociada implícitamente el concepto de aprendizaje. Dicho aprendizaje puede ser supervisado o no supervisado. El aprendizaje no supervisado se aplica en aquellos problemas en los que no hay clases definidas *a priori*, bien porque no es posible definir las, bien porque no se saben etiquetar los prototipos del conjunto de entrenamiento. En estos casos, el algoritmo de entrenamiento del clasificador agrupará los prototipos en un determinado número de clases o categorías basándose en las características de los mismos. En el aprendizaje supervisado, en cambio, se parte de un conjunto de entrenamiento etiquetado con c clases predeterminadas; el clasificador, una vez entrenado y ante una observación de test no vista previamente, le asignará una de entre las c clases. En el presente trabajo se ha empleado el aprendizaje supervisado.

El segundo enfoque agrupa las distintas metodologías en dos aproximaciones: la **estructural o sintáctica** y la **geométrica o estadística**. Cada una de ellas presenta sus ventajas e inconvenientes. En las siguientes secciones se abordan ambas aproximaciones.

2.1.1. Aproximación estadística

La aproximación estadística, también llamada geométrica, se basa en una teoría clásica y robusta como es la teoría de la decisión. Un objeto se representa en forma de d características o medidas y se trata como un punto en un espacio vectorial d -dimensional. El significado de las características viene dado, consecuentemente, por la posición de los vectores (puntos) en el espacio. El objetivo es escoger aquellas características que permiten distribuir los vectores de manera que las distintas clases o categorías ocupen regiones lo más compactas y disjuntas posible. La efectividad del espacio de representación es mayor cuanto más clara sea la separación entre las clases. De este modo, dado un conjunto de patrones de entrenamiento de cada clase, el objetivo es establecer unas fronteras de decisión en el espacio de características que permitan separar los patrones pertenecientes a las distintas clases. Los métodos estadísticos están basados en la teoría de la decisión estadística, según la cual, las fronteras de decisión son determinadas por las distribuciones de probabilidad de los patrones que pertenecen a cada clase. Dichas distribuciones pueden ser estimadas a partir de metodologías ampliamente conocidas [Devr 96, Duda 73].

Este enfoque es adecuado para resolver problemas en los que los datos puedan representarse de forma numérica sencilla y, más específicamente, si su representación en un espacio vectorial resulta natural. Básicamente, la aplicación de técnicas

estadísticas consiste en determinar la clasificación de un objeto en una determinada clase o región en función de su posición en el espacio.

Dado que la aproximación estadística ha sido la utilizada en el presente trabajo, será estudiada con mayor profundidad en apartados posteriores.

2.1.2. Aproximación estructural

En muchos problemas de reconocimiento de formas, los patrones que se presentan contienen una riqueza estructural importante, susceptible de ser modelada. Los métodos estructurales aprovechan este tipo de información y ofrecen una perspectiva jerárquica de estas estructuras. De este modo, un patrón es visto como una composición de subpatrones más simples, los cuales están constituidos, a su vez, por otros subpatrones [Fu 82]. Los patrones más elementales se denominan primitivas. El patrón o forma original se representa en términos de interrelaciones entre estas primitivas. Partiendo de este punto de vista, puede establecerse una analogía entre la estructura de los patrones y la sintaxis de un lenguaje: los patrones son vistos como frases de un lenguaje y las primitivas como alfabeto de este lenguaje. En este caso, los métodos de clasificación se denominan sintácticos. En los métodos sintácticos, las frases son generadas a partir de una gramática que tendrá que ser inferida para cada patrón a partir de la muestra de entrenamiento. Una frase de test será clasificada según la aceptación de una determinada gramática o directamente por traducción.

Los objetos del mundo real pueden ser codificados de diversas maneras, como cadenas, grafos, etc. El formato de las características, por tanto, goza de una notable libertad y su significado será definido por la codificación elegida. Normalmente, mediante un número pequeño de primitivas y reglas gramaticales, pueden ser descritas una gran cantidad de formas o patrones de cierta complejidad estructural. Los lenguajes regulares han sido, tanto por su sencillez como por su capacidad de describir clases de objetos, los más estudiados en reconocimiento de formas.

La implementación de una aproximación sintáctica comporta algunas dificultades que, básicamente, tienen relación con la detección de primitivas en patrones con ruido, así como con la inferencia de las gramáticas a partir de un conjunto de entrenamiento. Para dotarla de mayor potencia, la aproximación sintáctica puede ser complementada con formulación estocástica. En [Fu 82] se introduce esta idea bajo el concepto de gramáticas con atributos, las cuales unifican el reconocimiento de formas estructural y estadístico.

2.1.3. Esquema general de un sistema de RF inductivo supervisado

Un sistema de reconocimiento opera funcionalmente en dos modos: entrenamiento o aprendizaje y test o clasificación. En la Figura 2.1 se muestra un modelo

genérico de sistema de reconocimiento de formas. Cada una de las etapas mostradas tendrá unas características determinadas en función del sistema de reconocimiento que se desee implementar. En el Capítulo 4 se describe con detalle las etapas de preprocesamiento, obtención de características y clasificación, orientadas a un sistema de ayuda al diagnóstico de cáncer a partir de imágenes médicas y se exponen distintas aproximaciones para cada una de ellas. A continuación se describe, de modo más genérico, las funciones básicas de cada una de estas etapas.

Módulo de adquisición

En esta primera etapa se mide un objeto del mundo real mediante los sensores físicos apropiados, por ejemplo, una cámara de vídeo para la captura de imágenes o un micrófono para registrar la voz. En muchos casos esta información se presenta en forma de señal analógica, la cual puede ser preprocesada para mejorar su calidad o para aislar la parte útil de la misma. Una posterior digitalización de la señal permite que ésta pueda ser tratada por un computador, bien para almacenarla en algún tipo de soporte permanente, bien para utilizarla directamente como entrada en la etapa posterior. Se dice que la señal obtenida pertenece al espacio de representación primario.

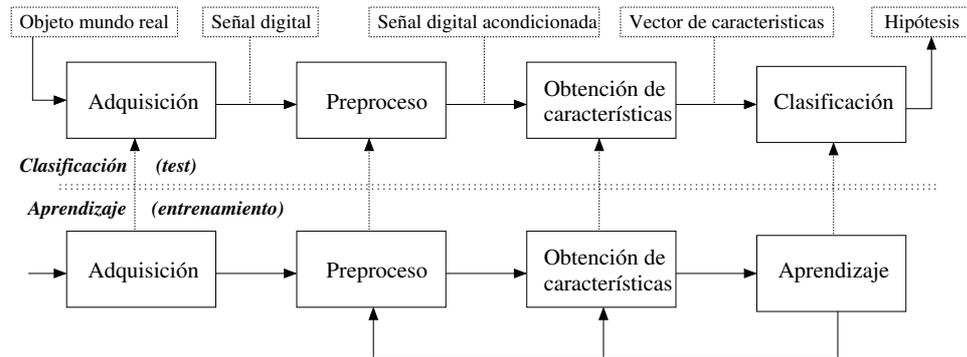


Figura 2.1: Modelo de sistema de reconocimiento de formas.

Módulo de preprocesamiento

La función del módulo de preprocesamiento, también llamado de tratamiento de la señal, es, básicamente, realzar y acondicionar el patrón de interés respecto del resto de la señal. En el caso particular de que la señal de entrada represente una imagen (tal es el caso del presente trabajo), serán operaciones habituales de esta etapa el escalado, cuantificación, filtrado, modificaciones del histograma, transformaciones geométricas, operaciones morfológicas, eliminación de ruido y segmentación. Además, cuando el módulo de preprocesamiento opera en modo de entrenamiento, se realizará el etiquetado del patrón para permitir su posterior identificación. Buena parte de estas operaciones son materia de un campo de trabajo particular como es el del tratamiento digital de imágenes, sobre el que existe extensa literatura [Jain 89, Gonz 99].

Módulo de obtención de características

Las características de un objeto son el resultado de aplicar ciertas funciones sobre las medidas obtenidas del objeto (señal digital acondicionada), que ayudan de alguna manera a distinguir la clase a la cual pertenece y que serán utilizadas como entrada al clasificador. La obtención de características incluye los procesos de extracción, selección y reducción de características. El objetivo de esta etapa es encontrar el conjunto de características que mejor representa el objeto de entrada, esto es, que minimice las diferencias intraclase y maximice las diferencias interclase. Cuando se trabaja en modo entrenamiento, es habitual realizar diversas iteraciones sobre este módulo para tratar de optimizar las prestaciones del sistema. En la fase de test se aplicarán los mismos métodos de extracción y/o selección escogidos durante el entrenamiento. La obtención de características es la parte menos sistemática del proceso y tiene una gran dependencia del tipo de tarea a realizar. Normalmente, la experiencia y la intuición del diseñador del sistema será de gran relevancia a la hora de resolver con éxito esta etapa.

Se han considerado tres tareas en el proceso de obtención de características: la extracción, que hace referencia a la transformación de la señal del dominio primario a un dominio más adecuado para tratar el problema; la selección, con la que se escogen aquellas características que aportan información valiosa para la identificación del objeto, desechando la información no discriminante o redundante; y la reducción del número de características, que contribuirá a reducir el coste computacional.

Módulo de aprendizaje

El clasificador se entrena tratando de particionar el espacio de características generado por el módulo anterior, con el objetivo de minimizar el error de clasificación entre los prototipos del conjunto de evaluación. La retroalimentación hacia los módulos anteriores permite al diseñador del sistema escoger, entre las distintas técnicas probadas, aquella que mejor rendimiento proporcione, atendiendo a los resultados obtenidos al clasificar el conjunto de evaluación.

Módulo de clasificación

En la fase de test, el clasificador toma el vector de características extraído del patrón de entrada y le asigna la clase con la que se encuentra una mayor similitud, de acuerdo al modelo inducido durante el entrenamiento. En ocasiones la salida del clasificador no se limita a la clase más similar y puede ofrecer múltiples hipótesis en forma de lista de n -tuplas conteniendo el identificador del objeto, una etiqueta de clase (hipótesis) y el nivel de fiabilidad asignado a la hipótesis.

2.2. Aproximación estadística al reconocimiento de formas

En el reconocimiento estadístico de formas, una forma o patrón se representa por un conjunto de d características o atributos en forma de vector d -dimensional. Además, se espera que los patrones de una misma clase se agrupen en zonas del espacio determinadas por las densidades dependientes de cada clase (denominadas *densidades condicionales de clase*) obtenidas a partir de un conjunto de muestras de entrenamiento. Ello posibilita la existencia de ciertas metodologías estadísticas que permiten la clasificación de nuevos patrones, lo que constituye la finalidad de cualquier sistema de reconocimiento.

Existen diversas estrategias a la hora de diseñar un clasificador estadístico, en función del tipo de información disponible sobre las densidades condicionales de clase. Si se dispone de densidades condicionales de clase completamente conocidas, se adopta la regla de decisión de Bayes óptima. Esta situación, sin embargo, no es habitual, por lo que las densidades deberán ser aprendidas a partir de un conjunto de muestras de entrenamiento. Si suponemos que las densidades condicionales de clase se ajustan a una forma determinada (por ejemplo, una o varias gaussianas multivariadas) pero algunos de sus parámetros son desconocidos (por ejemplo medias y/o matrices de covarianza), nos encontramos ante una aproximación paramétrica. En este contexto, la estrategia habitual para solucionar el problema consiste en reemplazar los parámetros desconocidos por sus valores estimados, lo que se conoce como clasificador empotrado (*plug-in*) de Bayes. Si no se presupone una forma determinada de las densidades condicionales de clase, nos encontramos ante una aproximación no paramétrica. En este caso, debe estimarse la función de densidad o bien construir directamente las fronteras de decisión a partir de las muestras de entrenamiento.

Otra dicotomía que se presenta en el reconocimiento estadístico es la que se refiere a los modelos de aprendizaje supervisado y no supervisado. La diferencia entre estos es que en el caso del aprendizaje supervisado las muestras de entrenamiento están etiquetadas (esto es, la clase a la que pertenecen es conocida), mientras que la etiqueta de clase (y en ocasiones el número de éstas) es desconocida en el caso no supervisado.

En la Tabla 2.1 se muestran las distintas aproximaciones al reconocimiento estadístico. A medida que nos desplazamos hacia abajo, la información disponible disminuye y, por tanto, aumenta la dificultad de la clasificación. En cierta manera, puede decirse que la mayoría de las aproximaciones intentan implementar la regla de decisión de Bayes. En las siguientes secciones se explica con detalle esta regla de decisión y se exponen brevemente los métodos paramétricos y no paramétricos

Densidades condicionales de clase	Conocidas	Teoría de la decisión de Bayes		
	Desconocidas	Aprendizaje supervisado	Paramétrico	Reglas óptimas
				Reglas <i>plug-in</i>
		No paramétrico	Estimación de densidades	
			Construcción de fronteras de decisión	
	Aprendizaje no supervisado	Paramétrico	Resolución de mixturas	
No paramétrico		Análisis de agrupamiento		

Tabla 2.1: Aproximaciones al reconocimiento estadístico de formas.

2.2.1. Clasificación probabilística: la regla de Bayes

La etapa de clasificación ha sido tradicionalmente el marco predominante de los estudios de reconocimiento de formas. En el conjunto de objetos a reconocer se establece una partición de manera que cada objeto pertenezca a una sola clase. De este modo, ante un nuevo objeto, el objetivo consistirá en determinar, a partir de su representación, la clase a la que pertenece.

El proceso de toma de decisión en el reconocimiento de formas estadístico se puede resumir de la siguiente manera: dado un patrón, éste ha de ser asignado (clasificado) en una de las c categorías w_1, w_2, \dots, w_c basándose en el vector de d valores de características $\mathbf{x} = x_1, x_2, \dots, x_d$. Se asume que las características tienen una *masa* o *densidad de probabilidad* (dependiendo de si son discretas o continuas), en forma de función condicionada a la clase, conocida como *densidad condicional de clase*. De este modo, un vector \mathbf{x} de un patrón que pertenezca a la clase w es visto como una observación o punto en el espacio d -dimensional escogido aleatoriamente de la función de probabilidad condicional de la clase $p(\mathbf{x}|w_i)$.

Consideremos, en primer lugar, que no disponemos de información de las características de la muestra a clasificar. La probabilidad de que una muestra cualquiera, de la cual no tenemos ningún tipo de información, pertenezca a la clase w_i es lo que conocemos como *probabilidad a priori* $P(w_i)$. Este valor está asociado a cada clase y no depende de la muestra en concreto. Es trivial estimar esta probabilidad por simple conteo, aunque para lograr una buena estimación será necesario disponer de un número suficiente de muestras correctamente etiquetadas, obtenidas

de un ámbito natural y de manera uniforme y aleatoria. En el caso de que no fuera posible tener ningún tipo de información sobre el objeto a clasificar, la decisión más razonable (aquella que minimiza el riesgo de error) sería asignarle la clase con probabilidad a priori más alta:

$$P(w_i) > P(w_j), \quad 1 \leq i, j \leq c, \quad i \neq j \Rightarrow \mathbf{x} \in w_i$$

Esta regla de decisión ciega es, evidentemente, poco útil y supone no hacer uso de ninguna característica del objeto que pueda contribuir a una clasificación más fiable. La información que aportan las características del objeto debería emplearse para maximizar la fiabilidad de la clasificación. En este sentido, resultaría muy útil disponer de un mecanismo que nos proporcionara, a partir de estas observaciones, información sobre la *probabilidad a posteriori* $P(w_i|\mathbf{x})$. La manera de obtener esta probabilidad es la característica diferenciadora de los distintos métodos de clasificación.

Sea $p(\mathbf{x}|w_i)$ la función de *densidad de probabilidad condicional* de \mathbf{x} para una determinada clase w_i . Supongamos que conocemos las probabilidades a priori $P(w_i)$ y las densidades condicionales $p(\mathbf{x}|w_i)$. Con esta información, la regla de Bayes permite calcular la probabilidad a posteriori como

$$P(w_i|\mathbf{x}) = \frac{p(\mathbf{x}|w_i)P(w_i)}{p(\mathbf{x})} \quad (2.1)$$

donde

$$p(\mathbf{x}) = \sum_{i=1}^c p(\mathbf{x}|w_i)P(w_i)$$

Una vez obtenida la probabilidad a posteriori de cada clase, se escogerá aquella clase que presente el mayor valor:

$$P(w_i|\mathbf{x}) > P(w_j|\mathbf{x}), \quad 1 \leq i, j \leq c, \quad i \neq j \Rightarrow \mathbf{x} \in w_i$$

y, teniendo en cuenta que en la ecuación 2.1 el denominador no depende de la clase, la regla de decisión de Bayes puede escribirse como

$$p(\mathbf{x}|w_i)P(w_i) > p(\mathbf{x}|w_j)P(w_j), \quad 1 \leq i, j \leq c, \quad i \neq j \Rightarrow \mathbf{x} \in w_i$$

Este criterio constituye la regla de decisión de Bayes de mínimo error, en la cual se basan, de algún modo, la práctica totalidad de los métodos de clasificación

estadísticos. Habitualmente la probabilidad a posteriori no se podrá calcular directamente, ya que no se conocerá con certeza la densidad de probabilidad condicional de cada clase $p(\mathbf{x}|w_i)$. Para ello será necesario estimar dichas probabilidades mediante un proceso de entrenamiento a partir de un conjunto de muestras.

Es posible que en una aplicación concreta el coste (o riesgo) en el que se incurre al clasificar incorrectamente una muestra en la clase w_i cuando en realidad pertenecía a la clase w_j no sea el mismo para todas las combinaciones de i y j . Por ejemplo, clasificar erróneamente una imagen que presenta tumor como sana tiene un riesgo mucho mayor que el caso contrario. Este concepto se puede formalizar en términos de una *función de coste* $L(w_i, w_j)$, también denominada *función de pérdida*, que cuantifica el coste de clasificar erróneamente un objeto de la clase w_i como perteneciente a la clase w_j . En estos casos resulta útil una extensión de la regla anterior que hace uso de esta información adicional y que, para el caso de dos clases, se conoce con el nombre de *regla de decisión de Bayes de mínimo riesgo*, la cual asigna el patrón x a la clase w_i para la cual el riesgo condicional sea mínimo:

$$R(w_i|\mathbf{x}) = \sum_{j=1}^c L(w_i, w_j)P(w_j|\mathbf{x})$$

2.2.2. Métodos de estimación paramétrica y no paramétrica

Como se ha dicho anteriormente, normalmente será necesario emplear alguna técnica que permita estimar las funciones de densidad condicional. Se puede distinguir entre aquellos procedimientos que no hacen uso de ningún tipo de suposición respecto de la naturaleza de las funciones de densidad a estimar y aquellos que asumen que estas densidades responden a una determinada distribución especificada paramétricamente. Los primeros se denominan métodos no paramétricos y los últimos métodos paramétricos.

En los métodos paramétricos el problema se reduce a estimar, a partir del conjunto de entrenamiento, los parámetros de la distribución, supuestamente conocida, que siguen las muestras de cada clase. En reconocimiento de formas se utilizan habitualmente dos métodos de estimación de los parámetros basados en distintos criterios de bondad de la estimación: estimación por *máxima verosimilitud* y estimación *bayesiana* [Duda 73]. El primero de ellos ve los parámetros como cantidades cuyos valores son fijos pero desconocidos. La mejor estimación en este caso es la que maximiza la probabilidad de obtener las muestras observadas en el conjunto de entrenamiento. Los métodos bayesianos, por contra, ven los parámetros como variables aleatorias con una distribución *a priori* conocida. La observación de las muestras de entrenamiento van modificando los valores de los parámetros iniciales.

En las aproximaciones no paramétricas, la función de densidad de cada clase (la cual da una idea de la probabilidad que en un punto del espacio de representa-

ción aparezca una muestra de esta clase) se estima localmente, esto es, llevando la cuenta de cuántos prototipos aparecen en cada región local del espacio.

En las siguientes secciones se exponen dos reglas de clasificación no paramétricas basadas en distancias y un clasificador paramétrico basado en modelos de Markov.

2.3. Clasificadores basados en distancias

Existen dos reglas ampliamente conocidas para la estimación no paramétrica de densidades mediante métodos de vecindad: las *ventanas de Parzen* y los *k-vecinos más cercanos*. A continuación se exponen estos dos métodos.

2.3.1. Ventanas de Parzen

En la aproximación de Parzen [Parz 62] se define una región \mathbb{R}_n como un hipercubo d -dimensional. Si h es la longitud de un lado del hipercubo, entonces su volumen¹ viene dado por:

$$V = h^d$$

Si se define un hipercubo unidad centrado en el origen, se puede calcular el número de muestras que quedan dentro del hipercubo con la ayuda de la siguiente *función ventana*:

$$K(z) = \begin{cases} 1 & \text{si } |z_j| \leq 1/2 \quad j = 1, \dots, d \\ 0 & \text{si no} \end{cases} \quad (2.2)$$

Dado un hipercubo de volumen V centrado en un punto x del espacio, cualquier punto del intervalo d -dimensional $(x - h/2, x + h/2)$ caerá dentro del hipercubo. De este modo, dado un número n de puntos distribuidos en el espacio y teniendo en cuenta que para los puntos x_i comprendidos en el volumen del hipercubo el valor $K((x - x_i)/h)$ será 1 y 0 para los que caen fuera, se puede calcular el número de puntos k que caen dentro del hipercubo mediante la expresión

$$k = \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right)$$

¹A pesar de que habitualmente se utiliza el término *volumen*, en el caso de espacios de más de 3 dimensiones sería más correcto hablar de *contenido*.

Ello permite hacer la siguiente estimación de la función de densidad condicional para una clase w_m

$$\hat{p}(x|w_m) = \frac{1}{n_m h^d} \sum_{i=1}^{n_m} K\left(\frac{x - x_i}{h}\right)$$

donde x_i , $i = 1, \dots, n_m$ son los puntos de la clase w_m y n_m es la talla del conjunto.

Debe notarse que si h es muy pequeña, el estimador degenerará en una colección de n_m picos agudos localizados en cada punto de la muestra. Por contra, si h es demasiado grande el estimador estará sobreesuavizado, y tendrá una función de densidad con una varianza muy pequeña. Parece lógico, además, que h debería estar en función de la talla de la muestra n_m ; para una muestra pequeña habría que forzar un suavizado mayor, esto es, el valor de h debería ser más grande para evitar los picos. Por contra, con una muestra grande se debería emplear una h más pequeña para evitar un suavizado excesivo.

La función ventana K también se conoce como función de núcleo o *kernel*. En el caso de definirse mediante la ecuación 2.2 se denomina *núcleo rectangular*. Dado que K puede resultar artificial en el sentido de que puntos muy próximos a los límites del hipercubo pueden contribuir con todo o nada a la estimación de la densidad, puede resultar conveniente emplear otro tipo de funciones que ponderen el peso de cada punto de acuerdo con su distancia a x . Un ejemplo puede ser la función normal

$$K(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

De acuerdo con lo que se ha visto, para estimar el valor de la función de densidad de probabilidad de una clase w_m en un punto x , se construye una región $L(x)$ alrededor de x . La masa de probabilidad de $L(x)$ se puede aproximar por $p(x|w_m)V$, donde V es el volumen de $L(x)$. De este modo, considerando una función de núcleo de tipo 0/1, si contamos el número de prototipos k_m que se encuentran en $L(x)$, de un total de n_m prototipos de la clase, obtenemos:

$$p(x|w_m)V = \frac{k_m}{n_m}$$

En la práctica, la magnitud que generalmente se define no es V sino h , de modo que, según una determinada métrica, se obtiene la región $L(x)$ alrededor de x . La elección de la h óptima para un problema determinado no es sencilla. En [Fuku 72] aparece una expresión que calcula el valor óptimo de h para un conjunto de muestras distribuidas normalmente. El valor obtenido crece exponencialmente con el número de dimensiones del espacio de características.

2.3.2. El clasificador de los k -vecinos más cercanos

En la aproximación de Parzen se fija V y se obtiene el número de prototipos k que quedan dentro del volumen por conteo. La regla de los k -vecinos más cercanos (k -NN) explora otra posibilidad, que es fijar k y obtener V de acuerdo al espacio ocupado por los k prototipos más cercanos a x . De este modo, $L(x)$ depende, para cada muestra, de la distribución del conjunto de prototipos, siendo menor en zonas de mayor densidad y mayor en zonas con menor número de prototipos. Podría decirse que se trata de una variante del método de ventanas de Parzen, en la que el tamaño de la hiperesfera es variable y ajustada automáticamente, según la densidad de prototipos. La función de núcleo asociada sería, en principio, uniforme, aunque en ocasiones se emplean funciones que dependen de la distancia. A continuación se hace un análisis formal de la regla de los k -vecinos más cercanos.

Para una función de densidad de probabilidad multivariada $p(x|w_m)$ de la clase w_m , la probabilidad de que un punto cualquiera x' caiga en la región local $L(x)$ es

$$\theta = \int_L p(x'|w_m) dx'$$

Si L es pequeña y V es su volumen, se puede hacer la siguiente aproximación

$$\theta \simeq p(x|w_m)V$$

resultando

$$p(x|w_m) \simeq \theta/V$$

donde θ/V será una aproximación suavizada de $p(x|w_m)$, esto es, el valor medio de $p(x'|w_m)$ en la región local L alrededor de x . Si L es suficientemente pequeña, entonces $p(x'|w_m)$ no se desviará mucho de $p(x|w_m)$. De este modo, si se estima θ se puede estimar directamente $p(x|w_m)$. Por tanto θ puede estimarse simplemente viendo qué proporción de los n_m prototipos de la muestra caen dentro de L . Si k es el número de puntos de la muestra de la clase w_m que caen en L , entonces

$$\hat{\theta} = k/n_m$$

lo que lleva a una estimación de la función de densidad de probabilidad en L

$$\hat{p}(x|w_m) = \frac{k}{n_m V} \quad (2.3)$$

Llegado a este punto se puede proceder de dos modos. Si se fija el volumen V y se contabiliza la cantidad de puntos que han caído dentro de L , estamos ante la aproximación de ventanas de Parzen explicada en la Sección 2.3.1. Esto provoca los inconvenientes ya mencionados de tener que estimar el parámetro h (radio del volumen V) y el hecho de que adoptar un valor de h igual para cualquier zona del espacio provocará, probablemente, una mala estimación de la función de densidad, con picos en regiones con baja densidad de prototipos y casi uniforme en zonas de alta densidad. Alternativamente se puede establecer V en función de los datos y, por tanto, variable según la zona del espacio en la que se defina. Esto puede conseguirse fijando k y determinando el volumen V necesario para abarcar los k puntos de la clase más cercanos a x . Si la densidad es alta, V será relativamente pequeño, mientras que en zonas de menor densidad, el volumen crecerá. El uso de este estimador se conoce como *método volumétrico de los k -vecinos más cercanos*.

Una desventaja del método volumétrico es que el estimador obtenido no es una función de densidad de probabilidad, ya que si se integra la Ecuación 2.3 sobre todo el espacio, se observa que no da la unidad sino infinito. Sin embargo, teniendo en cuenta que en la práctica lo que se pretende es utilizar el estimador como un clasificador capaz de discriminar entre clases para nuevas observaciones, se puede extender el método de la siguiente manera: dado un punto x y una combinación de n prototipos pertenecientes a distintas clases ($\sum_m n_m = n$), se puede definir la hiperesfera de volumen V que abarque exactamente los k puntos más próximos a x . Si se supone que entre estos k puntos hay k_m de la clase w_m , entonces se puede definir un estimador para la función de densidad para la clase w_m de la forma

$$\hat{p}(x|w_m) = \frac{k_m}{n_m V}$$

lo que es ligeramente distinto de la Ecuación 2.3, ya que el número de vecinos k seleccionados es distinto y por tanto V es distinto. Por otro lado, también se tienen otros estimadores obvios como son la probabilidad a priori de la clase $\hat{P}(w_m)$ y la probabilidad $\hat{p}(x)$

$$\hat{P}(w_m) = \frac{n_m}{n}$$

$$\hat{p}(x) = \frac{k}{nV}$$

A partir de estas probabilidades, y según la regla de decisión de Bayes, se puede estimar la probabilidad a posteriori como

$$\hat{p}(w_m|x) = \frac{\hat{p}(x|w_m)\hat{P}(w_m)}{\hat{p}(x)} = \frac{\frac{k_m}{n_m V} \frac{n_m}{n}}{\frac{k}{nV}} = \frac{k_m}{k}$$

lo que lleva a la regla de clasificación conocida como *votación entre los k -vecinos más cercanos* (k -NN)

$$x \in w_i \quad \text{si} \quad k_i = \max_m(k_m)$$

esto es, se clasifica x como perteneciente a la clase más votada entre los k vecinos. En resumen, en el método de votación se obtienen los k vecinos más cercanos a x de entre todas las muestras y se cuentan el número de vecinos pertenecientes a cada clase. La clase mayoritaria será la que se asigne a x , sin necesidad de aplicar explícitamente la regla de Bayes. En caso de empate se pueden aplicar diversas estrategias. Quizás la más rigurosa sea ir reduciendo k en una unidad hasta deshacer el empate. Otra posibilidad de menor coste consiste en asignar la clase del vecino más cercano.

La elección del parámetro k no es obvia. Se ha demostrado que en problemas de dos clases, cuando $k \rightarrow \infty$ y $\frac{k}{n} \rightarrow 0$, siendo n la talla del conjunto de entrenamiento, el error de la regla k -NN converge en probabilidad al error de Bayes [Ston 77].

La regla k -NN se puede ver como un intento de estimar las probabilidades a posteriori $p(w_i|x)$ a partir de las muestras i , por tanto, es fundamental aproximar adecuadamente los parámetros que regulan su comportamiento. Como acaba de explicarse interesa, por un lado, un valor de k suficientemente alto para obtener una estimación creíble y, por otro, una muestra suficientemente grande para que todos los vecinos seleccionados se encuentren cerca de x , lo cual asegurará que $p(w_i|x') \approx p(w_i|x)$. Ello fuerza a un compromiso de modo que k sea una pequeña fracción del número de muestras. Diversos autores han tratado de formular una elección adecuada del valor de k [Devi 82, Fuku 72], si bien la mayoría de autores recomiendan hacer una estimación empírica.

El clasificador de los k -vecinos más cercanos ha dado buenos resultados en muchas tareas de reconocimiento de formas. Sin embargo, plantea la desventaja de requerir un número de prototipos muy elevado para obtener buenas prestaciones. Manejar estos conjuntos extensos de prototipos lleva asociado, inevitablemente, un coste espacial y sobretodo temporal, nada despreciable, ya que los prototipos deben almacenarse y es necesario calcular, para cada punto de test, las distancias a todos los prototipos de la muestra. Ello hace que éste sea un método lento. En este sentido, se han propuesto diversos métodos que tratan de minimizar estos inconvenientes. Por un lado las técnicas basadas en la edición y condensado [Hart 68] tratan de reducir el número de prototipos sin que ello vaya en detrimento de la estimación de las funciones de probabilidad. Por otro lado se han propuesto distintos algoritmos de búsqueda rápida, tanto exacta como aproximada, dirigidos a reducir el coste temporal. En la Sección 4.6 se explica con detalle un método de búsqueda rápida aproximada basado en árboles k -dimensionales o *kd-trees*.

2.4. Modelos ocultos de Markov

Aunque en el campo del reconocimiento de formas los modelos de Markov han sido utilizados habitualmente en problemas de reconocimiento del habla, en los últimos años se han extendido también a otros dominios, como la visión por computador. En este apartado se ofrece una introducción general a los modelos de Markov y los algoritmos utilizados para su inferencia y estimación de probabilidades. En la Sección 4.7 se verá cómo se han empleado estos modelos en una tarea de análisis de imágenes médicas como la que nos ocupa.

Un modelo oculto de Markov o “*Hidden Markov Model*” (HMM) es una máquina de estados finitos en la cual interactúan entre sí dos procesos estocásticos bien definidos: uno de ellos permanece inobservable (de ahí el calificativo de oculto) y actúa en un segundo plano por detrás del otro (observable) que produce efectivamente la secuencia de observaciones de salida [Rabi 89]. El proceso que permanece oculto involucra un conjunto de estados conectados entre sí por medio de transiciones con probabilidades. Por otro lado, el proceso observable consta de un conjunto de observaciones de salida, cada una de las cuales puede ser emitida por cualquiera de los estados según una función de probabilidad asociada a cada uno de ellos.

Dependiendo de la naturaleza de las observaciones, se distinguen distintos tipos de HMMs. Si las observaciones son de naturaleza discreta (símbolos de un alfabeto), tal y como ocurre en el ejemplo planteado, se dice que los HMMs son discretos. Si por el contrario estas observaciones son de naturaleza continua, entonces se dice que los HMMs son continuos, con una función de densidad de probabilidad de emisión generalmente aproximada por una mixtura de gaussianas.

Dado un HMM, existen tres cuestiones básicas que cabe plantearse:

- ¿Cuál es la probabilidad de la secuencia observada dado un modelo, esto es, $P(O_1, O_2, \dots, O_T | M)$?
- ¿Cuál es la secuencia de estados que con mayor probabilidad genera una salida dada?
- Dado un conjunto de observaciones de entrenamiento, ¿cómo aprende el modelo los parámetros óptimos que describan lo mejor posible dicho conjunto de observaciones?

Estas tres cuestiones pueden resolverse, respectivamente, con los algoritmos *Forward*, *Viterbi* y *Backward-Forward* o *Baum-Welch*.

En las siguientes secciones se da una definición formal de los HMMs continuos y se trata con mayor detalle cada uno de los algoritmos mencionados anteriormente. Las definiciones formales de los HMMs están expuestas en [Jeli 98, Lee 89].

Puesto que los modelos utilizados en esta tesis son de naturaleza continua, nosotros únicamente expondremos las definiciones y formulaciones de los distintos algoritmos desde el punto de vista continuo.

2.4.1. Definición de HMM continuo

A continuación se presenta una definición formal de los HMMs continuos. Para evitar definir la probabilidad que presenta cada estado de ser estado inicial o final, en esta formulación se ha optado por definir un estado inicial único en el cual no se emite ningún vector, y un estado final único con similares características.

Formalmente, un HMM (continuo) \mathcal{M} es una máquina de estados finitos definida por la séxtupla (Q, I, F, X, a, b) [Youn 97], donde:

- Q es un conjunto finito de estados, que incluye un estado inicial $I \in Q$ y un estado final $F \in Q$.
- X es un espacio real d -dimensional de observaciones: $X \subseteq \mathbb{R}^d$.
- $a : (Q - \{F\}) \times (Q - \{I\}) \rightarrow [0, 1]$ es una función de distribución de probabilidad de transición entre estados, tal que:

$$\sum_{q_j \in (Q - \{I\})} a(q_i, q_j) = 1 \quad \forall q_i \in (Q - \{F\})$$

- $b : (Q - \{I, F\}) \times X \rightarrow [0, 1]$ es una función de densidad de probabilidad de emitir un vector $\vec{x} \in X$ en un estado $q_i \in Q$, tal que:

$$\int_{\vec{x} \in X} b(q_i, \vec{x}) d\vec{x} = 1 \quad \forall q_i \in (Q - \{I, F\})$$

En la definición de HMM dada, hay implícitos dos supuestos:

1. $a(q_i, q_j) = P(z_{t+1} = q_j | z_t = q_i)$ establece² que la probabilidad de una cadena de Markov en un estado dado q_j y en un instante $t+1$, depende sólo del estado q_i de la cadena de Markov en el tiempo t , y no depende de los estados visitados anteriormente en tiempos menores que t . Es decir:

$$P(z_{t+1} | z_1 \dots z_t) = P(z_{t+1} | z_t)$$

2. $b(q_i, \vec{x}) = p(x_t = \vec{x} | z_t = q_i)$ establece³ que la probabilidad de que \vec{x} sea emitida en el tiempo t depende únicamente del estado q_i en el tiempo t , y

² $z_t = q_i$ denota que el HMM se encuentra en el estado q_i en el tiempo t .

³ $x_t = \vec{x}$ denota que el HMM en el estado z_t emite \vec{x} en el tiempo t .

no depende ni de los vectores emitidos y ni de los estados visitados anteriormente en tiempos menores que t . Es decir:

$$p(x_t | x_1 \dots x_t, z_1 \dots z_t) = p(x_t | z_t)$$

Estos dos supuestos definen lo que se denomina un *HMM* de primer orden.

2.4.2. Probabilidad de emisión

Sea $\mathbf{x} = (\vec{x}_1 \vec{x}_2 \dots \vec{x}_T)$ con $\vec{x}_i \in \mathbb{R}^d$ una secuencia finita de vectores reales y sea $Z = \{\mathbf{z} = (z_1 z_2 \dots z_T) : z_k = q_i \in (Q - \{I, F\}), 1 \leq i \leq |Q| - 2\}$ un conjunto de secuencias de derivación de estados asociadas con la secuencia de símbolos \mathbf{x} . Entonces la probabilidad de que la cadena \mathbf{x} sea generada por el modelo \mathcal{M} es:

$$P(\mathbf{x} | \mathcal{M}) = \sum_{\mathbf{z} \in Z} \left(\prod_{i=1}^T a(z_{i-1}, z_i) b(z_i, \vec{x}_i) \right) a(z_T, F) \quad (2.4)$$

donde z_0 es igual al estado inicial I : $z_0 = I$.

Esta expresión, sin embargo, no permite un cómputo eficiente de la probabilidad debido a que el número de caminos posibles ($|Z|$) crece de forma exponencial. En cada tiempo $t = 1, 2, \dots, T$ se pueden tener N estados alcanzables, por lo tanto N^T operaciones. Existe un procedimiento más eficiente para calcular dicha probabilidad, denominado algoritmo *forward*.

2.4.3. Nomenclatura utilizada

De ahora en adelante, para simplificar la notación, utilizaremos la siguiente nomenclatura:

\mathcal{M}	conjunto de todos los parámetros que definen un HMM
N	número de estados del modelo excluyendo el inicial: $N = Q - 1$ donde Q es el conjunto de estados: $Q = \{q_0 = I, q_1, \dots, q_{N-1}, q_N = F\}$
M	número de componentes de la mezcla de gaussianas de cada estado
\mathbf{x}	una secuencia de vectores reales (observaciones)
T	número de vectores de la secuencia
x_t	un vector de la secuencia \mathbf{x} en el tiempo t : $x_t \in \mathbb{R}^d \quad 1 \leq t \leq T$
a_{ij}	probabilidad de transición del estado q_i al estado q_j : $a_{ij} \equiv a(q_i, q_j)$
$b_i(x)$	probabilidad de emisión del vector x en el estado q_i : $b_i(x) \equiv b(q_i, x)$
c_{jm}	coeficiente de peso de la componente de mezcla m del estado q_j
$\boldsymbol{\mu}_{jm}$	vector de medias para la componente de mezcla m del estado q_j
$\boldsymbol{\Sigma}_{jm}$	matriz de covarianza para la componente de mezcla m del estado q_j

Para modelar las probabilidades de emisión de vectores $x_t \in \mathbb{R}^d$ en cada estado del HMM, utilizaremos una mixtura de densidades de gaussianas. En este caso la función $b_j(x_t)$ es definida como:

$$b_j(x_t) = \sum_{m=1}^M c_{jm} b_{jm}(x_t)$$

donde

$$b_{jm}(x_t) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_{jm}|}} e^{-\frac{1}{2}(x_t - \mu_{jm})' \Sigma_{jm}^{-1} (x_t - \mu_{jm})}$$

cumpliéndose además que $\sum_{m=1}^M c_{jm} = 1$.

2.4.4. Algoritmo “Forward”

El algoritmo *forward* se utiliza para calcular, de forma eficiente, la probabilidad de que una secuencia de observaciones \mathbf{x} sea emitida por un modelo \mathcal{M} .

Sea $\alpha_j(t)$ con $0 < j < N$, la densidad de probabilidad de que un proceso de Markov se encuentre en el estado q_j en el tiempo t , y haya emitido la secuencia de vectores $x_1 x_2 \dots x_t$. Es decir, $\alpha_j(t) = P(x_1 x_2 \dots x_t, q_j)$. Esta probabilidad puede expresarse de forma recursiva como:

$$\alpha_j(t) = \begin{cases} a_{0j} b_j(x_1) & t = 1 \\ \left[\sum_{i=1}^{N-1} \alpha_i(t-1) a_{ij} \right] b_j(x_t) & 1 < t \leq T \end{cases}$$

con la condición inicial de $\alpha_0(1) = 1$. La probabilidad de que la secuencia \mathbf{x} sea emitida por el modelo \mathcal{M} en términos de $\alpha_j(t)$ es:

$$P(\mathbf{x}|\mathcal{M}) = P(x_1 x_2 \dots x_T|\mathcal{M}) = \alpha_N(T) = \sum_{i=1}^{N-1} \alpha_i(T) a_{iN}$$

La complejidad temporal para el cálculo de esta función es: $\mathcal{O}(|Q|^2 \times T)$. Si la topología del modelo \mathcal{M} es *izquierda-derecha* (topología lineal en la que una vez se abandona un estado ya no se puede regresar al mismo), entonces dicha complejidad se reduce a $\mathcal{O}(|Q| \times T)$.

2.4.5. Algoritmo “Backward”

Para el aprendizaje de un HMM se deben propagar las probabilidades en forma inversa. El algoritmo *backward* es la versión inversa del algoritmo *forward* y se define como sigue:

Sea $\beta_i(t)$ con $0 < i < N$, la densidad de probabilidad de que en un proceso de Markov se vaya a emitir la secuencia de vectores $x_{t+1} x_{t+2} \dots x_T$, sabiendo que en el instante t se halla en el estado q_i . Es decir, $\beta_i(t) = P(x_{t+1} x_{t+2} \dots x_T | q_i)$. Dicha función puede expresarse en forma recursiva como:

$$\beta_i(t) = \begin{cases} a_{iN} & t = T \\ \sum_{j=1}^{N-1} a_{ij} b_j(x_{t+1}) \beta_j(t+1) & 1 \leq t < T \end{cases}$$

con la condición inicial de $\beta_N(T) = 1$. La probabilidad de que la secuencia \mathbf{x} sea emitida por el modelo \mathcal{M} en términos de $\beta_i(t)$ es:

$$P(\mathbf{x} | \mathcal{M}) = P(x_1 x_2 \dots x_T | \mathcal{M}) = \beta_0(1) = \sum_{j=1}^{N-1} a_{0j} b_j(x_1) \beta_j(1)$$

La complejidad temporal para el cálculo de esta función es: $\mathcal{O}(|Q|^2 \times T)$. Si la topología del modelo \mathcal{M} es *izquierda-derecha*, entonces dicha complejidad se reduce a $\mathcal{O}(|Q| \times T)$.

2.4.6. Algoritmo “Viterbi”

El algoritmo de *Viterbi* obtiene el camino más probable en un HMM dada una secuencia de observación. Este algoritmo es una variante del *Forward* en el que se reemplaza un sumatorio por un máximo, como se muestra a continuación:

$$vit_j(t) = \begin{cases} a_{0j} b_j(x_1) & t = 1 \\ [\max_{i \in [1, N-1]} vit_i(t-1) a_{ij}] b_j(x_t) & 1 < t \leq T \end{cases}$$

con la condición inicial de $vit_0(1) = 1$. La probabilidad de Viterbi de la secuencia \mathbf{x} es:

$$vit_N(T) = \max_{i \in [1, N-1]} vit_i(T) a_{iN} \leq \sum_{i=1}^{N-1} \alpha_i(T) a_{iN} = \alpha_N(T)$$

El algoritmo de *Viterbi* es una cota inferior del algoritmo *Forward*. La complejidad temporal es $\mathcal{O}(|Q|^2 \times T)$. Si la topología del modelo \mathcal{M} es *izquierda-derecha*, entonces dicha complejidad se reduce a $\mathcal{O}(|Q| \times T)$.

2.4.7. Algoritmo “Backward-Forward” para entrenamiento de un HMM

El algoritmo “Backward-Forward” o “Baum-Welch” [Rabi 93] puede ajustar, a partir de un conjunto de muestras de entrenamiento, los parámetros que definen un HMM: a_{ij} , $b_i(x)$, c_{jm} , μ_{jm} y Σ_{jm} .

Este algoritmo, que puede verse como una instancia del algoritmo EM (“*Expectation - Maximization*”) [Demp 77], tiene garantizada su convergencia a un máximo local (por el criterio de máxima verosimilitud), de la verosimilitud de los parámetros de los HMMs para el conjunto de las muestras de entrenamiento.

Sea $E = \{\mathbf{x}^r = (x_1^r x_2^r \dots x_{T_r}^r) : x_k^r \in X \text{ para } 1 \leq k \leq T_r \wedge 1 \leq r \leq R\}$ un conjunto de R secuencias de vectores, utilizadas para estimar los parámetros de un HMM simple (aislado).

La fórmula básica para la reestimación de las probabilidades de transición a_{ij} es:

$$\hat{a}_{ij} = \frac{\sum_{r=1}^R \frac{1}{P_r} \sum_{t=1}^{T_r-1} \alpha_i^r(t) a_{ij} b_j(x_{t+1}^r) \beta_j^r(t+1)}{\sum_{r=1}^R \frac{1}{P_r} \sum_{t=1}^{T_r} \alpha_i^r(t) \beta_i^r(t)}$$

donde $0 < i < N$, $0 < j < N$ y $P_r = P(\mathbf{x}^r | \mathcal{M})$ es la probabilidad total de la r -ésima muestra del conjunto E .

Las transiciones desde el estado inicial I son reestimadas por:

$$\hat{a}_{0j} = \frac{1}{R} \sum_{r=1}^R \frac{1}{P_r} \alpha_j^r(1) \beta_j^r(1)$$

donde $0 < j < N$, y las transiciones al estado final F :

$$\hat{a}_{iN} = \frac{\sum_{r=1}^R \frac{1}{P_r} \alpha_i^r(T) \beta_i^r(T)}{\sum_{r=1}^R \frac{1}{P_r} \sum_{t=1}^{T_r} \alpha_i^r(t) \beta_i^r(t)}$$

donde $0 < i < N$.

Para un HMM con mezclas de gaussianas de M componentes, los parámetros c_{jm} , $\boldsymbol{\mu}_{jm}$ y $\boldsymbol{\Sigma}_{jm}$ en cada estado q_j se reestiman como sigue. Se define $L_{jm}^r(t)$ como la densidad de probabilidad que el vector $x_t^r \in \mathbb{R}^d$ esté generado por la m -ésima componente de la mezcla del estado q_j :

$$L_{jm}^r(t) = \frac{1}{P_r} U_j^r(t) c_{jm} b_{jm}(x_t^r) \beta_j^r(t)$$

donde

$$U_j^r(t) = \begin{cases} a_{0j} & \text{si } t = 1 \\ \sum_{i=1}^{N-1} \alpha_i^r(t-1) a_{ij} & \text{caso contrario} \end{cases} \quad (2.5)$$

Para el caso de una mezcla de gaussianas de una sola componente, la densidad de probabilidad de que el vector $x_t^r \in \mathbb{R}^d$ esté generado por dicha mezcla, es igual a la densidad de probabilidad de que éste haya sido emitido en el estado q_j . En este caso

$$L_{jm}^r(t) = L_j^r(t) = \frac{1}{P_r} \alpha_j^r(t) \beta_j^r(t)$$

Dadas las definiciones de arriba, la reestimación de los parámetros μ_{jm} , Σ_{jm} y c_{jm} puede expresarse en términos de $L_{jm}^r(t)$ como sigue

$$\hat{\mu}_{jm} = \frac{\sum_{r=1}^R \sum_{t=1}^{T_r} L_{jm}^r(t) x_t^r}{\sum_{r=1}^R \sum_{t=1}^{T_r} L_{jm}^r(t)}$$

$$\hat{\Sigma}_{jm} = \frac{\sum_{r=1}^R \sum_{t=1}^{T_r} L_{jm}^r(t) (x_t^r - \hat{\mu}_{jm})(x_t^r - \hat{\mu}_{jm})'}{\sum_{r=1}^R \sum_{t=1}^{T_r} L_{jm}^r(t)}$$

$$c_{jm} = \frac{\sum_{r=1}^R \sum_{t=1}^{T_r} L_{jm}^r(t)}{\sum_{r=1}^R \sum_{t=1}^{T_r} L_j^r(t)}$$

El orden de complejidad temporal es: $\mathcal{O}(R \times |Q|^2 \times T)$; pero si el modelo \mathcal{M} es de tipo *izquierda-derecha*, entonces es $\mathcal{O}(R \times |Q| \times T)$.

2.5. Redes neuronales

Las redes neuronales han estado relacionadas con los métodos estadísticos por diversos autores [Ande 90] ya que buena parte de los modelos más conocidos son implícitamente equivalentes o muy similares a métodos de reconocimiento estadístico. Las redes neuronales pueden ser vistas como sistemas de computación paralela, consistentes en un gran número de procesadores con gran cantidad de interconexiones. Los modelos de redes neuronales tratan de usar principios como aprendizaje, generalización, adaptabilidad, tolerancia a fallos y computación distribuida. Estos son plasmados en una red de grafos dirigidos ponderados en la que los nodos son neuronas artificiales y aristas dirigidas (con pesos) conectan la salida de una neurona con la entrada de otra. Las principales características de las redes neuronales son su capacidad de adaptabilidad para aprender relaciones complejas no lineales entrada-salida, el uso de procedimientos de entrenamiento secuenciales y la capacidad para adaptarse por sí mismas a los datos del problema.

La familia más común de redes neuronales en tareas de clasificación de patrones es la de las redes con conexiones hacia delante [Jain 97], la cual incluye las redes de perceptrón multicapa y función de base radial (RBF). Estas redes están organizadas en capas y disponen de conexiones unidireccionales entre dichas capas. Otro tipo de red muy popular es la de mapa auto-organizativo (SOM) o red de Kohonen [Koho 95]. El proceso de aprendizaje comprende la actualización de la arquitectura de la red y los pesos de las conexiones, para que la red pueda llevar a cabo eficientemente tareas específicas de agrupamiento o clasificación.

2.6. Máquinas de Vectores Soporte

Las Máquinas de Vectores Soporte (SVM) son un conjunto de métodos de aprendizaje supervisado pensados para realizar tanto tareas de clasificación binaria (aunque existen extensiones para trabajar con múltiples categorías) como de regresión. La idea fundamental se basa en aplicar técnicas de clasificación lineal a problemas de clasificación no lineales. Para ello se parte de un espacio de representación de características n -dimensional, el cual se mapea en un espacio de alta dimensionalidad. Posteriormente, en este espacio de alta dimensionalidad, se construye un clasificador lineal mediante la definición de un *hiperplano óptimo de separación* [Vapn 95, Hast 01].

Consideremos en primer lugar el caso en que las muestras de entrenamiento sean separables linealmente en su espacio original. Supongamos que el conjunto de entrenamiento $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)\}$ donde $\mathbf{x}_i \in \mathbb{R}^d$ e $y_i \in \{1, -1\}$, puede ser separado por un hiperplano $(\mathbf{w} \cdot \mathbf{x}) + b = 0$. El clasificador en este caso sería:

$$f(\mathbf{x}) = \text{sgn}(\mathbf{w} \cdot \mathbf{x} + b)$$

Se dice que el hiperplano $H : y = \mathbf{w} \cdot \mathbf{x} + b = 0$ es óptimo si, además, la distancia entre los vectores más cercanos al hiperplano es máxima. Para ello se definen dos nuevos hiperplanos H_1 y H_2 paralelos a H como:

$$H_1 : y = \mathbf{w} \cdot \mathbf{x} + b = 1$$

$$H_2 : y = \mathbf{w} \cdot \mathbf{x} + b = -1$$

de modo que no haya ningún vector \mathbf{x}_i entre H_1 y H_2 , y la distancia entre H_1 y H_2 (margen) sea máxima.

Las muestras más próximas al hiperplano óptimo de separación son conocidas como muestras críticas o *vectores soporte*, lo que da nombre a la SVM. Estas muestras son las que participan en la definición del hiperplano.

Teniendo en cuenta que en el espacio bidimensional la distancia de un punto $p = (x_0, y_0)$ a una recta $r = Ax + By + C = 0$ es

$$d(p, r) = \frac{Ax_0 + By_0 + C}{\sqrt{A^2 + B^2}},$$

de forma similar la distancia entre un punto de H_1 y el hiperplano H es

$$d(H_1, H) = \frac{\mathbf{w} \cdot \mathbf{x} + b}{\|\mathbf{w}\|} = \frac{1}{\|\mathbf{w}\|}$$

y, por tanto, la distancia entre H_1 y H_2 es

$$d(H_1, H_2) = \frac{2}{\|\mathbf{w}\|}$$

En consecuencia, para maximizar la distancia entre H_1 y H_2 debe minimizarse $\|\mathbf{w}\| = \mathbf{w}^T \mathbf{w}$, con la condición de que no haya puntos entre H_1 y H_2 , esto es:

$$\begin{aligned} (\mathbf{w} \cdot \mathbf{x}_i) + b &\geq 1 & \text{si } y_i = 1 \\ (\mathbf{w} \cdot \mathbf{x}_i) + b &\leq -1 & \text{si } y_i = -1 \end{aligned}$$

Estas dos condiciones pueden combinarse en:

$$y_i[(\mathbf{w} \cdot \mathbf{x}_i) + b] \geq 1 \quad i = 1, \dots, n$$

Por tanto, el problema de encontrar el hiperplano óptimo de separación para el caso en que las muestras sean linealmente separables puede formularse como:

$$\min_{\mathbf{w}, b} \left(\frac{1}{2} \mathbf{w}^T \mathbf{w} \right) \quad \text{con } y_i[(\mathbf{w}^T \mathbf{x}_i) + b] \geq 1$$

Hasta ahora se ha considerado que que las muestras de entrenamiento de las dos clases podían separarse linealmente. En caso de que esto no sea posible, se aplica una transformación Φ de modo que los puntos originales $\mathbf{x}_i \in \mathbb{R}^n$ son transformados a otro espacio de alta dimensión, en el cual puedan separarse linealmente. En este caso, el problema a resolver será

$$\min_{\mathbf{w}, b} \left(\frac{1}{2} \mathbf{w}^T \mathbf{w} \right) \quad \text{con } y_i[(\mathbf{w}^T \Phi(\mathbf{x}_i)) + b] \geq 1$$

Supongamos que

$$\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j) \equiv K(\mathbf{x}_i, \mathbf{x}_j)$$

esto es, que el producto escalar de dos puntos en el espacio de alta dimensión es equivalente a una *función núcleo* definida en el espacio original. En este caso no es necesario definir explícitamente la transformación Φ , sino que basta con definir la función núcleo.

Existen muchas funciones núcleo que pueden utilizarse de este modo. Las que suelen utilizarse con mayor frecuencia son:

- Lineal: $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$
- Polinomial: $K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i^T \mathbf{x}_j + r)^d, \gamma > 0$
- Función de Base Radial (RBF): $K(\mathbf{x}_i, \mathbf{x}_j) = e^{(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2)}, \gamma > 0$
- Sigmoide: $K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma \mathbf{x}_i^T \mathbf{x}_j + r)$

donde γ , r y d son parámetros de los distintos núcleos.

Otra extensión al problema de SVM consiste en considerar la posibilidad de que existan muestras con ruido en el conjunto de entrenamiento o, dicho de otro modo, permitir una *separación imperfecta* entre las muestras de ambas clases. En este caso se elimina la restricción de que no existan puntos entre H_1 y H_2 , aunque se penalizan los puntos que sobrepasan el hiperplano definido para su clase. El problema de la *separación perfecta* puede verse como una particularización de este en el que la penalización es $C = \infty$.

Para resolver el problema de la separación imperfecta se definen las variables $\xi_i \geq 0$ de manera que

$$\begin{aligned} (\mathbf{w} \cdot \mathbf{x}_i) + b &\geq 1 - \xi_i & \text{si } y_i = 1 \\ (\mathbf{w} \cdot \mathbf{x}_i) + b &\leq -1 + \xi_i & \text{si } y_i = -1 \end{aligned}$$

y se añade a la función objetivo el término de penalización:

$$\min_{\mathbf{w}, b, \xi} \left(\frac{1}{2} \mathbf{w}^T \mathbf{w} \right) + C \sum_{i=1}^n \xi_i \quad \text{con } y_i [(\mathbf{w}^T \Phi(\mathbf{x}_i)) + b] \geq 1 - \xi_i, \quad \xi_i \geq 0$$

En la Figura 2.2 se muestra como, el uso de un factor de penalización, puede ayudar a evitar el sobreentrenamiento.

La aproximación mediante SVM puede crear problemas cuando se trabaja con conjuntos de entrenamiento muy desequilibrados, en los que el número de muestras de una de las clases es muy superior al de la otra. Esto ocurre habitualmente cuando se trabaja con imágenes médicas, donde el número de muestras de la clase negativa (tejido sano) es mucho mayor que el de la clase positiva (tejido anómalo). En [Mori 99] se introduce una extensión al problema de la separación imperfecta, en la que se utilizan factores de penalización distintos para cada clase, lo que permite ajustar el coste de los falsos positivos y de los falsos negativos de manera independiente. En este caso, el problema de encontrar el hiperplano de separación se expresa del siguiente modo:

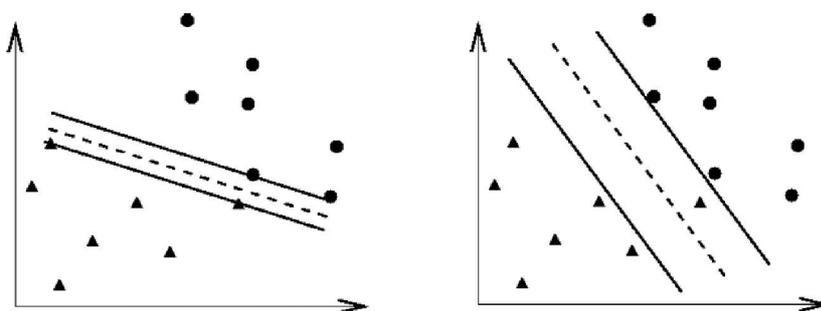


Figura 2.2: Muestras de entrenamiento y un clasificador de SVM sobreentrenado (izquierda) y las mismas muestras de entrenamiento con otro clasificador con separación imperfecta y el problema de sobreentrenamiento corregido(derecha) [Hsu].

$$\min_{\mathbf{w}, b, \xi} \left(\frac{1}{2} \mathbf{w}^T \mathbf{w} \right) + C_+ \sum_{i: y_i=1} \xi_i + C_- \sum_{i: y_i=-1} \xi_j$$

donde C_+ y C_- representan, respectivamente, los factores de penalización para la clase positiva y la negativa.

En [Vero 99] se muestra cómo se modifica el rendimiento de un clasificador de SVM en una tarea de diagnóstico de imágenes médicas, al cambiar la relación entre C_+ y C_- . En la Sección 6.4.8 de este trabajo se aborda esta misma situación. Emplear valores distintos para C_+ y C_- permite, no sólo corregir el desequilibrio existente entre las muestras de una y otra clase en el conjunto de entrenamiento, sino que además posibilita abordar la situación habitual existente en el campo de la medicina, donde un falso negativo tiene un riesgo mucho más elevado que un falso positivo.

2.7. Boosting

En ocasiones, construir una regla de clasificación que tenga una alta precisión puede ser una tarea de gran complejidad y, sin embargo, puede resultar fácil definir varias reglas sencillas que tengan una precisión moderada (entendiendo por moderado aquello que supera una clasificación aleatoria). Por ejemplo, una regla extremadamente sencilla para detectar tejido canceroso en una imagen médica po-

dría ser “clasifica todos los píxeles con nivel de intensidad mayor que 128 en la clase *cáncer*”. Esta sencilla regla, aun teniendo una precisión muy limitada, podría hacer predicciones significativamente mejores que una clasificación aleatoria.

Boosting es un método de clasificación que se basa en combinar varias reglas sencillas para formar una única regla más compleja y precisa. La idea se basa en la afirmación de que varias reglas sencillas, cada una de ellas con una precisión sólo ligeramente superior a una clasificación aleatoria, pueden combinarse para formar una regla de mayor precisión, siempre y cuando se disponga de un número suficiente de muestras de entrenamiento [Scha 90, Hast 01].

Para aplicar la técnica de *Boosting* se debe establecer, en primer lugar, un algoritmo de aprendizaje sencillo que genera lo que se conoce como clasificador “débil” o clasificador “base”. El algoritmo de *Boosting*, en un proceso iterativo, invoca repetidas veces este algoritmo de aprendizaje sencillo, con lo que crea diversos clasificadores base. Para el entrenamiento de los distintos clasificadores base emplea, en cada iteración, un subconjunto diferente de muestras de entrenamiento o, de forma más exacta, una distribución diferente de pesos sobre las muestras de entrenamiento. Finalmente, estos clasificadores base se combinan en un único clasificador que, supuestamente, será mucho más preciso que cualquiera de los clasificadores base por separado.

Para utilizar esta aproximación, hay tres cuestiones fundamentales que deben tenerse en cuenta: a) qué tipo de clasificador base utilizar, b) cómo debe escogerse la distribución de pesos para entrenar cada uno de los clasificadores base y c) cómo deben combinarse dichos clasificadores base en una única regla. Respecto a la primera cuestión, cualquier clasificador es válido, si bien debe tenerse en cuenta que emplear clasificadores demasiado complejos creará problemas de sobreentrenamiento. Respecto a la segunda cuestión, lo habitual es dar mayor peso a las muestras que han sido incorrectamente clasificadas por el clasificador base precedente; esto tiene el efecto de forzar a que el siguiente clasificador base que se infiera centre su atención en las muestras más difíciles de clasificar. En cuanto a la combinación de clasificadores base, simplemente se lleva a cabo una votación ponderada con las hipótesis dadas por cada clasificador. Para llevar a cabo esta votación ponderada, cada clasificador deberá tener un peso asociado, que se calculará teniendo en cuenta el error que haya cometido al clasificar las propias muestras de entrenamiento.

En función de los clasificadores base que se utilicen, las distribuciones que se empleen para entrenarlos y el modo de combinarlos, podrán crearse distintas instancias del algoritmo genérico de *Boosting*. Uno de los algoritmos de *Boosting* que ha tenido mayor éxito es el conocido como *AdaBoost* [Freu 77]. En la Figura 2.3 se muestra el pseudocódigo de este algoritmo.

El algoritmo toma como entrada un conjunto de muestras de entrenamiento $(x_1, y_1), \dots, (x_m, y_m)$ donde cada x_i pertenece a algún dominio X y cada y_i es

Algoritmo AdaBoost**Entada:** $(x_1, y_1), \dots, (x_m, y_m) : x_i \in X, y_i \in Y = \{-1, +1\}$ **Salida:** $H : X \rightarrow \{-1, +1\}$ **Inicialización:** $D_1(i) = 1/m$ **para** $t = 1, \dots, T$ Construir clasificador base $h_t : X \rightarrow \{-1, +1\}$ Escoger un peso $\alpha_t \in \mathbb{R}$ para el clasificador h_t

Actualizar:

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times \begin{cases} e^{-\alpha_t} & \text{si } y_i = h_t(x_i) \\ e^{\alpha_t} & \text{si } y_i \neq h_t(x_i) \end{cases}$$

$$= \frac{D_t(i)}{Z_t} \exp(-\alpha_t y_i h_t(x_i))$$

 donde Z_t es un factor de normalización.**finpara**

$$H(x) = \text{sign} \left(\sum_{t=1}^T \alpha_t h_t(x) \right)$$

Figura 2.3: Algoritmo AdaBoost.

la etiqueta de clase $\in \{-1, +1\}$ (en [Scha 02] se da una extensión para el caso de más de dos clases). El un proceso iterativo, se invoca repetidas veces a un segundo algoritmo de entrenamiento más sencillo, que crea un *clasificador base* h_t a partir de las muestras de entrenamiento x_i y de una distribución de pesos D_t . El objetivo de este clasificador base es minimizar el error

$$\epsilon_t = Pr_{D_t}[h_t(x_i) \neq y_i]$$

Una vez el clasificador base h_t ha sido creado, se elige un parámetro $\alpha_t \in \mathbb{R}$ que determina el peso que h_t tendrá en el clasificador final H . Para el caso particular de que la salida de h_t sea binaria ($\{+1, -1\}$), este valor se define como

$$\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \epsilon_t}{\epsilon_t} \right)$$

Con este valor α_t calculado, se actualiza la distribución de pesos D_t tal y como se especifica en el algoritmo mostrado en la Figura 2.3 y, tras T iteraciones, se calcula el *clasificador final* o *clasificador combinado* H .

Capítulo 3

Detección de lesiones en imágenes médicas

3.1. Introducción

En los últimos años se han publicado gran cantidad de trabajos orientados a la detección de lesiones en imágenes médicas, normalmente con el objetivo de conseguir un mejor diagnóstico a partir de las mismas. En este sentido, las técnicas de Diagnóstico Asistido por Ordenador o *Computer-Aided Diagnosis* (CAD) tratan de caracterizar y detectar patrones normales y anormales dentro de la imagen, con el fin de mejorar la precisión y consistencia diagnóstica obtenida por los expertos humanos.

Una vez digitalizada la imagen, es posible analizarla de modo distinto a como lo haría un observador humano, aplicando técnicas de visión por computador y reconocimiento de formas. Los sistemas de CAD analizan la imagen con el propósito de detectar regiones de interés y caracterizarlas cuantitativamente. Debe quedar claro que el objetivo final de estos sistemas no es el de realizar un diagnóstico definitivo de modo completamente automático. Básicamente los objetivos de un sistema de diagnóstico asistido por ordenador pueden resumirse en:

- Mejorar el rendimiento del experto responsable de interpretar la imagen (normalmente el radiólogo), dirigiendo la atención del mismo hacia las zonas de la imagen más sospechosas de contener alguna anomalía.
- Aportar una segunda opinión sobre un diagnóstico determinado.
- Proporcionar sistemas de autoevaluación y reciclaje para radiólogos con diversos grados de experiencia y de aprendizaje para residentes en periodo de formación.

Uno de los mayores problemas en la detección de lesiones en imágenes médicas viene dado por la subjetividad con que el experto humano analiza dichas imágenes. En muchos casos, la experiencia del profesional encargado de interpretar la imagen, es un factor determinante en el diagnóstico final. Normalmente se requiere una curva de aprendizaje larga para que estos profesionales alcancen unos niveles altos de fiabilidad. Sin embargo, aun en los casos en los que el profesional tiene un grado elevado de experiencia, en muchos tipos de imágenes la detección de lesiones y, en consecuencia, el diagnóstico, resulta una tarea complicada. Las causas de esta dificultad en el diagnóstico suelen deberse a:

- Escasa calidad de la imagen.
- Ocultación de objetos debido al paso de 3D a 2D.
- Lesiones difícilmente visibles al ojo humano debido a su reducido tamaño (microtexturas).
- Subjetividad en la interpretación.
- Cansancio y falta de concentración.

La visión por computador, junto con técnicas de reconocimiento de formas, son herramientas que nos pueden ayudar a detectar ciertos patrones dentro de la imagen y, consecuentemente, pueden suponer una gran ayuda a la hora de establecer un diagnóstico.

Los sistemas de CAD se han aplicado en gran cantidad de disciplinas dentro del campo médico, entre las que cabe destacar la mamografía [Llob 04, Llob 05], radiología de tórax, radiología vascular, radiología ósea, radiología digestiva, ecografía (mama [Riss 95, Kuo 02], hígado [Laye 90, Laye 91], tiroides, ojo, corazón, próstata [Pere 02, Llob 03, Llob 06]) y las aplicaciones a imágenes 3D [Robb 02, Fair 99].

En las siguientes secciones se profundiza en los dos tipos de lesiones que han sido objeto de estudio en esta tesis: el cáncer de próstata, diagnosticado a partir de imágenes ecográficas y el cáncer de mama, diagnosticado a partir de imágenes obtenidas por rayos X. En ambos casos, y en las correspondientes subsecciones, se presentan en primer lugar las técnicas y métodos que se utilizan actualmente para el diagnóstico precoz de este tipo de patologías, a continuación las limitaciones intrínsecas a los métodos de diagnóstico actual y, finalmente, una profunda revisión al estado del arte del diagnóstico asistido por ordenador. Por razones obvias, algunas de estas secciones se centran en aspectos y técnicas relacionadas con el campo de la medicina, lejos de las técnicas de visión por computador y reconocimiento de formas cuyo estudio es el objetivo principal de esta tesis. Sin embargo, esta pequeña incursión en el ámbito médico es necesaria para entender mejor el problema que se pretende abordar.

3.2. Aproximaciones a la detección de tumores en imágenes médicas

En los estudios publicados referentes a la detección de tumores en imágenes médicas se han venido utilizando aproximaciones muy diversas. Es difícil abstraer un diseño común para todas ellas, sin embargo, estas aproximaciones suelen ajustarse a un esquema similar, que se ajusta al esquema genérico de un sistema de reconocimiento de formas supervisado explicado en la Sección 2.1.3. En la Figura 3.1 se muestran las distintas fases que se abordan habitualmente durante la implementación de un sistema de ayuda al diagnóstico de cáncer.

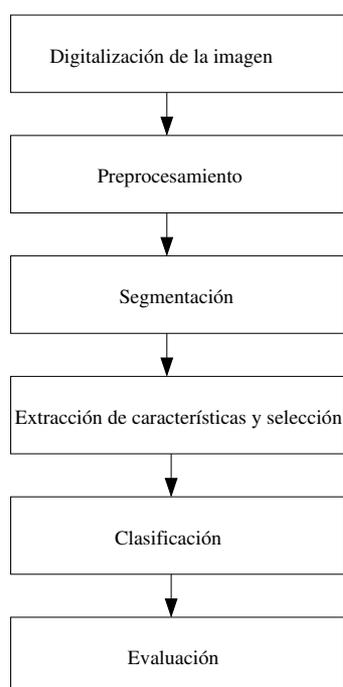


Figura 3.1: Etapas en el diseño de un sistema de ayuda al diagnóstico de cáncer por ordenador a partir de imágenes médicas [Chen 05].

Todo sistema de CAD comienza con la digitalización de la imagen adquirida (rayos X, ecografía o RMN, entre otras), tras lo cual es habitual realizar un preproceso que ayude a eliminar ruido de la misma y, especialmente, a mejorar el contraste y realzar las regiones de la imagen de mayor interés. A continuación, en la fase de segmentación, se detectan y localizan las regiones sospechosas, lo que define las llamadas Regiones de Interés o ROIs. No debe confundirse esta segmentación con el término de igual nombre que habitualmente se utiliza en el reconocimiento de formas y procesamiento de imágenes. En este caso la segmentación es una primera aproximación para aislar las regiones sospechosas de lo que claramente corresponde a tejido normal. En la siguiente fase, se extraen y seleccionan las ca-

racterísticas de las ROIs obtenidas en la fase anterior, para posteriormente proceder a su clasificación, otorgando a cada ROI un *score* o nivel de sospecha. Finalmente, es necesario emplear técnicas que permitan evaluar los resultados obtenidos, de manera que éstos puedan ser comparados de un modo objetivo. A continuación se explica con mayor detalle cada una de estas fases.

3.2.1. Preprocesamiento

La finalidad principal del preproceso es aumentar las diferencias entre las anomalías y el tejido sano. Los métodos más utilizados en este sentido son:

- a) La **modificación global del histograma**. Esto se consigue mediante distintas aproximaciones como son la **ecualización del histograma** (EH) [Nam 98], la **EH múltiplo** [Wong 98] donde el histograma se particiona en función de los picos que contenga, ecualizándose cada partición de forma independiente y la **expansión del histograma** [Nam 98, Wils 98] de modo que este abarque todo el rango de valores posibles.
- b) El **procesamiento local**, en el que se aplican técnicas muy diversas en ventanas locales de la imagen (por ejemplo ROIs) orientadas a resaltar las posibles anomalías que se encuentren en dichas ventanas locales [Nam 98, Wils 98, Kim 97, Brac 96, Wood 91, Pisa 98].
- c) El **procesamiento multiescala**, basado habitualmente en el uso de *wavelets* [Mall 92, Lu 94, Lain 95, Lain 94]. Para ello la imagen digital se transforma mediante el uso de *wavelets* y se modifican los coeficientes para realzar las anomalías. Finalmente se aplica la transformación inversa del *wavelet* para devolver la imagen a su dominio original.

3.2.2. Segmentación

La fase de segmentación trata de aislar las regiones sospechosas (ROIs) del resto de la imagen. Este proceso determinará la sensibilidad del sistema, esto es, su capacidad para detectar correctamente el tejido canceroso (el concepto de *sensibilidad* se explica con detalle en la Sección 4.8). La segmentación debería aislar el mayor número posible de anomalías, aunque entre las ROIs se encuentren regiones correspondientes a tejido sano (falsos positivos). Los falsos positivos deberán ser eliminados en una fase posterior. Atendiendo a su naturaleza, las aproximaciones a la segmentación pueden dividirse en tres grupos: técnicas clásicas, substracción de imágenes bilaterales y técnicas multiescala.

- a) **Técnicas clásicas:** la técnica clásica más sencilla es la **umbralización global**. Su funcionamiento se basa en que el pico generado en el histograma por

las regiones que contienen anomalías es distinto del pico generado por el tejido sano. Esta técnica, sin embargo, no tiene gran precisión en la detección de ROIs, por lo que la salida generada por esta umbralización se utiliza normalmente como entrada de algún otro proceso posterior. Una técnica algo más precisa consiste en la *umbralización local*. En este caso el valor de umbral se determina localmente para ventanas de menor tamaño, en función de los niveles de gris de los píxeles contenidos en la ventana. Una de las técnicas más populares para la segmentación de algunos tipos de tumor, caracterizados por presentar un nivel de brillo superior al del tejido circundante, es la conocida como *region growing*. La idea básica de este algoritmo es determinar un conjunto de semillas en la imagen y a continuación hacer crecer de manera iterativa dichas semillas, mediante la adición de píxeles vecinos cuyas características sean similares a la de la semilla. Cuando la región deja de crecer, se compara el nivel de intensidad medio de la misma con la intensidad de las regiones de alrededor, para determinar si se trata de una región sospechosa o no [Kall 92]. Las claves para el correcto funcionamiento de este algoritmo radican en el criterio de selección de semillas así como en el criterio utilizado para determinar si un píxel debe o no añadirse a la región. Una técnica muy similar es la conocida como *region clustering* [Papp 92]. En este caso las regiones se buscan directamente, sin necesidad de establecer semillas iniciales. El algoritmo de *k*-medias es una técnica de agrupamiento o *clustering* ampliamente utilizada. Otros algoritmos clásicos utilizados habitualmente para la segmentación de tumores son aquellos basados en la detección de bordes como los filtros de Sobel [Vito 96], Prewitt, Laplaciano, etc. Otra técnica ampliamente utilizada es el *emparejamiento de patrones* o *template matching*. Esta aproximación trata de segmentar las anomalías buscando en la imagen patrones que guarden cierta similitud con un conjunto de prototipos obtenidos previamente de una muestra de entrenamiento. Cuando el tamaño de los patrones (anomalías) a segmentar se desconoce, es necesario disponer de un conjunto de prototipos que abarque todas las escalas posibles, o analizar la imagen de test a distintas escalas [Llob 04, Llob 05, Cons 99].

- b) **Substracción de imágenes:** mediante la substracción de imágenes se pueden localizar diferencias entre dos imágenes médicas de un mismo paciente tomadas en distintos momentos (habitualmente cuando se lleva un seguimiento del paciente y se le realizan tests de imagen con cierta periodicidad) [Riss 95, Sall 99]. Para que esta aproximación genere resultados aceptables, es necesario realizar un alineamiento elástico entre ambas imágenes. Las diferencias encontradas entre una y otra imagen serán regiones sospechosas, aunque entre estas regiones puede haber un número importante de falsos positivos. Es necesario, por tanto, un posterior análisis basado en algún tipo de características, que ayude a reducir los falsos positivos. En los casos de órganos simétricos, como por ejemplo la mama, es posible también buscar diferencias entre las imágenes de cada órgano (por ejemplo mamografías del

pecho izquierdo y derecho) tomadas en la misma sesión.

- c) **Técnicas multiescala:** en ocasiones el tamaño de los tumores puede variar considerablemente. En este sentido, las técnicas multiescala pueden facilitar la detección de los mismos. Para ello, el método de detección escogido se aplica a diferentes escalas y se combina, de algún modo, el nivel de sospecha detectado en cada una de las escalas [Brak 99, Kars 96]. Por otro lado, mediante la transformación discreta de *wavelets* (DWT), también es posible hacer un análisis multiescala de la imagen [Li 99, Bruc 00].

3.2.3. Extracción de características y selección

Una vez segmentada la imagen, puede ser necesario extraer características de las distintas ROIs seleccionadas para determinar con mayor precisión el nivel de sospecha. El espacio de características puede ser muy grande y complejo, debido a la gran variabilidad que puede darse tanto en el tejido sano como en el canceroso. Sin embargo, no todas las características obtenidas son igualmente significativas. Usar un número excesivo de características puede incrementar la complejidad del clasificador y degradar el rendimiento del mismo. En [Yin 94] se hace un estudio extenso sobre distintas aproximaciones de extracción de características en mamografías. La extracción y selección de características es una etapa clave para la correcta detección de tumores, ya que el rendimiento del sistema de CAD depende más de la optimización de la extracción y selección de características que del método de clasificación [Chen 05].

Dependiendo del tipo de características que se extraigan de la imagen, el espacio de características puede dividirse en tres categorías: características de intensidad, características geométricas y características de textura.

- a) Las **características de intensidad** es la aproximación más simple. Básicamente consiste en utilizar los valores de niveles de gris de cada ROI para caracterizar la misma. Otras aproximaciones utilizan características que representen la diferencia entre el nivel de gris medio de la ROI y el nivel de gris medio de los píxeles que circundan la ROI [Brak 99, Heat 00b, Llob 05].
- b) Las **características geométricas** o morfológicas se basan en la forma de la ROI, calculándose a partir del área y borde de la misma (área, perímetro, convexidad, momentos de primer, segundo y tercer orden, oblicuidad, etc.).
- c) Las **características de textura** pueden obtenerse a partir de técnicas muy diversas, entre las que cabe destacar las *matrices de coocurrencia*, también llamadas *Matrices de Dependencia de Niveles de Gris* o *Space Gray Level Dependence Matrices* (SGLDM) de *vectores de estadísticas de diferencia de nivel de gris* (GLDS) o basarse en características de *Run Length Statistics* (RLS).

- c.1) Las SGLDM se calculan para una distancia d y ángulo θ , de modo que el elemento $P(i, j|d, \theta)$ de la matriz contiene la probabilidad con la que, dos píxeles cualesquiera a una distancia d y ángulo θ , tengan respectivamente niveles de gris i y j . En la Sección 4.2.3 se trata en profundidad este tipo de aproximación.
- c.2) Un vector GLDS representa el histograma de las diferencias absolutas de niveles de gris entre pares de píxeles separados por un cierto vector de desplazamiento [Wesz 76]. De estos vectores se extraen posteriormente una serie de estadísticos como homogeneidad, contraste, entropía, media, etc.
- c.3) Muchos tumores suelen tener texturas con una dependencia radial, debido al gradiente en la opacidad del tumor y a su propia forma. Para obtener características que tomen en consideración esta disposición radial, se puede utilizar la transformación conocida como **Rubber Band Straightening Transform** (RBST) [Sahi 01]. Esta transformación mapea la banda de píxeles alrededor del tumor (o de una ROI) en un plano cartesiano (región rectangular). De la transformación RBST se pueden extraer características basadas en RLS [Gall 75]. El elemento (i, j) de una matriz RLS $R_{\theta}(i, j)$ representa la frecuencia con que segmentos de j píxeles consecutivos con intensidad i aparecen en la dirección θ .

Independientemente del tipo de características obtenidas en primera instancia, se puede aplicar el **Análisis de Componentes Principales** (PCA) explicado en la Sección 4.3.3 para decorrelar las distintas características y reducir su número. Por otro lado, pueden utilizarse distintas aproximaciones de selección de características para escoger aquellas que realmente son relevantes o tienen mayor poder discriminatorio. En este sentido, las técnicas más utilizadas son la selección de características por etapas mediante el **Análisis Lineal Discriminante** (LDA) [Sahi 98, Sahi 01, Kupi 97, Chan 95] y los **Algoritmos Genéticos** (GA) [Sahi 98].

3.2.4. Clasificación

Una vez las características han sido extraídas y seleccionadas, estas deben ser clasificadas para determinar el grado de sospecha. En la literatura relacionada con el CAD de cáncer se referencian gran cantidad de clasificadores: LDA [Duda 73], redes neuronales [Bovi 00, Chen 94], árboles de decisión binarios [Zhen 01] o k -vecinos más cercanos [Llob 04, Llob 05], entre otros.

Cada clasificador puede tener su propia región del espacio de características donde su rendimiento es mejor, por lo que la combinación de distintos clasificadores puede mejorar significativamente el rendimiento global. Los clasificadores pueden combinarse en paralelo, en cascada o de forma jerárquica [Jain 00]. En [Cons 01] se combinan cinco clasificadores distintos: clasificador gaussiano,

red de función de base radial, Q -vector mediana, vecino más cercano y ventanas de Parzen hiperesféricas.

3.2.5. Evaluación

Hacer una comparación objetiva del rendimiento obtenido por diferentes sistemas de CAD puede ser un objetivo difícil, o incluso imposible cuando se utilizan bases de datos distintas. El primer paso, por tanto, para comparar el rendimiento de distintos sistemas, es utilizar la misma base de datos y los mismos conjuntos de entrenamiento y test en todos ellos.

Un método ampliamente utilizado para la comparación de clasificadores binarios en general y para la clasificación de tumores en particular, es el análisis mediante curvas ROC (*Receiver Operating Characteristic*), en las que se muestran los verdaderos positivos en función de los falsos positivos. En [Metz 84] se presenta un algoritmo para comprobar la significancia estadística de la diferencia entre dos curvas ROC.

Cuando el sistema a evaluar no es puramente de clasificación, sino de detección y localización, en el que puede observarse más de un tumor en la imagen, entonces el método más adecuado para la evaluación es el basado en curvas *Free-ROC* (FROC). En este caso, se muestra la relación entre los verdaderos positivos y el número medio de falsos positivos por imagen. En la Sección 4.8 se estudia en profundidad el análisis de resultados, tanto con curvas ROC como FROC.

3.3. Detección de cáncer de próstata

El cáncer de próstata es una de las neoplasias más frecuentes en el varón. En los países industrializados, actualmente el 10 % de los hombres es diagnosticado de cáncer de próstata en algún momento de su vida, y entre un 3 % y un 4 % muere a causa de esta enfermedad [Braw 99]. Debido a que la esperanza de vida crece de manera continua en los países industrializados, se puede prever que estos porcentajes aumentarán en el futuro, a menos que evolucionen las técnicas utilizadas tanto en la detección como en el tratamiento de esta enfermedad. Un factor de máxima importancia para tener éxito en la curación del cáncer de próstata, es la detección del mismo en sus primeras fases. Al igual que ocurre con otros tipos de cáncer, el de próstata sólo es curable, con ciertas garantías de éxito, cuando es detectado en sus fases incipientes. En este sentido, los sistemas de ayuda al diagnóstico, pueden ser una herramienta de gran utilidad en la detección de los tumores en sus primeras fases.

3.3.1. Métodos actuales en el diagnóstico del cáncer de próstata

Actualmente, los dos tests preliminares más utilizados para diagnosticar el cáncer de próstata son el tacto rectal o *Digital Rectal Examination* (DRE) y el análisis del antígeno específico de la próstata (PSA). Ambas pruebas pueden realizarse con un bajo coste económico, al tiempo que son poco agresivas para el paciente.

Mediante el tacto rectal, el urólogo palpa la glándula prostática en busca de durezas u otras anormalidades. La detección de masas sospechosas, texturas anormales o durezas, puede ser un indicador de la existencia de un tumor.

El PSA es una proteína producida en la próstata, cuya concentración en sangre puede aumentar cuando el cáncer está presente. Mediante un sencillo análisis, es posible determinar el nivel de PSA en la sangre. Un nivel de PSA más alto que lo esperado, puede significar la presencia de un tumor.

Estos dos tests son un primer indicativo de la posible existencia de un tumor, pero no son, ni mucho menos, concluyentes, tal y como se verá más adelante. Básicamente estos tests se utilizan como primera aproximación al diagnóstico, para decidir si conviene reforzar el diagnóstico con otras pruebas. En la práctica, cuando estas dos pruebas ofrecen un nivel de sospecha elevado, se inicia el estudio diagnóstico mediante algún método de imagen: Tomografía Axial Computerizada (TAC), Resonancia Magnética Nuclear (RMN) o ecografía transrectal (TRUS). Como se verá, ninguno de estos tests ofrecen unos niveles de sensibilidad y especificidad¹ suficientes, por lo que cualquier estudio diagnóstico requerirá, en su etapa final, un análisis histológico de tejido prostático mediante biopsia.

La tomografía computerizada es una técnica basada en rayos X que produce imágenes de distintas secciones del órgano analizado. Aunque esta técnica se ha empleado ampliamente en el diagnóstico de otros tumores, tiene menos utilidad en la detección del cáncer de próstata; en este caso, su uso queda relegado únicamente a examinar cómo de extendido se encuentra un tumor ya conocido.

La resonancia magnética, en lugar de utilizar rayos X, utiliza imanes y ondas de radio potentes para elaborar imágenes claras y detalladas de los tejidos estudiados. Al igual que la TAC, es un procedimiento no invasivo que produce imágenes de distintos planos o “cortes” de la próstata. La RNM puede, además, presentar vistas desde distintos ángulos. Las imágenes generadas por RNM tienen una gran claridad, sin embargo, un estudio comparativo entre RNM y TRUS demostró que la primera, aun siendo más sensible en la predicción de la enfermedad, es menos específica [Pres 96]. Por otro lado, otros trabajos no encuentran utilidad en esta exploración a la hora de modificar una actitud terapéutica ya establecida con otros criterios [Perr 96]. Por último cabría decir que esta exploración tiene un coste elevado, lo que dificulta utilizarlo masivamente como método de diagnóstico precoz.

¹Los conceptos de sensibilidad y especificidad se explican en la Sección 4.8.

La ecografía transrectal obtiene imágenes ecográficas de la próstata a través de un transductor de ultrasonidos que se inserta en el paciente a través del recto. Dicho dispositivo genera, en tiempo real, imágenes ecográficas de secciones de la próstata. Este tipo de imágenes ofrecen, al igual que ocurre con la RNM y la TAC, niveles de sensibilidad y especificidad muy limitados en cuanto al diagnóstico de cáncer de próstata se refiere. Sin embargo, la ecografía transrectal aporta la gran ventaja de permitir realizar una biopsia de forma dirigida. Para ello, junto con el transductor de ultrasonidos, se introduce una sonda por cuyo interior se inserta la aguja hueca con la que realizar la biopsia. La ecografía transrectal permite al urólogo ver el lugar exacto de la próstata donde se insertará la aguja. En la Figura 3.2 se muestra una imagen de la próstata obtenida mediante TRUS en la que se puede observar la trayectoria que seguiría la aguja de la biopsia. Cuando el urólogo desea realizar una biopsia de una región concreta de la próstata, orienta el transductor (junto con la aguja de la biopsia) para que dicha región quede sobre la línea de puntos. En ese momento activa un mecanismo que dispara la aguja en el interior de la glándula, lo que permite tomar una muestra de tejido prostático. Dicho tejido es posteriormente analizado en el laboratorio y se diagnostica, sin margen de error, si dicho tejido contiene o no células cancerosas. Dado el bajo coste de esta técnica y, sobre todo, debido a su capacidad para realizar biopsias dirigidas, la ecografía transrectal es, actualmente, el método más utilizado en el diagnóstico del cáncer de próstata. Aun así, tal y como se verá más adelante, esta técnica aún está lejos de producir diagnósticos de una gran fiabilidad.

3.3.2. Limitaciones actuales al diagnóstico precoz del cáncer de próstata

Los tests actuales utilizados en el diagnóstico precoz de cáncer de próstata tienen serias limitaciones en cuanto a capacidad diagnóstica se refiere, lo que se traduce en un importante número de pacientes a los que no se les detecta el cáncer tras la realización de varios tests (falsos negativos), así como en otra gran cantidad de pacientes que, no padeciendo la enfermedad, son sometidos a una serie de biopsias por existir una sospecha alta de presencia de cáncer (falsos positivos).

Por un lado, mediante el tacto rectal o DRE sólo es posible palpar la pared posterior de la glándula. Si bien la mayoría de los tumores se localizan precisamente en esta zona, cualquier tumor localizado en medio de la próstata o en la parte anterior, no sería detectado mediante esta prueba. Además, un tumor poco desarrollado sería difícilmente detectable con esta técnica.

Por otro lado, el análisis del PSA, aunque es un test bastante sensible (un cáncer de próstata aumenta generalmente el nivel de PSA) es poco específico (hay condiciones, no directamente relacionadas con el cáncer de próstata, que producen un incremento del PSA). En concreto, los niveles de PSA aumentan no sólo con la aparición de un tumor, sino también cuando se produce una inflamación de la

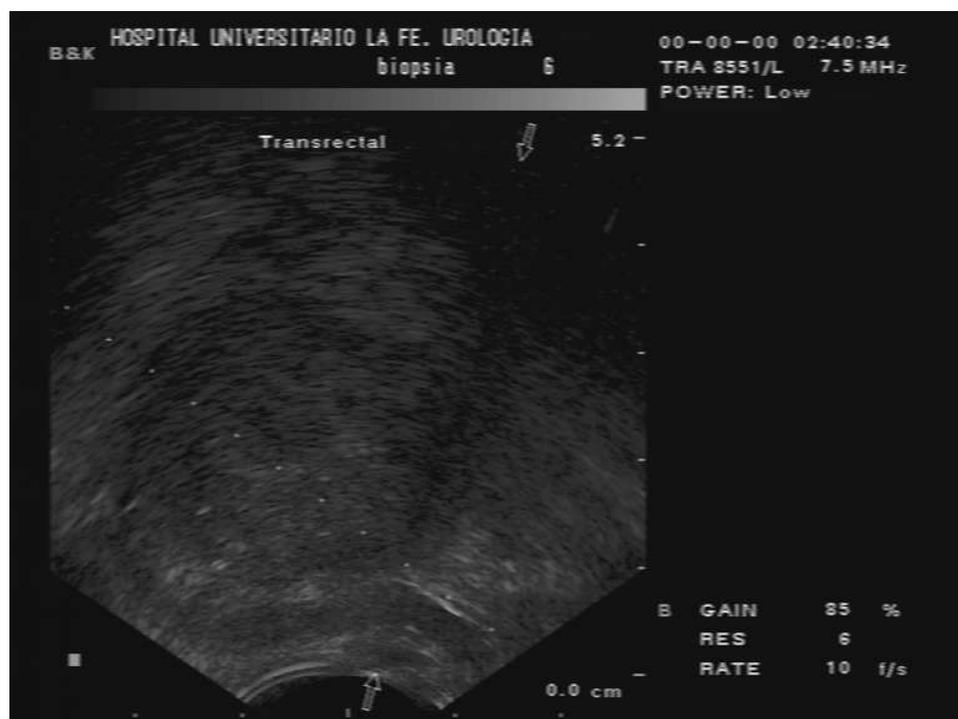


Figura 3.2: Imagen de ecografía transrectal tal y como se visualiza en el monitor del ecógrafo. La línea de puntos sobreimpresa indica la trayectoria que seguiría la aguja de la biopsia si se accionase el dispositivo de disparo en este instante.

próstata (prostatitis) o cuando simplemente el tamaño de la próstata aumenta debido a la edad, lo que genera un porcentaje elevado de falsos positivos. Estos falsos positivos ocurren principalmente en hombres a partir de 50 años. En este grupo de edad, el 15 % de los hombres tiene un nivel elevado de PSA (más de 4 ng/ml). De este 15 %, realmente sólo el 20 % padece de cáncer de próstata, mientras que el otro 80 % corresponde a pacientes cuyo PSA ha aumentado por causas distintas al cáncer de próstata (falsos positivos).

Debido a la inespecificidad de estos tests, cuando existe sospecha de cáncer, el diagnóstico debe de ser confirmado mediante análisis histológico (biopsia). Para ello, tal y como se ha explicado anteriormente, el urólogo se vale de la ecografía transrectal para dirigir la biopsia hacia las zonas de la próstata más sospechosas de contener células cancerosas. Sin embargo, esta técnica no está exenta de inconvenientes, ya que la inspección visual de la imagen de ultrasonidos no es de mucha ayuda en la localización de tejido canceroso. Por un lado, la escasa nitidez de una imagen ecográfica no permite ver con claridad los tejidos blandos (como lo es la glándula prostática). Por otro lado, ni siquiera los urólogos están todavía de acuerdo en cuáles son los patrones de imagen que caracterizan a un tejido afectado por células cancerosas. Como consecuencia de todo ello, debido a la incapacidad de

detectar zonas sospechosas en la imagen ecográfica, el urólogo rara vez dirige la punción basándose en las características de la imagen, y en su lugar acaba realizando una serie de biopsias uniformemente distribuidas a lo largo de la próstata. Algunos autores sugieren que el número de biopsias más adecuado para obtener un diagnóstico lo más fiable posible es entre 10 y 12, aunque en muchos casos acaban realizándose únicamente entre 5 o 6 biopsias debido a la intolerancia que causa en el paciente un elevado número de punciones. Algunos estudios han concluido que incluso utilizando la cantidad aconsejada de 12 punciones, aproximadamente el 10 % de los tumores no se detecta debido a que ninguna de las punciones alcanza el lugar exacto donde se localiza el cáncer [Stew 01, Djav 00].

Para aumentar el rendimiento en el diagnóstico de cáncer de próstata, básicamente se puede actuar en dos líneas bien diferenciadas:

- Mejorar el diagnóstico por la imagen en general
- Mejorar el diagnóstico de la imagen ecográfica transrectal y el procedimiento de biopsia

En los últimos años se han producido grandes avances en las distintas técnicas de imagen (TAC, RNM, TRUS). Sin embargo, ninguno de estos procedimientos ha conseguido aumentar la especificidad ni la sensibilidad en el diagnóstico y, por lo tanto, no son capaces de evitar la confirmación histológica mediante biopsia. Por otro lado, actualmente el único método de imagen que puede guiar y facilitar la realización de la biopsia es la ecografía transrectal. Por todo ello, cualquier técnica que pretenda aumentar el rendimiento diagnóstico del cáncer de próstata, debe ir en la línea de facilitar la interpretación de la ecografía transrectal de la próstata, con objeto de dirigir la biopsia hacia las áreas más sospechosas y, en caso de diagnóstico negativo, se pueda aseverar con menor margen de error la ausencia de cáncer de próstata. En este sentido, los sistemas de ayuda al diagnóstico pueden ser de gran ayuda a la hora de interpretar una ecografía transrectal y, por lo tanto, de dirigir la biopsia hacia las zonas más sospechosas. En la sección siguiente se hace una revisión del estado del arte, describiéndose los trabajos más relevantes relacionados con el diagnóstico de cáncer de próstata asistido por ordenador.

3.3.3. Diagnóstico del cáncer de próstata asistido por ordenador

Aunque los métodos utilizados actualmente para el diagnóstico de cáncer de próstata entraña grandes dificultades, tal y como se ha comentado anteriormente, han sido pocos los esfuerzos en materia de investigación dedicados al desarrollo de sistemas de CAD orientados a mejorar el diagnóstico de esta enfermedad.

Por otro lado, entre los escasos trabajos publicados en este área, la mayor parte de ellos han sido realizados con un número de muestras muy escaso, presentándo-

se en muchos casos errores metodológicos. Por ello, la validez de los resultados mostrados en algunas de estas publicaciones debe tomarse con cautela.

A continuación se presenta el estado del arte en el diagnóstico de cáncer de próstata asistido por ordenador mediante un breve resumen de los trabajos más relevantes en este campo.

- a) Uno de los primeros trabajos presentados sobre análisis de ecografía prostática lo publicó un grupo de investigadores de la Universidad de Viena en 1987 [Krat 88]. En dicho trabajo se utilizaron características locales basadas en texturas. Los resultados publicados mostraban tasas de sensibilidad y especificidad por encima del 80 %, sin embargo no se daban detalles de las condiciones sobre las que se habían realizado los experimentos.
- b) En 1993, un grupo francés de Villeurbanne realizó un trabajo en el que se utilizó una escasa muestra (37 imágenes), de las cuales se seleccionó, atendiendo a su homogeneidad, un mayor número de regiones de interés para el estudio [Bass 93]. Se utilizaron matrices de coocurrencia. Desconocemos si los resultados fueron evaluados o no con la misma muestra que se utilizó en la fase de entrenamiento. Se concluye diciendo que los resultados son buenos para tamaños de ventana grande, pero no en ventanas menores correspondientes a tumores de pequeño tamaño.
- c) El Centro para la Investigación de Cáncer de Alemania de Heidelberg, con el apoyo de urólogos de otras partes del país, ha desarrollado un sistema de identificación del cáncer de próstata en ecografía transrectal utilizando parámetros cognitivos de textura [Ever 96]. La muestra de este trabajo es muy limitada, 17 imágenes de ecografía (12 de ellas cancerosas) de las cuales extraen 197 regiones de interés. En su análisis presuponen la característica ecográfica del tejido a estudiar e identifican 5 ecoestructuras correspondientes a 5 tipos histológicos, donde sólo 1 de ellos corresponde a tejido canceroso. Con ello se presupone un único tipo de ecogenicidad en el cáncer de próstata, excluyendo cualquier otra posibilidad.

Con la imagen digitalizada realizan un procedimiento de extracción de características a lo largo de varios niveles sucesivos (mediante un proceso piramidal que busca imágenes similares al cáncer) y calculan conjuntos de parámetros de textura. Los parámetros utilizados son: media, varianza, momento isotrópico estructural y gradiente absoluto. Dichos parámetros se calculan a partir de las 197 regiones de interés que representan los 5 tipos de tejido definidos. Se utiliza la misma muestra para la fase de entrenamiento y para la evaluación de resultados, con lo que no es de extrañar que el sistema sea capaz de identificar correctamente los 12 tumores de la muestra, con una sensibilidad del 100 %.

- d) La compañía americana Biocomputer Research Inc. ha diseñado un software denominado “Image Metamorphosis” capaz de transformar la imagen ecográfica en escala de grises, en una señal en color con mayor capacidad de ser analizada por el ojo humano. Además extrae y realza los contornos donde hay mayor gradiente en la escala de grises para diferenciar más claramente los cambios en la ecoestructura.

Aunque en el trabajo se presenta alguna imagen de ecografía prostática transformada, no se presentan los resultados clínicos obtenidos con dicho sistema. En cualquier caso, esta tecnología únicamente transforma datos ecográficos en escala de grises, en imágenes más fácilmente identificables para el ojo humano, sin ningún tipo de intención diagnóstica en sí misma.

- e) Urólogos de la universidad de Kiel, en colaboración con el Departamento de Informática de la Universidad de Nevada, publicaron un trabajo en el que correlacionaban imágenes de ecografía transrectal con secciones histológicas de piezas de prostatectomía radical (extracción de la glándula prostática mediante cirugía). Utilizaron un sistema de análisis de redes neuronales artificial (ANNA - Artificial Neural Network Analysis) con muy buenos resultados [Loch 00, Loch 99]. Para crear la red emplearon descriptores de granularidad de la imagen en ventanas de 18×18 píxeles.

De un total de 553 muestras válidas, utilizaron un 10 % (correspondientes a 5 pacientes) para el proceso de aprendizaje y el resto para la fase de test. Consiguieron identificar correctamente el 71 % de los tumores hipoecoicos y el 97 % de los isoecoicos. Dicho sistema aún no ha sido desarrollado para su uso en tiempo real con intención de guiar la biopsia.

- f) W. Fair et al. presentaron en 1999 los resultados de su trabajo [Fair 99]. En dicho estudio se analizó una muestra de 137 pacientes y 664 biopsias mediante análisis espectral de radiofrecuencia (RF) procedente de la señal acústica. Se utilizaron distintos métodos de clasificación como redes neuronales y k -vecinos más cercanos. Los resultados obtenidos presentan un beneficio diagnóstico importante. Además se muestran reconstrucciones 3-D de la imagen sospechosa. No se detalla, sin embargo, el modo en que la muestra fue dividida en los correspondientes conjuntos de entrenamiento y test.

Estos trabajos se realizaron conjuntamente con el *Riverside Research Institute of New York* que ha trabajado intensamente con análisis espectral de la señal de radiofrecuencia en diversos tejidos [Fele 99, Fele 00, Fele 01, Lizz 97]. En estos estudios se extrajeron parámetros de la señal acústica de tejidos diagnosticados por biopsia, clasificándose posteriormente dichos parámetros mediante redes neuronales. Los resultados obtenidos, utilizando 664 biopsias de 137 pacientes, ofrecen un área bajo la curva ROC ² de 0.87,

²El análisis de resultados mediante curva ROC se explica con detalle en la Sección 4.8.

comparado con 0.64 obtenido mediante el estudio ecográfico convencional.

- g) En [Gart 98] se estudia la señal RF mediante Análisis de Potencia Espectral. Estudian parámetros acústicos cuantitativos, independientes del tipo de ecógrafo utilizado.

Parece ser que el estudio de la señal RF ofrece posibilidades alentadoras y existe en la actualidad gran interés por ello. No obstante, no es más que otro método de estudio de la ecoestructura tisular, siendo la correlación entre la imagen propiamente dicha y la señal de RF muy exacta.

- h) El trabajo que mayor eco ha tenido en relación con el CAD de cáncer de próstata ha sido el realizado por el grupo de la Universidad de Nijmegen en Holanda [Huyn 94] (curiosamente, tal y como se verá más adelante, en esta misma universidad desarrolla su actividad uno de los grupos de trabajo más activos relacionados con el CAD de cáncer de mama). En una población de 51 pacientes, con una metodología de adquisición de imagen y clasificación correcta, procedieron a analizar los parámetros de textura, calculados mediante matrices de coocurrencia (en la Sección 4.2.3 se explica en detalle el uso de este tipo de matrices). Concretamente se utilizaron descriptores de uniformidad, contraste, entropía, momento de diferencia inversa y correlación. Las matrices utilizadas se calcularon en función de las tonalidades de gris de los píxeles inmediatamente adyacentes (distancia= 1) e intermedios (distancia= 2) en ventanas de 3×3 mm. y en tres localizaciones distintas del área de biopsia: proximal, central y distal. Las 98 imágenes obtenidas fueron divididas en dos grupos (entrenamiento y test) de manera homogénea. Obtuvieron resultados espectaculares con una sensibilidad y especificidad del 80 y 88 % respectivamente. Sin embargo, se cometieron algunos errores metodológicos, lo que resta validez a los resultados obtenidos. Concretamente, la muestra utilizada en este trabajo se seleccionó artificialmente entre pacientes con una alta sospecha tumoral, lo que supone una edición de la muestra de test. Como veremos más adelante, los mismos autores obtienen resultados mucho menos alentadores cuando la muestra analizada proviene de una tarea real en la que no existe selección de pacientes.

Este mismo grupo ha presentado numerosas publicaciones referentes a su trabajo [Gies 94, Rose 95a, Rose 95b]. Entre otras, presentaron resultados relacionando la sospecha de carcinoma en imágenes de ecografía transrectal con los hallazgos en piezas de prostatectomía radical (extirpación de la glándula prostática) en 12 pacientes. En este caso obtuvieron una sensibilidad del 78 % con una especificidad del 50 % [Gies 95].

Otro trabajo con parecida metodología fue publicado por los mismos autores en relación con la capacidad diagnóstica de "prostatitis" (inflamación de la próstata) utilizando su sistema denominado AUDEX (Automated Urologic Diagnostic EXpert system). Obtuvieron una sensibilidad del 90 % con una especificidad del 64 %.

En una publicación más reciente [Beer 01], se objetiva la escasa correlación de su sistema AUDEX, con la confirmación histológica en piezas de prostatectomía radical. El sistema fue evaluado en 62 pacientes previamente a someterse a prostatectomía radical, encontrando una sensibilidad del 85 %, pero con una especificidad muy baja (18 %). Estos valores de sensibilidad y especificidad, tal y como se explica en la Sección 4.8, corresponden a un punto en la curva ROC prácticamente sobre la diagonal, lo que equivale a una clasificación aleatoria. Los autores concluyen diciendo que la aplicación de su sistema en una tarea real carece de utilidad clínica.

3.4. Detección de cáncer de mama

El cáncer de mama es, tras el de pulmón, el tipo de cáncer que tiene mayor índice de mortalidad entre las mujeres. En cuanto a incidencia, ocupa también el segundo lugar tras el cáncer de piel. La probabilidad de que una mujer desarrolle un cáncer de mama invasivo (tumor en estado avanzado) durante su vida está entre el 8 y el 13 % [Chen 05]. Aunque con mucha menos frecuencia, el cáncer de mama también se produce en los varones, a los que corresponde aproximadamente el 1 % de los cánceres de mama diagnosticados.

Al igual que ocurre con otros tipos de tumores, las posibilidades de curación del cáncer de mama aumentan espectacularmente si éste se detecta en sus primeras fases. Dependiendo de la severidad del cáncer (tamaño y localización) éste se clasifica en distintas fases o estadios. En el caso del cáncer de mama, estas fases van de la 0 (cáncer incipiente) hasta la *IV* (cáncer avanzado con metástasis). En la Tabla 3.1 se muestran los porcentajes de supervivencia tras los 5 primeros años para cada una de estas fases. Como puede observarse, las probabilidades de curación decrecen drásticamente a medida que aumenta la severidad del cáncer.

Tabla 3.1: Índice de supervivencia tras los 5 primeros años de haberse detectado un cáncer de mama (*Fuente: American Cancer Society*).

Fase	Supervivencia a 5 años vista (%)
0	100
<i>I</i>	98
<i>IIA</i>	88
<i>IIB</i>	76
<i>IIIA</i>	56
<i>IIIB</i>	49
<i>IV</i>	16

A la vista de los resultados, parece claro que el mejor método para combatir esta enfermedad es el diagnóstico precoz. Es por ello por lo que, cada vez con

mayor esfuerzo, los sistemas sanitarios de cada país en particular, intentan poner a disposición de la población los medios necesarios para el diagnóstico temprano. En este sentido, los programas de *screening* recomiendan a las mujeres mayores de 40 años realizarse anualmente una mamografía. Este test de imagen permite a los radiólogos detectar un número importante de tumores en sus primeras fases aunque, como se verá más adelante, el diagnóstico no tiene, ni mucho menos, una fiabilidad total.

3.4.1. Métodos actuales en el diagnóstico del cáncer de mama

Actualmente, los métodos más utilizados en el diagnóstico precoz del cáncer de mama son la autoexploración y la mamografía. Mediante la autoexploración sistemática, es posible detectar bultos u otros cambios en el seno, que puedan indicar la presencia de un tumor. Además de la autoexploración, se recomienda que las mujeres con factores de riesgo se realicen con cierta periodicidad una mamografía (normalmente cada 1 o 2 años).

La mamografía es un test de imagen en el que se utilizan rayos X de baja radiación y películas de alta resolución. Hoy en día este test es la técnica más efectiva, de bajo coste y alta sensibilidad para el diagnóstico del cáncer de mama [Chri 02]. Permite detectar entre un 85 y un 90 % de los tumores, antes incluso de que estos puedan detectarse por palpación. En la Figura 3.3 se muestra una mamografía en la que se ha resaltado una región afectada por un tumor maligno. Una anomalía puede ser detectada, bien porque en la propia mamografía se observe una región sospechosa, bien porque se encuentren diferencias significativas entre las mamografías de ambos pechos o entre mamografías de revisiones anteriores. De un modo u otro, cuando se detecta una anomalía, se realizan otros test de imagen para obtener un diagnóstico más preciso. Dichos test de imagen pueden consistir en una nueva mamografía más detallada (nuevos ángulos, o mayor detalle en ciertas áreas), una ecografía o, de manera menos frecuente, una resonancia magnética, o una termografía. En la Tabla 3.2 se muestran los distintos tipos de anomalías que pueden observarse en una mamografía y ser indicativos de la presencia de un tumor maligno.

La ecografía permite generar una imagen a partir del eco recibido de una secuencia de ultrasonidos. Una ventaja de la ecografía frente a la mamografía es que, la primera, puede tomarse prácticamente desde cualquier ángulo. Otra gran ventaja de esta técnica radica en que permite distinguir los quistes (tumores benignos formados por líquido) de las masas sólidas (normalmente nódulos cancerosos). En la mamografía, por contra, las masas y los quistes tienen un aspecto muy parecido. La desventaja de la ecografía es que, aunque ofrece un buen contraste, ofrece una resolución espacial muy pobre.

La Resonancia Magnética Nuclear (RMN) utiliza un potente campo magnético

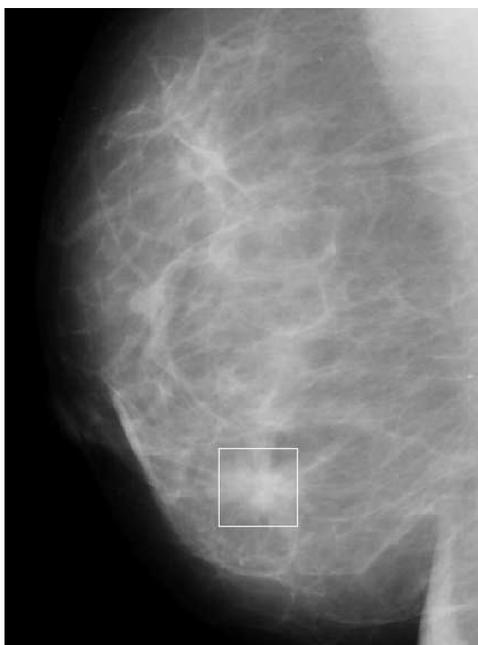


Figura 3.3: Ejemplo de mamografía. La región enmarcada en el cuadrado sobreimpreso corresponde a un tumor (en este caso una masa radial). Puede observarse el gran parecido que guarda con otras regiones de la imagen correspondientes a tejido sano.

Tabla 3.2: Clasificación de los distintos tipos de anomalías observables en una mamografía que pueden ser indicativos de la presencia de un tumor maligno [Suck 94].

Tipo	Apariencia
Microcalcificaciones	Agrupación de pequeños puntos de gran intensidad
Masas circunscritas	Región más o menos circular con alto brillo
Masas radiales	Región con alto brillo y forma estrellada
Otras masas poco definidas	Región con alto brillo y bordes poco definidos
Distorsiones	Distorsión de la estructura normal de la mama
Asimetrías	Diferencias entre la apariencia de ambas mamas

y ondas de radio. Los distintos tejidos absorben y reflejan las ondas de radio con distinta intensidad en función de su densidad y contenido en agua, de manera que a partir del eco producido por dichas ondas, es posible reconstruir una imagen muy precisa de los tejidos explorados. Adicionalmente, es posible inyectar al paciente un contraste para obtener imágenes de mayor calidad. La RMN permite modificar las direcciones del campo magnético y de las ondas de radio, para obtener vistas en cualquier plano. Mediante la RMN es posible detectar tumores de muy pequeño tamaño, que no se pueden observar mediante mamografía o ecografía. La RMN es

igualmente de gran utilidad a la hora de determinar el estadio en el que se encuentra un tumor. Además, en las mujeres que tienen algún implante o un pecho más denso, la RMN ofrece mejores resultados que la mamografía, ya que tanto la silicona como los tejidos más densos impiden el paso de los rayos X, ensombreciendo parte del tejido mamario. La gran desventaja de la RMN es su elevado coste y su baja especificidad (prácticamente no distingue los tumores malignos de los benignos).

La termografía permite detectar temperaturas anormales en ciertas regiones de la mama, lo que a su vez puede ser indicativo de la presencia de un tumor. Aunque actualmente la termografía clásica no está ampliamente aceptada como un método eficaz para la detección del cáncer de mama, la termografía computerizada es hoy en día objeto de estudio. Para realizarla utiliza una cámara con un sensor térmico que permite capturar una imagen digital basada en el calor irradiado por los distintos tejidos. Dicha imagen es posteriormente procesada por un ordenador que analiza los patrones de calor y resalta las áreas con alta sospecha de contener células cancerosas [Friz 03]. Con ello, el radiólogo puede decidir si es necesario algún otro test. La termografía computerizada es un método totalmente inocuo, ya que no requiere aplicar ningún tipo de radiación. Ofrece una alta sensibilidad, aunque menor especificidad. Algunos estudios han concluido que la termografía computerizada y la mamografía ofrecen resultados similares en cuanto a sensibilidad y especificidad se refiere [Head 99]. A pesar de ello, la mamografía sigue siendo hoy en día el método de imagen utilizado en primera instancia para la detección del cáncer de mama.

Independientemente del test de imagen que se utilice, si se confirma o revela una anomalía, se debe practicar una biopsia en la zona sospechosa para poder determinar si existe o no cáncer de mama. La biopsia consiste en tomar una muestra de tejido, para su posterior análisis al microscopio. Este análisis histológico permite revelar, con una fiabilidad del 100 %, si en el tejido analizado se encuentran o no células cancerosas. Para realizar la biopsia, normalmente se recurre de nuevo a la mamografía o a la ecografía, como ayuda para dirigir la aguja de la biopsia a la zona del pecho deseada.

3.4.2. Limitaciones actuales al diagnóstico precoz del cáncer de mama

La mamografía es uno de los métodos más fiables para la temprana detección de carcinomas. Sin embargo, es difícil para un radiólogo proporcionar un diagnóstico preciso, debido, principalmente, a la gran variabilidad que puede darse en la apariencia, tanto del tejido sano, como del canceroso. Estudios retrospectivos han demostrado que, en los programas actuales de *screening* de cáncer de mama, entre un 10 % y un 25 % de los tumores pasan inadvertidos a los ojos del radiólogo [Brak 98a, Wall 91]. Otros estudios confirman estos mismos datos al estimar que la sensibilidad alcanzada en la detección de tumores mediante la inspección visual de la mamografía por parte de un radiólogo no supera el 75 % [Chen 05].

Esta falta de precisión en el diagnóstico puede atribuirse, además de a la dificultad intrínseca que entraña la localización de tumores, tal y como acaba de explicarse, al cansancio acumulado tras largas sesiones de exploración radiológica, la subjetividad en el diagnóstico, la ocultación de las lesiones en el caso de pechos más densos o con implantes y la inexistencia de una búsqueda sistemática y exhaustiva.

Algunos estudios han demostrado que la habilidad para localizar patrones en una imagen que tienen cierta similitud con el fondo de la misma (tal y como ocurre con los tumores en una mamografía) requiere una búsqueda sistemática por toda la imagen [Trei 85]. Sin embargo, otros estudios que analizan el movimiento del ojo durante el análisis de una mamografía, han demostrado que el modo en que el radiólogo explora la imagen dista mucho de ser sistemático [Krup 96].

Por otro lado, tal y como se ha expuesto anteriormente en la Tabla 3.2, existen distintos tipos de anomalías, siendo algunas de ellas de más difícil localización. Por ejemplo, resulta más complicada la detección de masas que la de microcalcificaciones, ya que las primeras pueden tener características similares al tejido mamario normal. Las masas a menudo aparecen en áreas más densas de la mama, por lo que su apariencia se mimetiza aún más con el tejido normal. Además, sus bordes son más suaves que en el caso de las microcalcificaciones y sus formas más variadas: circunscritas, radiales, lobuladas o con otras formas no definidas. Por último, pueden presentar tamaños muy variados (entre 2 y 30mm de diámetro) [Chri 00].

De las biopsias realizadas tras un test de imagen sospechoso, entre el 65 y el 90 % resultan ser benignas (falsos positivos) [Chen 05, Chri 02]. Por otro lado, estudios retrospectivos han demostrado que entre un 10 y un 30 % de los tumores visibles no son detectados (falsos negativos) [Mart 79, Wall 91]. Un posible método para disminuir el porcentaje de errores cometidos consiste en realizar una doble inspección de la mamografía por parte de dos radiólogos [Heat 00a]. Sin embargo, proceder de este modo supondría duplicar los gastos de personal.

A la vista de estos resultados, parece interesante desarrollar sistemas de CAD que permitan realizar una segunda lectura de la mamografía y, consecuentemente, traten de reducir, tanto el número de falsos positivos, como el de falsos negativos. La combinación de un sistema de CAD junto con el conocimiento del experto humano puede mejorar de manera significativa la precisión en la detección de tumores. Algunos estudios han estimado que la sensibilidad conseguida por el radiólogo, sin el uso de un sistema de CAD, es del 80 %, mientras que con el CAD podría alcanzar hasta el 90 % [Doi 96].

3.4.3. Diagnóstico del cáncer de mama asistido por ordenador

Un modo de superar el problema generado por la escasez de radiólogos expertos en la interpretación de mamografías podría consistir en utilizar un ordenador

para que, de manera automática y sin supervisión, filtrara aquellas mamografías que fueran inequívocamente normales y dejar para el radiólogo únicamente aquellas que presentasen cierto grado de sospecha. Esto permitiría a los radiólogos concentrar sus esfuerzos en examinar únicamente un pequeño porcentaje de mamografías. Esta solución, sin embargo, no es posible actualmente, ya que requeriría la implementación de algoritmos con una sensibilidad extremadamente alta, capaces de detectar todos los posibles tipos de anomalías. Actualmente, tales niveles de sensibilidad están lejos de ser alcanzados.

Una segunda solución más realista consiste en utilizar la técnica denominada *prompting*. En este caso, el ordenador se utiliza para detectar potenciales anomalías en la imagen digital y llamar la atención del radiólogo hacia esas zonas sospechosas, marcando de alguna manera (por ejemplo mediante un círculo sobrepuesto) dichas regiones de la imagen. En esta segunda aproximación el ordenador se utiliza, no como sustituto del experto humano, sino como sistema de ayuda al diagnóstico (CAD).

En la Sección 3.3.3 se ha puesto de manifiesto la gran escasez de estudios dedicados a la investigación del CAD aplicado al cáncer de próstata. La realidad del cáncer de mama, sin embargo, es radicalmente opuesta. En este caso hay multitud de estudios publicados, revistas y congresos orientados única y específicamente a esta área de investigación, bases de datos de mamografías a disposición de la comunidad científica e incluso sistemas de CAD actualmente en explotación (aunque, como veremos más adelante, con una mejora de la capacidad diagnóstica aún por demostrar).

Durante los últimos 25 años se ha investigado ampliamente en el desarrollo de algoritmos que permitan la detección de anomalías en las imágenes de mamografía. Algunos de estos algoritmos son muy específicos para un tipo particular de anomalía, mientras que otros son métodos mucho más genéricos que permiten la detección de distintos tipos de patrones.

Los sistemas de CAD aplicados a la mamografía han demostrado resultados prometedores. Las publicaciones realizadas están relacionadas tanto con el estudio de algoritmos para la detección de los distintos tipos de anomalías, como con el posterior análisis de las mejoras que pueden aportar los sistemas de CAD respecto a la precisión diagnóstica conseguida por el experto humano. Algunos de estos estudios, sin embargo, dudan de la aplicabilidad práctica de los sistemas de CAD, por lo que, actualmente, todavía es tema de controversia entre los expertos la utilidad de estos sistemas [Kuo 02].

Uno de los primeros estudios que evaluó el efecto del marcado de zonas sospechosas en mamografías se centraba exclusivamente en la detección de microcalcificaciones [Chan 90]. Este estudio demostró el potencial diagnóstico ofrecido por un sistema de CAD; se consiguió una mejora significativa en la detección de microcalcificaciones mostrando una sensibilidad del 87% con una media de 4 falsos

positivos por imagen.

Sin embargo, un estudio posterior más realista demostró que el marcado de zonas sospechosas apenas mejora la capacidad diagnóstica del radiólogo cuando el número de falsos positivos por imagen es elevado [Hutt 96]. En este estudio, utilizando un algoritmo capaz de detectar correctamente el 90 % de los tumores, se observó que el rendimiento del radiólogo únicamente mejoraba cuando se marcaba, a lo sumo, un falso positivo cada 3 imágenes.

El sistema de CAD para mamografías más utilizado actualmente es el *R2 ImageChecker*, con más de 1,600 unidades vendidas por todo el mundo. Este sistema se ha evaluado en muchos centros de Estados Unidos y, más recientemente, en Europa [Roeh 98]. Es capaz de detectar correctamente el 98 % de las microcalcificaciones y el 72 % de las masas, con una media aproximada de 0.6 falsos positivos por imagen.

El *R2 ImageChecker* utiliza algoritmos originalmente diseñados por N. Karssemeijer y su equipo [Kars 96, Brak 98a, Brak 99, Brak 00]. Para la detección de masas utiliza una ventana con forma anular, de diámetro interno y externo igual a 6 y 32 mm respectivamente. Este anillo se mueve por toda la imagen y, mediante la transformación RBST explicada anteriormente, localiza segmentos de mayor intensidad que radian de un origen común (lo que se correspondería con una masa radial) o un gradiente con una intensidad menor cuanto más nos acercamos a la circunferencia exterior del anillo (lo que se correspondería con una masa circunscrita).

A pesar de los buenos resultados, hasta hace poco tiempo no se había podido demostrar una mejora en el rendimiento diagnóstico de los radiólogos asistidos por este sistema. Sin embargo, en una publicación más reciente [Bird 05] se asegura que con el uso de este sistema, los radiólogos fueron capaces de incrementar en un 7.4 % la detección de tumores. En este estudio se evaluaron 8682 pacientes, detectándose algún tipo de sospecha en 863 (9.9 %). Este grupo de sospecha se redujo a 181 tras una segunda exploración. Finalmente se confirmaron mediante biopsia 29 tumores malignos. Dos de estos tumores (6.9 %) habían pasado inadvertidos a los ojos del radiólogo pero fueron detectados por el sistema de CAD.

Otro sistema de ayuda al diagnóstico de cáncer de mama es el PROMAM (*PROMpting for MAMmography*). Este sistema marca correctamente el 94 % de las microcalcificaciones y el 73 % de las masas [Will 98]. No se deja claro, sin embargo, el número medio de falsos positivos por imagen. Este sistema se basa en los algoritmos desarrollados por el grupo de Karssemeijer para la detección de microcalcificaciones y en los algoritmos desarrollados por L. Miller del *Royal Observatory of Edinburgh* (ROE) para la detección de masas. Ningún estudio ha demostrado hasta el momento una mejora en la capacidad diagnóstica del radiólogo cuando utiliza este sistema de ayuda.

El sistema *Second Look* se encuentra actualmente bajo evaluación. En un estu-

dio que llevó a cabo la misma compañía propietaria del sistema (CADx Medical) realizado sobre 1,300 mamografías normales y otras 1,200 con algún tumor maligno confirmado por biopsia, se alcanzó una sensibilidad del 85% con una media de 1.4 falsos positivos por imagen. Sin embargo, en otro estudio más objetivo e independiente, estas tasas se reducen a una sensibilidad del 75.4% con 1.4 falsos positivos por imagen [Comp 01]. No se tienen datos sobre el efecto que el uso de este sistema tiene sobre el rendimiento diagnóstico de los radiólogos.

MedDetect es uno de los sistemas de análisis de imágenes médicas desarrollados más recientemente. Realiza, en tiempo real, extracción de características en un dominio invariante a la escala y a la rotación. Para el procesamiento de mamografías utiliza una arquitectura que divide el proceso en dos etapas. En una primera etapa, mediante el uso de *wavelets*, detecta y segmenta las ROIs sospechosas. En la segunda etapa, mediante el uso de una red neuronal, analiza las características extraídas de cada ROI para determinar la probabilidad de que la anomalía detectada sea realmente un tumor maligno. Este sistema ha sido desarrollado en la *University of South Florida* (USF) y utiliza los algoritmos desarrollados por L. Clarke y su equipo [Wood 91, Kall 92, Li 99].

Capítulo 4

Técnicas empleadas en la clasificación y detección de patrones en imágenes médicas

4.1. Introducción

Analizar imágenes médicas, ya sean ecografías, radiografías o representaciones de cualquier otra índole, puede ser una tarea complicada para un experto humano. Lo acertado del diagnóstico depende del entrenamiento acumulado, la experiencia y en muchos casos de criterios subjetivos. Incluso expertos con un nivel alto de entrenamiento pueden tener una gran variación en sus diagnósticos [Skaa 97]. El diagnóstico asistido por ordenador (CAD), mediante el empleo de técnicas de visión por computador y reconocimiento de formas, puede ayudar al experto humano en la interpretación de las imágenes médicas y permitir así un diagnóstico más preciso.

En este capítulo se revisan distintas técnicas de reconocimiento de formas empleadas en el análisis de imágenes médicas. Estas técnicas son evaluadas de forma experimental en capítulos posteriores. En la Sección 4.2 se detallan las distintas aproximaciones empleadas en la extracción de características. En la Sección 4.3 se explica el análisis de componentes principales (PCA) como técnica de extracción de características y de reducción de la dimensionalidad de las mismas. En la Sección 4.4 se explican los conceptos de características locales y globales. En la Sección 4.5 se propone un método de análisis multiescala para resolver el problema de la invarianza al escalado. En las Secciones 4.6 y 4.7 se explican técnicas de clasificación mediante k -NN y HMM respectivamente. Finalmente, en la Sección 4.8 se explica el uso de las curvas ROC como herramienta de evaluación del rendimiento de un clasificador binario.

4.2. Extracción de características

La información original contenida en una imagen digital es el color o nivel de gris existente en cada uno de los píxeles que conforman la misma. Normalmente, los equipos utilizados para la adquisición de imágenes médicas generan imágenes en escala de grises (radiografía, ecografía, resonancia magnética, tomografía) para lo que se utilizan habitualmente 256 niveles de gris distintos. Algunos test de imagen, como es el caso de la termografía, generan imágenes a color, de modo que distintos colores codifican o representan distintos rangos de temperatura. Sin embargo, esta codificación es únicamente un modo de presentación de la imagen más intuitiva al ojo humano, ya que cuantitativamente no aporta más información que una imagen en escala de grises. Por tanto, en caso de que una imagen de este tipo tuviera que ser analizada mediante un computador, bien podría codificarse igualmente mediante niveles de gris, ya que en ningún caso se utilizan más de 256 colores distintos. Consideraremos, por tanto, que no se produce ninguna pérdida de información si generalizamos la representación mediante niveles de gris para cualquiera de estas imágenes médicas. En cuanto al número de píxeles contenidos en cada imagen dependerá, obviamente, de la resolución utilizada durante la adquisición o digitalización de la misma.

Para poder modelar las distintas clases de nuestro problema (los distintos tipos de tejido, en el caso de imágenes médicas), se necesita de alguna técnica que sea capaz de extraer la información útil o relevante (características) de cada una de estas clases. En este sentido, un clasificador estadístico, como lo son los utilizados en este trabajo, requiere que cada clase sea representada mediante un conjunto de prototipos. Cada prototipo contiene los valores de n características de la clase que representa y puede expresarse como un punto en un espacio n -dimensional. Además, para que un posterior proceso de clasificación sea exitoso, el conjunto de características escogido debe de ser representativo en el sentido de que se minimicen las diferencias intraclase y se maximicen las diferencias interclase.

La extracción de características es, básicamente, un proceso de extracción de la información útil contenida en la señal a tratar (las imágenes, en nuestro caso). Los algoritmos de extracción de características, por tanto, crean nuevas características a partir de transformaciones y/o combinaciones de las características del conjunto original. La elección de un determinado método de extracción de características (o la combinación de varios de estos métodos) es una decisión que tendrá gran repercusión en las prestaciones del sistema.

En el caso de trabajar con imágenes médicas, y más concretamente con detección de lesiones, lo habitual será tratar de diferenciar únicamente entre dos clases: lesión y tejido normal o sano. Aunque tener que discriminar únicamente entre dos clases pueda parecer una reducción de la complejidad del problema, lo cierto es que éstas, especialmente la correspondiente a tejido sano, pueden tener apariencias muy dispares, por lo que lograr una buena representación de ambas puede llegar a

ser una tarea de una complejidad considerable.

En un sistema de reconocimiento de formas, la etapa de extracción de características tendrá que aplicarse, utilizando exactamente los mismos criterios y algoritmos, en dos fases distintas. En primer lugar, dado un conjunto de muestras (imágenes) de entrenamiento, tras el preproceso de las mismas, el algoritmo de extracción de características obtendrá un conjunto de prototipos para cada una de las clases que se desean modelar. Posteriormente, ante una muestra de test (imagen) no vista anteriormente, se obtendrán las características del mismo modo a como se hizo con las muestras de entrenamiento. Aplicando entonces alguna regla de decisión se estimará la clase a la que, con mayor probabilidad, pertenece la nueva observación.

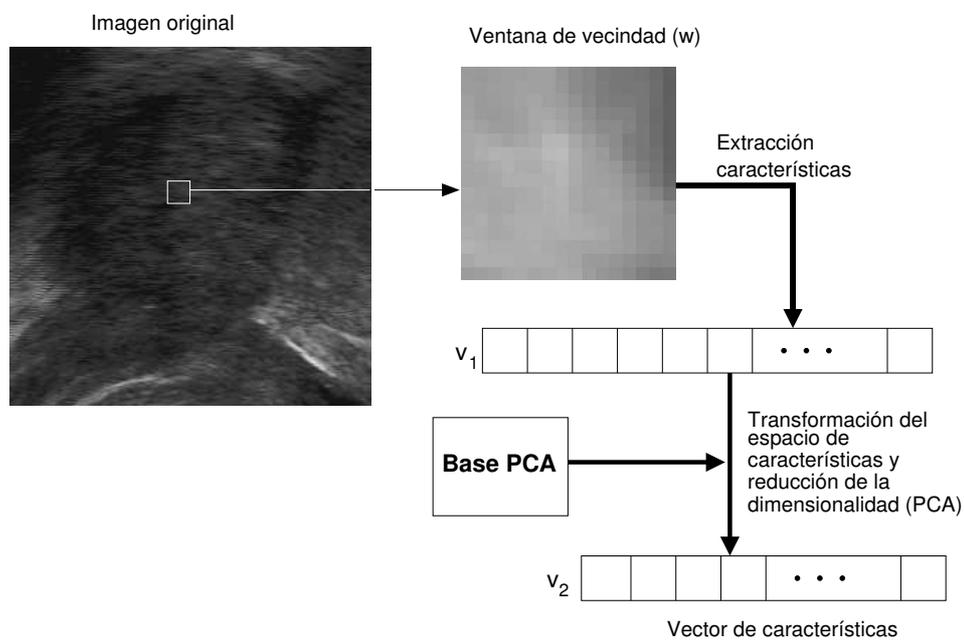


Figura 4.1: Esquema de extracción de características. Una ventana local w se desliza a lo largo de la imagen original, obteniéndose un vector de características v_1 a partir de los píxeles contenidos en w . Finalmente, mediante el uso de PCA, se obtiene un nuevo vector de características v_2 de menor dimensionalidad.

Para modelar las dos clases de nuestro problema (lesión y tejido sano), se obtienen los vectores de características de cada píxel de la imagen mediante una ventana de vecindad, tal y como se muestra en la Figura 4.1. A estos vectores de características se les asocia una etiqueta de clase, correspondiente a la clase de tejido que representan. Para obtener el vector de características de un píxel cualquiera de la imagen, se define una ventana de vecindad w centrada en dicho píxel. A continuación, aplicando localmente algún algoritmo de extracción de características a dicha ventana, se obtiene un vector de características v_1 representativo de los pí-

xeles contenidos en w . Finalmente, tal y como se explica en la Sección 4.3.3, se realiza una transformación del espacio de características sobre todos los vectores v_1 mediante PCA, y se generan vectores v_2 de menor dimensión.

En las siguientes secciones se presentan los distintos métodos de extracción de características utilizados en este trabajo. El primero de ellos y más sencillo se basa en utilizar directamente los niveles de gris de la imagen original. En el siguiente se emplea un filtro de Sobel para realzar los bordes o contornos de los objetos. En un tercer método se utilizan matrices de coocurrencia para obtener parámetros de texturas. Finalmente, los dos últimos métodos presentados (AFUM y SFUM) son métodos basados en el gradiente.

4.2.1. Niveles de gris

Para caracterizar los distintos píxeles de la imagen a analizar, lo más sencillo es utilizar directamente los valores de nivel de gris de dichos píxeles. Si se utiliza el método de extracción de características explicado en la sección anterior y mostrado en la Figura 4.1, entonces el vector de características asociado a cada píxel de la imagen estará compuesto por los niveles de gris de cada uno de los píxeles encerrados en la ventana de vecindad w .

La caracterización de una imagen mediante los niveles de gris de sus píxeles, aun siendo un método simple, ofrece buenos resultados cuando se combina con PCA.

4.2.2. Filtro de Sobel

Una forma habitual de mejorar o transformar una imagen digital es mediante la aplicación de operadores espaciales o filtros. Para calcular el nuevo valor de cada píxel de la imagen, dichos operadores tienen en cuenta, tanto el valor del propio píxel, como el de sus vecinos. Para ello, la imagen es transformada mediante convoluciones realizadas con unas matrices denominadas máscaras espaciales o *kernels*, de modo que cada píxel de la imagen resultante se obtiene mediante una combinación lineal de sus píxeles vecinos en la imagen original. Es precisamente la máscara, la que contiene los coeficientes de la combinación lineal.

La respuesta de estos operadores de convolución espacial dependerá, lógicamente, de los coeficientes utilizados en la máscara K . Un caso particular es el operador o filtro de Sobel, el cual mide el gradiente sobre una imagen y enfatiza aquellas regiones con una alta frecuencia espacial. Este filtro se utiliza típicamente para realzar los bordes de los objetos que se encuentran en una imagen.

Para implementar el filtro de Sobel se definen dos máscaras que detectan, respectivamente, los cambios horizontales y verticales en la imagen. Una de las más-

caras es, simplemente, la otra girada 90° . Este filtro puede ser visto como una aproximación sencilla al concepto de gradiente con suavizado. El suavizado en la dirección del eje x se realiza mediante la máscara de convolución $[1, 2, 1]$, mientras que el filtro derivativo se realiza mediante la máscara de convolución $[-1, 0, 1]$. De este modo, la máscara de convolución resultante de combinar las dos máscaras anteriores para detectar gradientes tanto en la dirección del eje x como en la del eje y será:

$$K_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad K_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

Ambas máscaras se pueden aplicar de forma separada a la imagen de entrada para producir medidas independientes del gradiente en el eje horizontal (G_x) y en el vertical (G_y). Dichas medidas pueden combinarse posteriormente para calcular tanto la magnitud del gradiente en cada punto como su orientación.

La magnitud del gradiente puede calcularse como:

$$|G| = \sqrt{G_x^2 + G_y^2}$$

aunque típicamente se calcula con menor coste computacional como:

$$|G| = |G_x| + |G_y|$$

La orientación del gradiente viene dada por

$$\theta = \arctan(G_y/G_x)$$

4.2.3. Matrices de Coocurrencia

En el análisis estadístico de texturas, la extracción de características se realiza a partir de la distribución estadística con la que se observa combinaciones de determinadas intensidades en posiciones relativas de la imagen. Según el número de puntos (píxeles) considerados en cada combinación, se obtienen estadísticos de primer orden, de segundo orden o superiores.

Las matrices de coocurrencia, también llamadas Matrices de Dependencia de Niveles de Gris o *Space Gray Level Dependence Matrices* (SGLDM) son un método para obtener características de texturas de segundo orden. Esta aproximación ha sido utilizada en muy diversas aplicaciones [Bass 93, Laye 90, Laye 91, Llob 03, Llob 05].

Básicamente, una matriz de coocurrencia es una matriz cuadrada, en la que el número de filas y columnas coincide con el número de niveles de gris (G) en la

imagen a analizar y donde cada elemento de la matriz $P(i, j)$ contiene la frecuencia relativa con la que dos píxeles de la imagen I , con intensidades i y j respectivamente y separados por una distancia d y ángulo θ , ocurren en una determinada vecindad. Dicho de otro modo, el elemento $P(i, j|d, \theta)$ contiene la probabilidad de que, dos píxeles cualesquiera a una distancia d y ángulo θ , tengan respectivamente niveles de gris i y j . Normalmente, para detectar posibles patrones de textura en la imagen, será necesario calcular varias matrices de coocurrencia, correspondientes a distintas combinaciones de (d, θ) .

En general, una textura suave genera matrices de coocurrencia con valores altos a lo largo de la diagonal, si el valor de d es pequeño en comparación con el tamaño de la textura. Ello es debido a que los pares de píxeles a una distancia d tendrán, en este caso, un nivel de gris similar. Por otro lado, si la imagen presenta macrotexturas y la distancia d es comparable a la escala de la textura, entonces los niveles de gris de píxeles separados por una distancia d serán bastante diferentes y, en consecuencia, los valores de la diagonal serán más bajos y aparecerán valores elevados en otros puntos de la matriz.

Cuando se trabaja con matrices de coocurrencia, si se utiliza una cantidad elevada de niveles de gris en la imagen, pueden producirse varias situaciones no deseables: por un lado podría darse la situación paradójica de que las matrices de coocurrencia fueran de mayor tamaño que la propia imagen original. Por otro lado, las matrices de coocurrencia pueden ser muy sensibles al tamaño de la ventana de vecindad cuando se trabaja con un valor elevado de niveles de gris. Finalmente, es recomendable que las matrices de coocurrencia tengan un grado alto de ocupación o, dicho de otro modo, que no sean excesivamente dispersas. Es evidente que cuanto mayor sea la cantidad de niveles de gris considerados, más dispersa será la matriz de coocurrencia.

Estos problemas podrían minimizarse reduciendo la cantidad de niveles de gris, o aumentando el tamaño de la ventana de vecindad. La primera solución produce una pérdida de precisión en los descriptores de textura, especialmente cuando se trata de analizar texturas de baja dinámica. La segunda solución causa incertidumbre si la textura cambia a lo largo de una ventana de vecindad demasiado grande. Un compromiso típico consiste en utilizar entre 16 y 20 niveles de gris, y ventanas de vecindad cuadradas de entre 30 y 50 píxeles de lado [Albr 95].

Las matrices de coocurrencia, tal y como se han descrito, no suelen utilizarse directamente como características. En su lugar, es habitual calcular una serie de estadísticos o descriptores de textura, calculados a partir de la matriz de coocurrencia. En este sentido, Haralick et al. propusieron, en su trabajo original [Hara 73], 14 descriptores. En el Apéndice A se detallan los descriptores que han sido utilizados en este trabajo.

Es importante observar que cada uno de estos descriptores pueden dar valores de magnitudes muy distintas, lo cual debe de tenerse en cuenta si se emplea un

clasificador basado en distancias. En este caso, al comparar dos vectores de características formados por varios de estos descriptores, aquellos de mayor magnitud tendrán un mayor peso a la hora de computar la distancia. Puesto que, a priori, no se tiene conocimiento de que unos descriptores sean más relevantes que otros, para evitar este inconveniente es necesario aplicar una normalización. De este modo se consigue que todos ellos tengan magnitudes similares y, consecuentemente, el mismo peso a la hora de calcular distancias.

4.2.4. AFUM

En [Heat 00b] se introduce un algoritmo denominado AFUM (*Average Fraction Under the Minimum*), útil para la detección de cierto tipo de tumores de pecho. La idea se basa en que la apariencia de dichos, vistos mediante una radiografía, consiste en una región con gran intensidad luminosa, rodeada de píxeles que van decreciendo en intensidad de forma gradual a medida que se alejan del centro del tumor.

Dado un píxel p_{ij} de la imagen a analizar, se calcula la fracción de píxeles a una distancia r_2 de p_{ij} que tienen un valor de intensidad menor que m , siendo m el valor de intensidad mínimo encontrado entre los píxeles encerrados en un círculo con centro en (i, j) y radio r_1 . Esta operación es denominada FUM (*Fraction Under the Minimum*). Dicha operación se realiza para distintos valores de r_1 y r_2 , con $r_2 > r_1$ y se calcula la media de todos los valores de FUM obtenidos (AFUM). Con esta operación se obtiene, por tanto, una única característica para cada píxel de la imagen, correspondiente al valor de AFUM calculado (Figura 4.2).

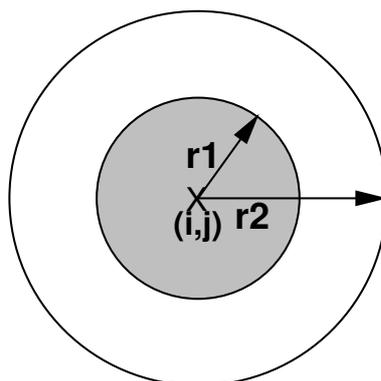


Figura 4.2: La fracción de píxeles pertenecientes a la circunferencia r_2 cuyo valor se encuentra por debajo del mínimo encontrado entre todos los píxeles del círculo r_1 se denomina FUM (*Fraction Under the Minimum*). Si se realiza esta operación para distintos valores de r_1 y r_2 y se calcula la media de todos ellos, da lugar al valor AFUM (*Average Fraction Under the Minimum*).

La ventaja de este algoritmo es que es invariante a rotaciones y no requiere el entrenamiento de parámetros.

4.2.5. SFUM

Aunque puede ser una ventaja caracterizar mediante un único valor cada píxel de la imagen, tal y como sucede en el algoritmo AFUM descrito anteriormente, es inevitable la pérdida de información que se produce al representar un conjunto de valores utilizando únicamente su media.

Para evitar esta pérdida de información, en este trabajo se ha propuesto una variante del algoritmo AFUM descrito. En lugar de calcular la media, se propone tener en cuenta el conjunto completo de valores FUM calculados, dando lugar a lo que hemos denominado SFUM (*Set of Fractions Under the Minimum*). En este caso, para cada píxel de la imagen analizada se genera un vector de características en lugar de un único valor.

La dimensión d del vector de características dependerá, obviamente, del rango de valores tomado para los parámetros r_1 y r_2 descritos en el apartado anterior. Si tomamos ventanas de vecindad cuadradas de tamaño T_w (con T_w impar), entonces los posibles valores de r_1 son $1, 3, 5, \dots, T_w - 2$. Para cada valor de r_1 , r_2 podrá tomar los valores $r_1 + 2, r_1 + 4, \dots, T_w$. Tendremos por tanto que

$$d = \sum_{i=1}^{(T_w-1)/2} i = \frac{1}{2} * \frac{T_w - 1}{2} * \left(\frac{T_w - 1}{2} + 1 \right) = \frac{T_w^2 - 1}{8} \quad (4.1)$$

4.3. Reducción de la dimensionalidad

En algunas ocasiones puede ocurrir que el número de características obtenido para representar los objetos de nuestro problema sea excesivo de cara a una posterior etapa de clasificación. Básicamente existen dos motivos principales por los que podría ser interesante trabajar con dimensionalidades más bajas (menor número de características): reducir el coste computacional (tanto espacial como temporal) y mejorar la tasa de aciertos. Cuanto menor sea la dimensionalidad de los puntos a clasificar, más rápido será el clasificador y menos memoria requerirá. Por otro lado, como se verá más adelante, un número de características excesivamente alto puede provocar, paradójicamente, una pérdida de precisión en el clasificador. Aunque, ciertamente, reducir el número de características puede implicar una pérdida de poder de discriminación y, consecuentemente, una reducción en la precisión del clasificador, también debe tenerse en cuenta que la inclusión de características redundantes o carentes de información discriminativa puede provocar que dos muestras cualesquiera acaben siendo similares si son codificadas con un número

excesivo de este tipo de características [Wata 85]. Es necesario, por tanto, encontrar el número de características óptimo que maximice la precisión del clasificador, al tiempo que minimice los costes computacionales a valores aceptables.

4.3.1. La maldición de la dimensionalidad

El funcionamiento de un clasificador depende, entre otros factores, de la relación entre el tamaño del conjunto de muestras de entrenamiento y el número de características de dichas muestras. Por ejemplo, una técnica sencilla como *table-lookup* que particiona el espacio de características en celdas y asocia a cada una de estas celdas una etiqueta de clase, requiere un número de muestras de entrenamiento exponencial respecto a la dimensión de dichas muestras [Bish 95]. Este fenómeno se conoce como “maldición de la dimensionalidad”.

Es sabido que la probabilidad de que una regla de decisión clasifique erróneamente una muestra de test no se incrementa con el número de características, siempre y cuando las densidades condicionales de cada clase sean totalmente conocidas (o, lo que sería equivalente, que el número de muestras de entrenamiento sea suficientemente extenso y representativo de las densidades subyacentes). Sin embargo, es difícil llegar a tener un conocimiento preciso de las densidades condicionales de clase, por lo que en la práctica se observa que tener un exceso de características puede degradar las prestaciones del clasificador si el número de muestras usado para diseñarlo es pequeño en relación al número de características [Jain 82, Raud 91, Raud 97]. La causa de este fenómeno radica en que a medida que crece el número de características, la dispersión de los puntos aumenta exponencialmente, con lo que los parámetros quedan peor estimados.

4.3.2. Métodos de selección de características

Tal y como se ha explicado, usar un número excesivo de características puede degradar las prestaciones del clasificador debido al fenómeno de la maldición de la dimensionalidad. Por otro lado, es evidente que un número escaso de características puede producir una pérdida del poder de discriminación y, consecuentemente, reducir igualmente el rendimiento del clasificador. Existe la necesidad, por tanto, de obtener un conjunto de características cuanto más reducido mejor, pero que al mismo tiempo recoja la información suficiente para caracterizar las distintas clases de nuestro problema.

Dado un conjunto de características extraído de una muestra de test, los métodos de selección de características tratan de detectar y rechazar aquellas características que presentan un bajo o nulo poder discriminatorio. Con ello, además de mejorar las prestaciones del clasificador, se consigue reducir costes computacionales.

Habría que aclarar la diferencia entre selección y extracción de características. El término selección de características se refiere a algoritmos que tratan de seleccionar el mejor subconjunto de características entre el conjunto de características de entrada previamente obtenidas. Por contra, los algoritmos de extracción de características son aquellos que crean nuevas características a partir de transformaciones y/o combinaciones de las características del conjunto original.

Algunos métodos, como por ejemplo el análisis de componentes principales (PCA), permite transformar las características originales en un nuevo conjunto de características. En este sentido, PCA es un método de extracción de características. Sin embargo, esta técnica también permite escoger del nuevo conjunto de características, aquellas que aportan mayor información. En este trabajo se ha utilizado PCA para conseguir un mejor y más reducido conjunto de características.

4.3.3. Análisis mediante componentes principales

El análisis mediante componentes principales (PCA) es una técnica estadística muy útil, que permite encontrar correlaciones en conjuntos de datos de alta dimensionalidad y proyectar dichos datos en otro sistema de coordenadas con el fin de decorrelar los datos.

Supongamos que, siguiendo alguno de los esquemas de extracción de características explicados en la Sección 4.2, hemos obtenido un conjunto de N prototipos representados por los vectores de características $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ con $\mathbf{x}_i \in \mathbb{R}^d$ ¹. A partir de este conjunto de datos podemos construir la matriz de covarianzas S_c del siguiente modo:

$$S_c = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)^T \quad (4.2)$$

donde $\mu \in \mathbb{R}^d$ representa la media de los vectores de características originales. S_c tiene dimensión $d \times d$ y es simétrica, por tanto tendrá d vectores propios $\{\phi_1, \phi_2, \dots, \phi_d\}$ con $\phi_i \in \mathbb{R}^d$, y d valores propios $\{\lambda_1, \lambda_2, \dots, \lambda_d\}$ que cumplen la relación:

$$S_c \phi_i = \phi_i \lambda_i \quad (4.3)$$

esto es,

$$S_c \Phi = \Phi \Lambda \quad (4.4)$$

¹Esto no es aplicable al método AFUM, el cual genera vectores de características 1-dimensionales y, por tanto, carece de sentido el análisis de componentes principales

donde,

Φ es la matriz de vectores propios $\{\phi_1, \phi_2, \dots, \phi_d\}$

Λ es la matriz de valores propios

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_d \end{bmatrix}$$

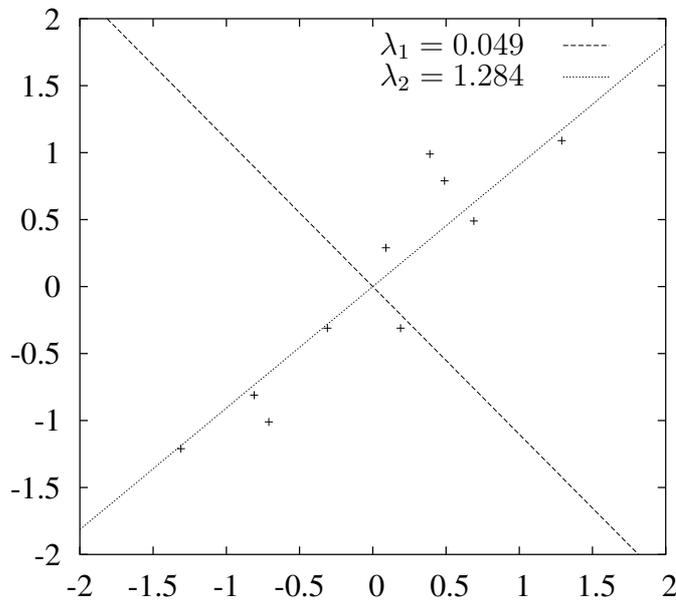


Figura 4.3: Representación de un conjunto de muestras bidimensionales. Las líneas punteadas representan los ejes de los vectores propios. El vector propio con mayor λ asociado coincide con la dirección que contiene la máxima varianza.

La matriz Φ define una transformación lineal ortogonal que no deforma el espacio original. Además, los vectores propios ϕ_i con mayores valores de λ_i asociado definen direcciones del espacio en los que la varianza de los datos es mayor. En la Figura 4.3 se muestra un ejemplo de un conjunto de vectores de características de dos dimensiones representados mediante puntos en un espacio bidimensional y de los vectores propios obtenidos de la matriz de covarianzas de dicho conjunto. Los puntos han sido previamente normalizados para que la media en cada una de las dimensiones sea cero. Puede observarse que el vector propio con mayor λ asociado representa la dirección del espacio con máxima varianza.

Si aplicamos a cada uno de los puntos \mathbf{x}_i del conjunto original la transformación

$$\mathbf{y}_i = \Phi^T \mathbf{x}_i \quad (4.5)$$

entonces los vectores de características resultantes \mathbf{y}_i (Figura 4.4) tendrán como matriz de covarianzas

$$\Phi^T S_c \Phi \quad (4.6)$$

Dado que Φ es ortogonal ($\Phi^T \Phi = I$) y usando (4.4), la nueva matriz de covarianzas será Λ , esto es, la covarianza entre dos dimensiones cualesquiera es cero. Con esta transformación se consigue decorrelar los valores de las nuevas dimensiones (que serán los nuevos vectores de características). Además, cada nueva dimensión, generada por un vector propio, tendrá como varianza el valor propio correspondiente.

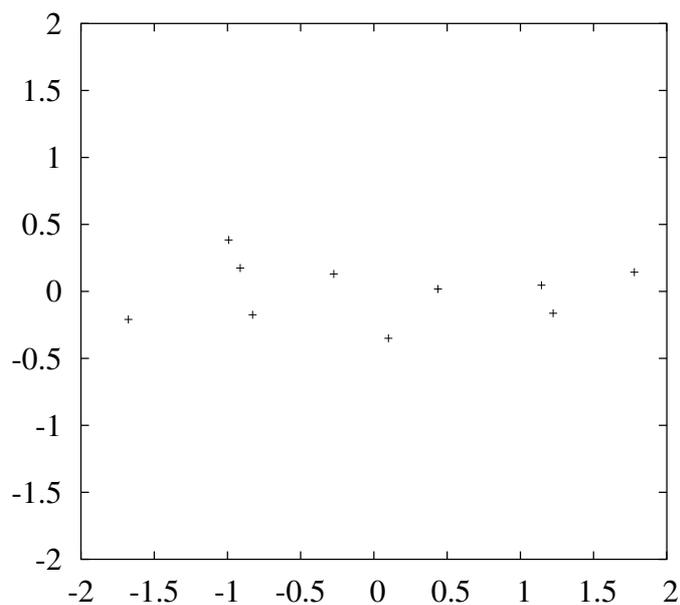


Figura 4.4: Resultado del conjunto de datos original (Figura 4.3) tras la transformación $\mathbf{y}_i = \Phi^T \mathbf{x}_i$. Los nuevos datos están completamente decorrelados (covarianzas igual a cero).

Si se define $\Phi_m = [\phi_1, \phi_2, \dots, \phi_m]$, siendo $\phi_1, \phi_2, \dots, \phi_m$ los m vectores propios con mayor λ_i dentro del conjunto Φ , se habrá escogido el subespacio lineal de m dimensiones que más cantidad de la varianza original de los datos explica. Habitualmente, el valor óptimo de m se escogerá de forma empírica.

4.4. Tamaño de la ventana de características

Sea T_w el tamaño de la ventana de vecindad w utilizada para la extracción de características (asumimos que ésta es cuadrada y T_w representa el número de píxeles de uno de sus lados). Sea T_o el diámetro, en píxeles, de los objetos o que deseamos detectar (tumores, en nuestro problema). Atendiendo a la relación entre T_w y T_o puede ocurrir que:

- a) $T_w < T_o$. Por ser la ventana utilizada para la extracción de características menor que el objeto a modelar, se obtienen características locales del objeto. En este caso, para tener una buena representación de todo el objeto, lo habitual es deslizar la ventana w por el interior de o y obtener, de este modo, varias representaciones locales del objeto.
- b) $T_w = T_o$. En este caso, por coincidir el tamaño la ventana utilizada para la extracción de características con el tamaño del objeto a modelar, se obtienen características globales del objeto.
- c) $T_w > T_o$. En este caso, la ventana de vecindad contiene, no sólo el objeto a representar, sino parte del fondo correspondiente a la clase negativa (tejido sano en nuestro problema). En esta situación, en general, no es deseable. Si tratáramos de modelar un objeto cualquiera en esta situación, para que dicho objeto quedase bien representado sería necesario modelar el mismo, no únicamente en todas las posibles variaciones del propio objeto, sino con todos los posibles fondos en cada una de sus variantes. En estas condiciones sería extremadamente complicado que el objeto quedara bien representado.

En un sentido estricto, se considera una representación global del objeto cuando T_w es exactamente igual a T_o . En este caso, el objeto queda representado por una única característica global. Cuando $T_w < T_o$, lo habitual será obtener distintas características locales del objeto colocando la ventana de características w en distintas posiciones dentro de la ventana o . Si deslizamos w a lo largo de o en todas las posiciones posibles, el número de características locales obtenidas será:

$$N = (T_o - T_w + 1)^2$$

Dicho de otro modo, podría definirse la ventana de muestreo m , de tamaño $T_m \times T_m$, como el área por la que se desplaza el centro de la ventana de características w . En el caso de que w no exceda nunca los límites de o , el número de píxeles contenidos en la ventana m coincide con el valor de N .

En la Figura 4.5 se muestra un esquema con la relación entre el tamaño del objeto y de la ventana de características. En este caso se ha fijado $T_o = 21$ y $T_w = 15$. Además se ha definido una ventana de muestreo de tamaño $T_m = 6$, con

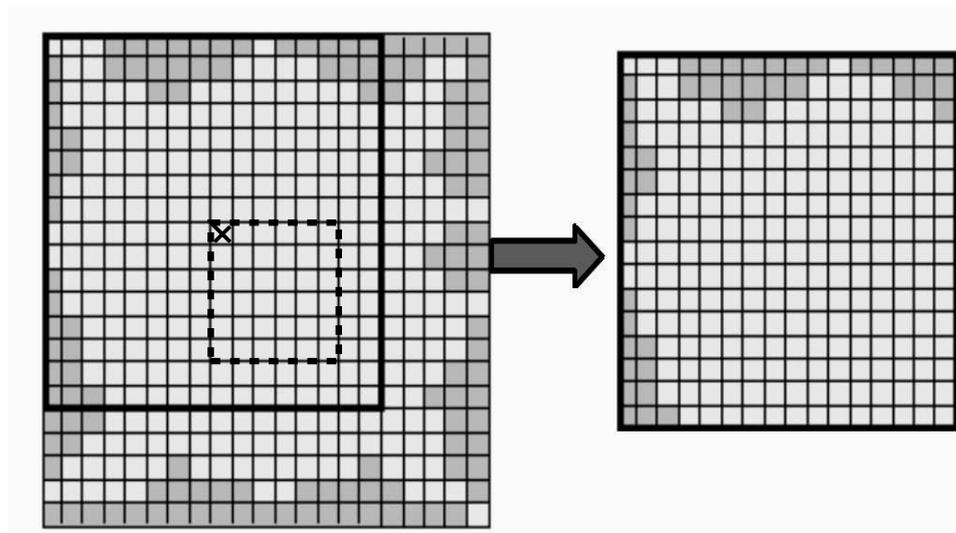


Figura 4.5: Relación entre el tamaño del objeto T_o y de la ventana de características T_w . En la izquierda se representa el objeto o , de mayor tamaño que w (resaltada mediante un recuadro negro). Desplazando el centro de w (marcado con una **X**) por el área de muestreo (recuadro punteado) se obtendrán distintos vectores de características de o (36 en este ejemplo).

lo que la ventana de características se desplazará por todas las posiciones posibles del tumor sin que sobrepase en ningún caso los límites del mismo.

Puede permitirse, por otro lado, que la ventana de características w sobrepase los límites del objeto o , esto es, que $T_m > T_o - T_w + 1$. En este caso se obtiene más información del borde del objeto. Sin embargo, como ya se ha comentado anteriormente, esta situación no será deseable si el fondo presenta apariencias muy diversas.

4.5. Esquema multiescala

Consideremos que se ha entrenado un modelo con distintos vectores de características, tanto de la clase positiva (*cáncer*) como de la negativa (*no-cáncer*), obtenidos a partir de una ventana de vecindad deslizante. Si se utiliza la regla de los k -vecinos, puede estimarse la la probabilidad a posteriori de que un píxel (x, y) de una imagen I pertenezca a la clase *cáncer*, por medio de la siguiente ecuación:

$$p(w_c|I, x, y) = \frac{\sum_{i \in s_c} \frac{1}{d(v_{xy}, n_i)}}{\sum_{i=1}^k \frac{1}{d(v_{xy}, n_i)}}, \quad (4.7)$$

donde:

- v_{xy} es el punto de test, esto es, el vector de características extraído de la imagen I en la posición (x, y) .
- $d(v_{xy}, n_i)$ es la distancia entre el punto de test y el vecino i -ésimo.
- s_c es el conjunto de subíndices de los prototipos pertenecientes a la clase *cáncer* entre los k vecinos más cercanos $n_1 \dots n_k$.

Este planteamiento implica que la invarianza al escalado debe abordarse en la fase de entrenamiento, incluyendo en el modelo muestras de tumores a distintas escalas. Sin embargo, ello implicaría que el tamaño de los tumores T_o en el modelo sería variable. Por otro lado, debe entenderse que el tamaño de la ventana de vecindad T_w debe mantenerse constante a lo largo de todo el proceso de entrenamiento y durante la posterior clasificación, con el objeto de ser coherentes en todo momento con el espacio de características empleado. En esta aproximación, por tanto, la relación T_w/T_o no sería siempre la misma, mezclando por tanto características locales y globales. Además, podría darse la situación poco deseable de que $T_w > T_o$, tal y como se ha explicado en la Sección 4.4.

Es necesario, por tanto, normalizar el tamaño de los objetos T_o a un valor constante K_o antes de incluir éstos en el modelo durante la fase de entrenamiento. En esta situación, nuestro nuevo modelo estimaría la probabilidad a posteriori de que, dado un píxel en la posición (x, y) de una imagen I escalada por un factor s , pertenezca a la clase *cáncer*. De modo similar a como se ha hecho anteriormente, esta probabilidad puede estimarse mediante la siguiente ecuación:

$$p(w_c|I, s, x, y) = \frac{\sum_{i \in s_c} \frac{1}{d(v_{sxy}, n_i)}}{\sum_{i=1}^k \frac{1}{d(v_{sxy}, n_i)}}, \quad (4.8)$$

donde:

- v_{sxy} es el punto de test, esto es, el vector de características extraído de la imagen I , escalada por un factor s , en la posición (x, y) .

- $d(v_{sxy}, n_i)$ es la distancia entre el punto de test y el vecino i -ésimo.
- s_c es el conjunto de subíndices de los prototipos pertenecientes a la clase *cáncer* entre los k vecinos más cercanos $n_1 \dots n_k$.

Dada una imagen de test, para escalar el tumor (si éste existe) a un tamaño igual al empleado en la fase de entrenamiento (K_o), es necesario aplicar a la imagen un factor de escala $s = K_o/T_o$. Obviamente, en la fase de test conocemos el parámetro K_o , pero no el tamaño real del tumor T_o . Este esquema requiere, por tanto, evaluar cada imagen de test con distintos factores de escala s con la intención de que, en caso de que exista un tumor en la imagen, éste aparezca en alguna de las escalas con tamaño K_o . Puesto que este proceso genera hipótesis distintas en cada escala, es necesario combinar de algún modo cada una de estas hipótesis para dar una única hipótesis final independiente de la escala. Considerando que $F_n = \{s_1, s_2, \dots, s_n\}$ es el conjunto de factores de escala evaluados, esta combinación de hipótesis puede realizarse mediante la regla del máximo, de la media y del producto, tal y como se detalla a continuación:

- **Máximo:** a cada punto (x, y) se le otorga la probabilidad máxima encontrada entre todas las escalas evaluadas.

$$p(w_c|I, x, y) = \max_{s_i \in F} (p(w_c|I, s_i, x, y)) \quad (4.9)$$

- **Media:** a cada punto (x, y) se le otorga la media de las probabilidades obtenidas en cada una de las escalas evaluadas.

$$p(w_c|I, x, y) = \frac{\sum_{s_i \in F} p(w_c|I, s_i, x, y)}{|F|} \quad (4.10)$$

- **Producto:** a cada punto (x, y) se le otorga una probabilidad calculada a partir del producto de las probabilidades obtenidas en cada una de las escalas evaluadas. Para conseguir que la probabilidad final sea cero sólo si la probabilidad en todas las escalas evaluadas es cero, y uno cuando la probabilidad en alguna de las escalas es uno, se calcula el producto del siguiente modo:

$$p(w_c|I, x, y) = 1 - \left(\prod_{s_i \in F} 1 - (p(w_c|I, s_i, x, y)) \right) \quad (4.11)$$

Finalmente, estas ecuaciones podrían optimizarse teniendo en cuenta la probabilidad a priori del tamaño del tumor $P(T_o)$. Dado que cuando realizamos el test con un factor de escala s_i estamos suponiendo que el tamaño real del tumor en la imagen de test es $T_o = K_o/s_i$, habría que dar mayor peso a aquellos factores de escala en los que $P(K_o/s_i)$ sea mayor. Con ello, las ecuaciones anteriores podrían optimizarse del siguiente modo:

■ **Máximo:**

$$\hat{p}(w_c|I, x, y) = \max_{s_i \in F} (p(w_c|I, \mathbf{s}_i, x, y) * P(K_o/s_i)) \quad (4.12)$$

■ **Media (suma ponderada):**

$$\hat{p}(w_c|I, x, y) = \sum_{s_i \in F} p(w_c|I, \mathbf{s}_i, x, y) * P(K_o/s_i) \quad (4.13)$$

■ **Producto:**

$$\hat{p}(w_c|I, x, y) = 1 - \left(\prod_{s_i \in F} 1 - (p(w_c|I, \mathbf{s}_i, x, y) * P(K_o/s_i)) \right) \quad (4.14)$$

El esquema multiescala presentado es necesario para detectar los distintos objetos que pueden aparecer en las imágenes a analizar (tumores en nuestro problema) pero, ¿qué tratamiento debe darse a la clase “no-objeto”? En este estudio, la clase “no-objeto” o clase negativa corresponde al tejido sano. Al contrario de lo que ocurre con los tumores, el tejido sano no puede acotarse y, por tanto, carece de sentido asociarle un tamaño. No parece adecuado, por tanto, emplear un esquema multiescala para la detección del tejido sano. Sin embargo, dado que las muestras de test necesariamente van a tener que analizarse a distintas escalas con el fin de detectar posibles tumores, es conveniente que el tejido sano esté correctamente modelado en cada una de esas escalas. Ello obliga a incluir en el modelo prototipos de la clase *no-cáncer* a distintas escalas.

En la Sección 6.4.5 se ofrecen resultados con distintos tamaños del conjunto de factores de escala F_n y para las distintas reglas de combinación de hipótesis expuestas.

4.6. Búsqueda rápida de los k -NN mediante *kd-trees*

En la Sección 2.3.2 se explicó la clasificación mediante la regla de los k -vecinos más cercanos. Tal y como se vio, esta regla requiere calcular las distancias

$$D(x, v_i), \quad 1 \leq i \leq N$$

esto es, la distancia entre el vector observado x y cada uno de los prototipos v_i del conjunto de entrenamiento P . Si suponemos que el coste de calcular la distancia D entre dos puntos es proporcional a la dimensión d del espacio, entonces el coste de buscar de forma exhaustiva el vecino más próximo sería $O(dN)$, siendo N el número de prototipos en P . La variante de calcular los k -vecinos más próximos en

lugar de únicamente el vecino más cercano puede computarse con el mismo coste, suponiendo que k es constante.

Uno de los requerimientos básicos para obtener buenas prestaciones con los métodos estadísticos no paramétricos, como es la regla de los k -NN, es disponer de un conjunto de prototipos extenso. En estas circunstancias, el coste asociado a la búsqueda de los k -NN puede ser excesivo. Este problema puede ser solucionado empleando métodos de búsqueda rápida del vecino más cercano. En este sentido se han propuesto diversos métodos y estructuras de datos para acelerar la búsqueda. Entre ellos cabe destacar, por su eficiencia y sencillez, la búsqueda rápida en kd -trees [Bent 75].

El kd -tree o árbol k -dimensional es un árbol de búsqueda binario que representa una subdivisión recursiva del universo d -dimensional en subespacios lineales disjuntos por medio de hiperplanos $(d-1)$ -dimensionales [Frie 77]. Estos hiperplanos de corte son perpendiculares a alguno de los ejes de coordenadas. Los subespacios definidos por una partición serán, por tanto, regiones de forma hipercúbica, denominadas habitualmente *rectángulos*. Los rectángulos asociados a los nodos hoja (nodos terminales del árbol) son conocidos como *buckets* o cubos. Cada partición o rectángulo ha de contener, al menos, un punto de datos. Los nodos interiores deben tener uno o dos descendientes y funcionan como discriminadores que guían la búsqueda. Si un nodo no tiene descendientes, entonces no es necesario asignarle un hiperplano de corte.

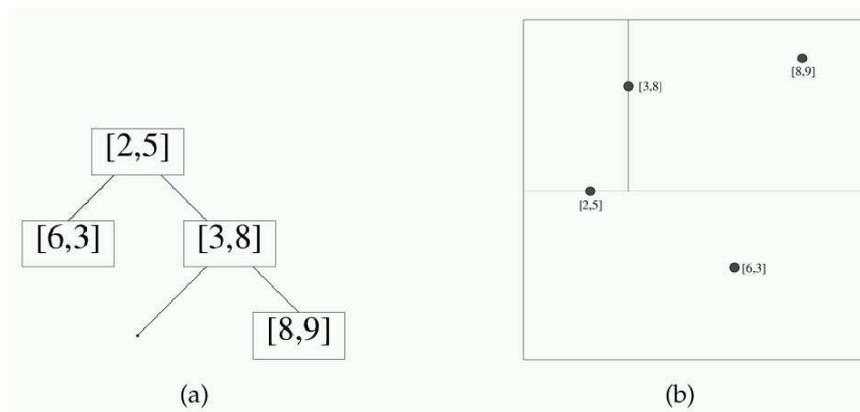


Figura 4.6: Ejemplo de una estructura de kd -tree clásica en un espacio de dos dimensiones. (a) Representación en forma de árbol. En el nodo raíz la coordenada de corte es $y = 5$. (b) Representación en forma de plano. Se pueden observar las divisiones ortogonales a los ejes y las partes del espacio donde se sitúan los puntos [Arla 04].

La Figura 4.6 muestra un ejemplo de estructura de kd -tree clásica en un espacio de dos dimensiones. Puede observarse que, dada una coordenada de corte en una dimensión (por ejemplo $y = 5$), todos los puntos con coordenada $y < 5$, están

representados en el subárbol de la izquierda, mientras que los puntos con coordenada $y > 5$ quedan representados en el subárbol de la derecha. El punto en cuestión queda representado por el nodo interior que determina la partición. Esta regla se aplica recursivamente a todos los subárboles que van creándose, hasta alcanzar la profundidad requerida en función del número de puntos que se quiera dejar en los nodos hoja (cubos). Si las particiones se realizan hasta las últimas consecuencias, en las hojas quedará un único punto. Sin embargo, es posible detener el proceso de particionar el espacio en un nivel previo, de manera que queden b puntos en cada hoja del árbol.

Búsqueda del vecino más cercano

En un *kd-tree*, la búsqueda del vecino más cercano a un punto de test comienza en el nodo raíz del árbol, el cual representa el espacio completo, y continúa descendiendo a través de sus ramas, escogiendo en cada nodo el subárbol que representa la región del espacio que contiene el punto de test. Cuando se llega a una hoja, se lleva a cabo una búsqueda exhaustiva entre los b puntos del cubo asociado a dicha hoja, seleccionándose el más cercano al punto de test. Sin embargo, es posible que entre las regiones próximas a la hoja alcanzada, haya puntos más próximos al punto de test que el seleccionado. Si no fuera así, la búsqueda podría resolverse con un coste logarítmico. En la Figura 4.7 se muestra un ejemplo que refleja esta situación. Se puede observar que el prototipo más cercano al punto de test (representado mediante una 'X') dentro del cubo, no es realmente el prototipo más cercano de todo el espacio. Sin embargo, una vez encontrado el prototipo más cercano dentro del cubo, queda claro que cualquier otro prototipo que pueda estar más próximo, debería encontrarse a una distancia menor que la existente entre el punto de test y el prototipo más cercano en el cubo, tal y como se muestra en el círculo dibujado alrededor del punto de test (y que denominaremos círculo B).

Es sencillo averiguar si existe o no la necesidad de explorar otros cubos adicionales al alcanzado en primer lugar. En este caso, el algoritmo ha de volver hacia atrás (*backtracking*) ascendiendo a los subárboles superiores e ir descendiendo de nuevo de manera recursiva, tantas veces como sea necesario hasta haber explorado todos los cubos que potencialmente puedan contener algún prototipo más próximo al actual. Este procedimiento puede ser implementado mediante un algoritmo de *ramificación y poda*.

El número de nodos y cubos visitados estará en función del tipo de algoritmo utilizado. Según las aproximaciones de Bentley et al. [Bent 80], los cubos a inspeccionar serían aquellos que intersectan con el círculo B , ya que dichos cubos podrán contener algún punto dentro de B . En [Arya 93] se sugiere una modificación a la propuesta anterior, denominada refinamiento de distancia, basada en actualizar el radio de B conforme van encontrándose prototipos más cercanos, lo que conlleva una importante reducción del número de nodos visitados.

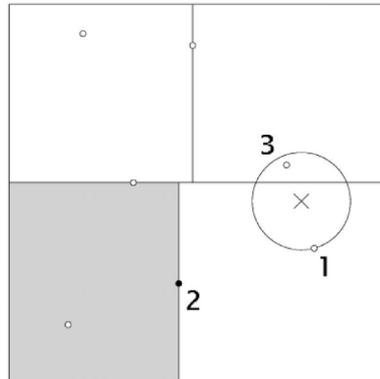


Figura 4.7: Justificación del proceso de *backtracking*. El prototipo 1 es el más cercano del cubo que contiene el punto de test (marcado con una 'X'), sin embargo el vecino más cercano es el punto 3. Para poder llegar hasta 3 es necesario subir hasta el nodo 2 (padre del 1), para después descender al 3 [Arla 04].

Coste de la búsqueda del vecino más cercano

El coste de la búsqueda en un kd -tree estará compuesto de dos términos, el descenso inicial hasta la hoja que representa la región en la que se encuentra el punto de test, más el coste del proceso de *backtracking*. Dado un kd -tree con N nodos, el coste medio asociado al descenso inicial en un árbol razonablemente equilibrado, es de $O(\log N)$. Por otro lado, en el peor de los casos, el proceso de *backtracking* puede llegar a suponer la inspección de la totalidad de los nodos del árbol. En este caso, el coste total de la búsqueda sería $O(N)$. Sin embargo, es de esperar que el coste del *backtracking* sea mucho menor en un árbol bien construido. El coste medio de este proceso es difícil de calcular ya que depende, entre otros, de la distribución de los prototipos y de la distribución de los puntos de test. En la Figura 4.8 se muestra cómo el número de hojas a visitar puede variar de manera importante, en función de la distribución de los prototipos.

El coste del proceso de *backtracking* estará en función, principalmente, de la dimensionalidad subyacente del conjunto de vectores empleados en la construcción del árbol y del grado de afinidad existente entre los puntos de test y los prototipos. El número de prototipos N no varía significativamente este coste [Spro 91, Moor 91, Zaka 96].

La búsqueda de los k -vecinos más cercanos es una extensión al algoritmo propuesto, que consiste en ampliar el radio de la hiperesfera de búsqueda B , de manera que abarque hasta el k -ésimo vecino más cercano en cada momento. El coste de este nuevo algoritmo es menor o igual que k veces el coste del algoritmo original.

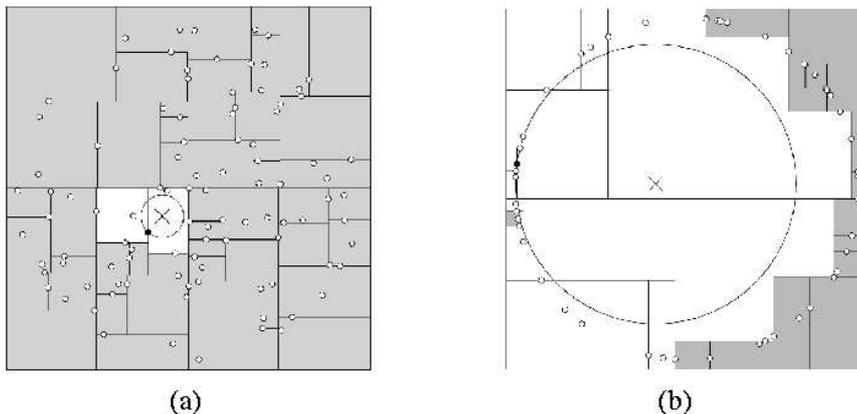


Figura 4.8: (a) Buena distribución de los prototipos. En la búsqueda de los k -vecinos más cercanos al punto de test 'X', únicamente habrá que visitar dos cubos. (b) El punto de test 'X' no se adecua a la distribución de los prototipos, por lo que será necesario visitar un número elevado de cubos [Moor 91].

Búsqueda aproximada de los k -NN

En muchos casos, no es necesario tener una garantía absoluta en la búsqueda del vecino más cercano a un punto de test, y basta con un algoritmo que encuentre, si no el más cercano, uno “suficientemente” próximo. En este sentido, se han propuesto una serie de algoritmos de búsqueda aproximada del vecino más cercano en un kd -tree.

En caso de que no sea necesaria una solución exacta, el proceso de *backtracking* descrito anteriormente, puede ser abortado cuando la mejor solución obtenida hasta el momento cumpla cierto criterio. En [Arya 98] se introduce el concepto del vecino más cercano $(1 + \epsilon)$ -aproximado. Un punto p es un vecino más cercano $(1 + \epsilon)$ -aproximado de q si la distancia de p a q no es superior a $(1 + \epsilon)$ veces la distancia de p a su vecino real r más cercano.

$$D(p, q) \leq D(p, r)(1 + \epsilon)$$

El algoritmo de búsqueda del vecino más cercano consiste en descender hasta la celda hoja que contiene el punto de test q (lo que se hace con un coste de $O(\log N)$). A continuación se realiza una búsqueda con prioridad, que consiste en enumerar las celdas hoja en orden ascendente de distancia a q e ir visitándolas ordenadamente para calcular la distancia de q al punto asociado a cada celda. En la Figura 4.9 se muestra un ejemplo del orden en que se van visitando las celdas adyacentes una vez se ha alcanzado una celda hoja. En cada momento se almacena cual es el punto más próximo a q . El algoritmo termina cuando la celda siguiente a analizar se encuentra a una distancia superior a $D(q, p)/(1 + \epsilon)$ (círculo punteado en la

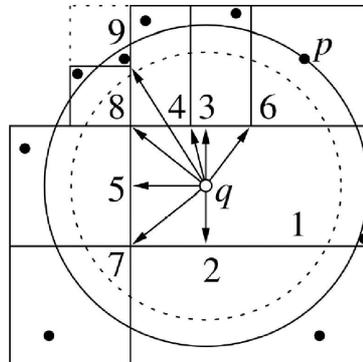


Figura 4.9: Búsqueda del vecino más cercano $(1+\epsilon)$ -aproximado de q . Se muestran las celdas más próximas visitadas por orden de distancia a q . Cuando se visita una celda cuya distancia es superior a $D(q, p)/(1+\epsilon)$, donde p es el punto más cercano encontrado hasta el momento, la búsqueda se detiene [Arya 98].

Figura 4.9). Podría ocurrir que exista algún punto m más próximo a q en alguna celda no visitada (por ejemplo el punto de la celda 9 en la Figura 4.9), sin embargo se asegura que p es el vecino más cercano $(1+\epsilon)$ -aproximado. Este punto p , como se ha dicho, estará como máximo a una distancia $(1+\epsilon)$ veces la distancia al vecino más cercano real.

La extensión de este algoritmo al problema de los k -vecinos más cercanos consiste, simplemente, en mantener una lista ordenada de los k -vecinos más cercanos a q encontrados durante la búsqueda. Cuando la distancia de q a la celda actual excede $D(q, p_k)/(1+\epsilon)$, donde p_k es el k -ésimo vecino más cercano, la búsqueda finaliza.

Los experimentos presentados en este trabajo en los que se ha utilizado un clasificador basado en k -vecinos, se han realizado usando el algoritmo de búsqueda aproximada expuesto.

4.7. Modelos de Markov aplicados a la detección de cáncer en imágenes de ecografía transrectal

Los modelos ocultos de Markov (HMM) emergieron originalmente en el dominio del reconocimiento del habla [Rabi 93], área en la que su uso sigue siendo muy común. En los últimos años, además, se ha experimentado un interés creciente en el área de la visión por computador. Concretamente, se han utilizado con éxito en tareas tan diversas como reconocimiento de texto manuscrito [Bazz 99] o reconocimiento de caras, entre otras. Dado que en los HMMs subyace una topología lineal o unidimensional, para adaptar los mismos de manera más flexible al campo

del reconocimiento de imágenes (problema basado habitualmente en un dominio de dos dimensiones), más recientemente se han estudiado extensamente modelos con topologías más complejas, como son los HMMs 2-D y los HMMs pseudo 2-D.

En este trabajo se han utilizado los modelos de Markov clásicos, con el objetivo de minimizar el problema de etiquetado explicado en la Sección 5.2.5. Para poder aplicar un modelo de Markov en una tarea determinada, es necesario calcular un vector de características en función de cierta variable independiente. Por ejemplo, en el caso de reconocimiento del habla, la señal acústica se divide en una secuencia de ventanas, obteniéndose un vector de características para cada una de estas ventanas. En este caso, la variable independiente es claramente el tiempo. En el caso de reconocimiento de texto manuscrito, la secuencia de vectores de características se calcula en función de una *posición horizontal* definida a lo largo de la línea o trayectoria que sigue el texto manuscrito.

Para poder aplicar los HMMs al caso que nos ocupa, será igualmente necesario establecer alguna variable, a partir de la cual obtener nuestros vectores de características. Para ello debemos observar que, en el problema del reconocimiento de cáncer de próstata en imágenes de ecografía transrectal, únicamente los píxeles que se corresponden con el tejido biopsiado (o cercanos al mismo) son de interés a la hora de entrenar un modelo. Dichos píxeles se encuentran dispuestos a lo largo de una línea recta cuya trayectoria coincide con la trayectoria que sigue la aguja de la biopsia al incidir sobre la glándula prostática. Esta trayectoria, cuyas coordenadas conocemos, será la variable independiente en función de la cual obtener los vectores de características. Para ello, se define una secuencia de ventanas de vecindad a lo largo de la trayectoria de la aguja. De cada una de estas ventanas, aplicando algún método de extracción de características (niveles de gris o SGLDM, en nuestro caso) se obtiene la secuencia de vectores de características necesaria para trabajar con los HMMs. En la Figura 4.10 se muestra el esquema de entrenamiento de los modelos de Markov. Dependiendo de que la secuencia de vectores de características provenga de una biopsia sana o de una cancerosa, dicha secuencia se utilizará para entrenar, respectivamente, el modelo de la clase *cáncer* o el de la clase *no-cáncer*. Dicho entrenamiento se lleva a cabo mediante el algoritmo *Backward-Forward* o algoritmo de *Baum-Welch* descrito en la Sección 2.4.7. Durante esta fase, los HMMs aprenden los parámetros que caracterizan a cada uno de sus estados.

Finalmente, dada una imagen perteneciente a una biopsia no vista anteriormente, se obtendrá la secuencia de vectores de características y se calculará la probabilidad con la que cada uno de los dos modelos genera dicha secuencia, obteniéndose un valor de confianza en función de dichas probabilidades.

En este trabajo se han utilizado HMMs continuos con topología izquierda-derecha, tanto para modelar la clase *cáncer* como la clase *no-cáncer*. Para ello, se han planteado las dos siguientes suposiciones:

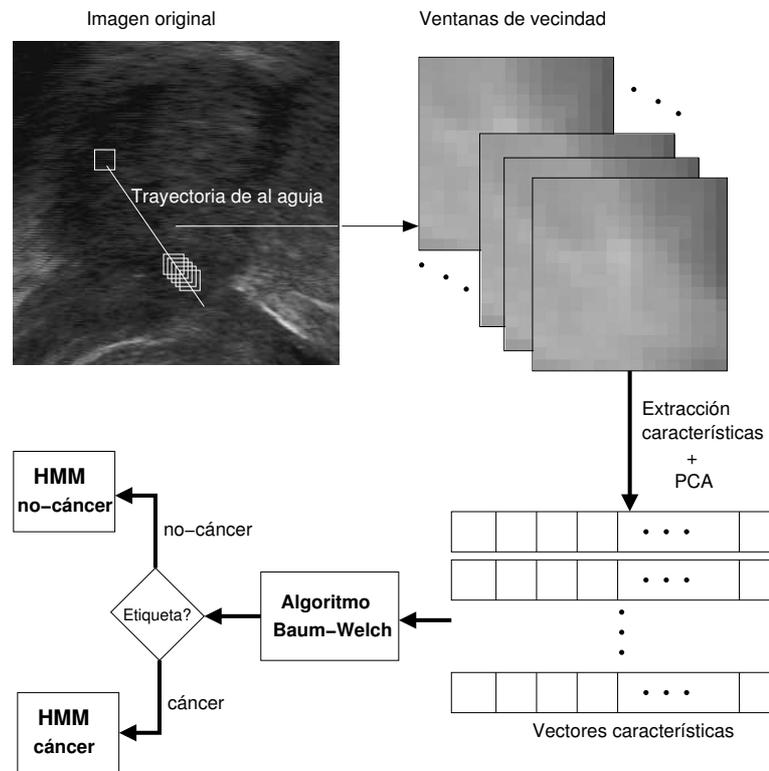


Figura 4.10: Entrenamiento de los modelos de Markov. Siguiendo la trayectoria definida por la aguja de la biopsia, se define una secuencia de ventanas de vecindad de las que se extraen los vectores de características. La secuencia de vectores de características se utiliza posteriormente para entrenar el HMM correspondiente.

- Cualquier tejido biopsiado contiene, como mucho, un tumor.
- Los tumores tienen una forma aproximadamente regular y convexa.

Estas restricciones permiten asegurar que, cuando se detecta un tumor en una biopsia, todos los píxeles correspondientes al tejido canceroso estarán conectados entre sí. Dicho de otro modo, dada una línea recta definida sobre la imagen ecográfica, no es posible encontrar píxeles correspondientes a tejido sano, que tengan a un lado y otro píxeles correspondientes a tejido canceroso. Debemos decir que estas hipótesis de partida no reducen la flexibilidad del modelo, ya que en una situación real es extremadamente difícil encontrarse con una configuración que no se adapte a la topología expuesta.

Basándonos en estas posibles secuencias de estados, lo razonable es crear un HMM que modele la clase *no-cáncer* mediante un modelo que tenga un único estado (*no-cáncer*), mientras que la clase *cáncer* se modelaría mediante un HMM de tres estados (*no-cáncer - cáncer - no-cáncer*). En la Figura 4.11 se muestra la

topología de los dos modelos propuestos.

En el caso de que el HMM que modela el tejido canceroso tuviera que dar cuenta de una imagen en la que no se observara tejido sano al principio o al final del tejido biopsiado, uno de los estados de *no-cáncer* del HMM (o ambos) se vería forzado a dar cuenta de una secuencia cancerosa. Esta secuencia, sin embargo, podría ser tan pequeña como un píxel, por lo que el coste de este desajuste sería mínimo.

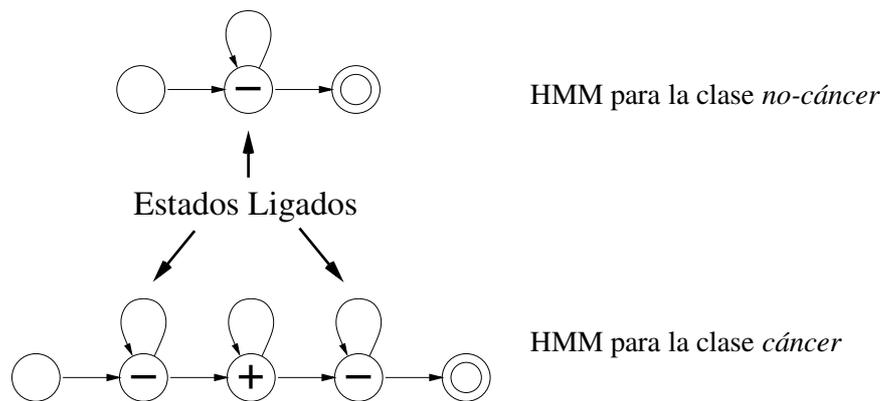


Figura 4.11: *HMMs* con topología lineal izquierda-derecha (incluyendo estados inicial y final) para las clases *no-cáncer* y *cáncer*. El HMM *no-cáncer*, con un único estado etiquetado con “-”, modela el tejido sano. El HMM *cáncer*, con tres estados etiquetados con “-”, “+” y “-” modela, respectivamente, una secuencia de vectores de características correspondientes a *no-cáncer*, *cáncer* y *no-cáncer*. Todos los estados etiquetados con “-” (*no-cáncer*) contienen exactamente los mismos parámetros, esto es, son estados compartidos.

Hay que hacer especial hincapié en que la secuencia de características que se consume/produce en cada estado del HMM es oculta. Dado un conjunto de entrenamiento, con distintos vectores de características, el HMM va especializando, iteración tras iteración, cada uno de sus estados en distintas subsecuencias de los vectores de características, sin requerirse que dichas subsecuencias tengan una etiqueta determinada. Esta característica de los HMMs se adapta perfectamente al problema que nos ocupa, ya que, tal y como se explicó en la Sección 5.2.5, desconocemos la etiqueta de cada uno de los píxeles que se encuentran en la trayectoria de la aguja, conociendo únicamente una etiqueta global para toda la secuencia.

Cada estado del HMM produce (o consume) vectores de características en función de ciertos parámetros o funciones probabilísticas; habitualmente mezclas de densidades de gaussianas. El número apropiado de gaussianas en cada estado depende, entre otros muchos factores, de la variabilidad de los vectores de características asociados a cada estado. Es lógico pensar que, cuanto mayor sea la complejidad de la realidad que se quiere modelar, mayor complejidad deberá tener el

modelo asociado. El número de densidades de Gaussianas debe ser ajustado de forma empírica en cada tarea.

4.8. Análisis de resultados mediante curvas ROC y FROC

En muchas aplicaciones de clasificación, para comparar el rendimiento de distintas aproximaciones, se utiliza habitualmente el error de clasificación (o su complementario, la tasa de aciertos). Sin embargo, esta medida no es apropiada para muchos otros casos. Imaginemos, por ejemplo, que deseamos obtener un listado de documentos relevantes relacionados con cierto criterio de búsqueda. Dicho listado puede contener gran cantidad de documentos, pero, en la práctica, sólo los 10 o 20 primeros son examinados por el usuario. Lo normal en estos casos es asignar un valor de idoneidad o *score* ρ a cada documento, para indicar cómo de relevante es dicho documento respecto del criterio de búsqueda. El objetivo a maximizar en un sistema como este no debería ser la tasa de aciertos (porcentaje de documentos que realmente guardan relación con el criterio de búsqueda), sino la correlación existente entre el valor ρ asignado a cada documento y la relevancia de dicho documento respecto al criterio de búsqueda.

De forma general, para un clasificador binario que asigne un valor real ρ a cada muestra de test, es más importante una buena correlación entre el valor de ρ asignado y la probabilidad de pertenencia a una u otra clase, que simplemente un error bajo de clasificación. Por otro lado, ofrecer el error de clasificación como medida de rendimiento de un clasificador binario, implica establecer un umbral τ a partir del cual decidir si la muestra se clasifica en una u otra clase ($\rho < \tau$ o $\rho \geq \tau$). Esto tiene el inconveniente de que, dados dos clasificadores, cada uno puede tener un comportamiento mejor que el otro para distintos valores de τ , lo que dificulta la comparación entre ambos clasificadores.

Por otro lado, en un clasificador binario puede ser interesante valorar, para cada una de las dos posibles clases, tanto el coste de clasificarlas incorrectamente como el beneficio de clasificarlas correctamente. Cuando se muestra el error de clasificación como medida de rendimiento, normalmente no se tienen en consideración estos costes, lo que invalida aún más utilizar dicha medida. Este problema podría paliarse calculando un error ponderado con los beneficios/costes de clasificar correcta/incorrectamente cada muestra de test (clasificación por mínimo riesgo). Sin embargo, en muchas aplicaciones, resulta difícil valorar cuantitativamente dichos costes y beneficios.

Las curvas ROC (*Receiver Operating Characteristic*) aportan notables ventajas frente al error de clasificación, por lo que su uso está ampliamente extendido en problemas de clasificación binarios. Un caso particular son las aplicaciones médicas. En dichas aplicaciones es habitual que, a partir del resultado de cierta prueba,

se deba emitir un diagnóstico que será, de forma genérica “paciente sano” o “paciente enfermo”.

En las siguientes secciones se explican los conceptos fundamentales sobre las curvas ROC, así como parámetros característicos de las mismas que pueden ser empleados como medida para comparar el rendimiento de distintos clasificadores.

4.8.1. Conceptos previos

Supongamos que hemos realizado un test a un paciente en el que se analiza la concentración en sangre de cierta sustancia, con el fin de determinar si el paciente ha desarrollado o no una determinada enfermedad. Para ciertas enfermedades el resultado de dicho test puede ser categórico, de modo que sea posible emitir un diagnóstico con una precisión prácticamente del 100 %. Sin embargo, en muchas otras situaciones, el resultado del test será un valor continuo ν en el que, por ejemplo, valores bajos de ν sean indicativos de que el paciente está sano, y valores altos de que se ha contraído la enfermedad. En este caso es necesario establecer un umbral o punto de corte τ para emitir uno de los dos diagnósticos posibles (sano / enfermo).

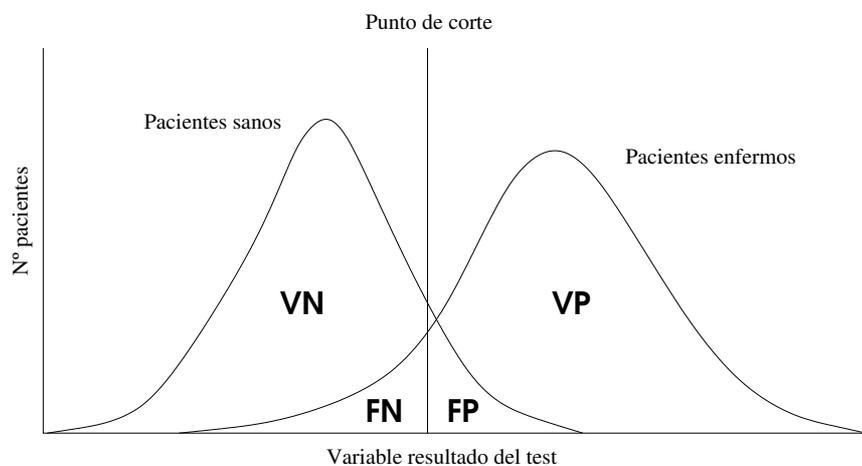


Figura 4.12: Distribución de los pacientes con y sin la enfermedad, en función del resultado de cierto test. Dado un umbral o punto de corte para la variable resultado del test, se muestran los verdaderos positivos (VP), verdaderos negativos (VN), falsos positivos (FP) y falsos negativos (FN).

Cuando consideramos los resultados de un test en particular sobre dos poblaciones, una con cierta enfermedad y otra sana, raramente se observa una separación perfecta entre ambos grupos. La distribución de los resultados, por el contrario, tendrá habitualmente un solapamiento. En la Figura 4.12 se muestra una hipotética distribución de probabilidad de las clases *sano* y *enfermo* respecto a la variable ν

de un determinado test. Debido al solapamiento existente entre la distribución de ambas poblaciones, se producirán inevitablemente errores de clasificación cuando emitamos un diagnóstico. En este caso, dado un umbral τ , la salida de nuestro clasificador será *paciente sano* si $\nu < \tau$ o *paciente enfermo* si $\nu \geq \tau$. En base a dicha clasificación, se definen los siguientes conceptos:

- **Verdadero Positivo (VP):** Paciente enfermo diagnosticado como enfermo.
- **Falso Positivo (FP):** Paciente sano diagnosticado como enfermo.
- **Verdadero Negativo (VN):** Paciente sano diagnosticado como sano.
- **Falso Negativo (FN):** Paciente enfermo diagnosticado como sano.

Si se desplaza el punto de corte τ hacia la derecha (valores mayores de ν) disminuyen los falsos positivos, pero aumentan los falsos negativos. Por el contrario, si se desplaza hacia la izquierda, aumentarán los falsos positivos al tiempo que disminuyen los falsos negativos.

Dado un clasificador (o un test clínico) y una muestra (paciente), definimos la siguiente notación:

- $T+$: resultado del test positivo
- $T-$: resultado del test negativo
- $E+$: paciente enfermo. Nótese que el número total de pacientes enfermos será $VP + FN$
- $E-$: paciente no enfermo (sano). Nótese que el número total de pacientes sanos será $VN + FP$

Podemos definir:

- **Fracción de Verdaderos Positivos (FVP):**

$$FVP = P(T+ | E+) = \frac{VP}{VP + FN}$$

- **Fracción de Verdaderos Negativos (FVN):**

$$FVN = P(T- | E-) = \frac{VN}{VN + FP}$$

- **Fracción de Falsos Positivos (FFP):**

$$FFP = P(T+ | E-) = \frac{FP}{VN + FP}$$

- **Fracción de Falsos Negativos (FFN):**

$$FFN = P(T - |E+) = \frac{FN}{VP + FN}$$

Es importante observar que la fracción de falsos negativos es 1 menos la fracción de verdaderos positivos

$$FFN = 1 - FVP$$

y de modo similar

$$FFP = 1 - FVN$$

- **Valor Predictivo Positivo (VPP):** es la probabilidad de que el paciente padezca la enfermedad cuando el test ha dado positivo.

$$VPP = P(E + |T+) = \frac{VP}{VP + FP}$$

- **Valor Predictivo Negativo (VPN):** es la probabilidad de que el paciente no padezca la enfermedad cuando el test ha dado negativo.

$$VPN = P(E - |T-) = \frac{VN}{VN + FN}$$

Los valores de VPP y VPN no deberían usarse en muestras en las que la probabilidad a priori de padecer la enfermedad haya sido controlada de forma artificial. Dichos valores no tendrían validez si para realizar el estudio se hubiesen utilizado, por ejemplo, pacientes sanos y enfermos en proporción uno a uno.

Habitualmente, en el campo de la medicina, se utilizan los términos:

- **Sensibilidad** para referirse a la Fracción de Verdaderos Positivos (*FVP*), esto es, la probabilidad de clasificar correctamente a un individuo enfermo. La sensibilidad, por tanto, mide la capacidad del test para detectar la enfermedad.
- **Especificidad** para referirse a la Fracción de Verdaderos Negativos (*FVN*), esto es, la probabilidad de clasificar correctamente a un individuo sano. La especificidad, por tanto, mide la capacidad del test para detectar los pacientes sanos.

Un test muy sensible será aquel que detecta con una alta probabilidad a todos los pacientes enfermos, aunque sea a costa de cometer más errores de falsos positivos. Por tanto, cuanto más bajo sea el umbral τ , más sensible será nuestro test.

Por contra, un test será más específico cuanto más pacientes sanos detecte, aunque sea a costa de cometer más errores de falsos negativos. Dicho de otro modo, cuanto más alto sea el umbral τ , más específico será nuestro test.

Dado un test, habrá que buscar un compromiso entre sensibilidad y especificidad. Para ello es importante valorar el coste que supone un error de falso positivo y un error de falso negativo. En algunas tareas este coste podrá medirse, de manera más o menos precisa, en términos económicos o en cualquier otra unidad de medida. Sin embargo, en el campo de la medicina, el coste que supone un error de falso positivo frente a un falso negativo es difícilmente mensurable. Un falso positivo habitualmente implica la realización de más pruebas, cuyo coste económico, por lo general, es mayor que el de la prueba que se realiza en primera instancia. Además, al paciente le ocasionará mayores molestias, a parte de la angustia que puede provocar el ser diagnosticado de una enfermedad que quizás no tenga. Por contra, un falso negativo ocasionará un retraso en la detección de la enfermedad (si es que llega a detectarse en algún momento) lo que dificultará el tratamiento de la misma. En resumen, mientras que el coste de un falso positivo se traduce en un mayor coste económico y en mayores molestias para el paciente, el coste de un falso negativo puede ser, en un caso extremo, la propia vida del paciente. Las unidades de medida en uno y otro caso son de naturaleza tan distinta, que resulta difícil, si no imposible, establecer mediante una ecuación matemática el lugar idóneo en el que establecer el umbral τ .

Una aproximación para decidir el umbral, consiste en evaluar la *utilidad* (U) del test en función de dónde se sitúe dicho umbral. Para ello se define:

- BVP: beneficio de verdadero positivo. Si la enfermedad tuviese un tratamiento eficaz y seguro, el beneficio de un verdadero positivo sería 1.
- CFN: coste de falso negativo.
- BVN: beneficio de verdadero negativo.
- CFP: coste de falso positivo.

El beneficio o coste en cada uno de estos casos dependerá de las consecuencias de la decisión y de las preferencias del paciente. Si somos capaces de evaluar cada una de estas variables, se puede calcular la utilidad esperada de la prueba como el promedio ponderado de los diferentes costes y beneficios, es decir:

$$U = p(E+)p(T+|E+)BVP - p(E)p(T-|E+)CFN - p(E-)p(T+|E-)CFP + p(E-)p(T-|E-)BVN$$

El umbral ideal sería aquél que maximizara esta función de utilidad. Sin embargo, como ya se ha comentado, resulta muy difícil, si no imposible, establecer

los costes y beneficios en cada uno de los casos descritos, por lo que en la mayoría de las tareas, el umbral acaba fijándose en base a criterios totalmente subjetivos.

4.8.2. La curva ROC

Una curva ROC muestra la fracción de verdaderos positivos (sensibilidad) en función de la fracción de falsos positivos ($1 -$ especificidad) en todos los umbrales τ posibles [Swet 79].

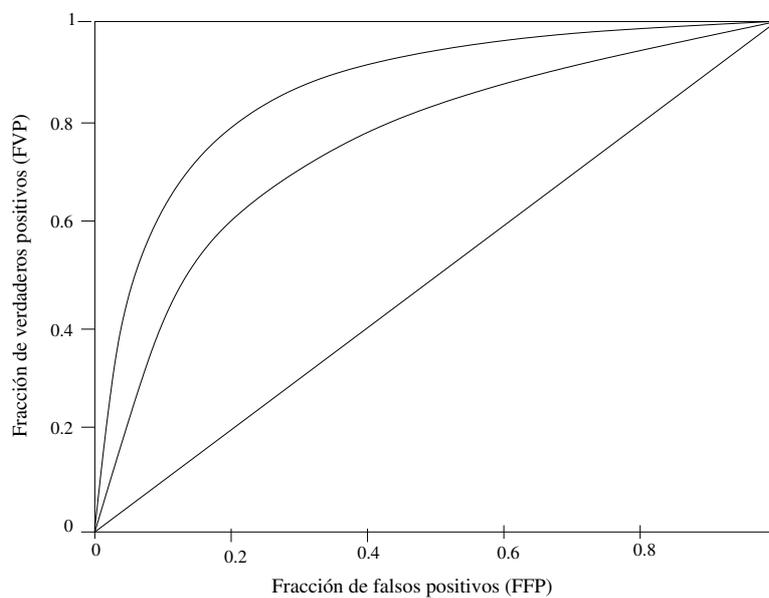


Figura 4.13: Ejemplo de dos curvas ROC. Cuanto más se aleja la curva ROC de la diagonal, mayor precisión tiene el test.

En la Figura 4.13 se muestra un ejemplo con la forma típica de dos curvas ROC. Si el umbral o punto de corte establecido para la variable del test ν es muy alto, entonces no tendremos apenas falsos positivos, pero tampoco seremos capaces de identificar apenas verdaderos positivos. Dicho de otro modo, tanto FVP como FFP estarán próximos a cero, por lo que nos encontraremos en la parte de la curva ROC más próxima al $(0, 0)$. A medida que bajemos el valor del umbral, el número de verdaderos positivos incrementará, pero también lo hará el número de falsos positivos, esto es, tanto FVP como FFP crecerán, con lo que iremos acercándonos al punto $(1, 1)$.

Obviamente, un buen test será aquel en el que la fracción de verdaderos positivos supere a la de falsos positivos para distintos umbrales. Lógicamente, cuanto mayor sea esta diferencia a favor de FVP, mejor será el test. Por tanto, cuanto más se aproxime la curva ROC al borde izquierdo y al borde superior de la gráfica, mayor precisión tendrá el test. Por contra, un clasificador que clasificase con un

criterio puramente aleatorio, produciría una curva ROC sobre la diagonal. Esto se puede entender con facilidad si se observa que, en este caso, para cualquier umbral establecido, la proporción entre muestras positivas y negativas por encima y por debajo del umbral τ sería la misma, independientemente de la probabilidad a priori de cada clase. Dicho de otro modo, $P(T+ | E+) = P(T- | E-)$, o lo que es lo mismo, $FVP = FFP$.

4.8.3. Comparación de curvas ROC. El área bajo la curva ROC.

Tal y como se ha visto en la sección anterior, cuanto más se acerca la curva ROC a la diagonal, menos útil es el test para discriminar entre dos poblaciones. Un modo sencillo de caracterizar cómo de cerca o lejos se encuentra la curva respecto a la diagonal, es simplemente calcular el área bajo la curva ROC. Cuanto más se aproxime el área a 0.5 (área por debajo de la diagonal), peor capacidad discriminativa tendrá el test. Por contra, cuanto más se aproxime a 1 (curva totalmente pegada a los bordes izquierdo y superior), mejor será el test.

La ventaja de utilizar el área bajo la curva ROC para determinar la bondad de un test, radica en que se trata de una medida no paramétrica. Es importante observar que, para construir la curva ROC, simplemente se ordenan todas las muestras en función del *score* ν obtenido en el test. A continuación, para distintos umbrales, se contabilizan las muestras positivas y negativas que hay por encima y por debajo de dicho umbral, calculándose así los valores de FVP y FFP. No es importante, por tanto, si en el ordenamiento de las muestras, éstas siguen una u otra distribución en función de ν . No importa, por tanto, la forma o la distribución de la población estudiada.

En [Hani 82] se explora el concepto de área bajo la curva ROC, demostrándose que dicha área equivale a la probabilidad de que, dadas dos muestras pertenecientes a cada una de las dos clases existentes, ambas muestras sean clasificadas correctamente.

4.8.4. Curvas FROC

Tal y como se ha explicado en las secciones anteriores, el uso de las curvas ROC es adecuado para evaluar el rendimiento de clasificadores binarios. Sin embargo, si el problema al que nos enfrentamos no es puramente de clasificación, sino de detección, entonces las curvas ROC no serán de utilidad.

Imaginemos una tarea en la que se deben detectar ciertas anomalías o lesiones en una imagen médica. En este caso no se trata simplemente de clasificar la imagen en su globalidad como *negativa* (sin lesión) o *positiva* (con una o más lesiones), sino de detectar, si las hay, un número desconocido de anomalías. Si en un problema de estas características consideramos que las muestras a clasificar son cada uno

de los píxeles que forman las imágenes de test y no las imágenes en sí, entonces nos encontraríamos ante un problema de clasificación binario, en el que cada píxel debería clasificarse como *negativo* o *positivo*. De este modo sería posible (aunque no conveniente) evaluar el rendimiento del clasificador mediante curvas ROC. El inconveniente de utilizar las curvas ROC en este caso radica en que el número de muestras de test negativas es desmesuradamente alto respecto al de positivas. Por ejemplo, en una muestra de test en la que se tuvieran 10 imágenes de 1000×1000 píxeles y entre todas ellas hubiese una única lesión de 30×30 píxeles, aproximadamente sólo 1 de cada 10000 muestras sería positiva. En este caso, probablemente, el mejor clasificador sería aquel que clasificara siempre con la clase negativa, ya que tendría una tasa de aciertos del 99.99%. Parece obvio que, en una tarea de estas características, no tiene sentido estudiar el rendimiento de un clasificador a nivel de píxel. Será necesario por tanto trabajar a nivel de imagen. En este caso el problema, como se ha visto, no será puramente de clasificación, sino de detección.

Un método para evaluar el rendimiento de un sistema de detección de objetos en imágenes, consiste en determinar el número de falsos positivos (FP) que se producen en cada imagen. Normalmente, el clasificador tendrá algún parámetro τ que permita variar la sensibilidad del mismo, de modo que podremos aumentar o disminuir la tasa de FP variando dicho parámetro (a consta, claro está, de aumentar o disminuir la tasa de verdaderos positivos).

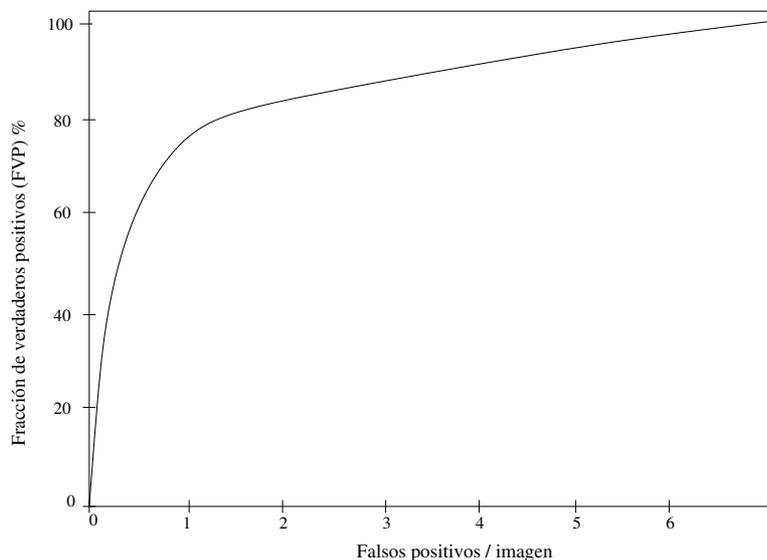


Figura 4.14: Ejemplo de curva FROC. Para un error medio de 2 falsos positivos por imagen, se detecta correctamente el 80 % de las muestras positivas

Una curva FROC (Free Response Operating Characteristic) muestra la relación entre sensibilidad (tasa de verdaderos positivos) y la tasa media de falsos positivos por imagen, para distintos valores del parámetro τ [Metz 96]. Cada valor de τ

produce un punto de operación en la curva. Este es un método muy apropiado para medir el rendimiento del sistema y encontrar un compromiso entre la tasa de aciertos y la de fallos. En la Figura 4.14 se muestra un ejemplo de una curva FROC. En este caso puede observarse que, cometiendo una media de 2 falsos positivos por imagen, se lograría detectar correctamente el 80 % de los objetos buscados.

4.8.5. Comparación de curvas FROC

Ya se ha visto que para la comparación de curvas ROC, un indicador que se utiliza habitualmente es el área bajo la curva. Cuando trabajamos con curvas FROC, sin embargo, no es trivial calcular el área que encierra la misma ya que el eje de abcisas no está acotado. En este caso, este parámetro debe calcularse normalizando el área bajo la curva FROC en función del rango en el eje de abcisas, de modo que el valor del área sea 1 cuando se consigue una tasa en los verdaderos positivos del 100 %, con cero falsos positivos por imagen. En [Brak 00] se calcula el área bajo la curva FROC utilizando una escala logarítmica en el eje de abcisas.

Capítulo 5

Aportaciones al diagnóstico de cáncer de próstata

5.1. Introducción

Tal y como se ha explicado en la Sección 3.3, para corroborar el diagnóstico de un posible cáncer de próstata, lo habitual es realizar una serie de biopsias ecodirigidas en distintas zonas de la glándula prostática, con el objeto de localizar células cancerosas mediante análisis histológico del tejido extraído. En la Figura 5.1 se muestra una imagen de la glándula prostática obtenida mediante ecografía transrectal. Debido a la escasa nitidez que ofrecen las imágenes obtenidas mediante esta técnica, resulta prácticamente imposible, incluso para un experto, localizar un hipotético tumor en dichas imágenes. Únicamente los tumores en un estado muy avanzado son fácilmente reconocibles en la imagen ecográfica.

En la práctica, llegado el momento de extraer tejido de la próstata para su posterior análisis histológico, el urólogo se guía por dos premisas:

- Biopsiar aquellas regiones que presentan una imagen ecográfica y/o tacto rectal sospechoso.
- Biopsiar aquellas regiones donde, en base a criterios puramente estadísticos, sea más probable la existencia de tumores malignos, tratando al tiempo de realizar las biopsias uniformemente distribuidas por toda la glándula.

En el primer caso se realiza la biopsia dirigiendo la punción hacia las zonas sospechosas (generalmente áreas hipoecoicas situadas en la zona periférica de la próstata). En el segundo caso, mucho más probable que el anterior, se dirige la biopsia según ciertas reglas empíricas, basadas en diversos estudios que tratan de

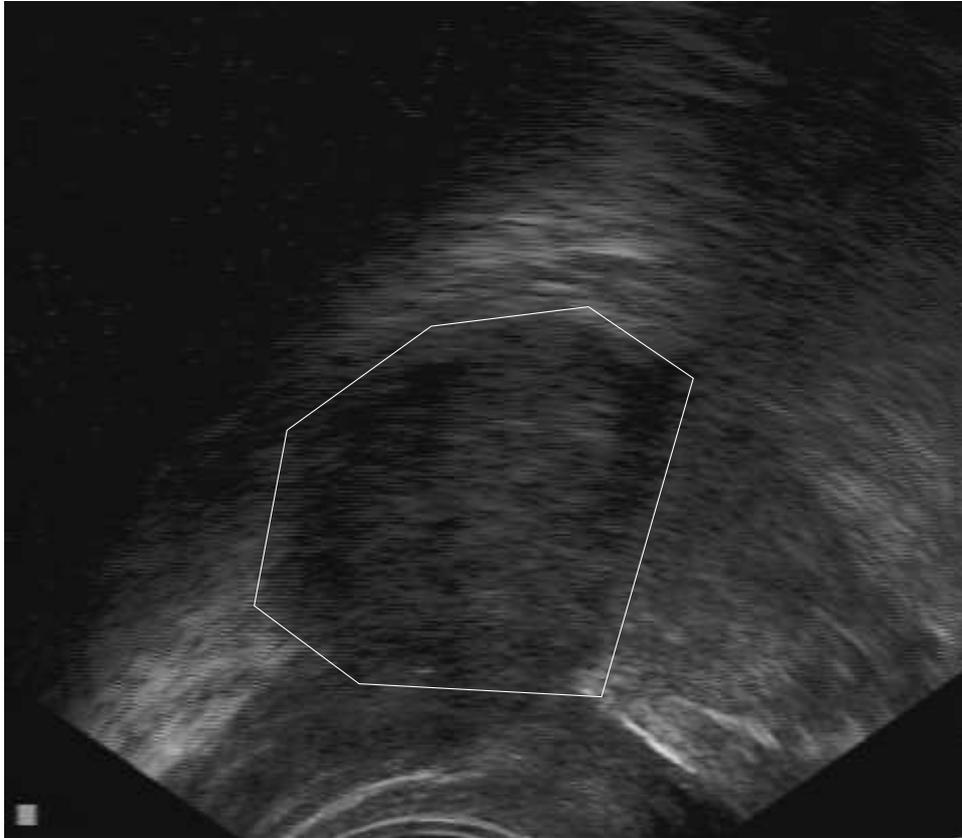


Figura 5.1: Imagen de ecografía transrectal. La región encerrada en el polígono corresponde a la próstata.

analizar las zonas de la próstata donde, de manera generalizada, se localizan habitualmente los tumores. Actualmente, dentro del campo de la urología, todavía es motivo de controversia las reglas que deben seguirse a la hora de dirigir la biopsia. Lo habitual es realizar varias biopsias con la técnica conocida como “en abanico”, iniciando las punciones en la base de la próstata y dirigiéndolas hacia las zonas laterales. Otra estrategia también ampliamente utilizada consiste en dirigir las biopsias de manera más lateral (punciones paralelas), para tratar de cubrir el máximo de la zona periférica.

Como puede observarse, el proceso de biopsiado se realiza prácticamente “a ciegas”, sirviendo las características de la imagen ecográfica únicamente (salvo casos excepcionales) para localizar los límites de la glándula prostática. Como consecuencia de las limitaciones inherentes a este proceso, es posible, y probable, que algunos tumores no se detecten, aun después de realizar varias biopsias. Algunos estudios han concluido que, realizando entre 10 y 12 punciones correctamente dirigidas a las zonas de riesgo, se puede llegar a diagnosticar más del 90 % de los

cánceres de próstata [Stew 01, Djav 00], lo que supone que aproximadamente el 10% de los tumores pasan inadvertidos tras la realización del test descrito. Además, resulta poco operativo realizar tal cantidad de punciones en una sola sesión, dado el dolor y molestia que supone.

Una posible manera de mejorar este proceso se basa en el uso de un Sistema de Ayuda al Diagnóstico Asistido por Computador (CAD), tal y como se explica en la Sección 3.3.3. La idea básica de un sistema de este tipo consiste en el análisis de la imagen ecográfica mediante ordenador, resaltando en dicha imagen las regiones que tienen una alta probabilidad de contener células cancerosas. De este modo, los médicos, en lugar de dirigir la biopsia siguiendo unas reglas genéricas previamente establecidas, podrían dirigir las biopsias hacia zonas cancerosas con mayor probabilidad de éxito, consiguiendo disminuir el riesgo de que un tumor pase inadvertido (disminución de falsos negativos) o mantener el rendimiento diagnóstico con un menor número de punciones (disminución de falsos positivos).

En este capítulo se presentan una serie de aportaciones al diagnóstico del cáncer de próstata asistido por ordenador. Las técnicas abordadas y los resultados obtenidos son fruto de un proyecto de investigación realizado conjuntamente entre la Universidad Politécnica de Valencia (UPV) y el Departamento de Urología del Hospital La Fe de Valencia.

El trabajo se dividió en tres fases claramente diferenciables: la primera fase consistió en la adquisición de un corpus específico para este trabajo. En dicho corpus quedaron registradas las imágenes de 1648 ecografías transrectales, correspondientes a otras tantas biopsias. En una segunda fase se evaluaron distintas técnicas para tratar de encontrar algún método que, a partir de una imagen de ecografía transrectal digitalizada no vista previamente, fuera capaz de discriminar entre tejido sano y tejido canceroso. En esta segunda fase se evaluaron dos métodos distintos de extracción de características (niveles de gris y matrices de coocurrencia), y dos clasificadores estadísticos, uno de ellos paramétrico (Modelos Ocultos de Markov) y el otro no paramétrico (k -vecinos más cercanos). En la tercera y última fase, utilizando la aproximación que mejores resultados ofreció en la etapa anterior, se realizó la simulación de un sistema de ayuda al diagnóstico asistido por ordenador, para evaluar en qué medida dicho sistema podría ayudar al experto humano a la hora de dirigir la biopsia con mayor éxito a zonas cancerosas de la próstata.

En el apartado de experimentación correspondiente a la segunda fase se clasificó, de forma automática (sin la intervención de un experto humano), un conjunto de imágenes del corpus adquirido. Las imágenes se clasificaron, no a nivel de píxel (ya que no se tiene certeza del etiquetado a este nivel, tal y como se explicará en la Sección 5.2.5), sino globalmente, otorgando a toda la imagen una probabilidad de pertenencia a una de las dos clases *cáncer* o *no-cáncer*. El objetivo de tales experimentos es el de evaluar los métodos de extracción de características y los clasificadores propuestos.

En el apartado de experimentación correspondiente a la tercera fase, utilizando la aproximación que mejores resultados ofreció en la etapa anterior, se realizó un postproceso a una serie de ecografías digitalizadas, para resaltar las regiones que tenían una alta probabilidad de contener células cancerosas. Para ello, en cada una de las imágenes ecográficas, se clasificó cada uno de sus píxeles, y aquellos que ofrecieron un nivel de confianza de pertenencia a la clase *cáncer* por encima de cierto umbral, se resaltaron en color rojo. Finalmente, para evaluar la utilidad clínica del sistema de CAD propuesto, se pidió a una serie de urólogos, con distinto grado de experiencia, que clasificaran cada una de las ecografías sin y con la ayuda del sistema, esto es, las ecografías originales, sin colorear, y las ecografías resultantes del postproceso, coloreadas por el sistema. El objetivo de esta segunda batería de experimentos era evaluar, en qué medida, el sistema de ayuda al diagnóstico propuesto, podría ayudar a un urólogo a detectar con mayor facilidad tumores en imágenes obtenidas mediante ecografía transrectal.

Las distintas técnicas empleadas en el desarrollo del sistema de CAD que se presenta en este capítulo han sido explicadas con detalle en el Capítulo 4. Concretamente, los dos métodos de extracción de características analizados (niveles de gris y matrices de coocurrencia) fueron explicados en la Sección 4.2, mientras que la búsqueda rápida de k -vecinos y la clasificación mediante HMMs fueron tratados, respectivamente, en las Secciones 4.6 y 4.7. En este capítulo, en la Sección 5.2 se detalla todo lo referente a la adquisición y puesta a punto del corpus. A continuación, en la Sección 5.3 se expone todo lo relacionado con los experimentos correspondientes a la fase de clasificación de imágenes, mientras que en la Sección 5.4 se detallan los experimentos correspondientes a la fase de evaluación de la utilidad clínica del sistema de CAD propuesto. Por último, en la Sección 5.5 se hace una valoración de los resultados obtenidos y se exponen las conclusiones finales.

5.2. Adquisición y puesta a punto de una base de datos de imágenes ecográficas de próstata

A todos los pacientes que sufren síntomas comúnmente asociados al cáncer de próstata (tacto rectal sospechoso y PSA elevado), se les practica de forma rutinaria una serie de biopsias dirigidas mediante ecografía transrectal, para poder confirmar de manera más precisa el diagnóstico inicial. Durante los meses de febrero de 2000 y febrero de 2001, fueron registradas digitalmente las ecografías transrectales realizadas a los pacientes del hospital La Fe de Valencia. Estas imágenes de ecografía, junto con diversos datos clínicos de cada paciente, forman la base de datos que ha sido utilizada en este trabajo.

La adquisición de toda esta información y su posterior puesta a punto, supuso un gran esfuerzo que consiguió reunir más de 1500 grabaciones de ecografía

transrectal correspondientes a más de 300 pacientes. Esta es, según nuestros conocimientos, la mayor base de datos utilizada en un trabajo de este tipo.

5.2.1. Población del estudio

En el estudio realizado se incluyeron pacientes pertenecientes al área de salud del Hospital Universitario la Fe de Valencia remitidos por sospecha de padecer cáncer de próstata. Estos pacientes dieron su consentimiento para que sus ecografías fueran registradas con ánimo de ser utilizadas en trabajos de investigación.

La biopsia prostática fue indicada por personal de distintos centros o servicios: consulta externa del hospital la Fe, sala de hospitalización o centro de especialidades. Esta variada procedencia de pacientes supone el estudio de una muestra de población heterogénea, en cuanto a motivo de consulta y manifestación clínica se refiere. No se estableció inicialmente ninguna limitación para ser incluido en este trabajo, por lo que, “a priori”, no se estableció ningún sesgo en relación con las características clínicas, cifras de PSA, edad u otras circunstancias que puedan ocasionar una preselección de candidatos. Se pretendió incluir inicialmente y de manera sistemática, todo tipo de paciente en el que estuviera indicada la biopsia (tacto rectal sospechoso y/o $PSA \geq 4ng/ml$).

Todos los datos necesarios para la identificación del paciente, identificación de la imagen ecográfica, características demográficas, clínicas, ecográficas, diagnóstico, etc., fueron recogidos en una base de datos creada específicamente para este trabajo.

Aunque inicialmente, tal y como se ha comentado, se registraron las ecografías realizadas a todos los pacientes en los que estaba indicada la biopsia, para este estudio se seleccionaron los casos de cáncer en estadio precoz, de difícil diagnóstico. Los pacientes con un tumor en estado avanzado fueron excluidos del estudio. Este tipo de tumores pueden ser diagnosticados sin ninguna dificultad por el urólogo y, en consecuencia, no aportaría ninguna ventaja un sistema de ayuda al diagnóstico que fuera capaz de detectar dichos tumores. Por otro lado, los tumores en estado avanzado pueden presentar texturas o características muy distintas a las que presentaría un tumor incipiente, por lo que la inclusión de dichas muestras podría incrementar la incertidumbre del clasificador. Los tumores en estado avanzado eliminados del estudio fueron seleccionados por un urólogo en base a los siguientes criterios:

- Cifras elevadas de PSA.
- Imagen ecográfica heterogénea, en la que se observa claramente un tumor que se extiende fuera de la glándula prostática.
- Tacto rectal muy sospechoso.

- Gran volumen tumoral, determinado por el número de cilindros (secciones de tejido extraídos en la biopsia) afectados.

5.2.2. Equipo de adquisición

Para la adquisición de las imágenes ecográficas se utilizó un PC convencional con tarjeta de adquisición de vídeo (*frame grabber*) conectada a la salida de vídeo de un ecógrafo Bruel&Kjaer modelo 3535. Se utilizó igualmente una sonda transrectal, modelo 8551 de $7.5MHz$. Dicha sonda presenta un canal de biopsia integrado en el transductor, que permite el paso de la aguja de punción-biopsia.

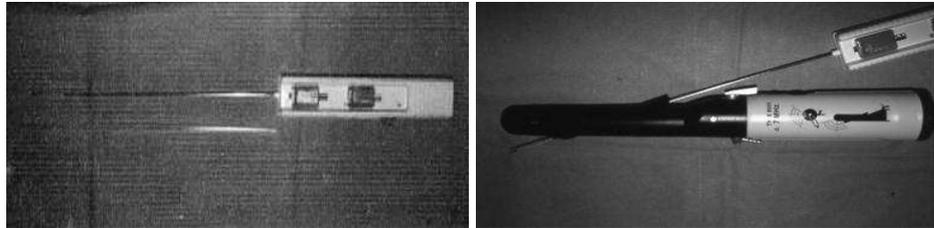


Figura 5.2: Pistola automática de punción para la biopsia (izquierda) y transductor transrectal con aguja de punción atravesando el canal de trabajo (derecha).

Todas las imágenes registradas se realizaron bajo las mismas condiciones de frecuencia, velocidad de adquisición, ganancia y escala, con el fin de poder establecer comparaciones entre las distintas imágenes.

El dispositivo de punción utilizado fue el modelo Microvasive© (Boston Scientific) 18 Gauge de $21cm$. de longitud. Este dispositivo se acciona automáticamente mediante una pistola, de modo que al disparar la misma, la aguja se introduce casi instantáneamente en la glándula prostática. El tejido biopsiado, consecuentemente, se corresponde exactamente con la zona visualizada en la pantalla en el instante inmediatamente anterior a la realización de la punción. La aguja utilizada para realizar la biopsia es hueca, de modo que al introducirse en la próstata, en su interior queda alojado un fino hilo de tejido. Dicho tejido es posteriormente enviado al laboratorio para realizar un análisis histológico y determinar si existen o no células cancerosas en el mismo. En la Figura 5.2 se muestra el dispositivo de punción utilizado para realizar la biopsia, así como el transductor del ecógrafo con dicho dispositivo de punción acoplado.

5.2.3. Grabación de la biopsia

Una vez introducida la sonda y escogida el área de la próstata a biopsiar, se comprueba que en la pantalla del ecógrafo aparece la información necesaria para la identificación de la biopsia. En la pantalla aparece también, mediante una línea

punteada sobreimpresa a la imagen ecográfica, la zona del tejido en la que incidirá la aguja en el momento que se accione el mecanismo de disparo correspondiente. En la Figura 5.3 se muestra la imagen de la ecografía tal y como se visualiza en la pantalla del equipo de ecografía. Se puede observar la línea punteada que indica la trayectoria que seguirá la aguja de la biopsia. Así mismo, se muestran los parámetros de configuración del ecógrafo (frecuencia, ganancia y velocidad de adquisición entre otros). El nombre del paciente, por motivos de privacidad, ha sido borrado de la imagen.

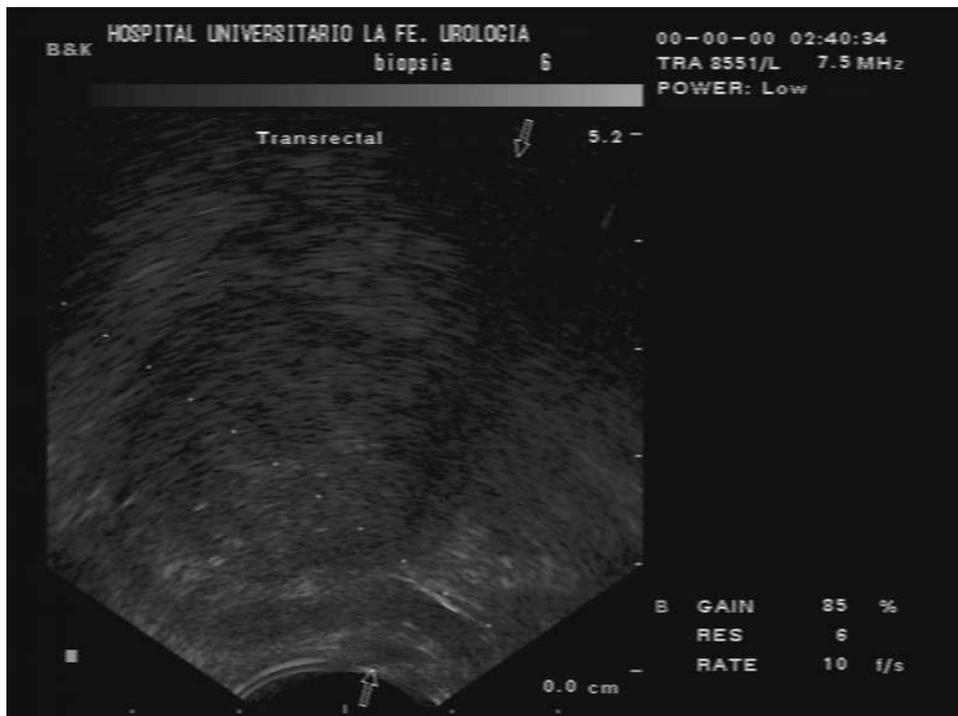


Figura 5.3: Imagen de ecografía transrectal tal y como se visualiza en el monitor del ecógrafo. La línea de puntos indica la trayectoria que seguiría la aguja de la biopsia.

Mediante un dispositivo de pedal, se inicia la grabación digital que, durante 10 segundos y a una frecuencia de 5 imágenes por segundo, almacena en el ordenador las imágenes tal y como se visualizan en la pantalla del ecógrafo. Durante estos 10 segundos que dura la grabación, la línea discontinua que indica la trayectoria desaparece y el ordenador comienza a emitir un pitido cada segundo. Al cuarto pitido aproximadamente, se introduce la aguja de punción y se acciona el disparo de la misma. Durante el tiempo que dura la grabación, se intenta no mover la sonda transrectal, de modo que, no sólo la imagen inmediatamente anterior a la punción, sino también sus predecesoras, coinciden con la zona biopsiada. En la Figura 5.4 se muestran dos imágenes de una misma ecografía. La imagen de la derecha co-

responde al preciso instante en que la aguja de la biopsia penetra en la próstata. Se puede observar el eco que dicha aguja genera en la imagen de ultrasonidos. La región más oscura, en cuyo interior se encuentra la aguja, corresponde precisamente a la glándula prostática. La imagen de la izquierda corresponde al instante anterior a la punción. En dicha imagen se ha sobreimpreso un rectángulo que se corresponde con la región de la próstata biopsiada.

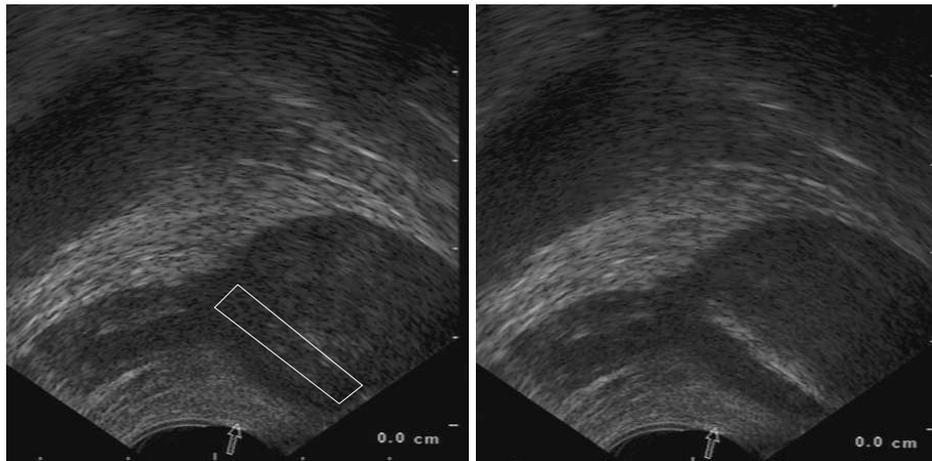


Figura 5.4: Imágenes de la misma biopsia tomadas 200ms antes de la punción (izquierda) y en el momento de la punción (derecha)

El ordenador, a través de su tarjeta de adquisición de vídeo, registra las imágenes con una resolución de 768×576 píxeles y una profundidad de 8 bits (256 niveles de gris). A cada paciente se le realizan normalmente entre 5 y 6 ecografías, correspondientes a otras tantas biopsias. La grabación de cada una de estas ecografías se almacena en un fichero distinto, que se identifica mediante el código del paciente y el número de biopsia. Paralelamente, en dos ficheros de texto, uno de pacientes y otro de biopsias, se almacenará, respectivamente, la información relevante de cada paciente y de cada biopsia. En la Sección 5.2.6 se ofrece una descripción detallada del contenido de estos ficheros.

5.2.4. Segmentación del tejido biopsiado

En un proceso posterior a la grabación, cada una de las ecografías es inspeccionada, con el fin de obtener la información necesaria para poder segmentar el área concreta de la imagen que corresponde con el tejido biopsiado. Debe de quedar claro que únicamente podemos dar una etiqueta de clase (de forma imperfecta, como se verá más adelante) a los píxeles de la imagen correspondientes al tejido analizado mediante biopsia, cuyo diagnóstico conocemos. El resto de la imagen, por tanto, no será útil a efectos de crear o entrenar un modelo de tejido canceroso (o de tejido sano), ya que desconocemos la clase a la que pertenecen sus píxeles.

Es necesario, por tanto, almacenar en algún lugar el área de la imagen que se corresponde con el tejido cuyo diagnóstico conocemos (tejido biopsiado). Para ello se almacena, por un lado, el número de frame en el que aparece la aguja en la imagen y, por otro, las coordenadas de la imagen correspondientes a los extremos de la aguja. Estas coordenadas se utilizarán para definir, en la imagen inmediatamente anterior a la punción, un rectángulo de ancho fijo, tal y como se ha mostrado en la Figura 5.4. Los píxeles de dicho rectángulo serán, por tanto, la parte de toda la ecografía digitalizada que será de mayor utilidad para nuestro estudio.

Por último, cabe decir que este proceso de inspección manual de las ecografías digitalizadas sirvió, además de para segmentar el área biopsiada tal y como se ha explicado, para desechar algunas grabaciones que, por distintos motivos, tenían algún defecto. Los defectos más habituales fueron: un ajuste incorrecto en la ganancia del ecógrafo, una grabación en la que no llega a visualizarse la aguja de la biopsia o grabaciones en las que no desaparece en ningún momento la línea discontinua indicativa de la trayectoria de la aguja.

5.2.5. El problema de la supervisión imperfecta

Los cilindros de tejido prostático extraídos en cada punción, deben de ser analizados en laboratorio para realizar el diagnóstico. Para ello, cada uno de estos cilindros, cuya sección se mide en décimas de milímetro, es cortado longitudinalmente en varios segmentos y teñido para su diagnóstico microscópico. Tras dicho estudio microscópico, se establece, no sólo si el cilindro contiene o no células cancerosas, sino el porcentaje de cilindro afectado por las mismas.

Durante el proceso de corte del cilindro, es fácil que el mismo se rompa. Esta situación no impide realizar el diagnóstico tal y como se ha descrito; sin embargo, impide establecer el lugar exacto del cilindro en el que se encontraban las células cancerosas (en caso de que las hubiera). Esto provoca que sea imposible transformar el tejido extraído en un mapa de etiquetas, en el que se asigne una etiqueta (*cáncer* o *no-cáncer*) a cada píxel del tejido analizado. Debemos conformarnos, por tanto, con tener una única etiqueta de clase para todos los píxeles de la zona biopsiada, a pesar de que no pertenezcan todos los píxeles a la misma clase. Este problema es lo que hemos denominado como “supervisión imperfecta”. En ocasiones, la zona afectada de cáncer puede representar tan solo el 5% del cilindro analizado y, sin embargo, la totalidad de los píxeles se etiquetarán con la clase *cáncer*. Este problema de etiquetado condicionará, sin duda, los resultados obtenidos en la fase experimental.

5.2.6. Características del corpus

Durante el periodo de adquisición de la base de datos, se consiguieron registrar un total de 303 sesiones, correspondientes a otros tantos pacientes. A cada paciente se le practicaron entre 5 y 6 biopsias, obteniendo al final un total de 1648 biopsias válidas (una vez descartadas las grabaciones defectuosas mencionadas en la Sección 5.2.4). De cada una de estas biopsias registradas, se seleccionaron 3 imágenes anteriores al instante de punción. En concreto, dado el instante t correspondiente a la imagen en la que se observa la introducción de la aguja, se seleccionaron las imágenes correspondientes a los instantes $t - 1$, $t - 3$ y $t - 5$, esto es, las imágenes correspondientes a $200ms.$, $600ms.$ y $1s.$ anterior a la punción (dado que la frecuencia de grabación fue de 5 imágenes por segundo). En total, por tanto, la base de datos adquirida consta de 4944 imágenes (1648 biopsias $\times 3$ imágenes/biopsia).

Debido a que la salida de vídeo del ecógrafo opera en modo entrelazado, para evitar ruido en la imagen ecográfica y a la vez mantener el ratio de la misma, las filas y columnas pares de cada imagen fueron descartadas. En consecuencia, la resolución de imagen con la que se trabajó finalmente fue de 384×288 píxeles.

Tal y como se ha explicado en la Sección 5.2.1, las biopsias que revelaron un cáncer en estado avanzado, fueron excluidas del estudio. En consecuencia, la base de datos finalmente utilizada en este trabajo, se compone de 4593 imágenes extraídas de 1531 biopsias, obtenidas a su vez de 289 pacientes diferentes. De las 1531 biopsias, 202 (aproximadamente el 13 %) corresponden a tejido canceroso y el resto a tejido sano.

Para cada biopsia registrada, se almacenó adicionalmente, en dos únicos ficheros de texto (uno de pacientes y otro de biopsias), la información asociada a cada paciente y a cada biopsia respectivamente. El fichero de pacientes almacena, en cada línea, la siguiente información:

- Código del paciente.
- Fecha de la exploración.
- Índice de PSA.

Por otro lado, en el fichero de biopsias se almacena en cada línea la información relativa a cada una de las biopsias practicadas. Concretamente:

- Código del paciente a quien corresponde la biopsia
- Código de la biopsia.
- Resultado del análisis histológico del tejido biopsiado (diagnóstico). Las etiquetas posibles de este campo son:

- BHP (Benign Prostate Hyperplasia)
- PROSTATITIS
- CANCER1 (cilindro con más del 40 % de tejido afectado y tumor en estado incipiente)
- CANCER2 (cilindro con menos del 40 % de tejido afectado y tumor en estado incipiente)
- CANCER3 (cilindro con casi el 100 % de tejido afectado y tumor en estado avanzado)

Tanto la hiperplasia como la prostatitis, son alteraciones de la próstata que producen un aumento en el índice de PSA, si bien no se corresponden con la existencia de un tumor maligno. En este estudio, las dos primeras etiquetas se fusionaron en una única clase (*no-cáncer*), las dos siguientes en otra clase (*cáncer*), mientras que las biopsias etiquetadas con CANCER3, tal y como se ha mencionado anteriormente, fueron descartadas del estudio.

- Número de frame en la ecografía en el que aparece la aguja de la biopsia.
- Coordenadas de los extremos de la aguja
- Porcentaje de tejido afectado

5.3. Experimentos con clasificación automática de imágenes

En una primera batería de experimentos se trató de evaluar las prestaciones de distintas aproximaciones de extracción de características y clasificación. Concretamente se probaron dos esquemas de extracción de características: niveles de gris y SGLDM. Estos esquemas de extracción de características han sido explicados en la Sección 4.2.

Para cada uno de estos esquemas de extracción de características se evaluó, en primer lugar, el clasificador k -NN. El clasificador k -NN es una técnica ampliamente utilizada en tareas muy diversas. Dicho clasificador había sido utilizado previamente en otras tareas de visión [Pere 00] con resultados muy satisfactorios, por lo que parecía adecuado y recomendable utilizarlo también en esta tarea en concreto.

Posteriormente, se evaluó un clasificador basado en HMMs. Los HMMs, a diferencia de los k -NN, es una técnica mucho más novedosa en el contexto de esta tarea. El uso de los HMMs en este trabajo se justifica por dos motivos principalmente: por una lado, la secuencialidad de los píxeles a lo largo de la trayectoria seguida por la aguja de la biopsia, con una clara relación de orden, invita a su uso.

Por otro lado, el empleo de esta aproximación, podría paliar el efecto negativo que provoca el problema del etiquetado o supervisión imperfecta descrito en la Sección 5.2.5. El clasificador HMM y su aplicación concretamente a esta tarea ha sido explicado en la Sección 2.4.

El objetivo básico de estos experimentos es determinar si alguno de los esquemas propuestos puede predecir la presencia de células cancerosas en una región determinada de una imagen obtenida mediante TRUS. Como se ha mencionado anteriormente, esta es una tarea extremadamente difícil, ya que ni siquiera un urólogo experimentado es capaz actualmente de realizar un diagnóstico preciso a partir de la imagen de ecografía transrectal. A ello se le añade el problema de la supervisión imperfecta descrito anteriormente. El objetivo en esta primera fase es diseñar un sistema que sea capaz de detectar, con la mayor precisión posible, la presencia de células cancerosas en la imagen ecográfica.

Nuestra esperanza es que ciertas relaciones entre píxeles a nivel local puedan ser representadas y reconocidas mejor por los algoritmos que por el ojo humano. Si esta capacidad de discriminación entre tejido sano y tejido canceroso se consiguiera, cabría plantearse abordar una segunda fase del trabajo, en la que se tratase de combinar las capacidades del experto humano con las del clasificador automático, con objeto de dirigir la biopsia con mayor precisión a regiones cancerosas de la próstata. Si el clasificador fuera capaz de aportar nuevas pistas al experto humano acerca de la presencia de tejido canceroso, cabría esperar un mayor éxito en la realización de la biopsia, lo que se traduciría en una disminución del índice de tumores que pasan inadvertidos (falsos negativos) tras los tests preceptivos. Dado el alto coste que supone un error de clasificación en este tipo de tareas, una mejora en este sentido, por pequeña que fuera, justificaría el uso de un sistema de ayuda al diagnóstico como el propuesto en un entorno real.

5.3.1. Conjuntos de entrenamiento y test

Para todos los experimentos descritos a continuación se utilizó el corpus presentado en la Sección 5.2. Las 4593 imágenes correspondientes a 1531 ecografías disponibles en la base de datos, se dividieron en dos conjuntos de entrenamiento y test del siguiente modo:

- Aproximadamente el 70 % de las imágenes se incluyó en el conjunto de entrenamiento, mientras que el 30 % restante pasó a formar parte del conjunto de test.
- Se evitó que imágenes de un mismo paciente se repartieran en conjuntos distintos, esto es, todas las imágenes de un paciente se incluyeron, bien en el conjunto de entrenamiento, bien en el de test.

- Tanto en el conjunto de entrenamiento como en el de test se conservó la probabilidad a priori observada en el corpus (13% para la clase *cáncer* y 87% para la clase *no-cáncer*).
- Teniendo en cuenta las restricciones anteriores, la elaboración de los conjuntos de entrenamiento y test se hizo de forma aleatoria.

En base a estas premisas, se elaboraron 10 particiones distintas de entrenamiento-test. Para reducir al máximo el intervalo de confianza de los resultados obtenidos, cada uno de los experimentos se realizó con cada una de las particiones elaboradas. Los resultados finales ofrecidos corresponden a la media de los resultados obtenidos en cada una de las particiones.

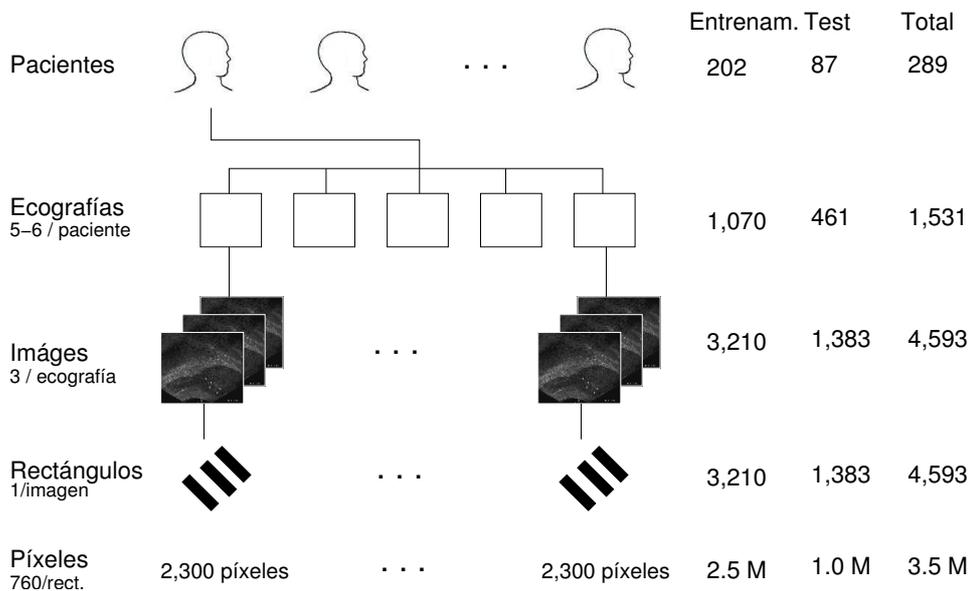


Figura 5.5: Partición del corpus en conjuntos de entrenamiento y test. Se dispone de 289 pacientes con 5 o 6 biopsias cada uno. Por cada biopsia se seleccionaron 3 imágenes de 384×288 píxeles, extrayéndose posteriormente por cada imagen un rectángulo de aproximadamente 760 píxeles (alrededor de $1/150$ de la imagen completa).

En la Figura 5.5 se muestra el tamaño de los conjuntos de entrenamiento y de test en base a número de pacientes, ecografías, imágenes y píxeles. Debemos aclarar que cuando se habla de “ecografía” nos referimos a la secuencia de imágenes registradas (50 aproximadamente) de una biopsia en particular. Puesto que a cada paciente se le practican entre 5 y 6 biopsias, habrá otras tantas ecografías de un mismo paciente. De las 50 imágenes que tiene cada ecografía, tal y como se explica en la Sección 5.2.6, se seleccionan únicamente 3 y de cada imagen se segmenta el rectángulo correspondiente al tejido biopsiado. Dado que las particiones

entrenamiento-test se realizaron con las restricciones descritas anteriormente, y teniendo en cuenta que el número de biopsias puede ser distinto para cada paciente, no todos los conjuntos tiene exactamente el mismo número de ecografías en cada una de las particiones, ni todas las ecografías el mismo número de píxeles, por lo que los valores ofrecidos son aproximados.

A la vista de estos valores, cabe resaltar el volumen de datos a tratar en cada uno de los experimentos. Puesto que, tal y como se ha mencionado, cada experimento se repitió con 10 pares distintos de conjuntos de entrenamiento-test, se puede deducir que la ejecución de un experimento supone la extracción de características de 30 millones de puntos y la clasificación de 15 millones de puntos. Para procesar este volumen de datos se requirió de un ordenador con gran capacidad, tanto de proceso como de almacenamiento primario y secundario.

5.3.2. Detalles de la experimentación

A continuación se describen los detalles de la fase experimental, mostrándose tanto los parámetros empleados en cada una de las aproximaciones de extracción de características propuestas (niveles de gris y matrices de coocurrencia) como los parámetros y ecuaciones utilizadas en cada uno de los clasificadores (k -NN y HMM). En la Sección 5.3.3 se ofrecen distintos resultados en función de dichos parámetros.

Extracción de características

a) *Niveles de gris*

La extracción de características basada en niveles de gris consiste en representar cada uno de los píxeles de la imagen en función de los niveles de gris de dicho píxel y de sus vecinos. Una cuestión fundamental de esta aproximación será, por tanto, el modo en que se define la vecindad de un píxel dado. En este trabajo, dicha vecindad se ha definido mediante una ventana cuadrada centrada en el píxel a representar. Se realizaron experimentos utilizando distintos tamaños de ventana, desde 8×8 hasta 25×25 píxeles, lo que supone vectores de características de dimensión entre 64 y 625.

La dimensionalidad de estos vectores fue reducida mediante el uso de Análisis de Componentes Principales (PCA), proyectando cada uno de los vectores de características en espacios de dimensionalidad entre 10 y 50. La base de proyección fue calculada a partir de un subconjunto de 250,000 vectores seleccionados aleatoriamente del conjunto de entrenamiento.

b) *Matrices de coocurrencia*

Para el cómputo de matrices de coocurrencia se requiere, como en el caso anterior, definir un contexto de vecindad a partir del cual extraer los des-

criptores de texturas. En este caso se probaron ventanas de características de 8×8 hasta 40×40 píxeles.

Por otro lado, tal y como se explica en la Sección 4.2.3, cuando se trabaja con matrices de coocurrencia, es conveniente trabajar con un número pequeño de niveles de gris en la imagen. En nuestro caso, dicho número se redujo, mediante cuantización, de 256 hasta 20.

Otra consideración a tener en cuenta es el número de matrices a calcular en cada ventana de vecindad. En este trabajo se combinaron 4 valores de distancia (1, 2, 3 y 4) con 4 ángulos distintos (0 , $\frac{\pi}{4}$, $\frac{\pi}{2}$ y $\frac{3\pi}{4}$), lo que genera 16 matrices de coocurrencia distintas. Para cada una de estas 16 matrices se calcularon los 11 descriptores detallados en la Sección 4.2.3, lo que genera vectores de características de dimensión $16 \times 11 = 176$.

Los vectores de características obtenidos, al igual que se hizo en el caso de niveles de gris, fueron proyectados en espacios de menor dimensión mediante el uso de PCA. Al igual que en el caso anterior, se probaron entre 10 y 50 componentes principales, obteniendo los mejores resultados con 30 componentes.

Clasificación

a) *k*-NN

Para esta tarea en concreto, se creó un modelo para las clases *cáncer* y *no-cáncer* que permitiera, posteriormente, clasificar un punto de test, aplicando la regla de los *k*-NN sobre dicho modelo. Para ello se empleó la búsqueda rápida aproximada descrita en la Sección 4.6. Para acelerar la búsqueda sin perder precisión de manera significativa, se trabajó con un valor de *epsilon* = 2. Por otro lado, se probaron distintos valores de *k* (desde 10 hasta 100), ofreciendo resultados sin diferencias significativas.

Para la creación del modelo se utilizaron los vectores de características extraídos del conjunto de entrenamiento que se detalla en la Sección 5.3.1. Se crearon modelos con características basadas, tanto en niveles de gris como en matrices de coocurrencia, con el fin de evaluar ambas aproximaciones.

Posteriormente, durante la fase de test, se obtuvieron los vectores de características de cada uno de los píxeles de la región de la imagen a evaluar (rectángulo de la biopsia). Cada uno de estos vectores de características se corresponde con un punto de test que se clasifica aplicando la regla de los *k*-vecinos sobre el modelo creado.

En lugar de clasificar cada punto de test en función de la clase más votada, se utilizó un valor de confianza de pertenencia a la clase *cáncer* calculado en función de las distancias a los vecinos más cercanos:

$$f_c = \frac{\sum_{i \in s_c} \frac{1}{d(p, n_i)}}{\sum_{i=1}^k \frac{1}{d(p, n_i)}}, \quad (5.1)$$

donde:

- f_c es el valor de confianza de pertenencia a la clase c .
- $d(p, n_i)$ es la distancia entre el punto de test y el vecino i -ésimo.
- s_c es el conjunto de subíndices de los prototipos pertenecientes a la clase c entre los k vecinos más cercanos $n_1 \dots n_k$.

Debe hacerse especial hincapié en que la clasificación mediante la regla de los k -vecinos, tal y como se ha descrito, se realiza a nivel de píxel. Sin embargo, en nuestro corpus no se dispone de un etiquetado a nivel de píxel, sino que existe una etiqueta única para toda la imagen (o más exactamente, para la región de la imagen correspondiente al tejido biopsiado). La clasificación a nivel de píxel es adecuada en la implementación de un sistema de ayuda al diagnóstico, ya que el objetivo en este caso es el de resaltar cada píxel de manera individual, en función de su probabilidad de pertenencia a la clase *cáncer*. Sin embargo, el objetivo de esta primera batería de experimentos no es evaluar la utilidad clínica del sistema de CAD, sino únicamente evaluar si los métodos propuestos son capaces o no de discriminar entre tejido sano y canceroso. Puesto que no se dispone de un etiquetado correcto a nivel de píxel, no parece adecuado calcular tasas de clasificación a dicho nivel; en su lugar haremos una clasificación a nivel de biopsia. Para obtener un valor de confianza único para toda la biopsia, se calculó de manera individual el valor de confianza para cada uno de los píxeles correspondientes al tejido biopsiado y se obtuvo la media aritmética de todos ellos.

Puesto que el problema abordado es un problema de dos clases, únicamente se calculó el valor de confianza de una de las clases (en este caso *cáncer*). Si establecemos un umbral T , una imagen se clasifica como *cáncer* si el valor de confianza obtenido es mayor que T . En este trabajo, en lugar de seleccionar un umbral específico, se calcularon las tasas de clasificación para distintos valores de T , generando de este modo una curva ROC, tal y como se describe en la Sección 4.8.

b) *HMM*

El clasificador HMM se utilizó tal y como se describe en la Sección 4.7. En este caso, en lugar de utilizar el algoritmo de Viterbi para clasificar cada secuencia de test en función del HMM que con mayor probabilidad genera dicha secuencia, se utilizó una función de confianza. Dicha función de confianza obtiene la probabilidad a posteriori de la clase *cáncer*, de modo muy

similar a la función de confianza utilizada en el clasificador de k -NN. En este caso, la función se define mediante la siguiente ecuación:

$$f_c = \frac{p_{\text{HMM}_c}(x|c)Pr(c)}{\sum_{\forall c'} p_{\text{HMM}_{c'}}(x|c')Pr(c')} \quad (5.2)$$

donde:

- f_c es el valor de confianza de pertenencia a la clase c .
- x es una secuencia dada de vectores de características.
- $p_{\text{HMM}_c}(x|c)$ es el valor obtenido mediante Viterbi sobre el camino más probable encontrado en HMM_c .
- $Pr(c)$ es la probabilidad *a priori* de la clase c .

Dado un valor de confianza de pertenencia a la clase *cáncer* y un umbral T , una imagen se considera cancerosa si dicho valor de confianza es mayor que T . Debe observarse que en este caso, a diferencia de lo que ocurre con el clasificador k -NN, el valor de confianza se calcula para toda la secuencia de píxeles correspondientes al tejido biopsiado, y no individualmente para cada píxel.

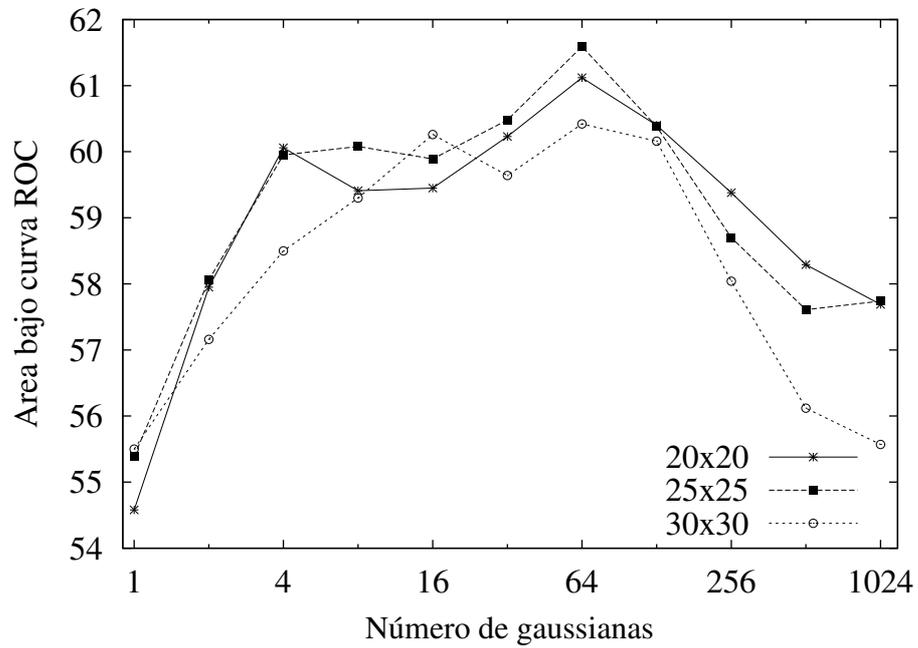
5.3.3. Resultados

A continuación se muestran los resultados obtenidos con los distintos esquemas de extracción de características y clasificadores propuestos.

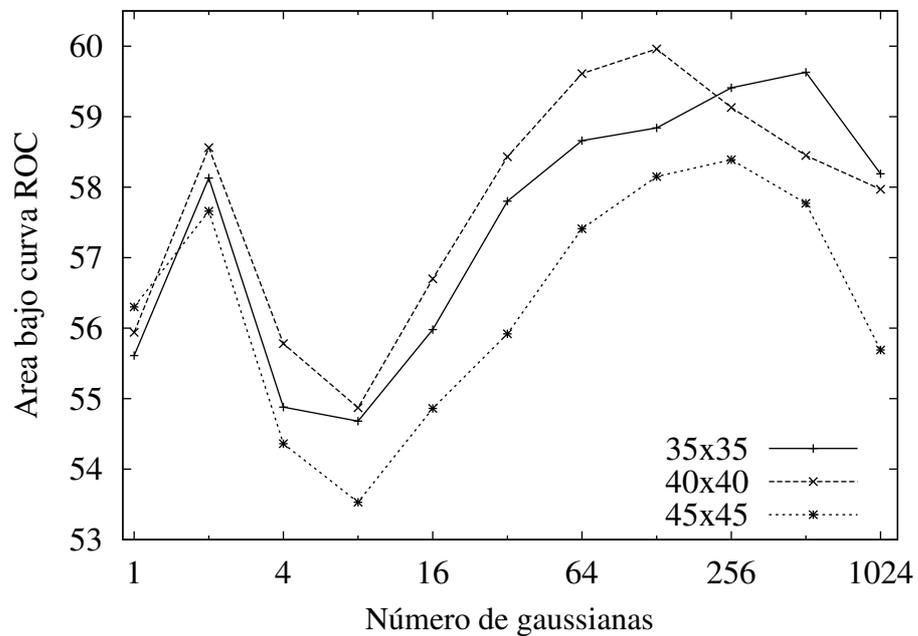
En primer lugar se muestra el área bajo la curva ROC obtenida con el uso de los HMMs bajo distintas condiciones. Concretamente, en las Figuras 5.6(a) y 5.6(b) se muestra el efecto de variar el número de gaussianas en cada estado del HMM, para distintos tamaños de ventana de vecindad. La Figura 5.6(a) corresponde a HMMs entrenados con características de niveles de gris, mientras que los resultados de la Figura 5.6(b) se obtuvieron a partir de HMMs entrenados con características de texturas (matrices de coocurrencia). Se observa que el número óptimo de gaussianas se encuentra entre 64 y 128. En la Figura 5.7 se muestra el efecto de variar el tamaño de la ventana de vecindad para distinto número de gaussianas en HMMs entrenados con matrices de coocurrencia.

En la Figura 5.8 se muestran los resultados obtenidos con el clasificador de k -vecinos más cercanos y matrices de coocurrencia, para distinto número de vecinos (k) y ventana de vecindad de 30×30 . Como puede observarse, el número de vecinos apenas afecta a los resultados obtenidos.

En la Figura 5.9 se muestran las curvas ROC correspondientes a los mejores resultados obtenidos con cada uno de los clasificadores, combinados con cada uno de



(a) Niveles de gris



(b) Matrices de coocurrencia

Figura 5.6: Áreas bajo la curva ROC obtenidas con un clasificador HMM. Se muestran resultados para distintos números de gaussianas y distintos tamaños de ventana de vecindad.

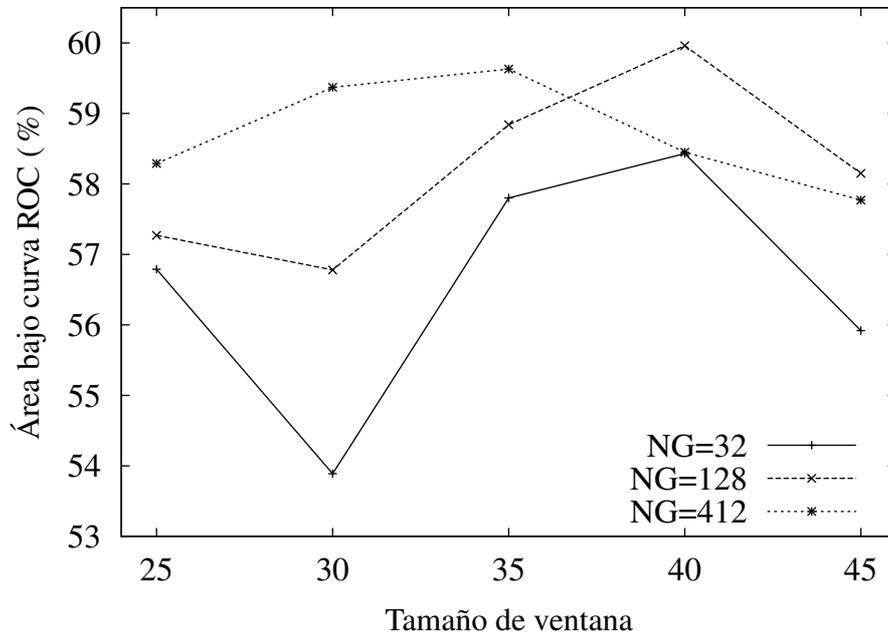


Figura 5.7: Áreas bajo la curva ROC obtenidas con matrices de coocurrencia y clasificador HMM. Se muestran resultados variando el tamaño de la ventana de vecindad, para distinto número de gaussianas.

los métodos de extracción de características propuestos. En el caso de los HMMs, dichos resultados se obtuvieron con modelos de 64 densidades de gaussianas por estado y ventana de vecindad de 25×25 para el caso de niveles de gris, y 128 densidades de gaussianas por estado y ventanas de vecindad de 40×40 para el caso de matrices de coocurrencia. En lo referente al clasificador k -NN, se utilizó, para ambos esquemas de extracción de características, un valor de $K = 40$ y ventanas de vecindad de 30×30 . En todos los casos se utilizaron 30 componentes principales.

Los HMMs obtuvieron un rendimiento ligeramente superior al clasificador k -NN cuando se trabajó con niveles de gris; sin embargo, no existen diferencias significativas entre ambos clasificadores cuando se trabaja con matrices de coocurrencia. La Tabla 5.1 muestra el área bajo la curva ROC obtenida, para el mejor de los casos, con cada uno de los clasificadores (k -NN y HMM) y cada uno de los esquemas de extracción de características (niveles de gris y SGLDM).

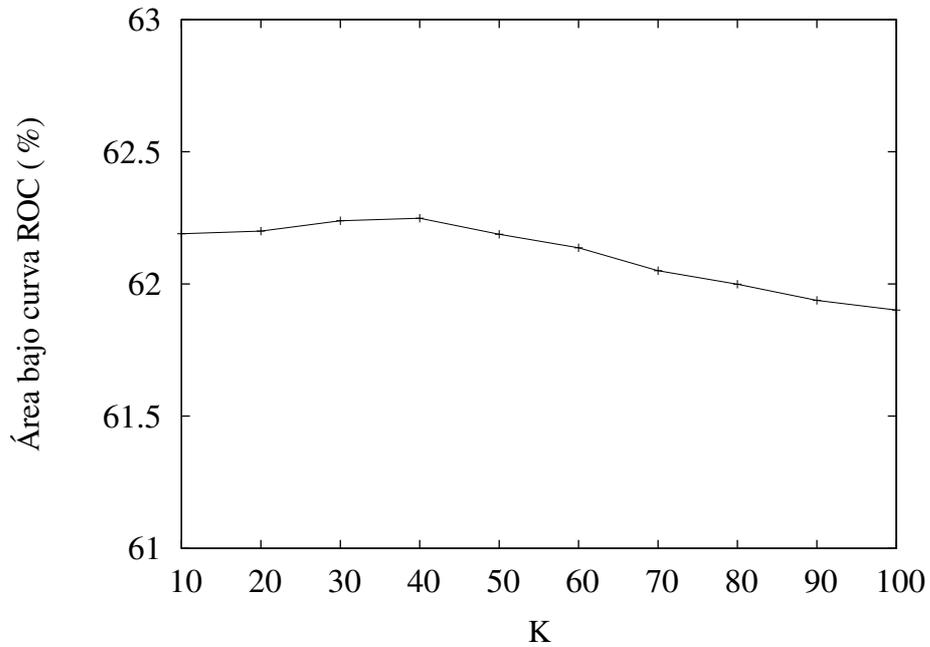


Figura 5.8: Áreas bajo la curva ROC obtenidas con matrices de coocurrencia y clasificador k -NN. Se muestran resultados para distintos valores de K y ventana de vecindad de 30×30 . En este caso los resultados se obtuvieron mediante *leaving-one-out* sobre todo el conjunto de datos.

Tabla 5.1: Áreas bajo la curva ROC con intervalos de confianza al 95% para los dos esquemas de extracción de características utilizados (SGLDM y niveles de gris) evaluados con los dos clasificadores presentados (k -NN y HMM)

	k -NN	HMM
SGLDM	$60.1 \pm 0.7\%$	$60.0 \pm 0.7\%$
Grisés	$59.7 \pm 0.7\%$	$61.6 \pm 0.7\%$

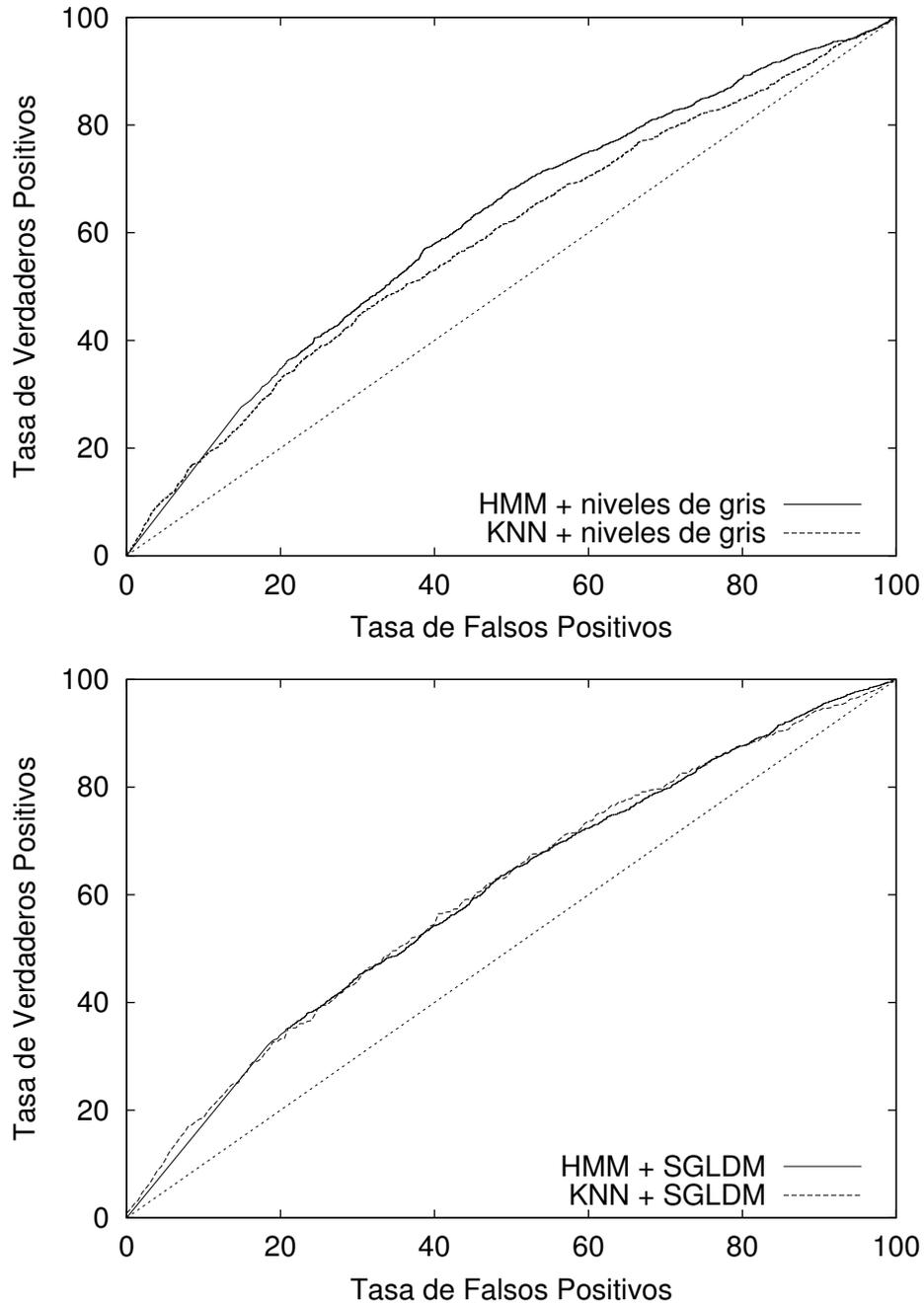


Figura 5.9: Curvas ROC correspondientes a los mejores resultados obtenidos con niveles de gris (arriba) y matrices de coocurrencia (abajo), tanto con el clasificador de k -NN como con HMM.

5.3.4. Conclusiones

Los resultados obtenidos muestran que, con las distintas aproximaciones presentadas, es posible discriminar entre células sanas y cancerosas a través de la imagen obtenida mediante ecografía transrectal. Dichos resultados, aunque modestos, son significativos.

Todas las aproximaciones presentadas ofrecen un rendimiento muy similar, con una mejora muy ligera, apenas significativa, en los HMMs cuando trabajan con niveles de gris.

A pesar de que los resultados no ofrezcan una elevada tasa de aciertos, vista la capacidad de discriminación entre tejido sano y canceroso, parece interesante abordar el problema de la ayuda al diagnóstico de cáncer de próstata asistido por ordenador.

5.4. Sistema de ayuda al diagnóstico

Los experimentos realizados con clasificación automática de imágenes ofrecieron unos resultados modestos pero significativos. El hecho de poder discriminar entre tejido sano y canceroso, aunque sea de un modo poco preciso, anima a plantear una segunda batería de experimentos para evaluar la mejora diagnóstica que podría ofrecer un sistema de ayuda por ordenador. Dado el elevado coste que supone (difícilmente mesurable por otro lado) un falso negativo en una tarea de esta índole, una pequeña mejora en las tasas de clasificación, supondría un alto beneficio.

A continuación se presenta una serie de experimentos que tienen por objeto evaluar, hasta qué punto, un sistema de ayuda al diagnóstico implementado con alguna de las aproximaciones propuestas anteriormente, puede ayudar a un urólogo a realizar un diagnóstico más certero y, en consecuencia, aumentar el éxito de la biopsia.

5.4.1. Experimentos

Para llevar a cabo los experimentos se simuló una herramienta clínica que, en tiempo real, resaltara en falso color en la pantalla del ecógrafo las áreas de la próstata con una alta probabilidad de contener células cancerosas. La idea básica es estimar si el rendimiento diagnóstico del urólogo aumenta cuando se añade el falso color a la imagen ecográfica. Para realizar tal simulación, se diseñó una aplicación que permite visualizar una serie de ecografías tal y como se verían en la pantalla del ecógrafo. Para ello, las ecografías fueron previamente procesadas para añadirles el falso color, tal y como se explicará más adelante. De este modo, el urólogo,

tras visualizar las ecografías, debe asociarles un valor de sospecha. Para evaluar la mejora que puede suponer el uso de una herramienta como la propuesta, este proceso se realizó, tanto con las ecografías originales, como con las ecografías procesadas por el sistema (con falso color).

De las 1531 ecografías existentes en el corpus, se seleccionaron la totalidad de las mismas correspondientes a biopsias cancerosas (204), y el mismo número de ecografías, seleccionadas de forma aleatoria, correspondientes a biopsias sanas. A diferencia de los experimentos explicados en la sección anterior, en los que sólo se utilizaron 3 imágenes de cada ecografía, en este caso se utilizó la secuencia completa de imágenes.

Para añadir el falso color, cada una de las imágenes se clasificó a nivel de píxel. Aquellos píxeles cuyo valor de confianza resultó estar por encima de cierto umbral T , se resaltaron en color rojo. El valor de T se escogió de forma empírica, llegando a un compromiso entre sensibilidad y especificidad. Concretamente se utilizó un valor de $T = 0.6$, produciendo valores de 57.2% de sensibilidad y 61.0% de especificidad. Valores de T más bajos aumentan la sensibilidad y disminuyen la especificidad, generando imágenes con un elevado porcentaje de píxeles coloreados, mientras que valores mayores de T producen el efecto contrario. En la Figura 5.10 se muestra una imagen de una de las ecografías tras el coloreado. Como puede observarse, se coloreó la imagen completa y no únicamente los píxeles correspondientes al cilindro biopsiado. Sin embargo, dado que únicamente se tiene certeza de la etiqueta de clase correspondiente a la región del cilindro biopsiado, dicha región se marcó mediante un rectángulo verde. De este modo, cuando los distintos urólogos clasificaron cada una de las ecografías, se les pidió que únicamente tuvieran en consideración la zona de la imagen comprendida dentro de dicho rectángulo. El resto de la imagen, por otro lado, puede ser de utilidad para tener una referencia del porcentaje de píxeles que habitualmente se resaltan en falso color para el valor de T utilizado.

Para el coloreado de las ecografías se utilizó el clasificador k -NN con niveles de gris. Aunque el clasificador basado en HMM ofreció resultados ligeramente superiores, dicho clasificador no es adecuado para trabajar a nivel de píxel, por lo que se desestimó su uso en esta fase. Por otro lado, la extracción de características mediante niveles de gris y matrices de coocurrencia ofrecieron resultados muy similares para el caso de k -NN. Sin embargo, se observó una mayor correlación en el coloreado de un mismo píxel en imágenes consecutivas cuando se utilizaron niveles de gris.

Como en el conjunto de test se incluyeron la totalidad de ecografías cancerosas disponibles en el corpus, durante la clasificación previa al coloreado se empleó un esquema de *leaving-one-out*, evitando así clasificar píxeles que hubieran sido vistos previamente en la fase de entrenamiento. En este proceso se eliminaron del conjunto de entrenamiento, no sólo los píxeles de la ecografía a clasificar, sino también los píxeles de otras ecografías pertenecientes al mismo paciente (lo que

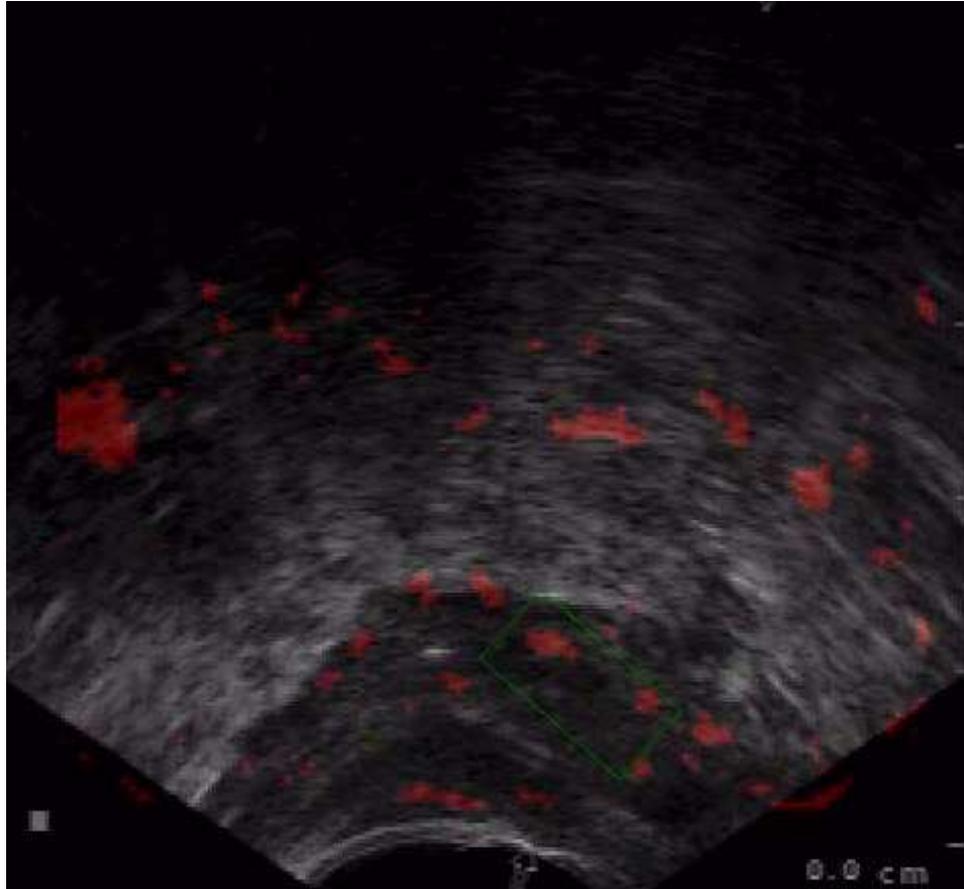


Figura 5.10: Imagen de ecografía transrectal con falso color. Las áreas rojas indican zonas con alta sospecha tumoral. El rectángulo verde delimita el tejido biopsiado. La zona más oscura en la parte inferior de la ecografía, corresponde a la glándula prostática.

podríamos denominar como *leaving-one-patient-out*).

Para realizar el test, se desarrolló una aplicación basada en web para evaluar el rendimiento que podría aportar un sistema de ayuda al diagnóstico como el propuesto. Dicha aplicación permite a los urólogos clasificar las 408 ecografías incluidas en cada uno de los dos conjuntos de test (ecografías originales/coloreadas), para posteriormente comparar las tasas de aciertos en uno y otro conjunto.

El urólogo comienza una sesión introduciendo su código de usuario y clave, y seleccionando el conjunto que desea clasificar (ecografías originales/coloreadas), tal y como muestra la Figura 5.11. La siguiente pantalla (Figura 5.12) muestra un formulario con enlaces a cada uno de los ficheros que contienen el vídeo de cada ecografía (originales o coloreadas, según la opción seleccionada en el paso previo)

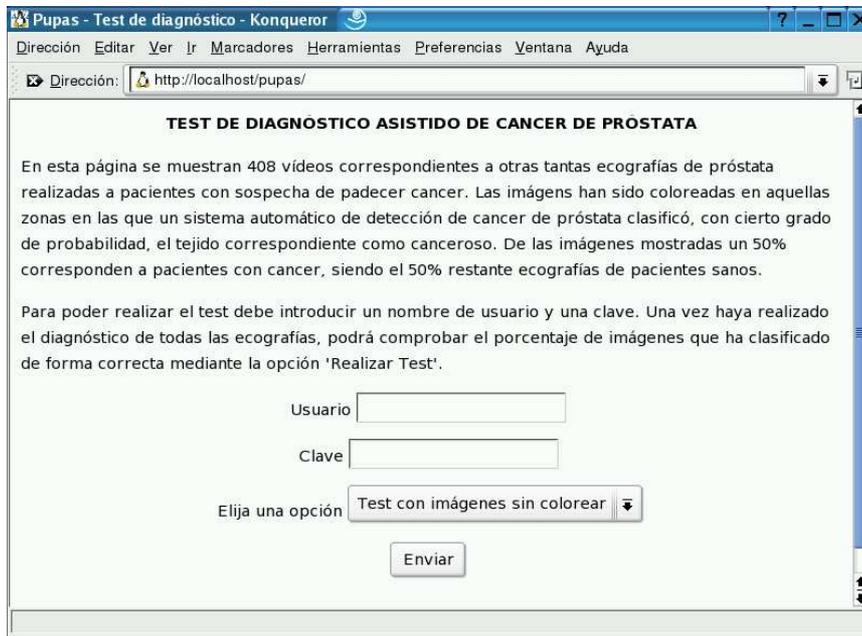


Figura 5.11: Pantalla de inicio del programa de evaluación del sistema de ayuda al diagnóstico.

y una serie de campos para indicar la sospecha tumoral de dichas ecografías. Tras reproducir el vídeo de la ecografía a través del enlace correspondiente, se debe establecer el nivel de sospecha estimado. En lugar de clasificar las ecografías en las clases *cáncer* y *no-cáncer*, se establecieron 6 niveles distintos de sospecha tumoral, para permitir de este modo realizar un análisis mediante curvas ROC. La escala utilizada fue 0 – 2 – 4 – 6 – 8 – 10, siendo 0 la sospecha nula y 10 la máxima sospecha. El campo *SD* (*sin diagnóstico*) indica que la ecografía correspondiente todavía no ha sido clasificada.

En cada uno de los dos conjuntos de ecografías a clasificar (originales y coloreadas), el orden en que se presentan las ecografías es distinto, por lo que cuando un urólogo clasifica las ecografías de un conjunto determinado, desconoce el nivel de sospecha que le asignó en el otro conjunto.

En las ecografías originales, la graduación de sospecha tumoral se llevó a cabo teniendo en cuenta la apariencia de dicha ecografía y basándose en el criterio y experiencia que cada urólogo poseyera. En el caso de ecografías con falso color, se consideró, no tanto la cantidad de píxeles coloreados dentro del área de punción, sino su agrupación en regiones que pudieran indicar con más probabilidad la existencia de un tumor.

En estos experimentos se valoró la sospecha de cáncer de próstata exclusivamente en función de la ecografía individual de una biopsia en concreto, en un corte

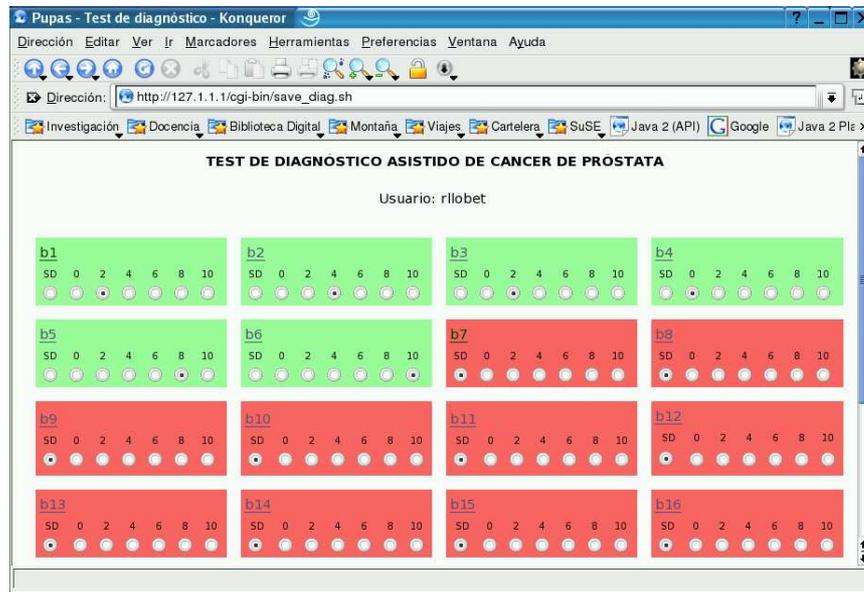


Figura 5.12: Evaluación del sistema de ayuda al diagnóstico. Tras seleccionar una de las biopsias del test y visualizar el vídeo correspondiente, el urólogo establece el valor de sospecha asociado a dicha biopsia, siendo 10 la sospecha máxima de cáncer. Posteriormente, mediante curvas ROC, se evalúa la capacidad diagnóstica del urólogo.

longitudinal, sin tener en cuenta otros datos como las características del tacto rectal o el valor de PSA. Aunque obviar esta información limita la capacidad diagnóstica, dicha limitación se produce exactamente igual en los dos conjuntos a evaluar (ecografías originales/coloreadas). Con ello, aunque la capacidad diagnóstica obtenida pueda ser algo menor a la que se obtendría en un sistema real, la comparación entre los dos conjuntos de test sigue siendo válida.

5.4.2. Resultados

Cuatro urólogos, todos ellos expertos en análisis de ecografía transrectal, y una quinta persona sin ningún tipo de experiencia en evaluación de imágenes de ultrasonidos (el autor de este trabajo), clasificaron tanto el conjunto de ecografías originales como el conjunto de ecografías con falso color.

A continuación se presentan los resultados obtenidos por cada uno de los cuatro urólogos que realizaron el test. En dichos resultados se muestra, de forma tabulada, los valores de porcentaje de aciertos, número total de verdaderos positivos (VP), verdaderos negativos (VN), falsos positivos (FP), falsos negativos (FN), sensibilidad y especificidad. Puesto que el nivel de sospecha se determinó mediante un valor en la escala 0 – 2 – 4 – 6 – 8 – 10, los resultados se presentan para los distintos umbrales posibles (1 – 3 – 5 – 7 – 9). Dado un umbral T y un nivel

de sospecha S , para calcular los distintos resultados, se consideran todas las ecografías con $S > T$ clasificadas como *cáncer* y todas las ecografías con $S < T$ clasificadas como *no-cáncer*. Dichos resultados se muestran, para cada urólogo y para cada conjunto de test (original y coloreado) en las Tablas 5.2 a 5.9.

Adicionalmente a los datos tabulados, se calcularon las curvas ROC obtenidas por cada urólogo en cada uno de los conjuntos de test. Dichas curvas se muestran en las Figuras 5.13 a 5.16.

A modo de resumen, en las Tablas 5.10 y 5.11 y en la Figura 5.17, se muestran los datos globales obtenidos conjuntamente por los cuatro urólogos que realizaron el test.

En la Tabla 5.12 y Figura 5.18 se muestran, respectivamente, las tasas de clasificación y curvas ROC obtenidas por una persona sin ningún tipo de experiencia en análisis de imágenes ecográficas.

Como puede observarse en los resultados presentados, el rendimiento diagnóstico de una persona sin experiencia previa en análisis de imágenes ecográficas, se aproxima mucho a la obtenida por urólogos expertos cuando se cuenta con la ayuda del sistema informático (ecografías con falso color).

En cuanto a los resultados obtenidos por los urólogos, puede observarse que el hecho de realizar la clasificación con el sistema de CAD, únicamente aporta una pequeña mejora en la precisión diagnóstica respecto a la conseguida cuando se trabaja sin dicho sistema. Incluso en uno de los cuatro urólogos que realizaron el test, la capacidad diagnóstica se merma cuando es asistido por el sistema de CAD. Ello hizo sospechar que los urólogos, al clasificar las imágenes coloreadas, podrían no haber tenido en cuenta su intuición y experiencia en el análisis de ecografías en escala de grises y haber utilizado únicamente la información aportada por el sistema de CAD para realizar la clasificación de las imágenes coloreadas.

Para confirmar este extremo, se realizó un último experimento en el que se tuviera en cuenta, tanto la experiencia del urólogo en clasificación de ecografías en escala de grises, como la información aportada por el sistema de CAD. Para ello se combinó, para cada ecografía, los valores de sospecha tumoral asignados en cada uno de los dos conjuntos de test. El valor combinado se obtuvo simplemente calculando la media aritmética de ambos valores. Como se muestra en la Figura 5.17 la curva ROC obtenida de este modo supera ligeramente a las curvas individuales de cada conjunto de test por separado. Esto confirma que el sistema de ayuda al diagnóstico es capaz de encontrar información en un espacio de características distinto al espacio de características con el que trabaja habitualmente el urólogo y, por tanto, añadir cierta información a la imagen, que el urólogo no es capaz de detectar en la imagen original.

Para finalizar, en la Tabla 5.13 se muestra un resumen con las áreas bajo la curva ROC obtenida en cada uno de los experimentos.

Urólogo 1

Tabla 5.2: Urólogo 1: resultados de clasificación para ecografías originales

Umbral	Aciertos	VP	VN	FP	FN	Sensibilidad	Especificidad
1	53.4 %	162	56	148	42	79.4 %	27.5 %
3	58.1 %	129	108	96	75	63.2 %	52.9 %
5	61.0 %	98	151	53	106	48.0 %	74.0 %
7	56.1 %	37	192	12	167	18.1 %	94.1 %
9	51.5 %	8	202	2	196	3.9 %	99.0 %

Tabla 5.3: Urólogo 1: resultados de clasificación para ecografías coloreadas

Umbral	Aciertos	VP	VN	FP	FN	Sensibilidad	Especificidad
1	57.6 %	176	59	145	28	86.3 %	28.9 %
3	59.3 %	130	112	92	74	63.7 %	54.9 %
5	58.1 %	97	140	64	107	47.5 %	68.6 %
7	58.8 %	61	179	25	143	29.9 %	87.7 %
9	52.9 %	21	195	9	183	10.3 %	95.6 %

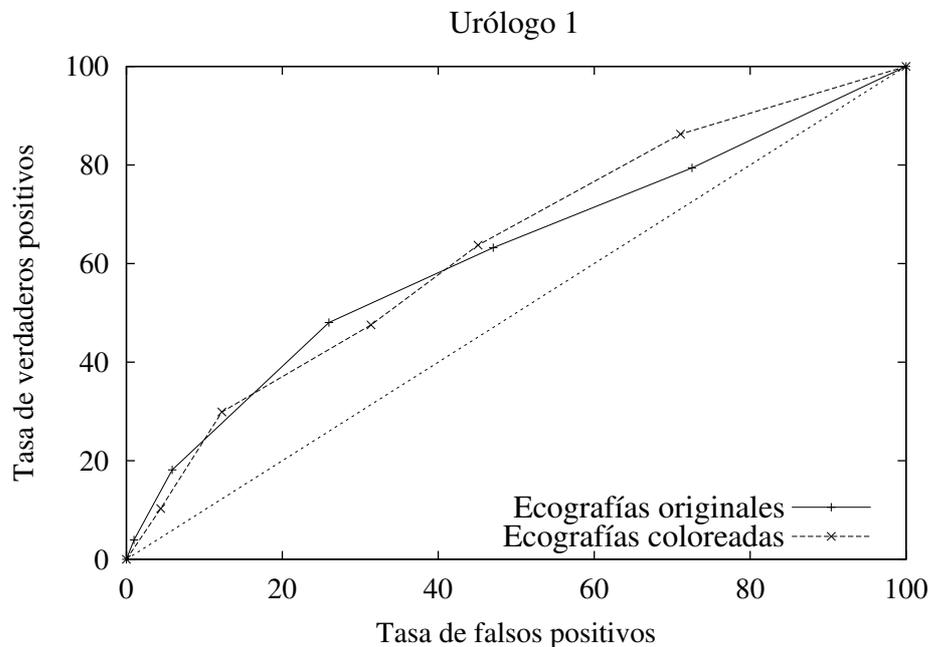


Figura 5.13: Curvas ROC del urólogo 1 correspondientes a la clasificación de ecografías originales y ecografías coloreadas. Las áreas bajo la curva ROC son, respectivamente, 61.7 % y 63.3 %.

Urólogo 2

Tabla 5.4: Urólogo 2: resultados de clasificación para ecografías originales

Umbral	Aciertos	VP	VN	FP	FN	Sensibilidad	Especificidad
1	51.5 %	181	29	175	23	88.7 %	14.2 %
3	60.3 %	92	154	50	112	45.1 %	75.5 %
5	55.6 %	44	183	21	160	21.6 %	89.7 %
7	52.7 %	14	201	3	190	6.9 %	98.5 %
9	50.5 %	2	204	0	202	1.0 %	100.0 %

Tabla 5.5: Urólogo 2: resultados de clasificación para ecografías coloreadas

Umbral	Aciertos	VP	VN	FP	FN	Sensibilidad	Especificidad
1	52.5 %	203	11	193	1	99.5 %	5.4 %
3	64.2 %	134	128	76	70	65.7 %	62.7 %
5	58.8 %	52	188	16	152	25.5 %	92.2 %
7	51.7 %	14	197	7	190	6.9 %	96.6 %
9	49.8 %	0	203	1	204	0.0 %	99.5 %

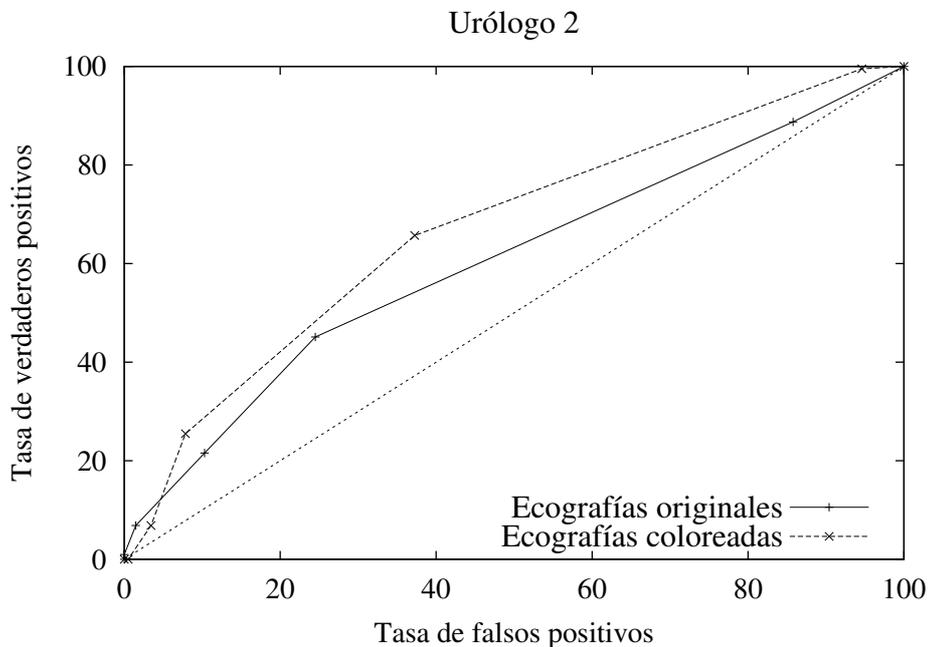


Figura 5.14: Curvas ROC del urólogo 2 correspondientes a la clasificación de ecografías originales y ecografías coloreadas. Las áreas bajo la curva ROC son, respectivamente, 60.5 % y 67.0 %.

Urólogo 3

Tabla 5.6: Urologo 3: resultados de clasificación para ecografías originales

Umbral	Aciertos	VP	VN	FP	FN	Sensibilidad	Especificidad
1	51.7 %	191	20	184	13	93.6 %	9.8 %
3	61.8 %	153	99	105	51	75.0 %	48.5 %
5	62.0 %	132	121	83	72	64.7 %	59.3 %
7	61.0 %	87	162	42	117	42.6 %	79.4 %
9	57.6 %	35	200	4	169	17.2 %	98.0 %

Tabla 5.7: Urólogo 3: resultados de clasificación para ecografías coloreadas

Umbral	Aciertos	VP	VN	FP	FN	Sensibilidad	Especificidad
1	51.5 %	201	9	195	3	98.5 %	4.4 %
3	58.3 %	156	82	122	48	76.5 %	40.2 %
5	57.4 %	149	85	119	55	73.0 %	41.7 %
7	64.0 %	110	151	53	94	53.9 %	74.0 %
9	51.5 %	12	198	6	192	5.9 %	97.1 %

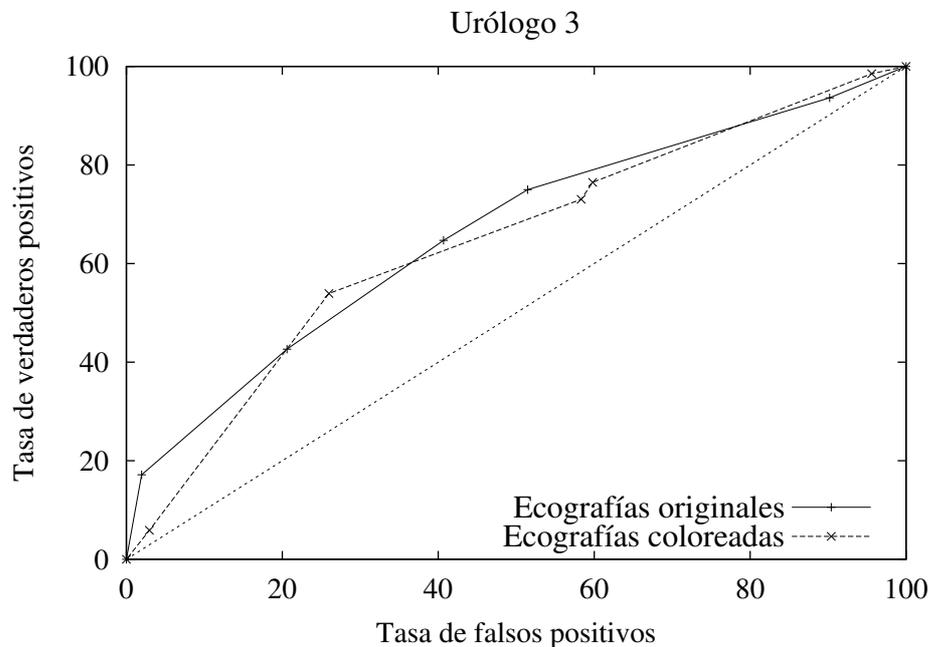


Figura 5.15: Curvas ROC del urólogo 3 correspondientes a la clasificación de ecografías originales y ecografías coloreadas. Las áreas bajo la curva ROC son, respectivamente, 66.2 % y 64.3 %.

Urólogo 4

Tabla 5.8: Urologo 4: resultados de clasificación para ecografías originales

Umbral	Aciertos	VP	VN	FP	FN	Sensibilidad	Especificidad
1	53.4 %	188	30	174	16	92.2 %	14.7 %
3	58.1 %	119	118	86	85	58.3 %	57.8 %
5	58.8 %	92	148	56	112	45.1 %	72.5 %
7	52.7 %	20	195	9	184	9.8 %	95.6 %
9	50.0 %	0	204	0	204	0.0 %	100.0 %

Tabla 5.9: Urólogo 4: resultados de clasificación para ecografías coloreadas

Umbral	Aciertos	VP	VN	FP	FN	Sensibilidad	Especificidad
1	55.6 %	184	43	161	20	90.2 %	21.1 %
3	61.5 %	138	113	91	66	67.6 %	55.4 %
5	57.8 %	95	141	63	109	46.6 %	69.1 %
7	56.9 %	43	189	15	161	21.1 %	92.6 %
9	50.0 %	1	203	1	203	0.5 %	99.5 %

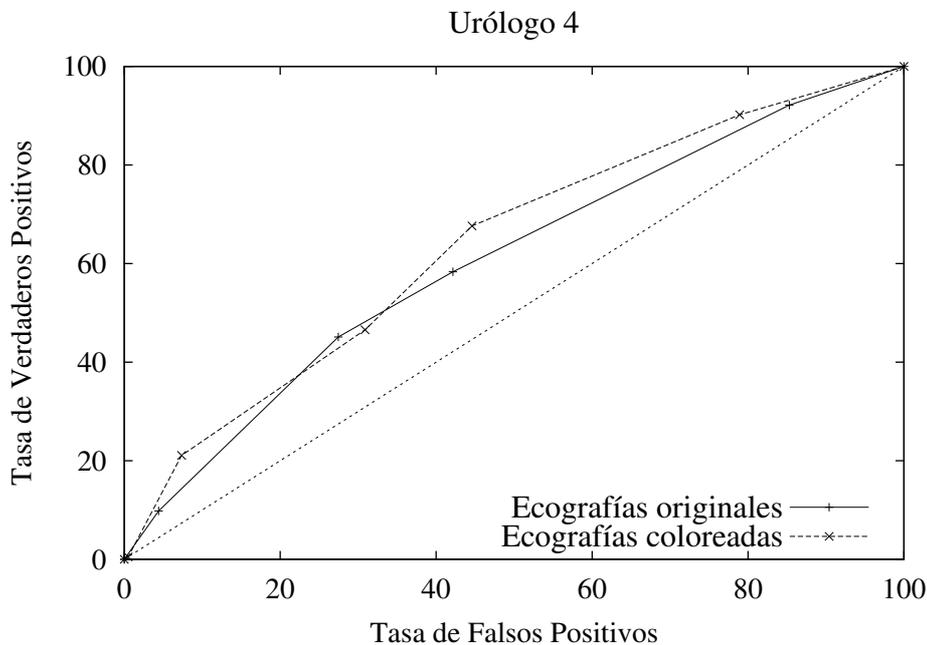


Figura 5.16: Curvas ROC del urólogo 4 correspondientes a la clasificación de ecografías originales y ecografías coloreadas. Las áreas bajo la curva ROC son, respectivamente, 60.7 % y 63.7 %.

Resultados combinados de todos los urólogos

Tabla 5.10: Resultados de clasificación para ecografías originales

Umbral	Aciertos	VP	VN	FP	FN	Sensibilidad	Especificidad
1	52.5 %	722	135	681	94	88.5 %	16.5 %
3	59.6 %	493	479	337	323	60.4 %	58.7 %
5	59.4 %	366	603	213	450	44.9 %	73.9 %
7	55.6 %	158	750	66	658	19.4 %	91.9 %
9	52.4 %	45	810	6	771	5.5 %	99.3 %

Tabla 5.11: Resultados de clasificación para ecografías coloreadas

Umbral	Aciertos	VP	VN	FP	FN	Sensibilidad	Especificidad
1	54.3 %	764	122	694	52	93.6 %	15.0 %
3	60.8 %	558	435	381	258	68.4 %	53.3 %
5	58.0 %	393	554	262	423	48.2 %	67.9 %
7	57.8 %	228	716	100	588	27.9 %	87.7 %
9	51.0 %	34	799	17	782	4.2 %	97.9 %

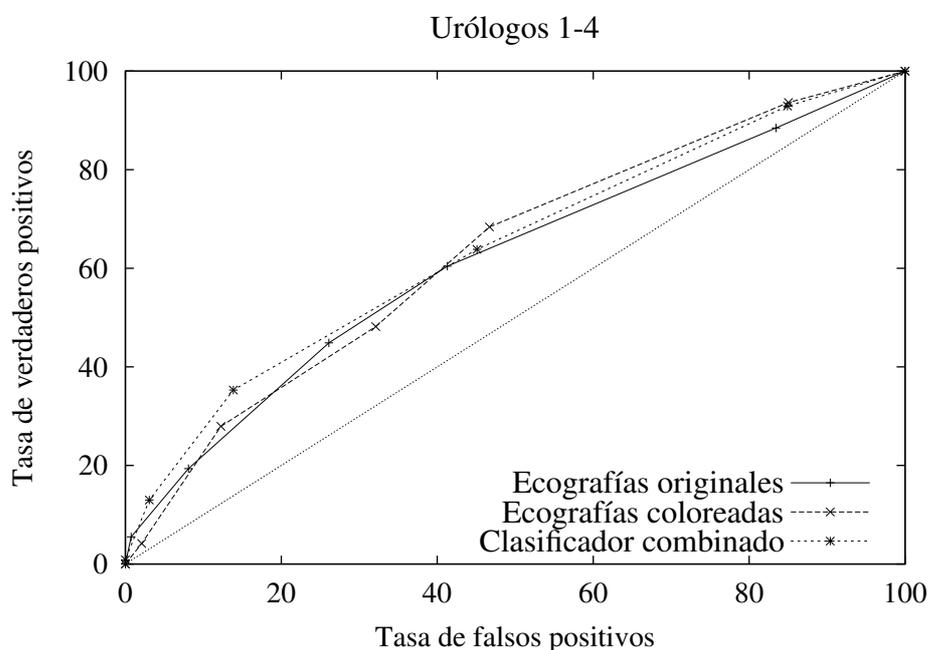


Figura 5.17: Curvas ROC correspondientes a la clasificación de ecografías originales, ecografías coloreadas y un clasificador combinado con ecografías originales y coloreadas. Las áreas bajo la curva ROC son, respectivamente, 61.7 %, 63.3 % y 64.1 %.

Persona sin experiencia previa en análisis de TRUS

Tabla 5.12: Resultados de clasificación para ecografías coloreadas obtenidos por una persona sin experiencia en análisis de ecografías

Umbral	Aciertos	VP	VN	FP	FN	Sensibilidad	Especificidad
1	51.5 %	200	10	194	4	98.0 %	4.9 %
3	55.4 %	167	59	145	37	81.9 %	28.9 %
5	59.3 %	109	133	71	95	53.4 %	65.2 %
7	56.1 %	50	179	25	154	24.5 %	87.7 %
9	51.5 %	8	202	2	196	3.9 %	99.0 %

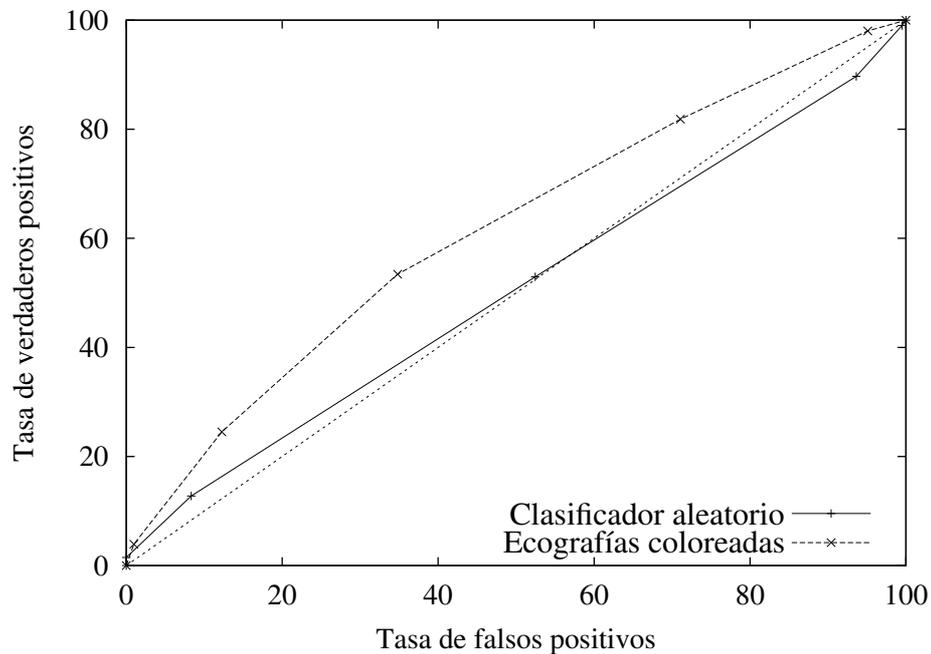


Figura 5.18: Curvas ROC obtenidas por una persona inexperta en exploración de ecografías, correspondientes a la clasificación de ecografías originales (realizada de un modo completamente aleatorio) y de ecografías coloreadas. Las áreas bajo la curva ROC son, respectivamente, 50.1 % y 61.3 %.

Resumen de resultados

Tabla 5.13: Resumen de áreas bajo la curva ROC en cada uno de los experimentos.
 (*) *En este caso la clasificación se hizo de forma aleatoria.*

Test	U1	U2	U3	U4	Todos	No-experto
Original	61.7 %	60.5 %	66.2 %	60.7 %	61.7 %	50.1 %*
Falso color	63.3 %	67.0 %	64.3 %	63.7 %	63.3 %	61.3 %
Combinado	—	—	—	—	64.1 %	—

5.5. Conclusiones

Tanto en este capítulo como en la Sección 3.3.2 se ha puesto de manifiesto las grandes dificultades que tienen actualmente los urólogos para detectar nódulos cancerosos en la próstata mediante la inspección visual de imágenes de ecografía transrectal. Estas dificultades son la causa principal de que aproximadamente el 10 % de los pacientes que se someten a una exploración de próstata y realmente tienen un tumor maligno, dicho tumor no sea detectado debido a que ninguna de las biopsias se realiza en el lugar exacto donde se encuentra.

En este capítulo se ha presentado y evaluado un sistema de ayuda al diagnóstico de cáncer de próstata, cuyo objetivo es paliar las dificultades intrínsecas de que adolecen las técnicas actuales de diagnóstico. Dicho sistema se basa en el análisis de ecografías transrectales digitalizadas y su posterior postproceso para resaltar, en la propia ecografía, las regiones con alta sospecha tumoral.

Para llevar a cabo el estudio, en primer lugar se ha procedido a la adquisición de un corpus de imágenes de ecografía transrectal y su posterior etiquetado. Estas imágenes son exactamente las mismas con las que trabajan los urólogos habitualmente. A continuación se ha realizado un estudio para evaluar, hasta qué punto, un sistema de clasificación automática podría ser capaz de segmentar el tejido canceroso en dichas imágenes. El objetivo de esta fase no es, de momento, implementar un sistema de ayuda al diagnóstico, sino únicamente evaluar distintas aproximaciones, tanto de extracción de características como de clasificación, para detectar cuál de ellas ofrece un mayor poder discriminatorio, así como valorar cuantitativamente dicha capacidad de discriminación.

Para ello se han evaluado dos técnicas distintas de extracción de características: niveles de gris y matrices de coocurrencia. Como técnica de clasificación se utilizó, en primera instancia, k -NN. Tanto la aproximación de niveles de gris como de matrices de coocurrencia ofrecieron resultados muy similares, con un área bajo la curva ROC (AUC) de 59.7 % y 60.1 % respectivamente, con un intervalo de confianza al 95 % del 0.7 % en ambos casos. Los resultados obtenidos demuestran que es posible discriminar entre tejido sano y canceroso en las imágenes de ecografía

transrectal, aunque dicho poder discriminatorio no sea excesivamente alto.

Posiblemente el escaso rendimiento de las técnicas propuestas se deba, tanto a la dificultad intrínseca de la tarea (tal y como se ha visto a lo largo de este y otros capítulos, ni siquiera un experto humano apenas es capaz de distinguir el tejido sano del canceroso), como al problema de la supervisión imperfecta apuntado en la Sección 5.2.5.

Para tratar de paliar el efecto de la supervisión imperfecta, se ha evaluado un segundo clasificador basado en HMMs continuos. Al igual que se hizo para el caso de k -NN, el clasificador basado en HMMs se ha evaluado con los dos esquemas de extracción de características propuestos. Los resultados obtenidos, en forma de área bajo la curva ROC, han sido del 61.6 % para el caso de niveles de gris y 60.0 % para el caso de matrices de coocurrencia, con un intervalo de confianza al 95 % del 0.7 % en ambos casos.

Como puede observarse, el clasificador HMM trabajando con niveles de gris muestra una muy ligera ventaja frente al resto, aunque a groso modo los resultados obtenidos en las distintas aproximaciones son muy similares. Posiblemente la causa de esta paridad en los resultados se deba a la imposibilidad de encontrar información suficiente en la imagen que permita sobrepasar holgadamente la cota obtenida en torno al 60 %. Este extremo, en cualquier caso, es difícil de confirmar.

A pesar de la limitación de los resultados obtenidos, dado que se ha confirmado la posibilidad de discriminar entre tejido sano y canceroso, aunque sea de forma discreta, se decide abordar una segunda fase para implementar un sistema de ayuda al diagnóstico y evaluar su utilidad clínica. Tal y como se ha explicado en secciones anteriores, 5 personas con distinto grado de experiencia han elaborado su propio diagnóstico a partir de las imágenes originales por un lado, y de las imágenes coloreadas por otro. Los resultados demuestran que 3 de los 4 urólogos que realizaron el test consiguieron mejorar su diagnóstico al utilizar el sistema de CAD propuesto, aunque dicha mejora sea muy sutil. En uno de los urólogos parece observarse una clara mejora, pasando de una AUC del 60.5 % al 67.0 %, sin embargo en otro de los urólogos el uso del sistema de CAD llega a empeorar su capacidad diagnóstica, pasando de una AUC del 66.2 % al 64.3 %.

Por otro lado, una persona sin ningún tipo de experiencia en inspección de ecografías, consigue, con la ayuda del sistema, una AUC del 61.3 %, tasa similar a la conseguida por un urólogo experto cuando trabaja sin el sistema de ayuda.

Puede concluirse que el sistema de ayuda al diagnóstico de cáncer de próstata propuesto, claramente ayuda a una persona inexperta en realizar un mejor diagnóstico, poniéndola en un nivel similar a urólogos con varios años de experiencia. Sin embargo, la utilidad clínica del sistema es escasa cuando se trabaja en un entorno real en el que, por razones obvias, la persona que lo utilizará debería ser un urólogo con cierta experiencia.

Si las imágenes de ecografía fueran más nítidas y el problema de la supervisión imperfecta pudiera resolverse, posiblemente las presetaciones del sistema propuesto aumentarían. En este sentido debe decirse que la tecnología relacionada con la imagen de ecografía ha mejorado enormemente en los últimos años. Actualmente, en muchos hospitales se dispone de equipos de ecografía con prestaciones muy superiores a las del empleado en este estudio. Hay que tener en cuenta, sin embargo, que una mejora en los equipos de adquisición, conllevaría un aumento de la capacidad diagnóstica, no sólo del sistema de clasificación automático, sino también del experto humano. Por tanto, aunque se mejorase el rendimiento del clasificador, a medida que el experto humano tiene mayor capacidad para elaborar correctamente su propio diagnóstico, más difícil resulta aportarle información que no conozca ya. Habría que evaluar, por tanto, en qué medida mejoran uno y otro sus tasas de clasificación y, sobretodo, si el sistema de clasificación automático es capaz de aportar información extra que el experto humano no sea capaz de observar por sí mismo, o que, aun teniendo la capacidad de observarla, no lo haga debido a otros motivos (fatiga, desmotivación, etc.).

Finalmente, es difícil valorar cualitativamente los resultados obtenidos teniendo en cuenta la dificultad de la tarea. Además, la imposibilidad de hacer un estudio comparativo con las aproximaciones presentadas por otros autores, hace aún más complicado evaluar la bondad de los algoritmos propuestos. Aunque todo parece apuntar a que la tarea abordada es extremadamente compleja y los resultados obtenidos difícilmente podrían mejorarse significativamente, sería gratuito afirmar tajantemente este extremo y concluir de este modo el trabajo. Es por ello por lo que se ha decidido aplicar estas mismas técnicas a otra tarea de CAD, donde las imágenes de partida son mucho más nítidas, el etiquetado es correcto y la posibilidad de comparar objetivamente los resultados obtenidos con lo publicado por otros autores, una realidad. En el siguiente capítulo se aborda el diagnóstico de cáncer de mama asistido por ordenador mediante técnicas muy similares a las utilizadas en el sistema de CAD que se ha presentado en este capítulo.

Capítulo 6

Aportaciones al diagnóstico de cáncer de mama

6.1. Introducción

Tal y como se explicó en la Sección 3.4.1, actualmente el método más utilizado para la detección precoz del cáncer de mama es, además de la autoexploración, la mamografía mediante rayos X. Esta técnica, a pesar de ofrecer una gran sensibilidad, puede dejar sin detectar entre un 10 % y un 25 % de los tumores, lo que provoca un retraso en el diagnóstico y su consiguiente complicación en el tratamiento [Brak 98a, Wall 91]. No debe olvidarse que el índice de curación de un tumor de mama detectado en sus primeras fases es de prácticamente el 100 %, decreciendo drásticamente este porcentaje a medida que el tumor crece, lo que corrobora la importancia de un diagnóstico precoz.

Las causas fundamentales en los errores de falsos negativos son atribuibles tanto a la dificultad intrínseca que conlleva analizar una mamografía como a factores humanos. Por un lado, la imagen de mamografía muestra grandes variaciones, tanto en el tejido sano como en el canceroso. Por otro lado, el cansancio acumulado de los radiólogos responsables de analizar las mamografías y la técnica de búsqueda no sistemática empleada por los mismos, provoca que algunos tumores que podrían detectarse con una exploración más cuidadosa, pasen inadvertidos.

Varios estudios (algunos de ellos descritos en la Sección 3.4.3) apuntan a que un método posible para mejorar el diagnóstico de cáncer de mama consiste en el uso de sistemas de ayuda al diagnóstico asistido por ordenador (CAD). La idea fundamental de estos sistemas de CAD consiste en marcar sobre la propia mamografía las zonas con alta sospecha tumoral, llamando de este modo la atención del radiólogo sobre dichas zonas.

En este capítulo se presentan una serie de aportaciones al diagnóstico de cáncer

de mama asistido por ordenador. Algunas de las técnicas empleadas son similares a las utilizadas en el capítulo anterior para la detección de cáncer de próstata. Otras son específicas para esta tarea en concreto. De un modo u otro, todas ellas han sido explicadas en el Capítulo 4.

Para la fase experimental del estudio se ha utilizado la base de datos de uso público MIAS [Suck 94]. Ello ha permitido comparar los resultados obtenidos con los logrados por otros autores que utilizaron esta misma base de datos [Heat 00b, Brak 98b, Sall 99].

Ya se expuso en el la Tabla 3.2 los tipos de tumores que pueden observarse en una mamografía. Debe quedar claro que el objetivo del estudio que se presenta en este capítulo es la detección y localización de todos los tipos de tumor descritos en dicha tabla excepto las microcalcificaciones. Los métodos propuestos en este trabajo no son válidos para la detección de este tipo de anomalías, debido a su peculiar apariencia, muy distinta del resto de tumores. Para poder detectar las microcalcificaciones habría que emplear otras técnicas, igualmente viables atendiendo a la gran cantidad de trabajos publicados al respecto [Chan 90, Roeh 98, Will 98, Nam 98, Wils 98, Wu 92], pero en cualquier caso distintas de las empleadas para el resto de anomalías. En términos generales, la detección de microcalcificaciones entraña menor dificultad que la detección de otros tipos de anomalías [Comp 01, Roeh 98, Will 98]. Por otro lado, otros trabajos previos con los que se pretende contrastar los resultados obtenidos, tampoco utilizaron en el conjunto de test ninguna mamografía que presentase microcalcificaciones.

Es necesario puntualizar que el problema abordado en este capítulo es ligeramente distinto al del diagnóstico de cáncer de próstata tratado en el capítulo anterior. En el caso del diagnóstico mediante TRUS no se tenía información sobre la localización de los tumores, conociéndose únicamente si una imagen contenía o no tejido canceroso. En esas condiciones, por tanto, para evaluar el rendimiento del sistema, nos conformábamos con clasificar cada imagen (o más exactamente, el rectángulo de la imagen correspondiente al tejido biopsiado) en una de las dos posibles clases (*cáncer* o *no-cáncer*). En el caso de las mamografías, no se conoce únicamente si la mama presenta o no células cancerosas, sino también dónde se encuentra exactamente el tumor (o tumores) en caso de que exista. Al sistema de CAD, en este caso, debemos exigirle mayor precisión que al presentado en el capítulo anterior, de modo que muestre las regiones sospechosas correctamente localizadas. Ya no se trata únicamente de un problema de clasificación de imágenes, sino de detección y localización de tumores.

Para abordar el problema de la detección, la tarea puede dividirse en dos fases: en primer lugar se calcula el nivel de sospecha de cada píxel de la mamografía en base a ciertas características obtenidas a partir de una ventana de vecindad centrada en dicho píxel, tal y como se explicó en la Sección 4.2. Con ello se define lo que podríamos llamar un “mapa de sospecha”. En una segunda fase, se realiza un postproceso sobre dicho mapa de sospecha para agrupar píxeles próximos con

alta sospecha tumoral en lo que denominaremos “regiones sospechosas”. Combinando de algún modo el nivel de sospecha individual de cada uno de los píxeles que forman parte de una región sospechosa, se calcula un nivel de sospecha de la región completa. Obviamente, el éxito de esta segunda fase vendrá condicionado por la precisión con la que se hayan estimado los niveles de sospecha de cada píxel individualmente.

Atendiendo a las dos fases anteriormente expuestas, el trabajo presentado en este capítulo se ha dividido igualmente en dos partes. En primer lugar, para simplificar la tarea, se ha partido de un conjunto de regiones de interés (ROIs) extraídas de las mamografías. Cada ROI viene determinada por un único píxel (el correspondiente al centro de la misma). Colocando una ventana de vecindad sobre el píxel que representa la ROI, se obtienen las características de la misma. Algunas de las ROIs utilizadas corresponden a tejido canceroso y otras a tejido sano. En esta primera parte debe clasificarse cada ROI en una de las dos clases posibles (*cáncer* o *no-cáncer*). Con ello la tarea se reduce a un problema de clasificación y no de detección. La precisión de la solución propuesta puede evaluarse mediante curvas ROC. La finalidad de esta primera fase es encontrar los algoritmos que mejores prestaciones ofrecen a la hora de estimar el nivel de sospecha individual de cada píxel de la mamografía, obviando la parte de postproceso del mapa de sospecha anteriormente mencionado y la consiguiente búsqueda de regiones sospechosas. Los esquemas de extracción de características evaluados (explicados en la Sección 4.2) son: niveles de gris, matrices de coocurrencia, filtro de Sobel, AFUM y SFUM. Como clasificador se utilizó la regla de los k -vecinos más cercanos con búsqueda aproximada, explicada en la Sección 4.6, así como Máquinas de Vectores Soporte (SVM) y *Boosting* introducidos en las Secciones 2.6 y 2.7 respectivamente.

En una segunda etapa, empleando el esquema de extracción de características y los parámetros que mejores resultados ofrecen en la etapa anterior, se evalúa un sistema de detección de tumores a partir de la imagen completa de mamografía. En este caso ya no se trata de un problema puramente de clasificación, sino de detección, de modo que en cada imagen pueden aparecer cero, uno o más tumores. Para ello, tal y como se ha explicado anteriormente, es necesario determinar las regiones con alta sospecha tumoral a partir del mapa de sospecha. El rendimiento del sistema en este caso se evalúa mediante curvas FROC.

A lo largo de este capítulo se detalla, en la Sección 6.2, la base de datos de mamografías utilizada. A continuación, en la Sección 6.3, se explica el entrenamiento del modelo, utilizado tanto en el problema de clasificación de ROIs como en la tarea final de detección de tumores en mamografías completas. En la Sección 6.4 se muestran los experimentos relativos a la primera etapa: selección de las ROIs y resultados de clasificación de dichas ROIs con cada uno de los esquemas de extracción de características propuestos. Finalmente, en la Sección 6.5, se aborda la segunda fase, relativa a la detección de tumores en imágenes completas.

6.2. Corpus

Existen diversas bases de datos de mamografías de carácter público que permiten evaluar el rendimiento de un test de diagnóstico (ya sea asistido por ordenador o tradicional) mediante un análisis retrospectivo.

En la Tabla 6.1 se detallan algunas de las bases de datos de mamografías digitalizadas actualmente disponibles. Todas ellas están debidamente supervisadas y se especifica, de algún modo, las regiones o píxeles correspondientes a tejido canceroso.

Tabla 6.1: Bases de datos de mamografías actualmente disponibles

Nombre	Descripción
DDSM	<i>Digital Database for Screening mammography</i> . Base de datos creada por el Hospital General de Massachusetts, la Universidad de Florida y <i>Sandia National Laboratories</i> . Fuente: http://marathon.csee.usf.edu/Mammography/Database.html
MIAS	Base de datos creada por la <i>Mammographic Image Analysis Society</i> , Reino Unido. Fuente: http://www.wiau.man.ac.uk/services/MIAS/MIASweb.html . Existe una versión reducida que contiene las mismas imágenes que la versión original, pero reducidas a un tamaño de 1024×1024 píxeles, disponible en http://www.wiau.man.ac.uk/services/MIAS/MIASmini.html .
LLNL/UCSF	Base de datos creada por Lawrence Livermore National Laboratories (LLNL) y el Departamento de Radiología de la Universidad de California en San Francisco (UCSF). Fuente: mammo-db-help@llnl.gov
Nijmegen	Base de datos creada por el <i>National Expert and Training Centre for Breast Cancer Screening</i> y por el Departamento de Radiología de la Universidad de Nijmegen (Holanda). Fuente: nico@mbfys.kun.nl

En este trabajo se ha empleado la base de datos recopilada por la *Mammographic Image Analysis Society* (MIAS) [Suck 94]. Esta base de datos consta de 322 mamografías correspondientes al pecho izquierdo y derecho de 161 pacientes. De las 322 mamografías, 115 presentan uno o más tumores, con un total de 123 tumores. Las 207 restantes corresponden a pacientes sanas. Los tumores se clasifican en 6 categorías: calcificaciones (CALC), masas circunscritas (CIRC), masas radiales (SPIC), distorsiones estructurales (ARCH), asimetrías (ASYM) y masas poco definidas (MISC), tal y como se describió en la Tabla 3.2. La Tabla 6.2 muestra el número de imágenes y de tumores en la base de datos MIAS para cada una de las categorías definidas.

	Imágenes	Tumores
CALC	25	30
CIRC	23	25
SPIC	19	19
ARCH	19	19
ASYM	15	15
MISC	14	15
Total	115	123
NORM	207	-

Tabla 6.2: Tamaño de la base de datos MIAS desglosada por tipo de tumor. Para cada tipo de tumor se muestra el número de imágenes en las que aparece dicho tumor (columna central) y número total de tumores (columna derecha). La última fila muestra las imágenes normales (sin ningún tipo de tumor).

El etiquetado de los tumores se define mediante las coordenadas del centro y el radio aproximado del mismo. Puede deducirse, por tanto, que el etiquetado no es del todo preciso puesto que los tumores no tienen una forma circular exacta (y en ocasiones ni siquiera aproximada). Otras bases de datos, como la DDSM, ofrecen un etiquetado a nivel de píxel mucho más preciso.

Cada mamografía tiene una resolución de 5 píxeles/mm. y un tamaño de 1024×1024 píxeles, con una profundidad de 8 bits por píxel (256 niveles de gris).

Desafortunadamente, no están definidos en esta base de datos unos conjuntos “estándar” de entrenamiento y test, lo que dificulta realizar un estudio comparativo con otros resultados previamente publicados. Sin embargo, en algunos trabajos se detalla con exactitud qué imágenes de la base de datos fueron usadas en el conjunto de test [Heat 00b, Brak 98a, Sall 99]. En nuestro estudio emplearemos los conjuntos utilizados en los trabajos citados anteriormente, con el fin de poder comparar los resultados obtenidos.

Como se ha comentado anteriormente, en el estudio realizado se descartaron las mamografías con calcificaciones debido, entre otros motivos, a que estas imágenes tampoco fueron utilizadas en [Heat 00b, Brak 98a, Sall 99] y ya se ha comentado que uno de nuestros objetivos es, precisamente, hacer un estudio comparativo de nuestras técnicas con las publicadas en estos trabajos. Por otro lado, una de las mamografías etiquetada como CIRC (masa circunscrita) no tiene definidas las coordenadas de la anomalía, por lo que también fue descartada. Finalmente, la base de datos usada en este trabajo consta de 296 mamografías, 89 de las cuales tienen algún tipo de tumor.

6.3. Entrenamiento del modelo

Para todos los métodos de extracción de características presentados en la Sección 4.2 excepto AFUM, se necesita crear un modelo con prototipos de las clases *cáncer* y *no-cáncer*, para la posterior clasificación de una muestra de test mediante la regla de los k -vecinos más cercanos. Dicho modelo fue creado a partir de los vectores de características obtenidos de las 296 mamografías que componen la base de datos empleada en este trabajo. Los vectores de características se calcularon a partir de una ventana de vecindad o ventana de características que se desliza a lo largo de la imagen de mamografía, tal y como se ha explicado en la Sección 4.2. Durante la fase de test se evitó que en el modelo empleado para clasificar la muestra a evaluar apareciese, no sólo dicha muestra, sino cualquier otra perteneciente a alguna de las mamografías de la misma paciente.

6.3.1. Invariantes a rotación y escalado

En cualquier tarea de clasificación de imágenes, es importante que los métodos utilizados sean invariantes a las distintas transformaciones que pueden sufrir los objetos a clasificar (traslación, rotación, escalado, reflexión o diversas deformaciones). En la tarea que nos ocupa, los tumores pueden verse afectados por la rotación y el escalado.

Básicamente existen tres aproximaciones para conseguir la invarianza a las distintas transformaciones: a) Utilizar características que sean inherentemente invariantes a la transformación deseada; b) Incluir en el modelo, durante la fase de entrenamiento, características del objeto en todas sus transformaciones posibles; c) Analizar la muestra de test en todas sus transformaciones posibles y establecer una regla que combine el resultado obtenido en cada una de esas transformaciones.

La invarianza a la traslación no plantea ningún problema, ya que ésta queda resuelta por el modo de extracción de características mediante una ventana deslizante. En cuanto al escalado y rotación, de los métodos de extracción de características propuestos en la Sección 4.2, ninguno de ellos es inherentemente invariante al escalado y únicamente los métodos AFUM y SFUM son invariantes a la rotación. En consecuencia, de las tres alternativas presentadas anteriormente, sólo las dos últimas son viables con las aproximaciones propuestas.

El problema de la invarianza a la rotación puede ser fácilmente abordado durante el entrenamiento del modelo. Para ello, la ventana de vecindad de la cual se obtienen las características, no sólo debe desplazarse a lo largo de los distintos píxeles que componen la mamografía, sino que para cada uno de ellos debe rotarse a distintos ángulos.

La invarianza al escalado, por contra, no tiene una solución tan inmediata. La

solución en este caso pasa por implementar un esquema de multiescalado, consistente en incluir en el modelo los tumores normalizados a un tamaño fijo, y evaluar las muestras de test a distintas escalas. Esta técnica ha sido explicada en la Sección 4.5.

En base a lo expuesto, durante el entrenamiento del modelo se hicieron rotaciones de los tumores con intervalos de 6 grados. Dichas rotaciones no se realizaron en las aproximaciones basadas en AFUM y SFUM por ser éstas inherentemente invariantes a la rotación. Además, los tumores se normalizaron a un tamaño constante.

Respecto a la invarianza a escalado y rotación en el tejido sano, se siguió un esquema similar al empleado con las muestras positivas. Aunque la cantidad de muestras negativas es suficientemente grande como para considerar que los patrones de tejido sano aparecen en prácticamente todos los ángulos posibles, se obtuvieron características de dichas muestras a distintos ángulos (en este caso cada 10 grados), aunque no todas ellas se incluyeron en el modelo. Tal y como se explica más adelante, el algoritmo de *bootstrapping* empleado es el que toma la decisión de incluir o no en el modelo los prototipos obtenidos en cada rotación. En cuanto al escalado del tejido sano, puesto que no es posible normalizar el mismo a un tamaño fijo, tal y como se hace con los tumores, es necesario incluir prototipos del mismo en el modelo a distintas escalas. Esto es así ya que en la fase de test nos veremos obligados a evaluar cada muestra a distintas escalas y conviene que el tejido sano esté bien representado en cada una de ellas.

6.3.2. *Bootstrapping*

La clase *no-cáncer* cuenta con una gran muestra, ya que la mayor parte del tejido encontrado en las mamografías corresponde a tejido sano. La apariencia del tejido sano, sin embargo, puede presentar grandes variaciones, aunque también es cierto que dado el tamaño de la muestra, todos los patrones de dicho tejido podrían quedar representados en el modelo. Sin embargo, utilizar la muestra entera para la creación del modelo, generaría un número de prototipos difícil de manejar. Por otro lado, escoger de forma arbitraria un subconjunto de la muestra de tejido sano podría implicar que la clase *no-cáncer* no quedase representada en todas sus variantes. Una técnica utilizada habitualmente para evitar el uso de un conjunto de entrenamiento desmesuradamente grande sin perder representatividad es la conocida como *bootstrapping*. La idea básica de esta técnica consiste en ir añadiendo selectivamente nuevos prototipos al modelo, al tiempo que este se va creando [Rowl 98].

En nuestro caso se creó inicialmente un modelo con las características de todas las regiones que contienen algún tipo de tumor (prototipos de la clase *cáncer*) y con un número similar de prototipos de la clase *no-cáncer*. En segundo lugar, las

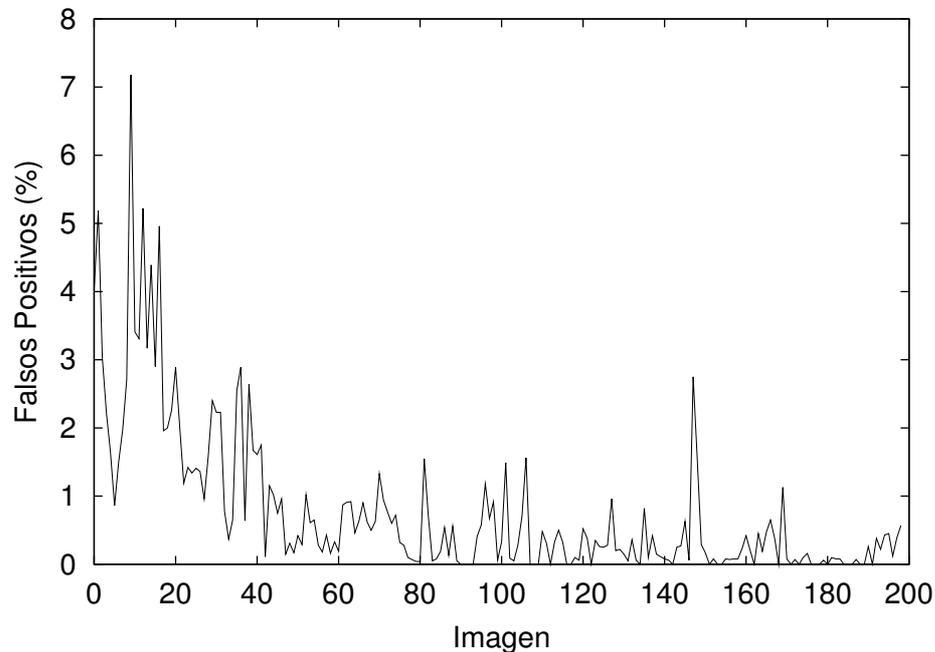


Figura 6.1: Porcentaje de falsos positivos (prototipos añadidos al modelo) detectados en cada una de las imágenes analizadas durante el proceso de *bootstrapping*. El gráfico corresponde a un modelo inicializado con 12,000 muestras negativas y 2,200 positivas.

mamografías sanas que no fueron empleadas en el paso previo de inicialización del modelo, se utilizaron para incorporar nuevos prototipos al modelo mediante la técnica de *bootstrapping* mencionada anteriormente. Para ello, en sucesivas etapas, se calcularon los vectores de características de una mamografía diferente y se clasificaron mediante la regla de los k -vecinos utilizando el último modelo aprendido (o el modelo inicial para la primera etapa). A continuación, únicamente aquéllos vectores clasificados como *cáncer* (esto es, los falsos positivos) se añadieron al modelo.

La Figura 6.1 muestra el porcentaje de muestras negativas añadidas al modelo para cada imagen analizada durante el proceso de *bootstrapping*. Como es de esperar, a medida que avanza el proceso, la clase negativa va quedando mejor representada y, consecuentemente, el número de muestras necesarias para mejorar el modelo va decreciendo. Puede observarse, sin embargo, que este decremento no es monótono, de modo que en algunas imágenes el número de muestras añadidas al modelo es más alto de lo esperado. Esto ocurre cuando se procesa una mamografía cuya apariencia es significativamente distinta de las observadas en mamografías previas.

6.4. Experimentos de clasificación con regiones de interés

En esta sección se aborda la primera parte del estudio explicada en la introducción de este capítulo: evaluación de distintas aproximaciones a la extracción de características y optimización de parámetros mediante clasificación de regiones de interés.

El conjunto de test utilizado en esta primera parte está compuesto por una serie de regiones de interés extraídas de la base de datos MIAS. El conjunto de entrenamiento lo forma la práctica totalidad de la base de datos, tal y como se ha explicado en la sección anterior. Se tiene en cuenta, no obstante, que cuando se evalúa una ROI de una mamografía determinada, ninguna muestra de esa mamografía o de otra perteneciente a la misma paciente sea utilizada en el conjunto de entrenamiento.

En los siguientes apartados se define con más exactitud qué es una región de interés y qué criterios se han seguido para su selección. A continuación se estudian de forma experimental algunos conceptos introducidos anteriormente: preproceso de la imagen, *bootstrapping*, PCA, multiescalado, tamaño de la ventana de características, así como los distintos esquemas de extracción de características propuestos.

En todos los casos se empleó un clasificador basado en k -vecinos y búsqueda rápida aproximada mediante *kd-trees*, con $\epsilon = 2$. El resto de parámetros fueron estudiados de forma experimental. Para cada uno de los experimentos que se presenta en los siguientes apartados se especifica el valor o valores de los parámetros empleados, así como el método de extracción de características utilizado. La nomenclatura que se ha utilizado para referirse a los distintos parámetros y métodos de extracción de características es la siguiente:

	Descripción	Valores posibles
CAR	Método de extracción de características	NG (niveles de gris), SGLDM Sobel, AFUM, SFUM
NCP	Número de componentes principales	1 – 50
T_w	Tamaño de la ventana de características	5 – 25
K_o	Tamaño del tumor normalizado	20 – 35
T_m	Tamaño del área de muestreo	5 – 16
k -NN	Número de vecinos	3 – 200
ESC	Número de escalados	1 – 18
RCH	Regla de combinación de hipótesis	Máximo, media, producto

En los siguientes apartados se muestran diversos resultados para distintos valores de estos parámetros. Los resultados se presentan en forma de curva ROC o de área bajo la curva ROC. Cada gráfica mostrada lleva asociada una tabla en la que se detalla el valor de cada parámetro utilizado para ese experimento. El método que mejores resultados ofrezca en esta primera parte del estudio, será el que se utilice en la segunda parte (detección de tumores en imágenes completas).

6.4.1. Definición de las regiones de interés

Definimos una región de interés (ROI) como una región cuadrada establecida sobre una imagen de mamografía. Cada ROI se identifica mediante las coordenadas de su píxel central. El tamaño de la ROI coincide con el tamaño de la ventana de características T_w .

Para obtener las características de una ROI se utiliza una ventana de vecindad (o ventana de características), la cual se sitúa en el píxel que define la ROI. A partir de la información contenida en dicha ventana de vecindad y siguiendo alguno de los esquemas propuestos en la Sección 4.2, se extraen las características de la ROI.

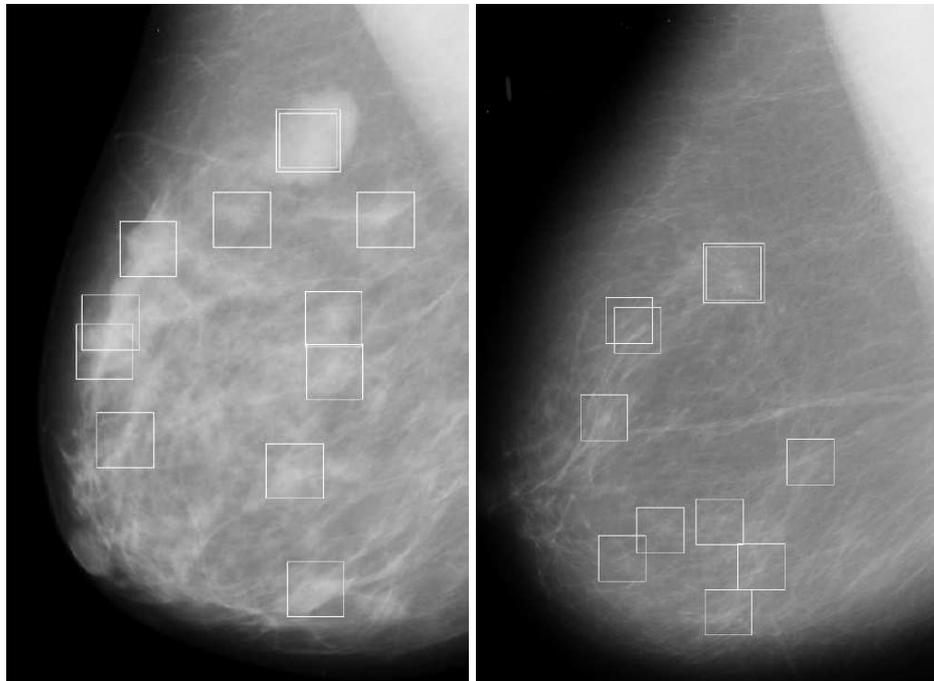
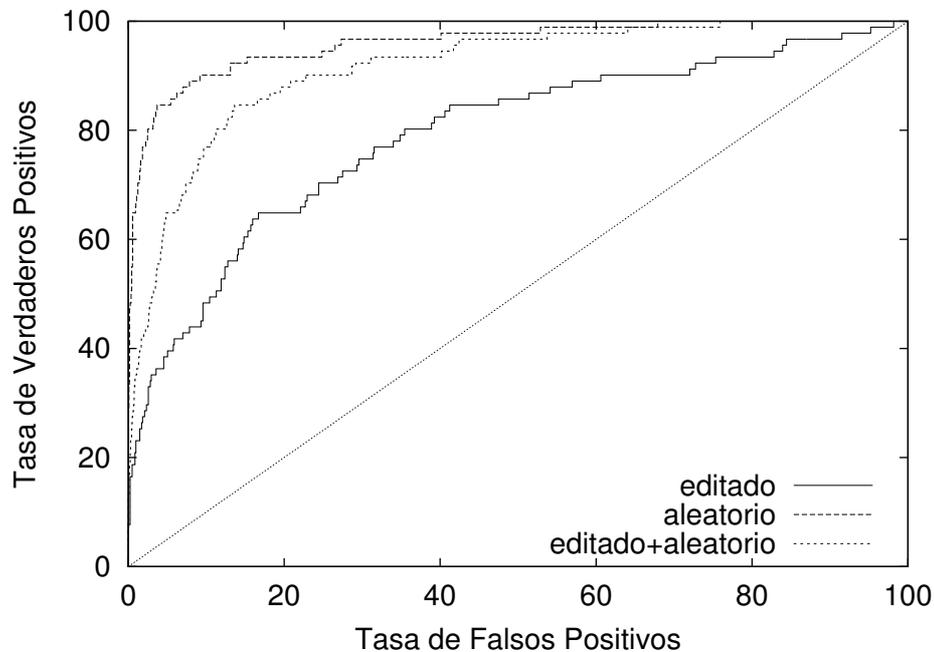


Figura 6.2: Regiones de interés seleccionadas en una mamografía con una masa circunscrita (izquierda) y con una distorsión estructural (derecha). El doble recuadro indica la región que contiene el tumor, mientras que el resto de recuadros corresponde a regiones de interés seleccionadas de modo supervisado.

Para realizar los experimentos expuestos en esta sección, correspondientes a la primera fase del estudio, se creó un conjunto de test formado por 3399 ROIs. De ellas, 91 corresponden a cada uno de los tumores (excepto calcificaciones) definidos en la base de datos, 820 corresponden a regiones sanas seleccionadas de modo supervisado, escogidas por su parecido visual con algún tipo de tumor, y las 2488 restantes corresponden también a tejido sano, pero en este caso seleccionadas aleatoriamente. En la Figura 6.2 se muestra un ejemplo de las ROIs seleccionadas de modo supervisado en dos mamografías, así como las ROIs correspondientes al

tumor. A partir de estas 3999 ROIs, se definieron 3 conjuntos de test:



CAR	NCP	T_w	K_o	T_m	k -NN	ESC	RCH
NG	20	16	20	5	30	10	Media

Figura 6.3: Curvas ROC obtenidas con distintos conjuntos de test. Las áreas bajo la curva ROC son: 0.79 para el conjunto editado, 0.96 para el aleatorio y 0.92 para el editado+aleatorio.

- **Conjunto editado:** contiene las 820 ROIs sanas seleccionadas de modo supervisado por su parecido a un tumor, más las 91 ROIs cancerosas.
- **Conjunto aleatorio:** contiene las 2488 ROIs sanas seleccionadas aleatoriamente, más las 91 ROIs cancerosas.
- **Conjunto editado+aleatorio:** contiene las 3399 ROIs (3308 sanas y 91 cancerosas).

Todos los experimentos se realizaron con los 3 conjuntos de test descritos anteriormente. En la Figura 6.3 se muestran las curvas ROC obtenidas con cada uno de los conjuntos de test. El experimento corresponde a un esquema de extracción de características basado en niveles de gris. Como es de esperar, los resultados obtenidos con el conjunto editado son significativamente peores que los obtenidos con el conjunto aleatorio.

Para el resto de experimentos expuestos a continuación, se mostrarán únicamente los resultados obtenidos con el conjunto de test que contiene la totalidad de las ROIs (editado+aleatorio). Puesto que este conjunto contiene ROIs que han sido seleccionadas manualmente en base a criterios no aleatorios, debe quedar claro que los resultados mostrados en esta sección son útiles únicamente a efectos de comparar las distintas aproximaciones planteadas, así como para observar el efecto de variar los distintos parámetros del estudio, pero no para valorar en términos absolutos la bondad de las técnicas analizadas. La Sección 6.5, en la que se realizan experimentos con imágenes completas sin que exista una preselección de las mismas, servirá para tener una idea más precisa y real de las prestaciones de los algoritmos propuestos.

6.4.2. Preproceso

En la etapa de preproceso la imagen se normaliza a nivel local (en cada una de las ventanas de vecindad) de modo que el valor medio de los niveles de gris sea 0 y la desviación típica 1. Con ello se consigue invarianza a los rangos de niveles de gris.

En la Figura 6.4 se muestran 3 ROIs (generadas artificialmente a partir de la misma imagen) con distintas intensidades, y esas mismas ROIs tras el proceso de normalización. Puede observarse que, aunque existen diferencias significativas en las 3 ROIs originales, estas diferencias apenas son apreciables tras la normalización.

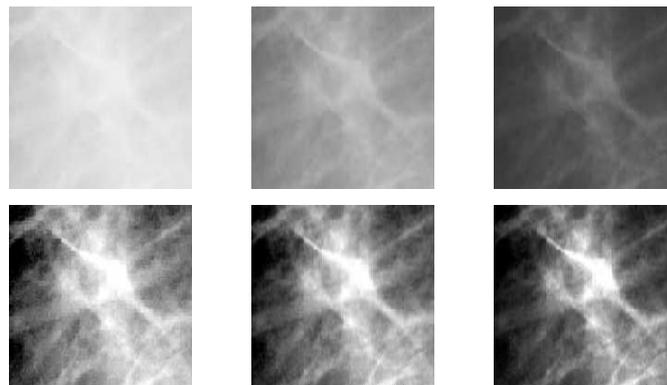
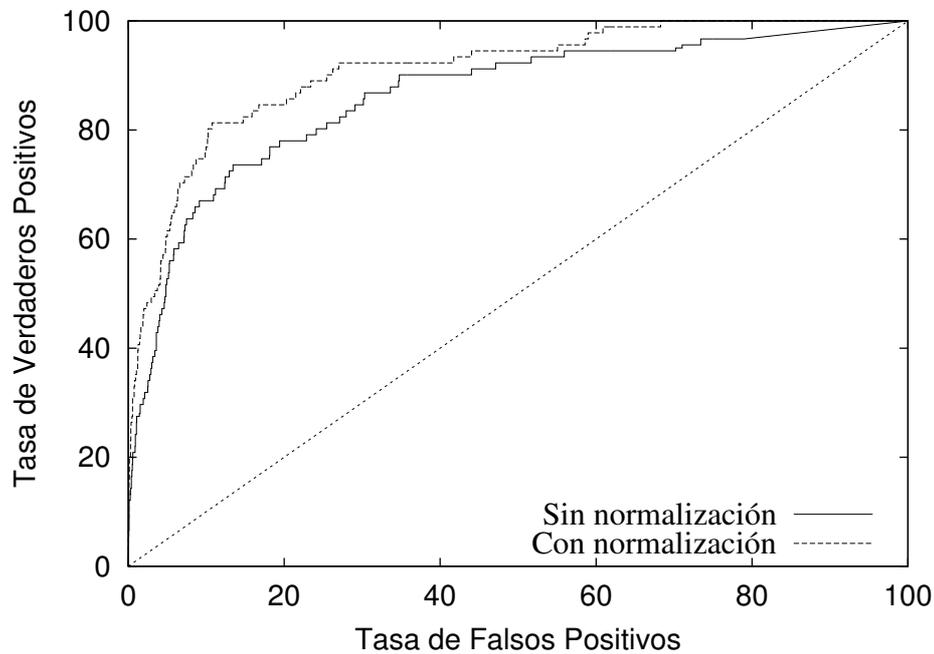


Figura 6.4: Ejemplo de una ROI con los niveles de gris alterados de manera artificial (arriba) y las mismas ROIs normalizadas (abajo).

En la Figura 6.5 se muestran las curvas ROC obtenidas con y sin normalización. Puede comprobarse la mejora conseguida cuando se incluye la etapa de normalización.



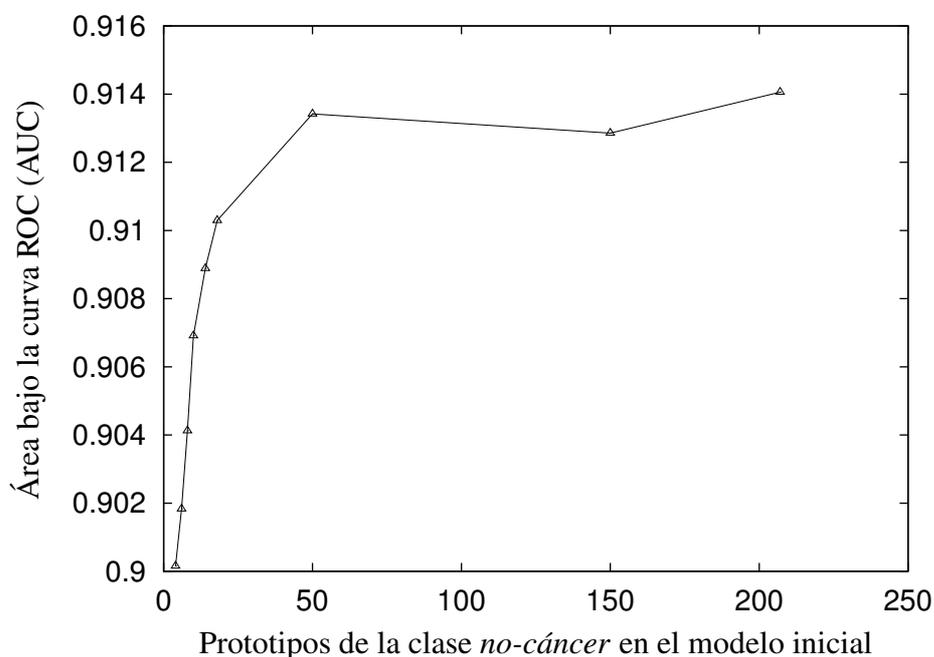
CAR	NCP	T_w	K_o	T_m	k -NN	ESC	RCH
NG	20	16	20	5	200	10	Media

Figura 6.5: Curvas ROC obtenidas sin normalizar los niveles de gris durante el pre-proceso y con normalización. Las áreas bajo la curva ROC son, respectivamente, 0.86 y 0.91.

6.4.3. Bootstrapping

Ya se ha explicado que el proceso de *bootstrapping* se define un modelo inicial con prototipos de una y otra clase. Dependiendo de los prototipos utilizados en la inicialización, el proceso de *bootstrapping* evolucionará de un modo diferente. En la Tabla 6.3 se muestra el efecto de variar el número de prototipos de la clase negativa (*no-cáncer*) en el modelo inicial. Lógicamente, cuantos menos prototipos tenga el modelo inicial, mayor número de éstos será necesario añadir durante el proceso de *bootstrapping* para que la clase quede bien representada. Sin embargo, el tamaño final del modelo es menor cuanto menor sea el modelo inicial. En la tabla presentada puede observarse que el área bajo la curva ROC obtenida no disminuye sustancialmente cuando se utilizan modelos con menor número de prototipos.

La Figura 6.6 muestra, en forma de gráfica, los valores de área bajo la curva ROC presentados en la tabla anterior, en función del número de de muestras empleadas en la inicialización del modelo. Puede observarse que a partir de una inicialización con un número razonable de muestras, las prestaciones del sistema



CAR	NCP	T_w	K_o	T_m	k -NN	ESC	RCH
SFUM	20	21	26	6	200	10	Media

Figura 6.6: Variación de los resultados de clasificación en función del número de prototipos utilizados en la inicialización del modelo.

no mejoran por el hecho de disponer de modelos con un número mayor de muestras.

Los resultados presentados corresponden a un esquema de extracción de características SFUM. La aproximación SFUM, como se ha visto, es invariante a rotaciones. Con otras aproximaciones sería necesario incluir en el modelo características obtenidas con distintas rotaciones, lo que incrementaría considerablemente el número total de prototipos del modelo. Concretamente, en este experimento el número total de muestras evaluadas durante el proceso de *bootstrapping* fue aproximadamente de 275,000. En otras aproximaciones que no son inherentemente invariantes a rotaciones, este número asciende a 5 millones, aunque, como acaba de exponerse, únicamente un pequeño porcentaje de las mismas pasa a formar parte del modelo final.

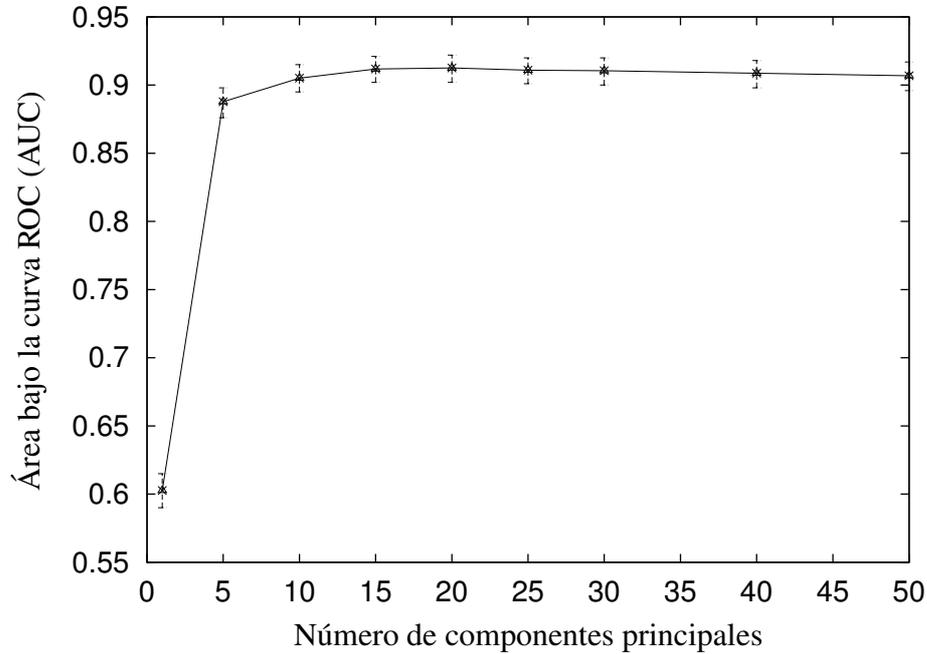
Imágenes iniciales	Muestras iniciales	Muestras añadidas	Total muestras	AUC
4	6K	2K	8K	0.900
6	9K	1.7K	11K	0.902
8	12K	1.6K	14K	0.904
10	14K	1.5K	16K	0.907
14	20K	1.3	22K	0.909
18	26K	1.1K	27K	0.910
50	58K	0.4K	59K	0.913
150	200K	48	200K	0.913
207	275K	0	275K	0.914

Tabla 6.3: Efecto de variar el número de prototipos empleados en el modelo inicial, antes de aplicar *bootstrapping*. En cada columna se muestra, respectivamente, el número de imágenes sanas utilizadas en la inicialización del modelo, el número de prototipos extraídos de dichas imágenes, el número total de prototipos añadidos durante el proceso de *bootstrapping*, el número de prototipos negativos (clase *no-cáncer*) del modelo final, y el área bajo la curva ROC obtenida en una tarea de clasificación usando dicho modelo. El número de prototipos positivos (clase *cáncer*) en todos los casos fue 2.200.

6.4.4. PCA

Como se ha explicado en capítulos anteriores, el espacio de características obtenido inicialmente mediante las distintas aproximaciones expuestas, es transformado y reducido mediante el uso de PCA. La precisión con la que queda representada cada una de las clases dependerá, entre otros factores, del número de componentes principales escogido. El hecho de utilizar un mayor número de componentes (esto es, de características) no implica necesariamente que las clases de nuestro problema queden mejor representadas. Por un lado, en la transformación PCA, únicamente las direcciones del espacio asociadas con valores propios altos presentan una gran varianza, mientras que las direcciones del espacio o componentes con un valor propio bajo aportarán poca información y, en muchos casos, únicamente ruido. Por otro lado, utilizar un número elevado de características conlleva al problema de la maldición de la dimensionalidad presentado en la Sección 4.3.1.

Para comprobar el efecto de utilizar un mayor o menor número de componentes principales, se realizaron una serie de experimentos variando dicho número. En la Figura 6.7 se muestran las áreas bajo la curva ROC obtenidas para distinto número de componentes principales. El experimento se realizó empleando un esquema de extracción de características basado en niveles de gris. Como puede observarse, con tan solo 10 componentes ya se alcanzan tasas de clasificación muy cercanas al máximo conseguido, el cual se logra con 20 componentes. A partir de aquí, el uso de un mayor número de componentes no aporta mayor rendimiento al clasificador y produce, incluso, un ligero empeoramiento.



CAR	NCP	T_w	K_o	T_m	k -NN	ESC	RCH
NG	1 – 50	16	20	5	200	10	Media

Figura 6.7: Área bajo la curva ROC en función del número de componentes principales utilizado.

En la Figura 6.8 se muestra una representación gráfica, en forma matricial, de los 15 primeros vectores propios de la base de transformación PCA, calculados directamente a partir de niveles de gris. Para lograr esta representación cada vector propio de dimensión $d = T_w^2$ se ha representado en una matriz de tamaño $T_w \times T_w$. Dada una ventana de vecindad definida por un píxel cualquiera, las 15 primeras características de dicho píxel quedarían definidas mediante la convolución de la ventana de vecindad con cada una de las matrices representadas en esta figura. Como puede observarse, la primera característica representará el gradiente en x , la segunda el gradiente en y , la tercera un gradiente aproximadamente radial, desde el centro hacia la periferia, etc. Ello explica, como se verá más adelante, que una técnica sencilla como el uso de niveles de gris seguida de PCA, genera resultados muy similares a otras técnicas basadas en gradientes.

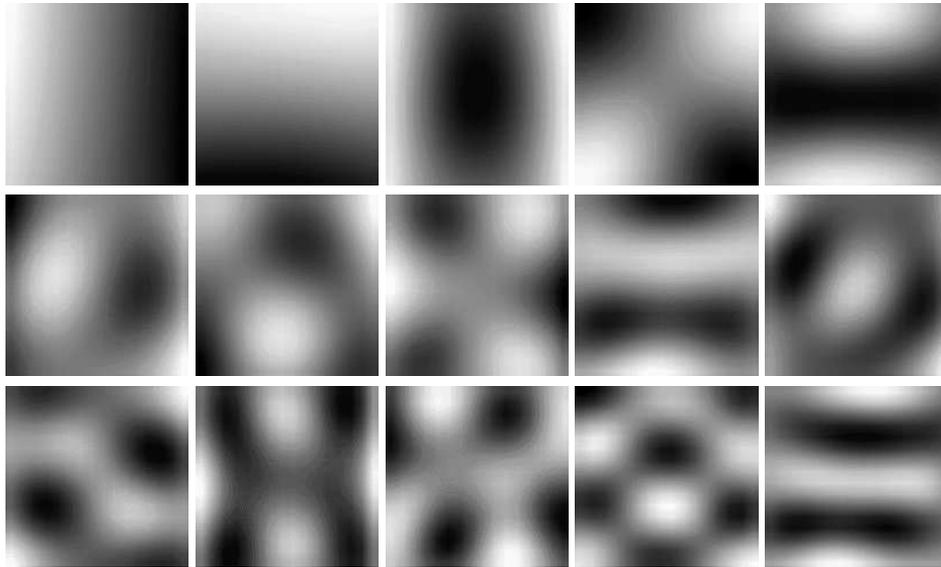


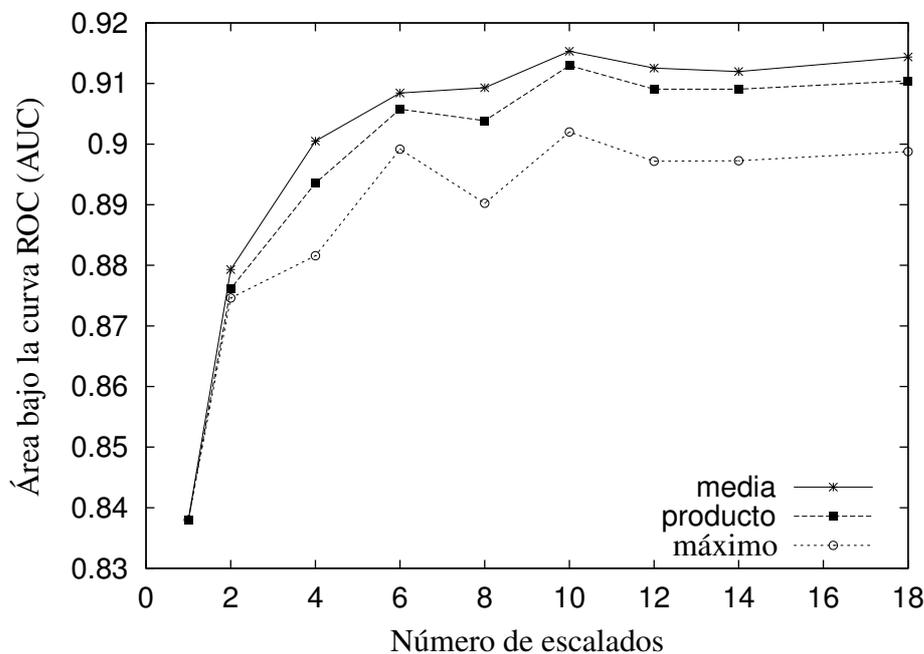
Figura 6.8: Representación gráfica de las 15 primeras componentes principales calculadas a partir de un esquema de extracción de características basado en niveles de gris.

6.4.5. Multiescalado

En la Sección 4.5 se explicó la necesidad de evaluar cada muestra de test a distintas escalas para conseguir una invarianza al escalado. Igualmente se expuso la necesidad de combinar las probabilidades a posteriori obtenidas en cada una de las escalas, para tener una única probabilidad independiente de la escala.

En este apartado se muestran los resultados de una serie de experimentos destinados a evaluar el efecto de aplicar un mayor o menor número de escalados a la muestra de test, así como de emplear distintas reglas de combinación de hipótesis (parámetros ESC y RCH). Algunos de los resultados presentados se muestran para distintos valores de k -NN, por lo que adicionalmente se comprueba el efecto de variar este parámetro. Los experimentos se realizaron con un esquema de extracción de características basado en niveles de gris.

En la Sección 4.5 se definió $F_n = \{s_1, s_2, \dots, s_n\}$ como el conjunto de factores de escala aplicados a una muestra de test. Para encontrar el tamaño adecuado del conjunto F_n , se realizaron experimentos con valores de n entre 1 y 18. Cada uno de los factores de escala s_i se distribuyó entre s_{min} y s_{max} , de manera que s_i/s_{i-1} fuera constante dentro de un mismo conjunto F_n . Los valores de s_{min} y s_{max} , similares en todos los conjuntos, se establecieron en función del tamaño de los tumores existentes en el corpus. En el caso de F_1 , el único factor de escala aplicado corresponde al de mayor probabilidad a priori.

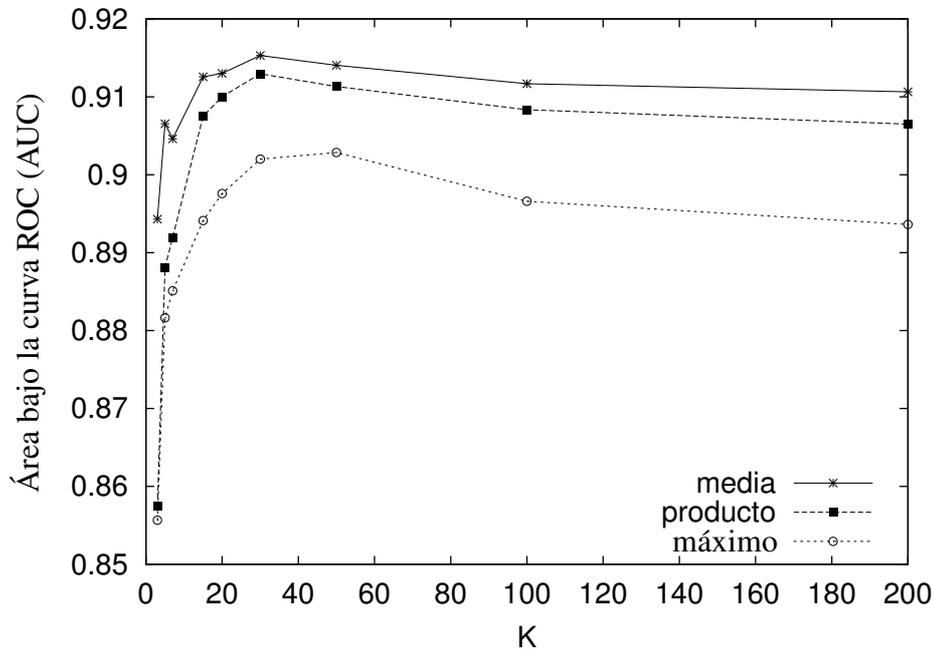


CAR	NCP	T_w	K_o	T_m	k -NN	ESC	RCH
NG	20	16	20	5	200	1 – 18	Med., Máx., Prod.

Figura 6.9: Área bajo la curva ROC obtenida variando el número de escalados empleado en la fase de test y para distintas reglas de combinación de hipótesis.

En la Figura 6.9 se muestra el área bajo la curva ROC obtenida en función del número de escalados aplicados a la muestra de test (ESC). Puede observarse que los mejores resultados se alcanzan cuando la muestra de test se evalúa en 10 escalas distintas. Con un mayor número de escalados los resultados no varían significativamente. Es interesante observar que, con un solo escalado, se obtienen resultados aceptables. La razón de ello puede deberse, por un lado, a que la clase negativa está modelada en las distintas escalas, por lo que una muestra de test perteneciente a esta clase puede detectarse correctamente evaluándola en una sola escala. Por otro lado, las características de gradiente obtenidas mediante PCA presentan cierta invarianza al escalado. Además, los tumores denominados como “masas radiales” (SPIC) también tienen cierto grado de invarianza al escalado.

En cuanto a las tres reglas de combinación de hipótesis presentadas en la Sección 4.5, puede observarse que la que mejores resultados ofrece es la de la media, mientras que el máximo ofrece el peor rendimiento. Debido a que existe cierta invarianza al escalado, tal y como se ha comentado anteriormente, una muestra positiva (*cáncer*) normalmente se detecta, no en una única escala, sino en un rango consecutivo de escalas, mientras que una muestra negativa puede ocasionalmente



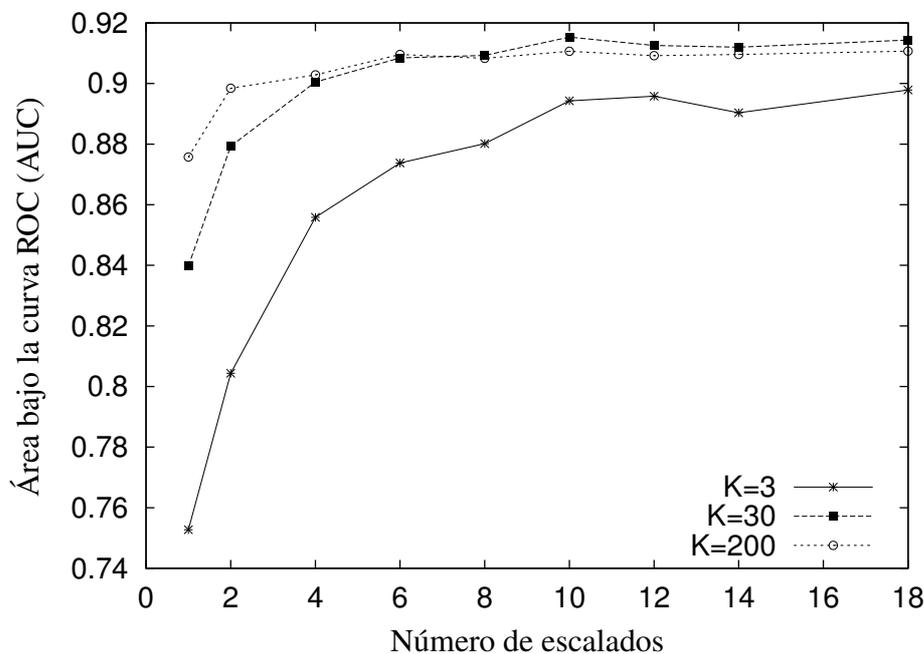
CAR	NCP	T_w	K_o	T_m	k -NN	ESC	RCH
NG	20	16	20	5	3 – 200	10	Med., Máx., Prod.

Figura 6.10: Área bajo la curva ROC para distintos valores de K y distintas reglas de combinación de hipótesis, obtenidas con 10 factores de escala distintos.

dar un nivel de sospecha elevado en una escala determinada. Es más difícil, sin embargo, que dicha sospecha se mantenga elevada a lo largo de varias escalas. En estas condiciones, la regla de la media puede detectar un número similar de tumores a los detectados con la regla del máximo, pero cometiendo menos errores de falsos positivos.

En la Figura 6.10 se muestra el área bajo la curva ROC para cada una de las reglas de combinación de hipótesis presentadas (RCH), en función del número de k -vecinos empleado en el clasificador (k -NN). Nuevamente se observa que la regla de la media es la que mejores resultados ofrece. Por otro lado, el número de vecinos óptimo se establece en torno a $k = 30$. En este caso los experimentos se realizaron evaluando la muestra de test a 10 escalas distintas.

En la Figura 6.11 se muestra el área obtenida bajo la curva ROC para distinto número de k -vecinos, en función del número de escalados empleado. En este caso se utilizó la media para combinar las probabilidades a posteriori obtenidas en cada escala. Un detalle interesante a observar es que el efecto negativo que puede producir aplicar un número reducido de escalados a la muestra de test, se puede paliar utilizando un valor de k elevado. De este modo, con $k = 3$ se acusa mucho



CAR	NCP	T_w	K_o	T_m	k -NN	ESC	RCH
NG	20	16	20	5	3, 30, 200	1 – 18	Media

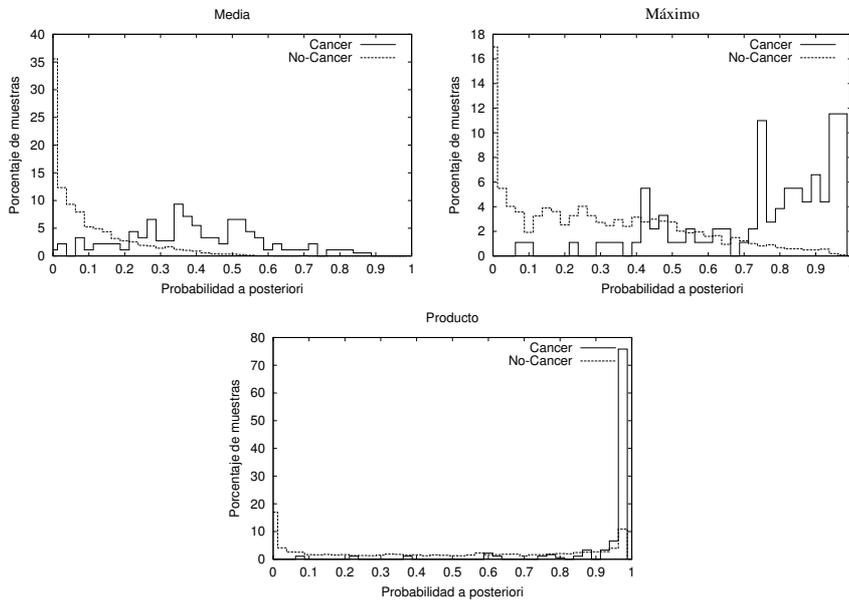
Figura 6.11: Área bajo la curva ROC obtenida variando el número de escalados empleado en la fase de test y para distinto número de k -vecinos.

la diferencia de utilizar un mayor o menor número de escalados, mientras que con $k = 200$ se minimiza dicha diferencia.

Finalmente, en la Figura 6.12 se muestra, en forma de histograma y para distintos intervalos de probabilidad a posteriori, el porcentaje de muestras de test de una y otra clase que se clasifica en dichos intervalos.

Puede observarse que, en general, la regla de la media otorga probabilidades bajas a las muestras de la clase *no-cáncer*. Ello se debe a que estas muestras, pese a que en ocasiones pueden confundirse con un tumor en alguna de las escalas, normalmente tendrán una probabilidad a posteriori baja en la mayoría de las escalas. Por contra, las muestras positivas generalmente obtienen una probabilidad a posteriori elevada, no sólo en la escala que les corresponde, sino también en las escalas cercanas. En consecuencia, al aplicar la regla de la media se obtiene una probabilidad final que, si bien es más baja de lo deseado, en general sigue siendo mayor que la asignada a las muestras negativas.

La regla del máximo, como es de esperar, otorga probabilidades altas a las muestras positivas. Sin embargo, como ya se ha mencionado, el tejido sano puede



CAR	NCP	T_w	K_o	T_m	k -NN	ESC	RCH
NG	20	16	20	5	200	10	Med., Máx., Prod.

Figura 6.12: Histogramas de los valores de probabilidad a posteriori para cada clase (*cáncer* y *no-cáncer*) y para cada regla de combinación de hipótesis (media, máximo y producto).

confundirse con tumores en algunas escalas, especialmente con factores de escala altos (esto es, cuando la ventana de vecindad es muy pequeña en relación a la imagen). En estos casos, lógicamente, la regla del máximo no será adecuada.

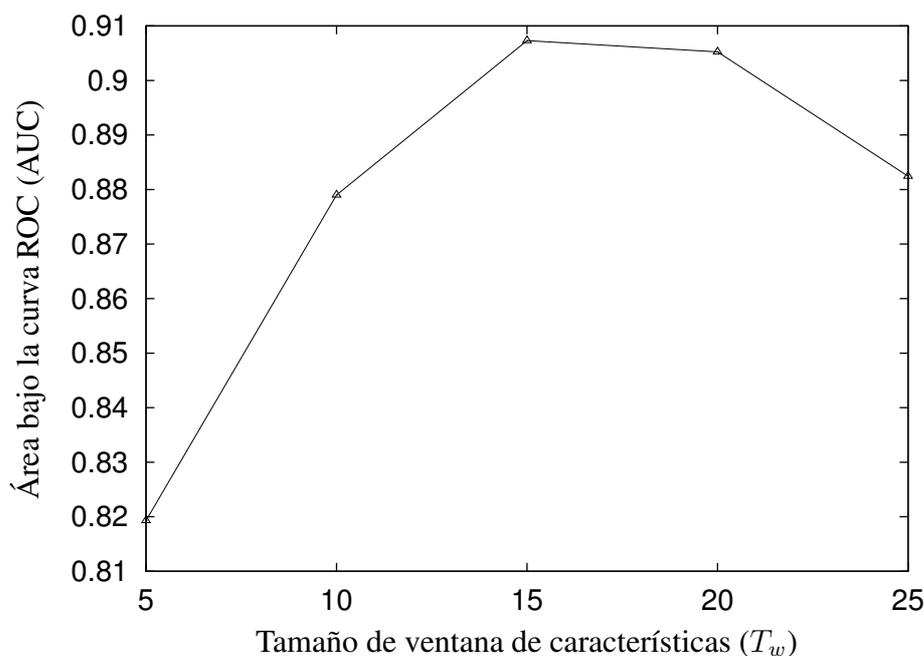
Por último, la regla del producto asigna correctamente una probabilidad muy elevada a la mayor parte de las muestras positivas. Sin embargo, las muestras negativas no obtienen mayoritariamente una probabilidad baja.

Para tratar de aprovechar al máximo la información que aporta el valor de probabilidad a posteriori obtenido en cada una de las escalas, se realizaron algunos experimentos tomando dichas probabilidades como características para entrenar un segundo clasificador. A continuación, utilizando muestras de test no empleadas en el entrenamiento de ninguno de los clasificadores, se evaluaron éstas a distintas escalas del modo habitual. Finalmente, las probabilidades obtenidas en cada escala se utilizaron como una nueva muestra de test para la segunda etapa de clasificación. Sin embargo, el empleo de este segundo clasificador no supuso finalmente ninguna mejora respecto a aplicar directamente la regla de la media.

6.4.6. Tamaño de la ventana de características y área de muestreo.

En este apartado se evalúa el efecto de variar la relación entre el tamaño de la ventana de características (T_w) y del tamaño del tumor tras su normalización a una escala fija (K_o), explicado en la Sección 4.4. Además se analiza el efecto de variar el tamaño del área de muestreo (T_m) presentado en esa misma sección.

En la Figura 6.13 se muestra el efecto de variar la relación entre T_w y K_o . En estos experimentos se utilizó un valor constante de $K_o = 20$ y se varió T_w entre 5×5 y 25×25 . Independientemente del valor de T_w , para entrenar el modelo de la clase *cáncer* la ventana de características w se deslizó por los 81 píxeles centrales del tumor, esto es, se empleó una ventana de muestreo de tamaño $T_m = 9 \times 9$ (en la Figura 4.5 de la Sección 4.4 se explica gráficamente el significado de este parámetro).

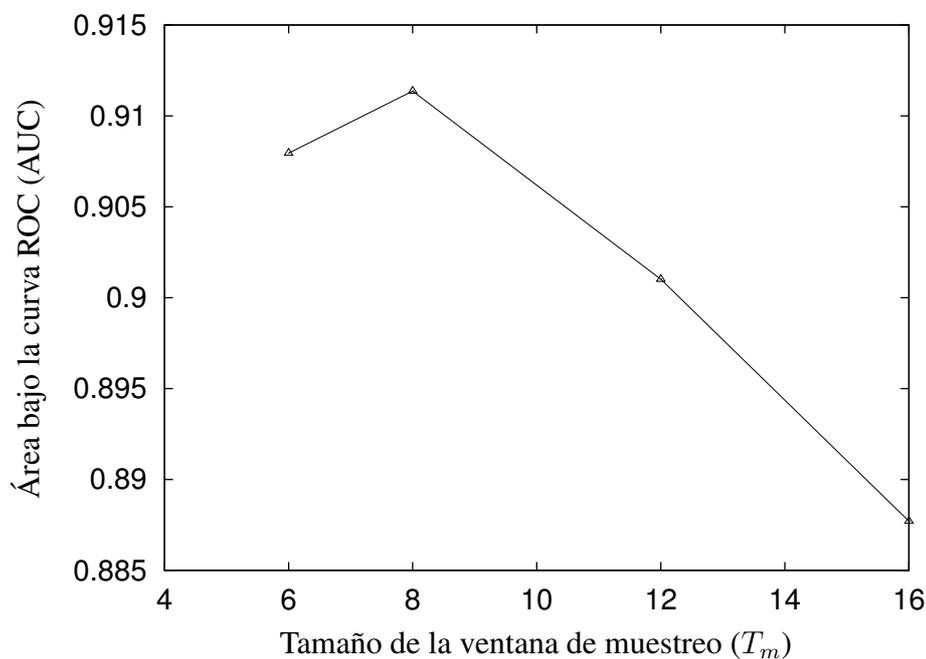


CAR	NCP	T_w	K_o	T_m	k -NN	ESC	RCH
NG	20	5 – 25	20	9	30	10	Media

Figura 6.13: Área bajo la curva ROC para distintos tamaños de ventana de características (T_w) para un tamaño constante de normalización de los tumores de $K_o = 20$.

Como puede observarse, cuando la ventana de características es excesivamente pequeña en relación al tamaño del tumor, la precisión del clasificador es menor. En estas condiciones, las características obtenidas representan regiones muy locales.

Dado que un tumor, analizado a este nivel de detalle, presenta un aspecto bastante homogéneo, con unas características muy similares a las que pueden encontrarse a nivel local en el tejido sano, parece lógico pensar que este tipo de representación limita la capacidad discriminativa entre un tipo de tejido y otro. Cuando la ventana de características es ligeramente menor que el tamaño del tumor, es posible recoger características del tumor que lo diferencian del tejido sano. En este caso, tal y como se observa en la gráfica, se obtienen los mejores resultados. Por último, como ya se predijo en la Sección 4.4, los resultados empeoran cuando la ventana de características supera en tamaño al tumor.



CAR	NCP	T_w	K_o	T_m	k -NN	ESC	RCH
NG	20	15	20	6 – 16	30	10	Media

Figura 6.14: Área bajo la curva ROC en función del tamaño de la ventana de muestreo T_m , para valores de $T_w = 15$ y $K_o = 20$. El número de píxeles que la ventana de características w excede de los límites del tumor o es $n = T_m - (K_o - T_w + 1)$, esto es, 0 píxeles para $T_m = 6$ y 10 píxeles para $T_m = 16$.

Para estudiar la relación entre K_o , T_w y T_m se realizó un segundo experimento en el que se fijaron los valores de K_o y T_w y se varió el tamaño del área de muestreo T_m . En la Figura 6.14 se muestran los resultados del experimento. En este caso se normalizaron los tumores a un tamaño $K_o = 20$ y se fijó el tamaño de la ventana de características en $T_w = 15$. En estas condiciones, se realizaron distintos experimentos, permitiéndose que w excediera los límites de o entre 0 y 10 píxeles, esto es, se emplearon ventanas de muestreo con $6 \leq T_m \leq 16$. Puede observarse

que los mejores resultados se obtienen cuando la ventana de características sobrepasa ligeramente (2 píxeles) los límites del tumor, sin embargo las prestaciones se degradan sensiblemente cuando dicha ventana sobrepasa en mayor medida a la del tumor.

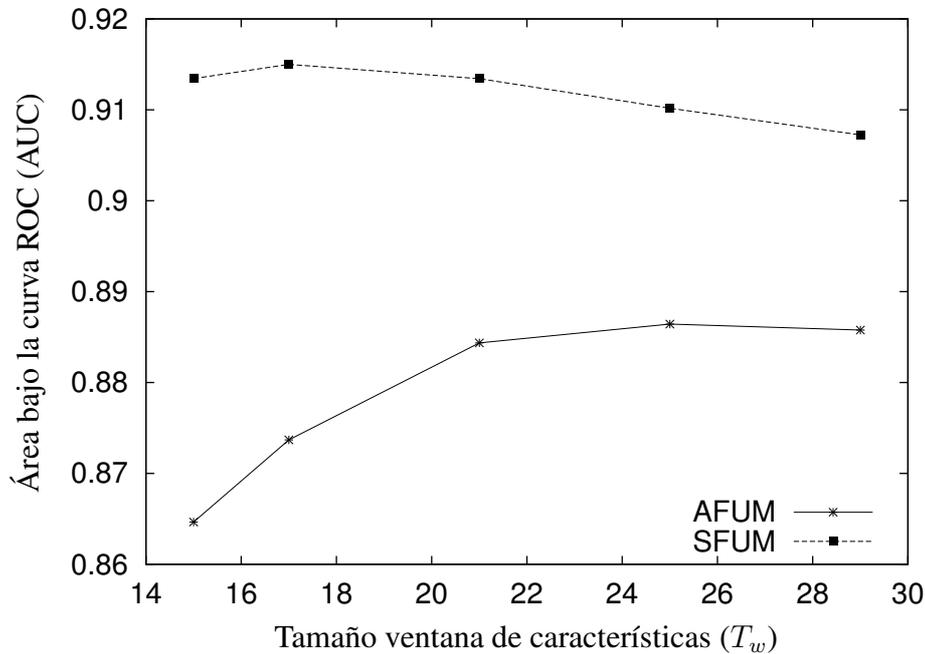
6.4.7. Aproximaciones de extracción de características

Hasta ahora, los resultados expuestos en este capítulo provienen de experimentos realizados con un esquema de extracción de características basado en niveles de gris junto con PCA. Adicionalmente se realizaron algunos experimentos similares a los presentados hasta el momento, con otros esquemas de extracción de características y se llegó a conclusiones similares respecto a los valores óptimos de los distintos parámetros empleados: NCP= 20, k -NN= 30, ESC= 10 y RCH=Media. El valor óptimo del área de muestreo T_m se establece de modo que la ventana de características w sobrepase ligeramente (2 píxeles aproximadamente) al tumor o . En cuanto al tamaño ideal de la ventana de características (T_w), se establece en 25 píxeles para el método AFUM y en torno a 16 para el resto de métodos de extracción de características.

En primer lugar, en la Figura 6.15 se comparan las aproximaciones AFUM y SFUM, para distintos tamaños de la ventana de características. Como puede observarse, el método SFUM propuesto supera claramente al AFUM. Además SFUM alcanza su rendimiento óptimo con un tamaño de ventana de características menor que el requerido para el caso de AFUM ($T_w = 17$ frente a $T_w = 25$). No hay que olvidar, sin embargo, que AFUM es un método mucho más sencillo que no requiere el entrenamiento de un modelo y, por tanto, más eficiente en términos computacionales.

Uno de los objetivos de esta primera fase del trabajo es, como ya se ha mencionado anteriormente, comparar distintos esquemas de extracción de características. En esta sección se hace un estudio comparativo de las distintas aproximaciones propuestas y explicadas en la Sección 4.2: niveles de gris, SGLDM, filtro de Sobel, AFUM y SFUM.

En la Figura 6.16 se muestran las mejores curvas ROC obtenidas con cada uno de los métodos de extracción de características propuestos, así como los valores de área bajo la curva. Los resultados obtenidos muestran que la aproximación basada en niveles de gris y SFUM superan claramente al resto de aproximaciones. Por otro lado, estas dos aproximaciones ofrecen resultados muy similares. La combinación de un método de extracción de características directo como es los niveles de gris, junto con el uso de PCA, demuestra ser una buena aproximación para este tipo de tarea. Al aplicar PCA, no sólo se tienen en cuenta individualmente los niveles de gris en cada píxel, sino ciertas relaciones entre los mismos como puede ser cambios de gradientes, tal y como se ha explicado en la Sección 6.4.4. Ello explica



CAR	NCP	T_w	K_o	T_m	k -NN	ESC	RCH
AFUM ¹ , SFUM	20	15 – 29	29 – 34	6	200	10	Media

Figura 6.15: Áreas bajo la curva ROC obtenidas para los métodos AFUM y SFUM, variando el tamaño de la ventana de características T_w .

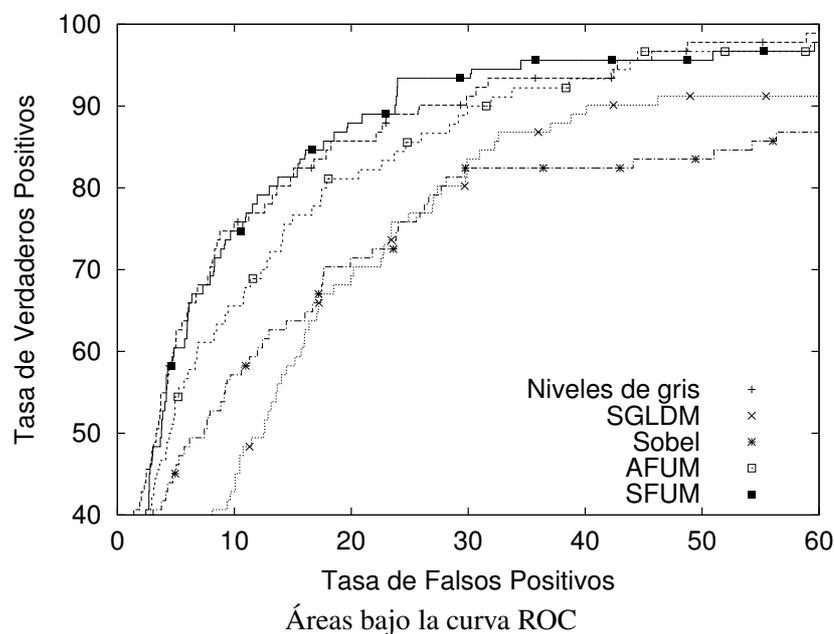
la similitud en los resultados obtenidos con niveles de gris y con SFUM. El método SFUM, sin embargo, presenta cierta ventaja frente a los niveles de gris, si tenemos en cuenta que el primero es invariante a rotaciones. Ello supone que los modelos basados en niveles de gris, para obtener un rendimiento similar al conseguido con SFUM, necesitan contener un mayor número de prototipos.

Por otro lado, la aproximación AFUM, a pesar de ser un método sencillo no supervisado, que no requiere entrenamiento alguno, ofrece unos resultados aceptables para esta tarea. La aproximación SFUM, sin embargo, mejora el rendimiento de AFUM tal y como era de suponer.

Los métodos basados en texturas como SGLDM o en detección de bordes como el filtro de Sobel, aun siendo técnicas ampliamente utilizadas y que han dado buenos resultados en otras tareas, para este problema en concreto ofrecen peores resultados que el resto de aproximaciones estudiadas.

Finalmente se realizaron dos experimentos con características combinadas. Por un lado se combinaron características SFUM con niveles de gris y por otro SFUM

¹Los valores de NCP, K_o , T_m y k -NN no son aplicables a la aproximación AFUM.



Método	Niveles gris	SGLDM	Sobel	AFUM	SFUM
AUC	0.91	0.81	0.81	0.89	0.91

CAR	NCP	T_w	K_o	T_m	k -NN	ESC	RCH
NG	20	16	20	6	30	10	Media
AFUM	-	25	-	-	-	10	Media
SFUM	20	21	26	6	30	10	Media
Sobel	20	16	20	5	30	10	Media
SGLDM	20	16	20	6	30	10	Media

Figura 6.16: Curvas ROC obtenidas con cada uno de los métodos de extracción de características propuestos y área bajo la curva ROC.

con SGLDM. Para ello se obtuvieron, en primer lugar, las características a combinar y, posteriormente, se aplicó PCA conjuntamente. Puesto que, en ambos casos, las características que se combinaron son de distinta naturaleza y, por tanto, pueden tener magnitudes muy distintas, antes del cálculo de PCA se normalizaron todas ellas. De este modo se consigue evitar que unas características puedan tener mayor peso que otras a la hora de calcular k -vecinos mediante distancia euclídea.

En la Tabla 6.4 se muestra el área bajo la curva ROC obtenida individualmente con cada esquema de extracción de características, así como con los esquemas combinados. Como puede observarse, SFUM+niveles de gris apenas mejora los resultados obtenidos de forma individual por ambos esquemas. Posiblemente, debido al uso de PCA, el espacio de características que representa cada uno de estos esquemas es muy similar, por lo que la combinación de ambos tipos de carac-

terísticas no aporta nueva información al modelo. En cuanto a la combinación de SFUM+SGLDM, los resultados obtenidos son un término medio entre lo conseguido individualmente con cada una de estas aproximaciones. Puede concluirse que, en esta tarea, el uso de descriptores de textura no es capaz de aportar información distinta de la conseguida mediante SFUM.

Método	SFUM	N.Gris	SGLDM	SFUM+N.Gris	SFUM+SGLDM
AUC	0.91	0.91	0.81	0.92	0.88

Tabla 6.4: Áreas bajo la curva ROC obtenidas individualmente con las aproximaciones SFUM, niveles de gris y SGLDM, y de forma combinada con SFUM+niveles de gris y SFUM+SGLDM.

6.4.8. Comparación con otros clasificadores

Para evaluar el rendimiento de otros clasificadores binarios, se realizaron una serie de experimentos en los que se emplearon Máquinas de Vectores Soporte (SVM) y *Boosting*. Para los experimentos realizados con SVM se utilizó el software SVM^{light} [Joac 99]. En el caso de *Boosting* los experimentos se realizaron con el paquete *Weka* [Freu 96] y con *BoosTexter* [Scha 00].

A continuación se muestran los resultados obtenidos con cada uno de estos clasificadores y se compara con el rendimiento obtenido mediante k -NN. En todos los casos se utilizó SFUM como aproximación para la extracción de características.

Corpus

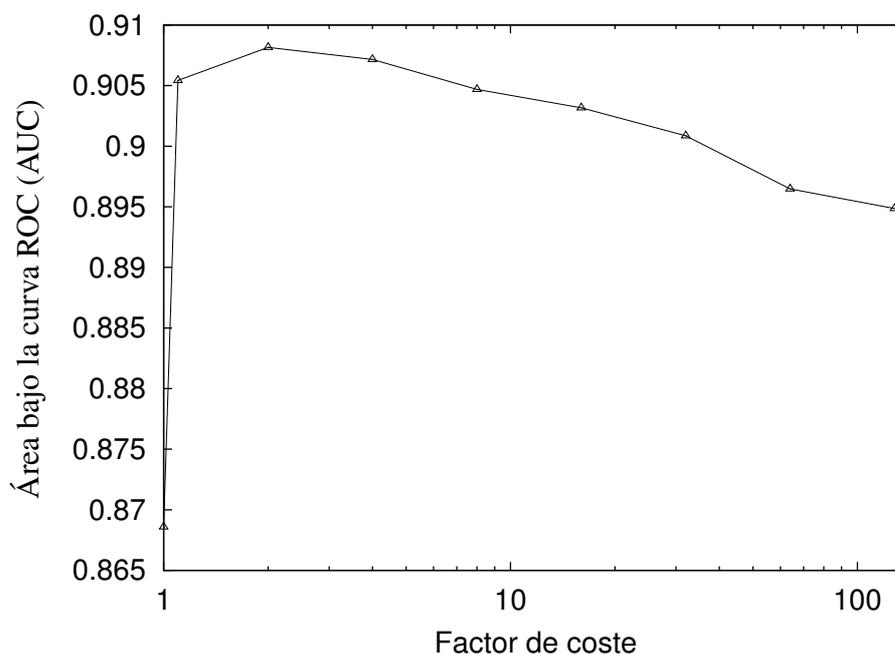
Hasta ahora, en los experimentos con k -NN se había utilizado la totalidad de la base de datos para el entrenamiento y 3399 ROIs para el test. Ello obligaba a que cuando se evaluaba una ROI de una mamografía determinada, debía descartarse del conjunto de entrenamiento cualquier otra muestra de esa mamografía o de otra perteneciente a la misma paciente. Cuando se trabaja con k -NN es relativamente sencillo descartar muestras del conjunto de entrenamiento, sin embargo este proceso resulta muy costoso cuando se trabaja con SVM o *Boosting* ya que, por un lado, es necesario crear un modelo nuevo cada vez y, por otro, el entrenamiento del modelo tiene un coste computacional mucho mayor con este tipo de clasificadores.

Como consecuencia de todo ello, para que esta batería de experimentos fuese viable, se definió un conjunto de entrenamiento con el 60 % de las ROIs y otro de test con el 40 % restante, manteniendo las probabilidades a priori similares en ambos conjuntos y teniendo en cuenta que mamografías de una misma paciente no apareciesen en conjuntos distintos.

Máquinas de Vectores Soporte (SVM)

Tal y como se expuso en la Sección 2.6, el rendimiento obtenido mediante SVM puede mermarse cuando se trabaja con conjuntos de entrenamiento no balanceados, en los que el número de muestras de una clase es muy superior al de la otra. En el corpus empleado para estos experimentos, el 95 % de las muestras corresponde a la clase *no-cáncer*. Una alternativa para corregir este desajuste consiste en emplear un factor de penalización distinto para cada clase [Mori 99], tal y como se explicó en la Sección 2.6.

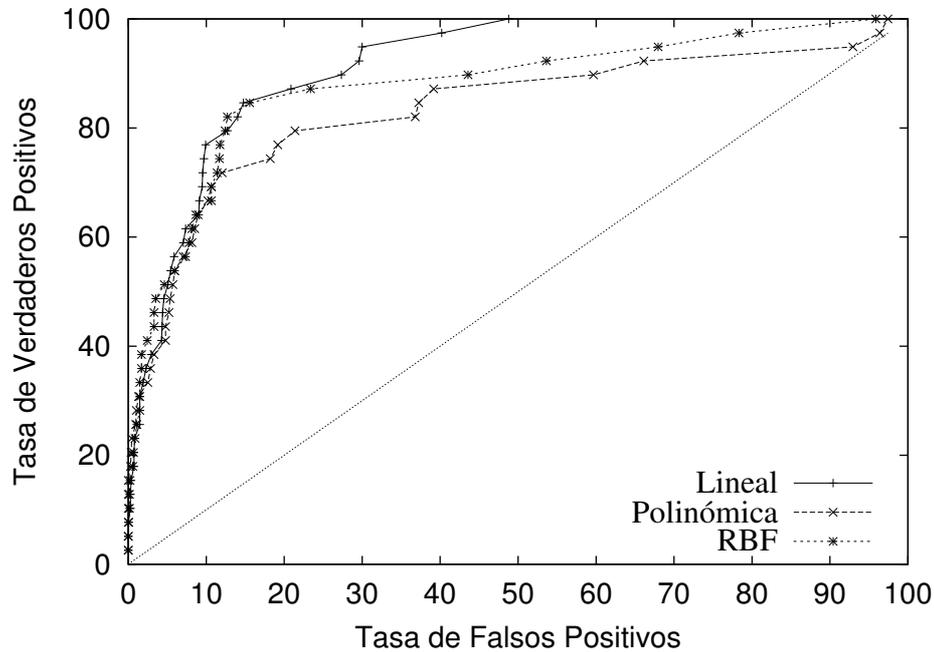
En la Figura 6.17 se muestra el área bajo la curva ROC obtenida con distintos factores de coste y con una *función de núcleo* lineal. El factor de coste indica la relación existente entre la penalización aplicada a los errores de falso negativo y la aplicada a los errores de falso positivo. Como puede observarse, si se utiliza la misma penalización para las muestras positivas que para las negativas (factor de coste = 1), el rendimiento obtenido es claramente inferior. Con un factor de coste = 2 se obtienen los mejores resultados.



CAR	NCP	T_w	K_o	T_m	ESC	RCH
SFUM	21	26	6	6	10	Media

Figura 6.17: Áreas bajo la curva ROC obtenidas con SVM para distintos factores de coste.

En la Figura 6.18 se muestran las curvas ROC obtenidas para distintas funcio-



CAR	NCP	T_w	K_o	T_m	ESC	RCH
SFUM	21	26	6	6	10	Media

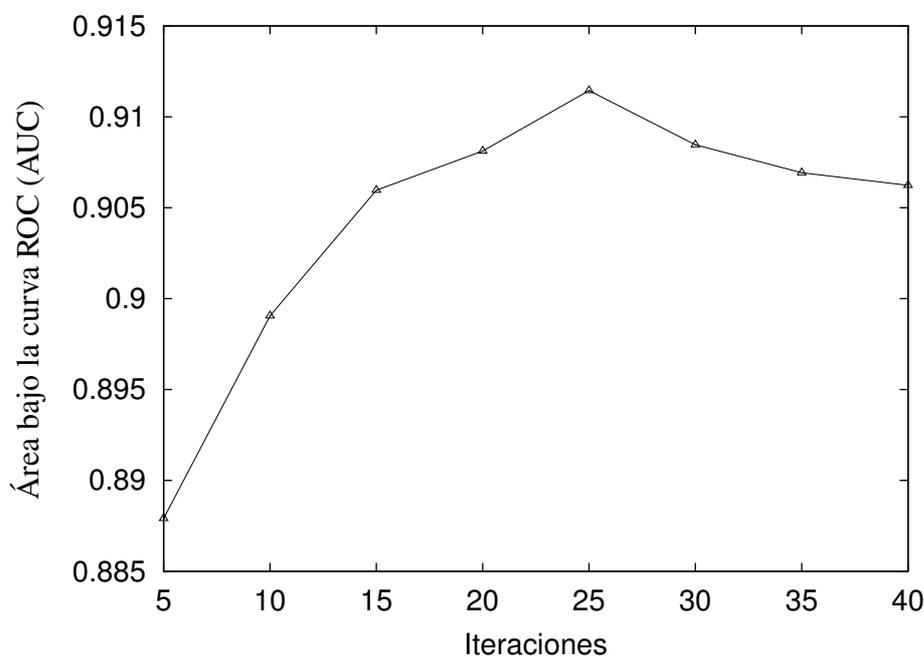
Figura 6.18: Curvas ROC obtenidas con SVM y distintas funciones de núcleo, con un factor de coste 2. Las AUC correspondientes son Lineal: 0.91, Polinómica: 0.82 y RBF :0.87.

nes de núcleo (lineal, polinómica y RBF). En este caso la función lineal es la que mejores resultados ofrece debido, posiblemente, a la pérdida de generalización que puede producirse al utilizar funciones de núcleo más complejas.

Boosting

Los experimentos de *Boosting* se realizaron con la herramienta *Weka* que implementa, entre otros, el algoritmo *AdaBoost* [Freu 77] explicado en la Sección 2.7. Como clasificador base se empleó un árbol de decisión con únicamente dos hojas (*Decision Stump*). Para crear este tipo de árboles, dada una característica f_i y un umbral T , todas las muestras de entrenamiento con $f_i \leq T$ se añaden en una de las hojas, mientras que el resto de muestras se añaden a la otra hoja.

Una vez decidido el clasificador base a emplear, el único parámetro que debe ajustarse en el algoritmo de *AdaBoost* es el número de iteraciones a realizar. En la figura 6.19 se muestra el área bajo la curva ROC obtenida en función del número de iteraciones empleado. En este caso, los mejores resultados se obtienen con 25 iteraciones.



CAR	NCP	T_w	K_o	T_m	ESC	RCH
SFUM	21	26	6	6	10	Media

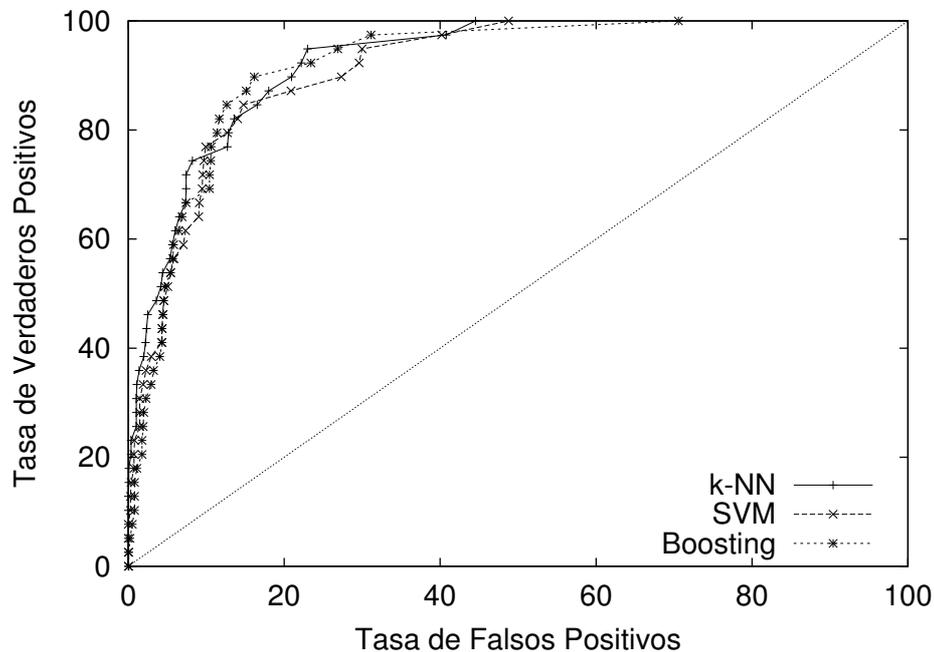
Figura 6.19: Áreas bajo la curva ROC obtenidas con el clasificador *Boosting* para distinto número de iteraciones.

Se realizaron también experimentos utilizando el paquete *BoosTexter* [Scha 00]. Aunque este software está inicialmente pensado para clasificación de texto, también admite características continuas como las empleadas en esta tarea. Los resultados obtenidos fueron ligeramente inferiores a los conseguidos con el algoritmo de *AdaBoost*.

Comparación entre k -NN, SVM y *Boosting*

A continuación, a modo de resumen, se muestra un estudio comparativo del rendimiento obtenido por los distintos clasificadores evaluados: k -NN, SVM y *Boosting*. En la Figura 6.20 se muestran las curvas ROC obtenidas en cada caso, con los parámetros que mejores resultados han ofrecido (factor de coste 2 y función de núcleo lineal para el caso de SVM, y 25 iteraciones para el caso de *Boosting*). En el caso de k -NN se realizaron de nuevo los experimentos con los conjuntos de entrenamiento y test definidos para esta tarea, y se empleó un número de vecinos $k = 30$.

Como puede observarse, no se aprecian diferencias significativas entre los dis-



Áreas bajo la curva ROC			
Clasificador	<i>k</i> -NN	SVM	<i>Boosting</i>
AUC	0.921	0.908	0.911

CAR	NCP	T_w	K_o	T_m	ESC	RCH
SFUM	21	26	6	6	10	Media

Figura 6.20: Curvas ROC obtenidas con *k*-NN, SVM y *Boosting*.

tintos clasificadores. Por otro lado debe decirse que si bien el coste computacional de clasificar una muestra de test es muy bajo en los tres clasificadores evaluados, el tiempo requerido para el entrenamiento cuando se emplea *Boosting* y SVM (especialmente si se trabaja con funciones de núcleo no lineales) es mucho mayor que el requerido para crear el *kd-tree* utilizado en *k*-NN.

También debe resaltarse que en este tipo de tareas de ayuda al diagnóstico, además de realizar una clasificación con la mayor precisión posible, es importante obtener una buena correlación entre el nivel de sospecha asignado a cada muestra de test y la presencia o no de células cancerosas en dicha muestra. En este sentido, *k*-NN ofrece una estimación de la probabilidad a posteriori, tal y como se ha explicado en la Sección 2.3.2, cosa que no ocurre, en principio, con SVM y *Boosting*. Estos clasificadores restringen su salida a los valores $+1$, -1 o, en el mejor de los casos, ofrecen un nivel de confianza que no está estrechamente correlacionado con la probabilidad a posteriori. Existen, sin embargo, técnicas más recientes que tratan precisamente de estimar la probabilidad a posteriori empleando este tipo de

clasificadores [Soll 02, Tao 05], las cuales sería interesante evaluar.

6.4.9. Conclusiones

En esta primera fase del estudio se ha definido un conjunto de test consistente en 3399 ROIs, 91 de las cuales corresponden a algún tipo de tumor. El objetivo es comparar las distintas técnicas propuestas, de modo que se escogerá aquella que ofrezca mejores tasas de clasificación para afrontar el problema de detección abordado en la segunda fase.

Se ha demostrado la conveniencia de normalizar los niveles de gris localmente en la ventana de características, como paso previo a la extracción de características. Con ello se consigue la invarianza en los rangos de nivel de gris observados y, consecuentemente, mejorar las tasas de clasificación.

Se ha expuesto igualmente la conveniencia de aplicar la técnica de *bootstrapping* durante el entrenamiento del modelo, con el fin de reducir el tamaño del mismo, sin merma en las tasas de clasificación.

El uso de PCA permite reducir la dimensionalidad del espacio de características, sin que ello suponga una pérdida en el rendimiento del clasificador. Se ha demostrado cómo, partiendo de un espacio de características de dimensión $d = 225$ (ventana de características de 15×15) se puede llegar a trabajar con tan solo 15 o 20 características y al tiempo mantener, o incluso mejorar, las tasas de clasificación.

En aquellas aplicaciones en las que no se tiene un conocimiento previo del tamaño del objeto a detectar, se requiere de algún mecanismo que contemple la invarianza al escalado. En esta tarea en particular, el tamaño de los tumores puede variar considerablemente, lo que dificulta la detección de los mismos. Este problema se ha abordado mediante la implementación de un esquema multiescala. En este caso se ha comprobado que, por lo general, evaluar la muestra de test a 10 escalas distintas genera resultados óptimos. Por otro lado, el número de escalados puede reducirse considerablemente sin que ello suponga una pérdida de rendimiento, siempre y cuando el valor de k durante la búsqueda de k -vecinos se mantenga alto.

El esquema de multiescalado propuesto requiere combinar de algún modo las hipótesis obtenidas en cada escala. En este caso, la regla de la media se ha presentado como la idónea.

Se ha evaluado también la relación entre el tamaño de la ventana de características T_w y el tamaño del tumor normalizado K_o , con objeto de determinar el funcionamiento de un sistema basado en características locales frente a otro basado en características globales. Teniendo en cuenta que en esta tarea no es posible determinar con absoluta precisión el tamaño del tumor y, por tanto, la frontera entre lo que podemos considerar como características locales y globales es un tanto di-

fusa, puede concluirse que utilizar características excesivamente locales perjudica el rendimiento del sistema. El motivo de ello es que un tumor no presenta, a nivel local, un aspecto que lo diferencie claramente del tejido sano. Dichas diferencias se hacen más patentes, por contra, cuando se trabaja a nivel global o, al menos, con una ventana local que abarque la mayor parte del tumor.

En relación a las distintas aproximaciones de extracción de características evaluadas, SFUM y niveles de gris (ambas combinadas con PCA) son las que mejores resultados ofrecen. AFUM, a pesar de ser una técnica extremadamente sencilla, ofrece también buenos resultados, aunque significativamente peores que los conseguidos con las dos anteriores.

Por último se ha comparado el rendimiento obtenido con los clasificadores k -NN, SVM y *Boosting* y se ha observado que no existen diferencias significativas entre ellos.

6.5. Experimentos de detección con imágenes completas

Hasta aquí se han estudiado diversas aproximaciones a la clasificación de regiones de interés en una de las dos posibles clases de nuestro problema: *cáncer* y *no-cáncer*. Cada región de interés se representaba únicamente mediante su píxel central, por lo que la tarea de clasificación abordada se reducía a clasificar dicho píxel en una de las dos clases posibles.

En esta sección se aborda la segunda fase del problema, consistente en detectar tumores a partir de una imagen completa de mamografía. Para ello será necesario, en primer lugar, clasificar individualmente cada uno de los píxeles que forman la imagen, empleando alguna de las aproximaciones estudiadas en la fase previa. Esto genera un mapa de sospecha que recoge la estimación de la probabilidad a posteriori de cada píxel de la mamografía. A continuación, a partir del mapa de sospecha, habrá que decidir qué regiones de la imagen corresponden a tejido sano y qué otras a algún tipo de tumor.

El sistema de ayuda al diagnóstico propuesto se evalúa mediante el uso de curvas FROC (Sección 4.8.4). En ellas se muestra el porcentaje de tumores que el sistema es capaz de detectar (verdaderos positivos) en función del número medio de errores de falso positivo cometidos por imagen. Los resultados obtenidos se comparan con otros resultados publicados con anterioridad por otros autores que emplearon la misma base de datos.

En las siguientes secciones se explica, en primer lugar, el algoritmo empleado para establecer las regiones sospechosas a partir del mapa de sospecha. A continuación se explican brevemente algunas aproximaciones publicadas en trabajos previos y cuyos resultados han sido comparados con los obtenidos en este trabajo

y se detallan, asimismo, los conjuntos de test utilizados. Por último se muestran los resultados obtenidos y las conclusiones de esta parte del trabajo.

6.5.1. Detección de regiones sospechosas

Como se ha dicho anteriormente, para poder localizar posibles tumores en una mamografía, en primer lugar se obtiene el nivel de sospecha individual de cada píxel de la imagen. Una primera y sencilla aproximación para marcar en la mamografía las regiones con sospecha tumoral consiste en establecer un umbral T y considerar que cada píxel (x, y) de una imagen I con $p(\text{cáncer}|x, y) > T$ define, por sí mismo, una región sospechosa. Este algoritmo trivial presenta el inconveniente de generar una tasa muy elevada de falsos positivos. Además no tiene en cuenta la distribución de los píxeles con alta sospecha tumoral, de modo que no se refuerza el nivel de sospecha cuando aparece una región con gran densidad de píxeles sospechosos. Parece conveniente, por tanto, refinar este proceso mediante el agrupamiento de píxeles cercanos entre sí y con alta sospecha tumoral, en regiones sospechosas.

Para poder definir regiones sospechosas a partir del mapa de sospecha es necesario:

- Establecer un umbral a partir del cual se considera que un píxel tiene una “alta sospecha tumoral”.
- Establecer un criterio de vecindad que permita agrupar píxeles sospechosos en regiones sospechosas.
- Establecer el nivel de sospecha de la región completa en función del nivel de sospecha individual de los píxeles que la forman.
- Establecer el centro de la región sospechosa.

El algoritmo utilizado en este trabajo para definir las regiones sospechosas se basa en dos parámetros T y d que establecen, respectivamente, el umbral a partir del cual un píxel se considera sospechoso y la distancia empleada como criterio de vecindad.

Para establecer las regiones sospechosas se seleccionan todos los puntos de la imagen que hayan dado un nivel de sospecha $p(\text{cáncer}|x_i, y_i) \geq T$ y se agrupan en regiones, de modo que todos los puntos de una región tengan al menos un vecino a una distancia $\leq d$ en la misma región y su distancia a cualquier punto de otra región sea mayor de d . El centro de la región se calcula promediando la posición de cada uno los puntos que la forman, de manera ponderada con la probabilidad a posteriori (nivel de sospecha) de los mismos. Es decir, si definimos una región R mediante un conjunto de n puntos (x_i, y_i)

$$R = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\},$$

entonces

$$(x_c, y_c) = \frac{\sum_{i=1}^n (x_i, y_i) p(w_c | x_i, y_i)}{\sum_{i=1}^n p(w_c | x_i, y_i)}$$

donde (x_c, y_c) representa el centro de R y w_c la clase *cáncer*.

Finalmente es necesario establecer el nivel de sospecha s de cada región en función del nivel de sospecha individual de los puntos que la forman. Para ello se pueden utilizar distintos criterios:

- **Máximo:** a cada región R se le otorga la probabilidad a posteriori máxima encontrada entre los puntos que la componen.

$$s(R) = \max_{(x_i, y_i) \in R} p(w_c | (x_i, y_i)) \quad (6.1)$$

- **Media:** a cada región R se le otorga la media de las probabilidades a posteriori de los puntos que la componen.

$$s(R) = \frac{1}{n} \sum_{(x_i, y_i) \in R} p(w_c | (x_i, y_i)) \quad (6.2)$$

donde n = número de píxeles sospechosos en R .

- **Suma:** a cada región R se le otorga la suma de las probabilidades a posteriori de los puntos que la componen.

$$s(R) = \sum_{(x_i, y_i) \in R} p(w_c | (x_i, y_i)) \quad (6.3)$$

6.5.2. Trabajos previos

La base de datos MIAS utilizada en este trabajo es de uso público pero, desafortunadamente, no establece una partición “estándar” para los conjuntos de entrenamiento y test. Es por ello que, aunque hay una cantidad importante de trabajos publicados que han utilizado esta base de datos, resulta difícil comparar los resultados de los distintos estudios por la falta de un criterio común a la hora de establecer el conjunto de imágenes utilizado para evaluar el rendimiento de las técnicas empleadas. En algunas de estas publicaciones, sin embargo, se detalla con precisión el conjunto de test utilizado.

Para contrastar la bondad de los algoritmos propuestos en este trabajo, se han escogido aquellos trabajos de relevancia en los que se detalla el conjunto de test empleado y se han comparado los resultados obtenidos. A continuación se hace un breve resumen de las aproximaciones empleadas en dichos trabajos:

Heath & Bowyer [Heat 00b]:

En este trabajo se emplea el algoritmo AFUM descrito en la Sección 4.2.4 para calcular el nivel de sospecha individual de cada píxel de la mamografía (mapa de sospecha). A continuación, cada máximo local encontrado en el mapa de sospecha, se marca como punto central de una región sospechosa y se otorga a dicha región un nivel de sospecha igual al valor AFUM del máximo local. Finalmente, si una región sospechosa se encuentran a una distancia inferior a 5mm de otra región con mayor nivel de sospecha, la primera es eliminada.

Brake & Karssemeijer [Brak 98b]:

En esta publicación se analizan tres métodos distintos de detección de tumores en mamografías. Todos ellos están basados en la detección de regiones más o menos circulares con un brillo relativamente alto, a partir de características locales de la imagen. Los dos primeros métodos constituyen técnicas clásicas de análisis de imágenes, mientras que el tercero supone una nueva aportación de este trabajo. A continuación se describe brevemente cada uno de estos métodos:

Filtro Laplaciano:

Se basa en convolucionar la imagen con un filtro definido mediante una función de media cero, con valores positivos en el centro y negativos en los extremos. Concretamente se utiliza la segunda derivada de una gaussiana, también llamada Laplaciano de la Gaussiana (LoG). Por razones computacionales, este filtro se implementa como una Diferencia de Gaussianas (DoG).

Emparejamiento de patrones (template matching):

Se asume un modelo de tumor y, mediante una ventana local que se desliza a lo largo de la mamografía, se calcula la correlación entre la ventana local y el modelo asumido. Para modelar un tumor de radio R se utiliza la siguiente función:

$$T(x, y) = \begin{cases} 0 & \text{si } x^2 + y^2 \geq R^2 \\ R^2 - x^2 - y^2 & \text{si } x^2 + y^2 < R^2 \end{cases}$$

Análisis de la orientación del gradiente:

Se parte de la idea, ya descrita, de que la apariencia de un tumor en una mamografía consiste en una región más o menos circular con un brillo relativamente alto. En estas condiciones, si se obtiene un mapa de gradientes de la mamografía, un tumor aparecerá como una región circular con muchos gradientes apuntando hacia el centro. El método consiste en calcular un nivel de sospecha para cada píxel i de la

mamografía del siguiente modo: se determina si el gradiente de cada píxel j a una distancia $rmin \leq d(i, j) \leq rmax$ está o no orientado hacia el centro. Un píxel se considera que está orientado hacia el centro si el ángulo θ entre la recta que une i, j y la orientación del gradiente es menor que $R/d(i, j)$, donde R es una constante. Se demuestra que es posible estimar el número de puntos N_p que están orientados hacia el centro cuando la imagen no presenta una estructura determinada, así como su varianza $\sqrt{N_p(1-p)}$, siendo p la probabilidad de que un píxel situado entre $rmin$ y $rmax$ esté orientado hacia el centro. En consecuencia, si se contabiliza el número de píxeles n que realmente apuntan hacia el centro, se puede estimar el nivel de sospecha de i como:

$$s_i = \frac{n - N_p}{\sqrt{N_p(1-p)}}$$

Los valores óptimos de R , $rmin$ y $rmax$ encontrados durante la fase de entrenamiento son, respectivamente, 4, 2 y 12mm.

En los tres métodos propuestos se utiliza una red neuronal para mapear las características obtenidas en un nivel de sospecha, incluso cuando únicamente existe una característica. En este caso la red funciona como una tabla de mapeado o *look-up table*. Finalmente se comenta que se aplica un umbral para segmentar regiones sospechosas a partir del nivel de sospecha individual de cada píxel y que posteriormente se calculan nuevas características de dichas regiones para eliminar falsos positivos, pero no se dan detalles de este proceso.

La última aproximación propuesta (análisis de orientación del gradiente) es la que mejores resultados obtiene. Los resultados presentados en esta tesis, por tanto, serán comparados únicamente con esta última aproximación.

Sallam & Bowyer [Sall 99]:

El método propuesto en este trabajo consiste en identificar diferencias entre mamografías de una misma paciente. Las imágenes pueden ser secuencias temporales de mamografías correspondientes al mismo pecho o mamografías del pecho izquierdo y derecho obtenidas en la misma sesión. En primer lugar se emplea un algoritmo que permite alinear las dos mamografías mediante deformaciones no uniformes, para lo que es necesario encontrar correspondencias entre una y otra imagen. Una vez alineadas las imágenes se hace una operación simple de resta para obtener la imagen diferencia. Dado que los tumores tienen normalmente un nivel de brillo mayor que el tejido sano, para resaltar las diferencias de las regiones con mayor sospecha tumoral, el valor de nivel de gris se utiliza para ponderar la diferencia. Finalmente, mediante umbralización de la imagen diferencia, se segmentan las regiones sospechosas.

6.5.3. Conjuntos de evaluación

Para poder comparar las aproximaciones propuestas en este trabajo con las presentadas en las publicaciones citadas anteriormente, se utilizaron los mismos conjuntos de test empleados en dichas publicaciones. En la Tabla 6.5 se muestran los códigos de mamografía de la base de datos MIAS incluidos en cada uno de los conjuntos de test evaluados (SET1 y SET2). En [Brak 98b] se utilizó el conjunto SET1, en [Sall 99] se empleó el SET2, mientras que en [Heat 00b] se realizaron experimentos tanto con SET1 como con SET2. En la Tabla 6.6 se muestra el número de mamografías de cada conjunto, agrupadas por tipo de tumor (los distintos tipos de tumor fueron detallados en la Sección 6.2). Como puede observarse, el conjunto SET2 es más completo, no sólo porque tiene un mayor número de mamografías, sino porque contiene tumores de todos los tipos (nótese que en SET1 no existen tumores de tipo ASYM).

SET1 (60 mamografías)	SET2 (228 mamografías)
003-004, 006-009, 011 014, 016, 018, 020 022-024, 026-029, 031 033-045, 058, 115, 117 120, 124-125, 130, 134 141, 144, 148, 155, 158 170-171, 178-179, 181 184, 186, 202, 206 264-265, 267, 270-271 274	001-014, 017-018, 023-024 029-068, 071-078, 081-090 093-174, 189-208, 229-230 243-244, 257-294, 297-304

Tabla 6.5: Mamografías de la base de datos MIAS utilizadas en cada uno de los dos conjuntos test evaluados (SET1 y SET2).

	SET1		SET2	
	Mamografías	Tumores	Mamografías	Tumores
CIRC	4	4	13	15
SPIC	8	8	12	12
ARCH	10	10	19	19
ASYM	0	0	14	14
MISC	8	9	12	13
Total tumores	30	31	70	73
NORM	30	-	158	-
Total	60	31	228	73

Tabla 6.6: Distribución de las mamografías, por tipo de tumor, en los conjuntos de test SET1 y SET2. La columna *Mamografías* muestra el número total de imágenes, mientras que la columna *Tumores* muestra el número de tumores, teniendo en cuenta que algunas imágenes pueden contener más de un tumor.

6.5.4. Resultados

A continuación se presentan los resultados obtenidos y se comparan con los publicados en [Heat 00b, Brak 98b, Sall 99]. Para mantener la consistencia con el criterio empleado en dichas publicaciones a la hora de computar los verdaderos y falsos positivos, este cálculo se hizo como sigue: un tumor que tenga al menos el centro de una región sospechosa dentro de sus límites, se computa como un verdadero positivo. Si el centro de más de una región sospechosa cae dentro del mismo tumor, se computa como un verdadero positivo y cero falsos positivos. Cualquier región cuyo centro queda fuera de un tumor se contabiliza como un falso positivo.

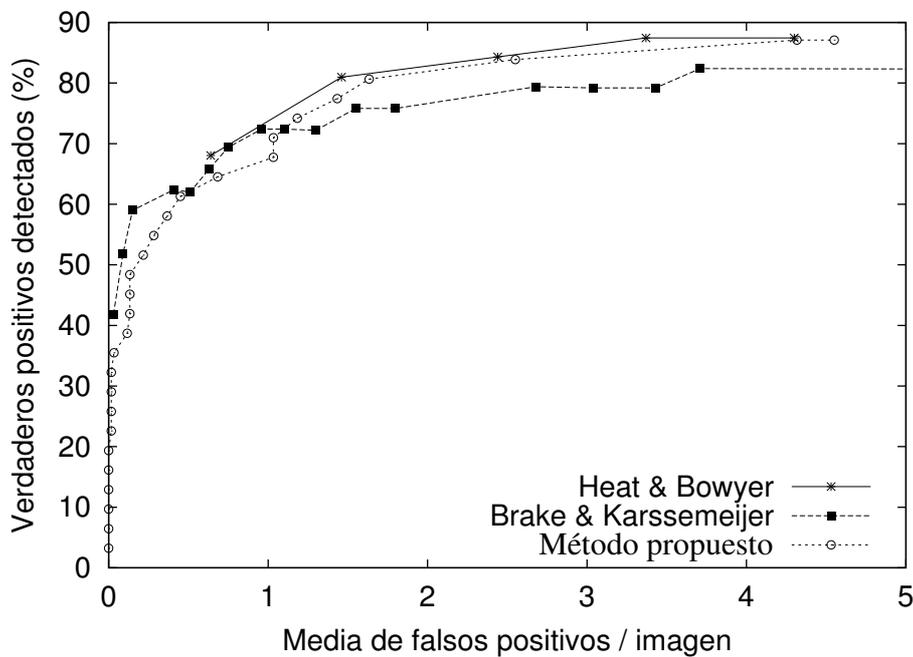


Figura 6.21: Resultados obtenidos al evaluar con el conjunto de test SET1 el método propuesto y los métodos presentados en [Heat 00b, Brak 98b].

En la Figura 6.21 se muestran los resultados obtenidos con el conjunto de test SET1 y se comparan con los publicados en [Heat 00b, Brak 98b]. En la Figura 6.22 se muestran los resultados obtenidos con el conjunto SET2 y en este caso se comparan con los publicados en [Heat 00b, Sall 99]. En ambos casos se utilizó la aproximación de extracción de características SFUM con los parámetros detallados en la Figura 6.16 mostrada en la Sección 6.4.7. Para definir las regiones sospechosas mediante el algoritmo presentado en la Sección 6.5.1 se estableció un umbral $T = 0.5$ para seleccionar los píxeles sospechosos y una distancia $d = 2$ para establecer el criterio de vecindad. Para computar el nivel de sospecha de cada región se utilizó la regla de la suma. La regla de la media y el máximo fueron evaluadas en experimentos previos y produjeron unas tasas de falsos positivos ligeramente

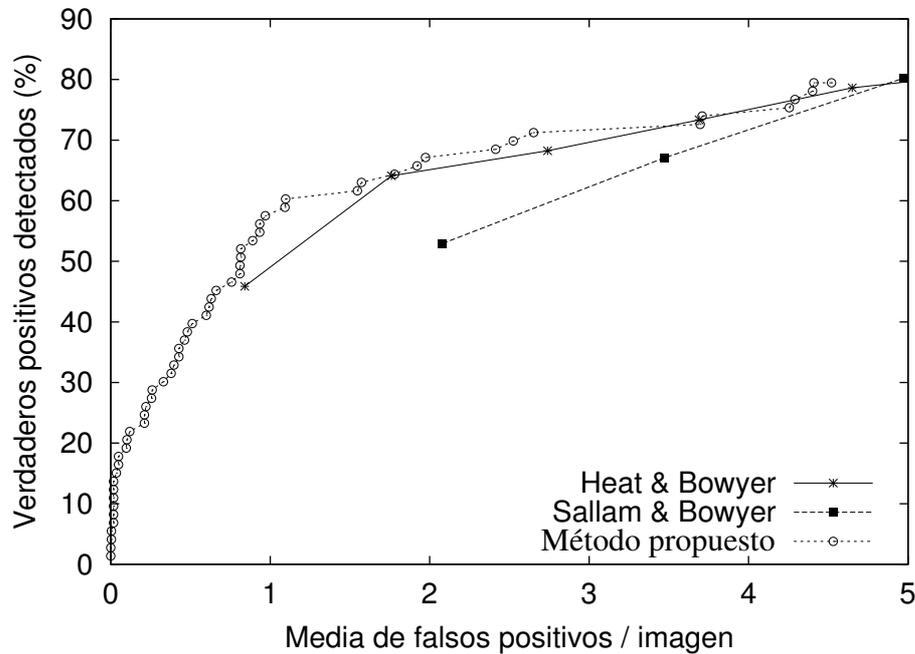


Figura 6.22: Resultados obtenidos al evaluar con el conjunto de test SET2 el método propuesto y los métodos presentados en [Heat 00b, Sall 99].

superiores.

Las curvas FROC se han limitado a un rango entre 0 y 5 falsos positivos por imagen, al entenderse que esta es la parte más útil de la curva. Como ya se adelantó en capítulos anteriores, ofrecer resultados con un número elevado de falsos positivos no ayuda al experto humano a realizar un mejor diagnóstico [Hutt 96]. Aunque en la Figura 6.21 se observa que el porcentaje de verdaderos positivos se estabiliza cuando se alcanza la media de 4 falsos positivos por imagen, no ocurre lo mismo con los resultados mostrados en la Figura 6.22. En este caso, a partir de 5 falsos positivos por imagen, tanto en el método propuesto como el presentado por *Heath & Bowyer* se estabiliza la tasa de verdaderos positivos, mientras que en la aproximación presentada por *Sallam & Bowyer*, sorprendentemente, la tasa de verdaderos positivos decrece significativamente.

Como puede observarse, la aproximación propuesta ofrece mejores resultados que las presentadas en [Brak 98b] (SET1) y en [Sall 99] (SET2). Respecto a los resultados publicados en [Heat 00b], la aproximación propuesta obtiene tasas similares o ligeramente inferiores cuando se evalúa el sistema con el conjunto SET1, y tasas superiores en la mayoría de los puntos de la curva cuando se utiliza el conjunto SET2. Conviene remarcar que el conjunto SET2 es más completo debido a su mayor tamaño, a que contiene mamografías de todos los tipos de tumores (el conjunto SET1 no contiene tumores de tipo “asimetrías”) y a que contiene una mayor

proporción de mamografías sanas, tal y como ocurre en una tarea real.

6.5.5. Conclusiones y desarrollos futuros

En esta segunda parte del capítulo se han empleado los algoritmos descritos en la primera parte y se ha aplicado el esquema de extracción de características SFUM. Se ha presentado además un algoritmo de agrupamiento para la detección de regiones sospechosas a partir del mapa de sospecha.

Los resultados obtenidos igualan o superan a los publicados por otros autores, lo que demuestra que las técnicas empleadas son totalmente adecuadas para abordar el problema de detección de tumores en mamografías digitales.

Actualmente, de entre los sistemas comerciales de ayuda al diagnóstico de cáncer de mama, el más extendido es el *R2ImageChecker* [Roeh 98, Bird 05]. Aunque no se desvela con detalle los algoritmos empleados por este sistema, si se sabe que fueron diseñados originalmente por N. Karssemeijer y su equipo. Una de las aproximaciones comparadas en este estudio se debe precisamente a este equipo [Brak 98b], lo que pone de manifiesto la relevancia de los trabajos empleados para contrastar los resultados conseguidos.

Como desarrollo futuro, debería evaluarse si el uso de un sistema como el propuesto mejora del rendimiento diagnóstico de un radiólogo. Para ello, como fase previa, tal y como se hizo con el sistema de ayuda al diagnóstico de cáncer de próstata presentado en el Capítulo 5, habría que marcar las regiones sospechosas de un conjunto de mamografías y contrastar el rendimiento diagnóstico de un grupo de radiólogos al analizar las mamografías originales y las marcadas. En la fase final, debería analizarse el rendimiento del sistema en una tarea real mediante un análisis prospectivo.

Capítulo 7

Conclusiones finales y desarrollos futuros

7.1. Conclusiones finales

En esta tesis se han presentado una serie de aproximaciones a la detección de tumores en imágenes médicas. Concretamente, se han abordado los problemas de detección de tumores de próstata en imágenes de ecografía transrectal y de tumores de mama a partir de mamografías obtenidas por rayos X. El objetivo ha sido la búsqueda de aproximaciones que permitan optimizar el rendimiento de estos sistemas de detección, con la finalidad de desarrollar sistemas de CAD que ayuden al experto humano a realizar diagnósticos más precisos.

Para lograr los objetivos se han ido abordando los diversos problemas que aparecen en cada una de las etapas que requiere un sistema de CAD. Las técnicas empleadas en las dos líneas seguidas (próstata y mama) han sido muy similares, aunque no siempre las mismas, debido a la distinta naturaleza de los problemas. Fundamentalmente, en el problema del cáncer de mama, se tiene un mayor conocimiento acerca del patrón que se desea detectar, lo que abre el abanico de técnicas a emplear.

En las Secciones 5.5 y 6.5.5 se han presentado, respectivamente, las conclusiones particulares de las tareas de detección de tumores de próstata y de mama abordadas en esta tesis. A continuación se hace una recapitulación de los aspectos más destacables.

Métodos:

En el problema de detección de tumores de próstata se ha realizado la adquisición y supervisión de más de 1.500 ecografías transrectales (TRUS). Esta es la mayor base de datos de imágenes de TRUS utilizada hasta ahora en un trabajo de

este tipo. En la fase de extracción de características se han evaluado dos aproximaciones basadas, respectivamente, en niveles de gris y en descriptores de textura, en ambos casos combinado con PCA. En la etapa de clasificación se han evaluado los clasificadores k -NN y HMM. Para el primero de ellos se han empleado técnicas de búsqueda rápida aproximada mediante *kd-trees*. El segundo, más novedoso en este ámbito, ha sido empleado en un intento de minimizar el problema de la supervisión imperfecta apuntado en la Sección 5.2.5. Para evaluar los resultados, se han empleado curvas ROC. Finalmente, para evaluar la mejora diagnóstica cuando se utiliza un sistema de CAD, se ha desarrollado una aplicación que se ejecuta a través de la *web*, la cual permite la clasificación de imágenes médicas y la comparación de los resultados obtenidos con y sin la ayuda del sistema.

En el problema de detección de tumores de mama se ha trabajado con una base de datos pública, lo que permite comparar los resultados obtenidos. En las fases de preprocesamiento y extracción de características se ha comprobado la importancia de normalizar los niveles de gris a nivel local y se han evaluado nuevos métodos de extracción de características, además de los empleados en el problema de la próstata. En la fase de entrenamiento se han empleado técnicas de *bootstrapping* para reducir la complejidad del modelo sin perder representatividad. La clasificación se ha realizado mediante k -NN con búsqueda rápida aproximada, de forma similar a como se hizo para el caso de cáncer de próstata, y se han evaluado otros dos clasificadores, concretamente SVM y *Boosting*. Se ha aplicado además un esquema multiescala para conseguir la invarianza al escalado. Finalmente, en la etapa de postproceso, se han aplicado técnicas de agrupamiento para definir regiones sospechosas y reducir el número de falsos positivos. Para evaluar los resultados, además de las curvas ROC, se han empleado curvas FROC.

Resultados:

En lo referente al CAD de cáncer de próstata, y a pesar de la dificultad de la tarea, se han conseguido obtener características de la imagen que permiten, de manera significativa, discriminar entre tejido sano y canceroso. Concretamente, en una tarea de clasificación de imágenes de TRUS, se ha obtenido un área bajo la curva ROC del 61.6% con un intervalo de confianza al 95% de 0.7, lo que demuestra que el sistema presentado tiene cierta capacidad discriminativa. El uso de modelos de Markov y de características de niveles de gris, combinadas con PCA, han ofrecido los mejores resultados, aunque, en general, las diferencias entre las distintas aproximaciones de extracción de características y de clasificación no son importantes.

En una segunda fase del problema se ha desarrollado un prototipo de ayuda al diagnóstico de cáncer asistido por ordenador, con el objetivo de evaluar en qué medida puede mejorarse el rendimiento diagnóstico al emplear una herramienta de este tipo. En los resultados se observa una ligera mejora en el diagnóstico cuando se trabaja con la ayuda del sistema de CAD, aunque dicha mejora no llega a ser significativa.

Por término medio un urólogo es capaz, con el uso del sistema de CAD propuesto, de aumentar los valores de sensibilidad del 87.5 % al 91.6 % para una especificidad en ambos casos del 18 %¹, lo que supone mejorar aproximadamente en un 5 % los valores de sensibilidad. Los únicos valores reportados hasta el momento en un proceso real con pacientes no seleccionados fueron una sensibilidad del 85 % y especificidad del 18 % [Beer 01], si bien la comparación de estos resultados con los presentados en esta tesis debe realizarse con la necesaria cautela, teniendo en cuenta que los corpora utilizados en cada experimento fueron distintos.

Por otro lado, una persona sin experiencia previa en análisis de TRUS consigue, con la ayuda del sistema, un área bajo la curva ROC (AUC) del 61.3 % (o sensibilidad del 90.7 % y especificidad del 18 % para un punto concreto de la curva ROC), lo que supone una clara mejora a las tasas conseguidas cuando no se cuenta con la ayuda del sistema de CAD. Esta AUC alcanzada demuestra que una persona inexperta puede conseguir un rendimiento similar al obtenido por urólogos expertos.

Por otro lado, una persona sin experiencia previa en análisis de TRUS consigue, con la ayuda del sistema, un área bajo la curva ROC (AUC) del 61.3 % (o sensibilidad del 90.7 % y especificidad del 18 % para un punto concreto de la curva ROC), lo que supone una clara mejora a las tasas conseguidas cuando no se cuenta con la ayuda del sistema de CAD. Esta AUC alcanzada demuestra que una persona inexperta puede conseguir un rendimiento similar al obtenido por urólogos expertos.

En el caso de CAD de cáncer de mama, la posibilidad de discriminar entre tejido sano y canceroso es mucho mayor que en el problema del cáncer de próstata, tal y como lo demuestran los resultados obtenidos. En primer lugar se realizaron distintos experimentos con regiones de interés, con el objetivo de optimizar en la medida de lo posible cada una de las etapas del sistema de CAD.

Gracias a estos experimentos se ha podido comprobar que se obtiene una mejor clasificación cuando, previo a la extracción de características, se normalizan los niveles de gris en la propia ventana de características. Se ha demostrado, asimismo, que mediante el uso de PCA es posible reducir el número de características sin perder prestaciones en la clasificación. Mediante el uso de *bootstrapping* se ha conseguido reducir por 10 el número de prototipos de la clase negativa con unos resultados de clasificación prácticamente idénticos. El problema de la invarianza al escalado se ha abordado mediante el uso de un esquema multiescala. Se han obtenido los mejores resultados calculando el nivel de sospecha como la media de los niveles de sospecha encontrados en 10 escalas distintas. Se ha evaluado también la relación entre el tamaño de la ventana de características y el tamaño del tumor tras su normalización a un tamaño fijo. Se ha comprobado que una ventana

¹Con objeto de poder comparar los resultados para un mismo nivel de especificidad, estos valores se han obtenido por interpolación de los resultados mostrados en las Tablas 5.10 y 5.11.

de características ligeramente menor que el tumor a modelar ofrece los mejores resultados. Ventanas de menor tamaño pueden generar prototipos con mayor similitud interclase, mientras que ventanas mayores generarán una mayor dispersión en los prototipos de una misma clase. Respecto a las técnicas de extracción de características, se ha comprobado que SFUM y niveles de gris, en combinación con PCA son, de entre los métodos evaluados, los más eficientes. Finalmente se han comparado los clasificadores k -NN, SVM y *Boosting*, sin encontrarse diferencias significativas en los resultados obtenidos.

En los experimentos realizados con imagen completa, se han comparado los resultados obtenidos con otros publicados anteriormente. Para ello se han seleccionado trabajos de relevancia que emplearon una misma base de datos. Algunos de los autores de dichos trabajos han participado en el desarrollo del *R2ImageChecker*, el sistema de CAD de cáncer de mama más utilizado actualmente [Roeh 98, Bird 05]. Los resultados obtenidos en este trabajo han igualado o incluso superado a los publicados con anterioridad, lo que demuestra la valía de las aproximaciones propuestas. El sistema presentado es capaz de detectar, por sí solo, entre un 70 y un 80% de los tumores, con una media de 2 errores de falso positivo por imagen. No se ha evaluado, sin embargo, la ayuda que esto pueda reportar a un radiólogo a la hora de elaborar un diagnóstico.

7.1.1. Publicaciones relacionadas con esta Tesis

La investigación llevada a cabo durante la realización de esta tesis, ha producido las publicaciones que se citan a continuación:

- R. Llobet, J.C. Pérez-Cortés, A.H. Toselli and A. Juan. “Computer-aided Detection of Prostate Cancer”, *International Journal of Medical Informatics*, in press., 2006.
- R. Llobet, R. Paredes and J.C. Pérez-Cortés. “Comparison of Feature Extraction Methods for Breast Cancer Detection”. In: *Proceedings of the 2nd Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA)*, pp.495–502, Estoril (Portugal), 2005.
- R. Llobet and J. C. Perez-Cortes. “Breast Cancer Detection in Digitized Mammograms Using Non-Parametric Methods”. In: *Proceedings of the 2nd International Conference on Advances in Biomedical Signal and Information Processing (MEDSIP)*, pp. 281-287, Sliema (Malta), 2004.
- R. Llobet, A.H. Toselli, J.C. Pérez-Cortés and A. Juan. “Computer-aided Prostate Cancer Detection in Ultrasonographic”. In: *Proceedings of the 1st Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA)*, pp.411-419, Puerto de Andratx (Mallorca, Spain), 2003.

7.2. Desarrollos futuros

Tal y como se ha expuesto en las conclusiones, el sistema de CAD de cáncer de próstata presentado permite mejorar ligeramente el rendimiento diagnóstico del experto humano, pero no de modo significativo. Una pequeña mejora en el sistema de CAD podría suponer que su utilidad clínica sea real, lo que supondría un enorme beneficio. A raíz de lo expuesto en el Capítulo 5, parece claro que para lograr tal mejora lo primero que debería abordarse es la adquisición y supervisión de un nuevo corpus. Los transductores empleados actualmente para obtener imágenes de ecografía tienen unas prestaciones muy superiores a las del utilizado en este trabajo. Trabajar con un ecógrafo de última generación permitiría obtener imágenes mucho más nítidas, en las que debería ser más sencillo encontrar características que permitan la discriminación entre distintos tipos de tejido. Por otro lado, debería resolverse el problema de la supervisión imperfecta comentado en la Sección 5.2.5. Si durante la supervisión se consiguiera localizar los tumores con precisión, además de la evidente ventaja que supondría entrenar el sistema con una muestra correctamente etiquetada, podrían emplearse técnicas impensables en las condiciones actuales y que sí han sido utilizadas en la tarea de cáncer de mama. Por ejemplo, podría abordarse el problema de la invarianza al escalado o a la rotación, o podrían realizarse experimentos que tuviesen en cuenta la relación entre el tamaño de la ventana de vecindad y el tamaño del tumor. En resumen, si se dispusiera de unas imágenes más nítidas y se lograra solucionar el problema de la supervisión imperfecta, probablemente se conseguirían mejores prestaciones.

Además, de forma general, sería interesante explorar otras aproximaciones a la extracción de características. En este sentido, las características SIFT (*Scale Invariant Feature Transform*) permiten detectar y extraer descriptores que son razonablemente invariantes a cambios en iluminación, rotación, escalado, pequeños cambios en el punto de vista y ruido en la imagen [Lowe 04].

En la tarea de cáncer de mama sería interesante entrenar el sistema con muestras procedentes de otras bases de datos. En este sentido, la base de datos DDSM (*Digital Database for Screening Mammography*) [Heat 98] cuenta con un gran número de mamografías y dispone de un etiquetado preciso a nivel de píxel.

Finalmente, debería evaluarse en qué medida se modifica el diagnóstico dado por el experto humano, cuando éste utiliza el sistema de CAD de cáncer de mama presentado. Para ello, con la misma aplicación utilizada en la tarea de cáncer de próstata, se debería solicitar a un grupo de radiólogos que clasificasen cierto número de mamografías con y sin marcado de regiones sospechosas, y comprobar si realmente el marcado permite elaborar un mejor diagnóstico.

Apéndice A

Cálculo de estadísticos en matrices de coocurrencia

En la Sección 4.2.3 se ha presentado una aproximación a la extracción de características basada en la obtención de descriptores de textura a partir de matrices de coocurrencia. A continuación se muestran las ecuaciones empleadas para obtener cada uno de estos descriptores.

Notación

- P representa la matriz de coocurrencia.
- G es el número de niveles de gris considerado.
- μ es la media de P .
- P_x y P_y son las probabilidades marginales de las columnas y filas de P respectivamente.

$$P_x(i) = \sum_{j=0}^{G-1} P(i, j), \quad i = 0, 1, \dots, G - 1$$

$$P_y(j) = \sum_{i=0}^{G-1} P(i, j), \quad j = 0, 1, \dots, G - 1$$

- μ_x, μ_y, σ_x y σ_y son, respectivamente, las medias y las desviaciones estándar de P_x y P_y .

$$\mu_x = \sum_{i=0}^{G-1} i \sum_{j=0}^{G-1} P(i, j) = \sum_{i=0}^{G-1} i P_x(i)$$

$$\begin{aligned}\mu_y &= \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} jP(i, j) = \sum_{j=0}^{G-1} jP_y(j) \\ \sigma_x^2 &= \sum_{i=0}^{G-1} (i - \mu_x)^2 \sum_{j=0}^{G-1} P(i, j) = \sum_{i=0}^{G-1} (P_x(i) - \mu_x(i))^2 \\ \sigma_y^2 &= \sum_{j=0}^{G-1} (j - \mu_y)^2 \sum_{i=0}^{G-1} P(i, j) = \sum_{j=0}^{G-1} (P_y(j) - \mu_y(j))^2\end{aligned}$$

- P_{x+y} y P_{x-y} son, respectivamente:

$$\begin{aligned}P_{x+y}(k) &= \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} G-1P(i, j), \quad i + j = k \\ P_{x-y}(k) &= \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} G-1P(i, j), \quad |i - j| = k\end{aligned}$$

A partir de esta notación, se definen los siguientes descriptores de texturas obtenidos de la matriz de coocurrencia P .

- **Homogeneidad, Segundo Momento Angular:**

$$SMA = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (P(i, j))^2 \quad (\text{A.1})$$

El segundo momento angular es una medida de homogeneidad de la imagen. Una imagen homogénea contiene pocos niveles de gris, por lo que producirá una matriz de coocurrencia con pocos valores distintos de cero, aunque relativamente altos. En consecuencia, la suma de los cuadrados será un valor elevado.

- **Contraste:**

$$CONTR = \sum_{n=0}^{G-1} n^2 \left(\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j) \right), \quad |i - j| = n \quad (\text{A.2})$$

Esta característica es una medida de contraste o de cantidad de variaciones locales de intensidad dentro de la imagen. Dichas variaciones de intensidad en la imagen producen matrices de coocurrencia con valores más elevados cuanto más alejados estén de la diagonal. En la ecuación anterior, el factor de ponderación n es más elevado para aquellas posiciones (i, j) que más se alejen de la diagonal.

■ **Correlación:**

$$CORRE = \frac{\left(\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} ijP(i,j) \right) - (\mu_x\mu_y)}{\sigma_x\sigma_y} \quad (A.3)$$

Esta característica es una medida de dependencia lineal de niveles de gris entre píxeles de la imagen separados por una distancia d y ángulo θ específicos.

■ **Suma de los Cuadrados, Varianza:**

$$VAR = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i - \mu)^2 P(i,j) \quad (A.4)$$

Esta característica da pesos relativamente altos a los elementos que difieren del valor medio de P .

■ **Homogeneidad local, Momento diferencial inverso:**

$$MID = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{1}{1 + (i - j)^2} P(i,j) \quad (A.5)$$

Esta medida, al igual que el segundo momento local, también depende de la homogeneidad de la imagen. Debido al factor de ponderación $(1 + (i - j)^2)^{-1}$, esta característica obtendrá pequeñas contribuciones de $P(i,j)$ cuando $i \neq j$ (puntos de P relacionados con imágenes poco homogéneas) y contribuciones mayores de la diagonal de P (puntos relacionados con imágenes homogéneas). El resultado, por tanto, será un valor elevado de MID para imágenes homogéneas y viceversa.

■ **Promedio Acumulado:**

$$PA = \sum_{i=0}^{2G-2} iP_{x+y}(i) \quad (A.6)$$

■ **Entropía Acumulada:**

$$EA = - \sum_{i=0}^{2G-2} P_{x+y}(i) \log(P_{x+y}(i)) \quad (A.7)$$

■ **Varianza Acumulada:**

$$VA = \sum_{i=0}^{2G-2} (i - S)^2 P_{x+y}(i) \quad (A.8)$$

donde: $S = EA(P)$

■ **Entropía:**

$$ENTROPIA = - \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j) \log(P(i, j)) \quad (\text{A.9})$$

Las imágenes poco homogéneas (aquellas en las que P tiene gran cantidad de elementos $\neq 0$), tendrán un valor alto de entropía, mientras que las imágenes más homogéneas (matriz con valores elevados cerca de la diagonal y ceros lejos de la misma) tendrán una baja entropía.

■ **Varianza Diferencial:**

$$VARD = \frac{G^2 \sum_{i=0}^{G-1} P_{x-y}(i)^2 - \left(\sum_{i=0}^{G-1} P_{x-y}(i) \right)^2}{G^4} \quad (\text{A.10})$$

■ **Entropía Diferencial:**

$$ENTD = - \sum_{i=0}^{G-1} P_{x-y}(i) \log(P_{x-y}(i)) \quad (\text{A.11})$$

Apéndice B

Imágenes de los corpora utilizados

En este apéndice se muestran, a modo de ejemplo, algunas imágenes de ecografía transrectal y de mamografía, empleadas en este trabajo. En el caso de las imágenes de ecografía transrectal se resalta, mediante un recuadro, la zona correspondiente al tejido biopsiado, y en el pie de figura se comenta si presenta o no tejido canceroso. En caso de presencia de un tumor, no se especifica exactamente en qué parte del tejido biopsiado se encuentran las células cancerosas por desconocerse este dato (Sección 5.2.5). En el caso de las mamografías, cuando existe un tumor, se especifica la posición de éste en la imagen mediante un recuadro.



Figura B.1: Imagen de ecografía transrectal con hiperplasia benigna.



Figura B.2: Imagen de ecografía transrectal con hiperplasia benigna.

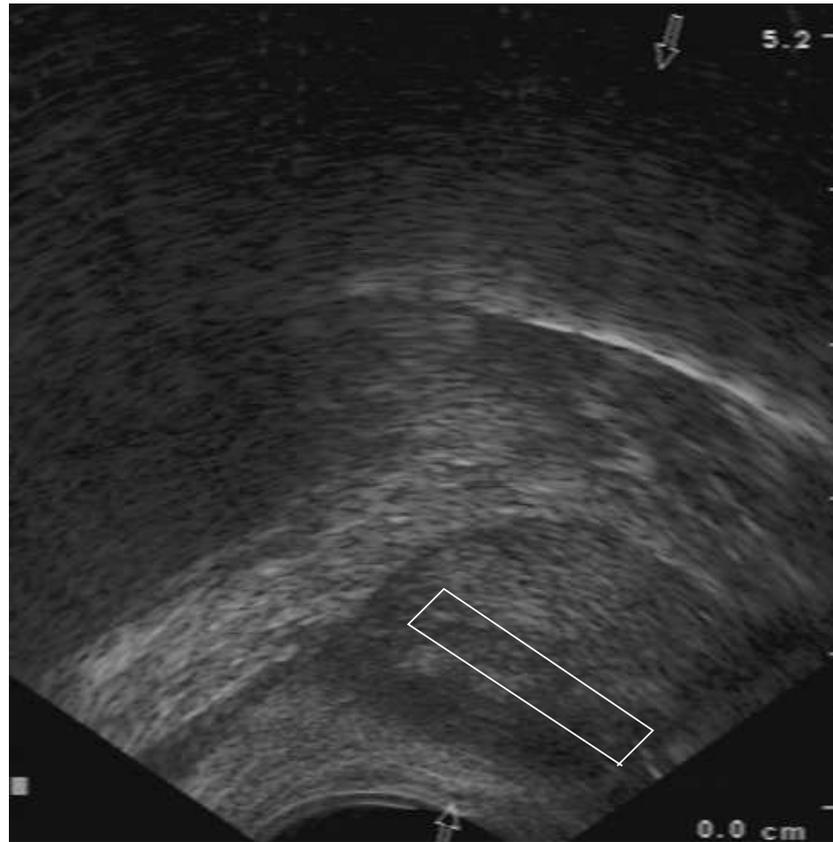


Figura B.3: Imagen de ecografía transrectal con tejido canceroso.

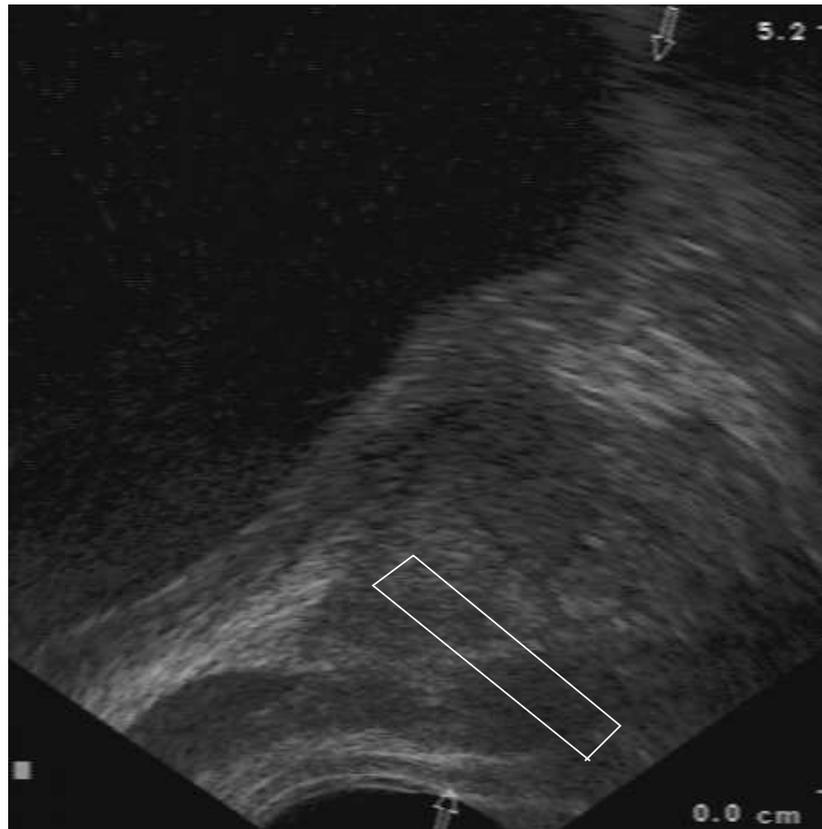


Figura B.4: Imagen de ecografía transrectal con tejido canceroso.

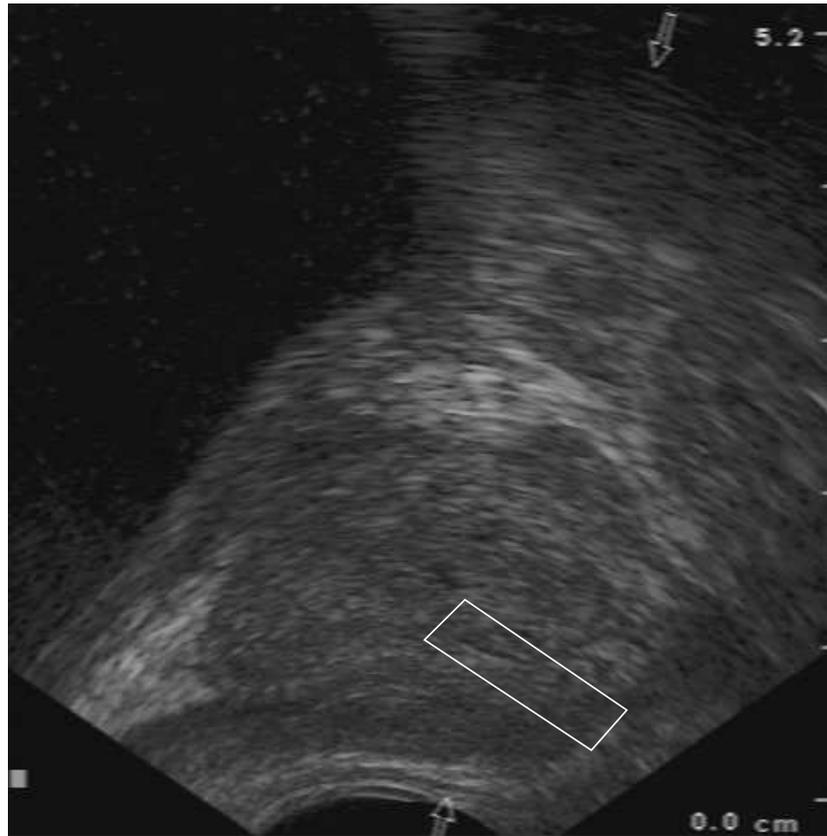


Figura B.5: Imagen de ecografía transrectal con hiperplasia benigna.



Figura B.6: Imagen de ecografía transrectal con hiperplasia benigna.

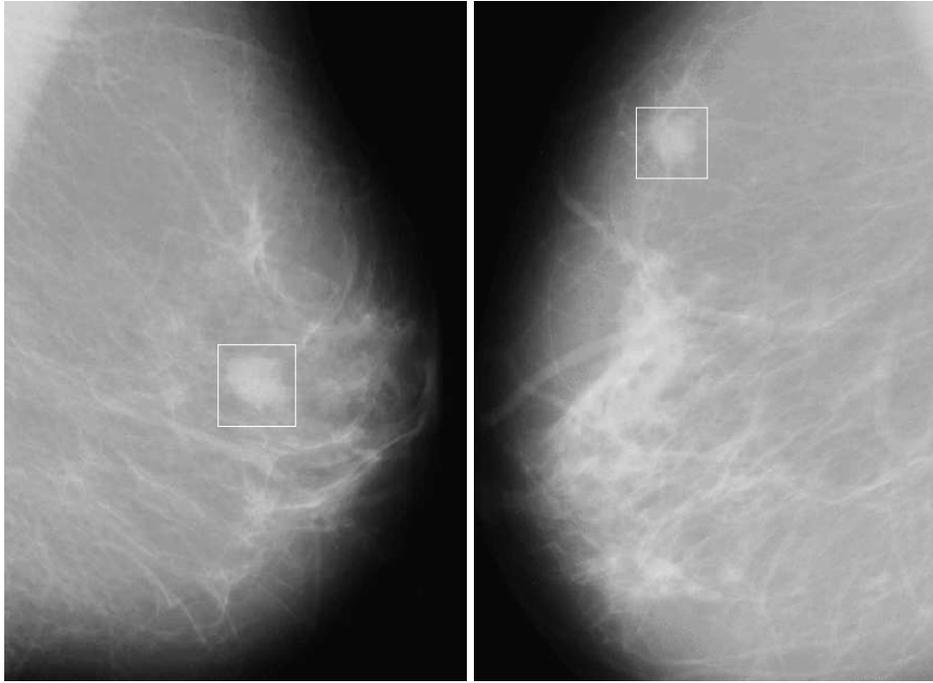


Figura B.7: Mamografías con tejido canceroso.

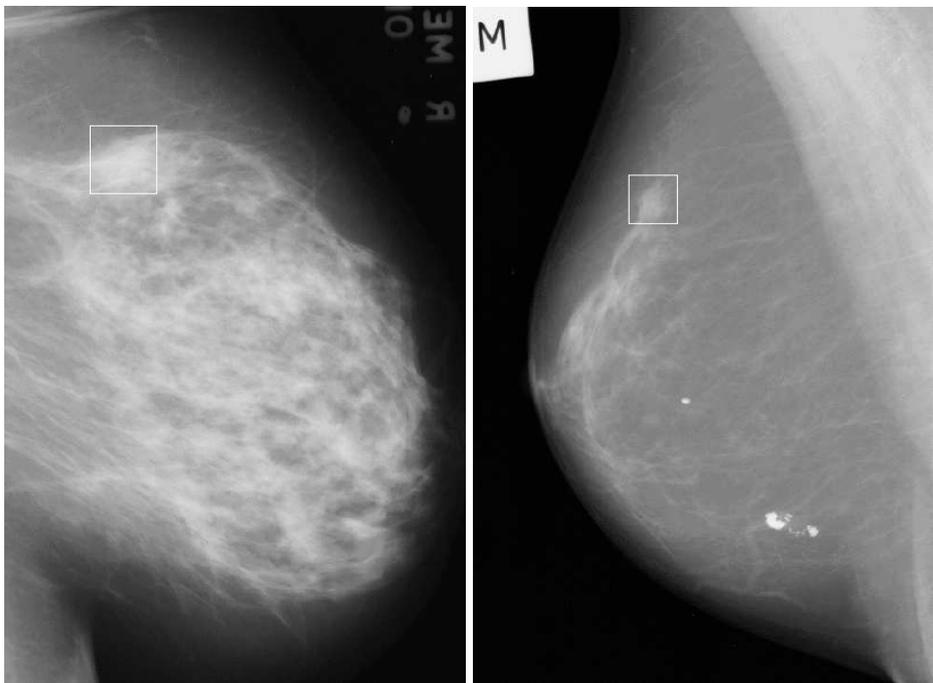


Figura B.8: Mamografías con tejido canceroso.

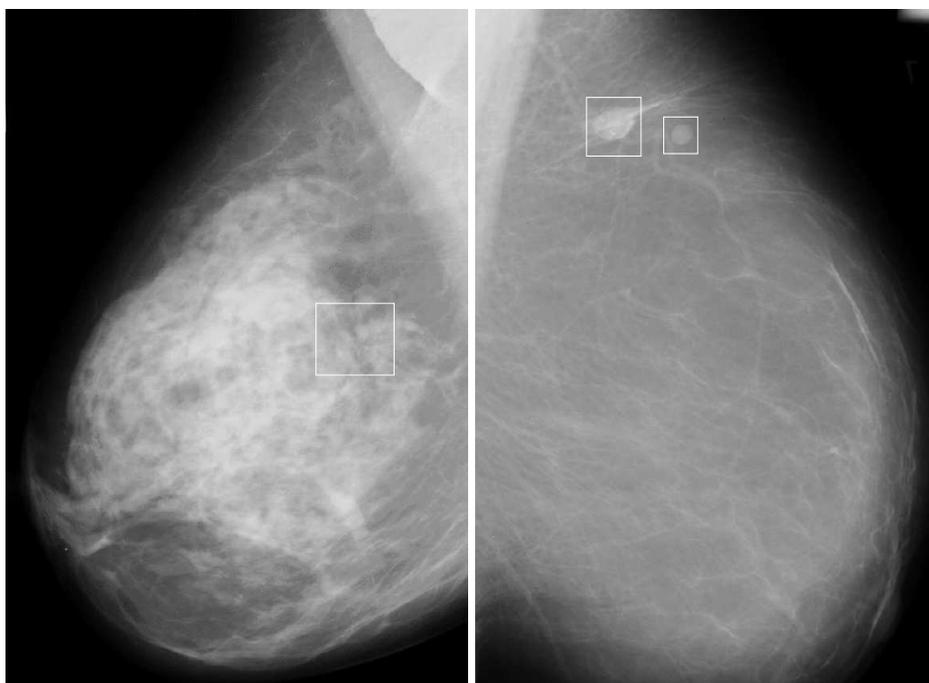


Figura B.9: Mamografías con tejido canceroso.

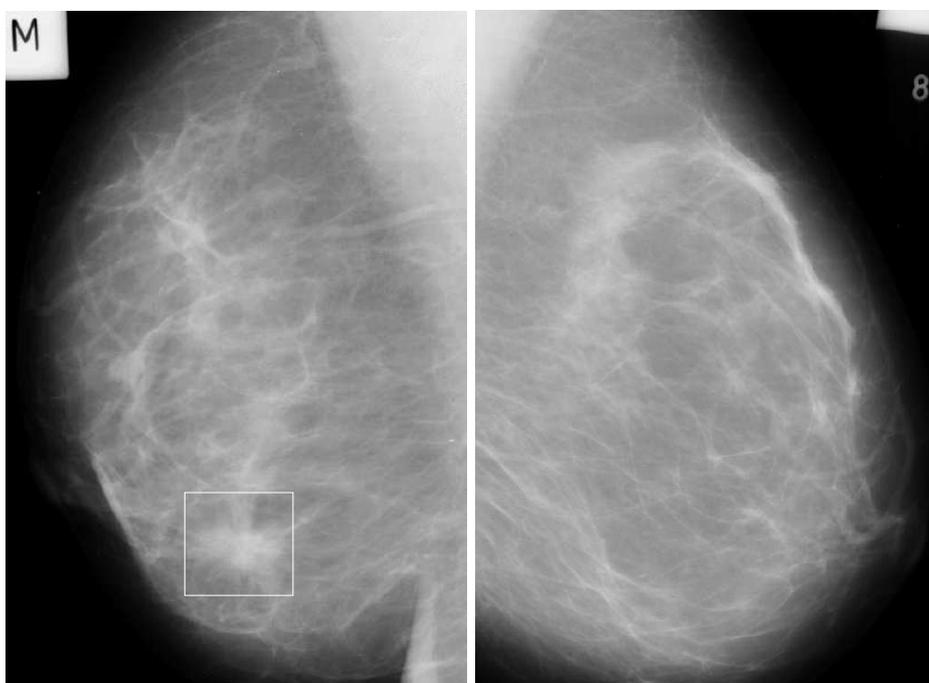


Figura B.10: Mamografía con tejido canceroso (izquierda) y sana (derecha).

Bibliografia

- [Albr 95] F. Albrechtsen. “Statistical Texture Measures Computed from Gray Level Cooccurrence Matrices”. November 1995.
- [Ande 90] J. Anderson. *Neurocomputing 2*. MIT Press, Cambridge, MA, 1990.
- [Arla 04] J. Arlandis. *La transformació contínua de la distància. Estudi i aplicació a un sistema OCR*. PhD thesis, Dep. Informàtica de Sistemes i Computadors, València, Spain, mar 2004. PhD thesis.
- [Arya 93] S. Arya and D. Mount. “Algorithms for Fast Vector Quantization”. In: *Proc. Data Compression Conference*, pp. 381–390, Snowbird, UT, USA, 1993.
- [Arya 98] S. Arya *et al.* “An optimal algorithm for approximate nearest neighbor searching”. *Journal of the ACM*, Vol. 45, pp. 891–923, 1998.
- [Bass 93] O. Basset, J. Mestas, *et al.* “Texture analysis of ultrasonic images of the prostate by means of cooccurrence matrices”. *Ultrasonic Imaging*, Vol. 15, pp. 218–237, 1993.
- [Bazz 99] I. Bazzi, R. Schwartz, and J. Makhoul. “An Omnifont Open-Vocabulary OCR System for English and Arabic”. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 21, No. 6, pp. 495–504, June 1999.
- [Beer 01] H. Beerlage, R. Aarnink, *et al.* “Correlation of transrectal ultrasound, computer analysis of transrectal ultrasound and histopathology of radical prostatectomy specimen”. *Prostate Cancer Prostatic Dis.*, Vol. 4, No. 1, pp. 56–62, 2001.
- [Bent 75] J. Bentley. “Multidimensional binary search trees used for associative searching”. *Communications of the ACM*, Vol. 18, pp. 509–517, 1975.
- [Bent 80] J. Bentley, B. Weide, and A. Yao. “Optimal Expected-Time Algorithms for Closest Point Problems”. *ACM Trans. on Mathematical Software*, Vol. 6, pp. 563–580, 1980.

- [Bird 05] R. Birdwell, P. Bandodkar, and D. Ikeda. "Computer-aided Detection with Screening Mammography in a University Hospital Setting". *Radiology*, Vol. 236, No. 2, pp. 451–457, 2005.
- [Bish 95] C. Bishop. *Neural networks for pattern recognition*. Oxford university Press, 1995.
- [Bovi 00] K. Bovis, S. Singh, J. Fieldsend, *et al.* "Identification of masses in digital mammograms with MLP and RBF nets". In: *Proc. of the IEEE-INNS-ENNS International Joint Conference on Neural Networks Com.*, pp. 342–347, 2000.
- [Brac 96] D. Braccialarghe and G. Kaufmann. "Contrast enhancement of mammographic features: a comparison of four methods". *Optical Engineering*, Vol. 35, No. 1, pp. 76–80, 1996.
- [Brak 00] G. Brake, N. Karssemeijer, *et al.* "An automatic method to discriminate malignant masses from normal tissue in digital mammograms". *Phys. Med. Biol.*, Vol. 45, No. 10, pp. 2843–2857, 2000.
- [Brak 98a] G. te Brake and N. Karssemeijer. "Automated detection of breast carcinomas that were not detected in a screening program". *Radiology*, Vol. 207, pp. 465–471, 1998.
- [Brak 98b] G. M. te Brake and N. Karssemeijer. "Comparison of three mass detection methods". In: *Digital Mammography*, pp. 119–126, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.
- [Brak 99] G. Brake and N. Karssemeijer. "Single and multiscale detection of masses in digitized mammograms". *IEEE Trans. Med. Imaging*, Vol. 18, No. 7, pp. 628–639, 1999.
- [Braw 99] M. Brawer and R. Kirby. *Fast Facts: Prostate Specific Antigen*. Health Press, 1999.
- [Bruc 00] L. M. Bruce and N. Shanmugam. "Using neural networks with wavelet transforms for an automated mammographic mass classifier". In: *Proc. of the 22nd Annual EMBS Int. Conference*, pp. 985–987, 2000.
- [Chan 90] H. Chan, K. Doi, *et al.* "Improvement in radiologists' detection of clustered microcalcifications on mammograms". *Investigative Radiology*, Vol. 25, pp. 1102–1110, 1990.
- [Chan 95] H. Chan, D. Wei, M. Helvie, *et al.* "Computer-aided classification of mammographic masses and normal tissue: linear discriminant analysis in texture feature space". *Phys. Med. Biol.*, Vol. 40, pp. 857–876, 1995.

- [Chen 05] H. Cheng, X. Shi, R. Min, *et al.* “Approaches for automated detection and classification of masses in mammograms”. *Pattern Recognition*, p. in press., 2005.
- [Chen 94] S. Cheng, H. Chan, *et al.* “Classification of mass and non-mass regions on mammograms using artificial neural networks”. *J. Imaging Sci. Technol.*, Vol. 38, No. 6, pp. 598–603, 1994.
- [Chri 00] I. Christoyianni, E. Dermatas, *et al.* “Fast detection of masses in computer-aided mammography”. *IEEE Signal Process.*, Vol. 17, No. 1, pp. 54–64, 2000.
- [Chri 02] I. Christoyianni, A. Koutras, *et al.* “Computer aided diagnosis of breast cancer in digitized mammograms”. *Computerized Medical Imaging and Graphics*, Vol. 26, pp. 309–319, 2002.
- [Comp 01] “Computer aided detection in mammography”. January 2001.
- [Cons 01] A. Constantinidis, M. Fairhurst, *et al.* “A new multi-expert decision combination algorithm and its application to the detection of circumscribed masses in digital mammograms”. *Pattern Recognition*, Vol. 34, No. 8, pp. 1527–1537, 2001.
- [Cons 99] A. Constantinidis, M. Fairhurst, F. Deravi, *et al.* “Evaluating Classification Strategies for Detection of Circumscribed Masses in Digital Mammograms”. In: *Proc. of 7th International Conference on Image Processing And Its Applications*, pp. 435–439, 1999.
- [Demp 77] A. Dempster, N. Laird, and D. Rubin. “Maximum Likelihood from Incomplete Data via the EM Algorithm”. *Journal of the Royal Statistical Society (B)*, Vol. 1, No. 39, pp. 1–38, 1977.
- [Devi 82] P. Devijver and J. Kittler. *Pattern recognition: a statistical approach*. Prentice Hall, 1982.
- [Devr 96] L. Devroye. *Automatic Pattern Recognition: A Study of the Probability of Error*. Springer-Verlag, 1996.
- [Djav 00] B. Djavan, A. Zlotta, *et al.* “Is one set of sextant biopsies enough to rule out prostate cancer? Influence of transition and total prostate volumes on prostate cancer yield”. *European Urology*, Vol. 38, No. 2, pp. 218–224, 2000.
- [Doi 96] K. Doi. “Computer-aided diagnosis: potential usefulness in diagnostic radiology and telemedicine”. In: *Proceedings of National Forum '95*, pp. 9–13, 1996.
- [Duda 73] R. Duda and P. Hart. *Pattern classification and scene analysis*. Wiley, 1973.

- [Ever 96] H. Evers. "Identification of prostate carcinoma in transrectal ultrasound images using cognitive texture parameters". In: *Proc. of ISCB-GMDS*, Heidelberg, 1996.
- [Fair 99] W. Fair, H. Tsai, *et al.* "Advanced 2D and 3D ultrasonic imaging in diagnosis of prostate cancer". In: *AUA TAM*, Dallas, 1999.
- [Fele 00] E. Feleppa, W. Fair, T. Liu, *et al.* "Three-dimensional ultrasound analyses of the prostate". *Mol Urol.*, Vol. 4, No. 3, pp. 133–139, 2000.
- [Fele 01] E. Feleppa, W. Fair, H. Tsai, *et al.* "Spectrum-analysis and neural networks for imaging to detect and treat prostate cancer". *Ultrason Imaging*, Vol. 23, No. 3, pp. 135–146, 2001.
- [Fele 99] E. Feleppa, W. Fair, H. Tsai, *et al.* "Progress in Two-Dimensional and Three-Dimensional Ultrasonic Tissue-Type Imaging of the Prostate Based on Spectrum Analysis and Nonlinear Classifiers". *Mol Urol.*, Vol. 3, No. 3, pp. 303–310, 1999.
- [Freu 77] Y. Freund and R. Schapire. "A decision-theoretic generalization of on-line learning and an application to boosting". *Journal of Computer and System Sciences*, Vol. 55, No. 1, pp. 119–139, 1977.
- [Freu 96] Y. Freund and R. Schapire. "Experiments with a new boosting algorithm". In: *Proc. International Conference on Machine Learning*, pp. 148–156, Morgan Kaufmann, San Francisco, 1996.
- [Frie 77] J. Friedman, F. Baskett, and L. Shustek. "An algorithm for finding best matches in logarithmic expected time". *ACM Trans. Math. Software*, Vol. 3, pp. 209–226, 1977.
- [Friz 03] M. Frize, C. Herry, *et al.* "Processing of Thermal Images to Detect Breast Cancer: Comparison with Previous Work". In: *IEEE-EMBS/BMES*, pp. 23–26, 2003.
- [Fu 82] K. Fu. *Syntactic pattern recognition and applications*. Prentice Hall, 1982.
- [Fuku 72] K. Fukunaga. *Introduction to statistical pattern recognition*. Academic Press, Boston, second edition, 1990. [First edition, 1972].
- [Gall 75] M. Galloway. "Texture classification using gray level run length". *Comput. Graphics Image Process.*, Vol. 4, pp. 172–179, 1975.
- [Gart 98] T. Gartner, M. Zacharias, K. J. KV, *et al.* "Equipment-independent ultrasound tissue characterization of testis and prostate". *Ultrasound Med Biol.*, Vol. 38, No. 5, pp. 424–433, 1998.

- [Gies 94] R. Giesen, A. Huynen, J. de la Rosette, *et al.* “The reliability of computer analysis of ultrasonographic prostate images: the influence of inconsistent histopathology”. *Ultrasound in Medicine and Biology*, Vol. 20, No. 9, pp. 871–876, 1994.
- [Gies 95] R. Giesen, A. Huynen, R. Aarnink, J. de la Rosette, *et al.* “Computer Analysis of Transrectal Ultrasound Images of Prostate for Detection of Carcinoma: Prospective Study in Radical Prostatectomy Specimens”. *The Journal of Urology*, Vol. 154, No. 4, pp. 1397–1400, 1995.
- [Gonz 99] J. Gonzalez. *Visión por computador*. ITP Paraninfo, 1999.
- [Hanl 82] J. Hanley and B. McNeil. “The meaning and use of the area under receiver operating characteristic (ROC) curve”. *Radiology*, Vol. 143, pp. 29–36, 1982.
- [Hara 73] R. M. Haralick *et al.* “Textural features for image classification”. *IEEE Trans. SMC*, Vol. 3, No. 6, pp. 610–621, 1973.
- [Hart 68] P. Hart. “The condensed nearest neighbor rule”. *IEEE Transactions on Information Theory*, Vol. 14, pp. 515–516, 1968.
- [Hast 01] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer-Verlag, 2001.
- [Head 99] J. Head, C. Lipari, *et al.* “Comparison of mammography and breast infrared imaging: sensitivity, specificity, false negatives, false positives, positive predictive value and negative predictive value”. In: *IEEE-EMBS/BMES*, p. 1116, 1999.
- [Heat 00a] M. Heath, K. Bowyer, and D. Kopans. “Computer-aided diagnosis of breast lesions in Medical images”. *Computing in Science and Engineering*, Vol. 2, pp. 39–45, 2000.
- [Heat 00b] M. D. Heath and K. W. Bowyer. “Mass Detection by Relative Image Intensity”. In: *Proceedings of the 5th International Workshop on Digital Mammography (IWDM-2000)*, pp. 219–225, Medical Physics Publishing, 2000.
- [Heat 98] M. Heath, K. Bowyer, D. Kopans, *et al.* “Current status of the Digital Database for Screening Mammography”. In: *Digital Mammography*, pp. 457–460, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.
- [Hsu] C.-W. Hsu, C.-C. Chang, and C.-J. Lin. “A Practical Guide to Support Vector Classification”.

- [Hutt 96] I. Hutt. *The Computer-aided Detection of of Abnormalities in Screening Mammograms*. PhD thesis, University of Manchester, Manchester, U.K., 1996. PhD thesis.
- [Huyn 94] A. Huynen, R. Giesen, *et al.* “Analysis of Ultrasonographic Prostate Images for the Detection of Prostatic Carcinoma: the Automated Urologic Diagnostic Expert System”. *Ultrasound in Medicine and Biology*, Vol. 20, No. 1, pp. 1–10, 1994.
- [Jain 00] A. Jain, R. Duin, and J. Mao. “Statistical pattern recognition: a review”. *IEEE Trans. Pattern Anal. Mach. Intel.*, Vol. 22, No. 1, pp. 4–37, 2000.
- [Jain 82] A. Jain and B. Chandrasekaran. *Dimensionality and sample size considerations in pattern recognition practice*. PR Krishniah and LN Kanal, 1982.
- [Jain 89] A. Jain. *Fundamentals of digital image processing*. Prentice Hall, Englewood Cliffs, NJ, 1989.
- [Jain 97] A. Jain and D. Zongker. “Feature selection: evaluation, application, and small sample performance”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 2, pp. 153–158, February 1997.
- [Jeli 98] F. Jelinek. *Statistical Methods for Speech Recognition*. MIT Press, 1998.
- [Joac 99] T. Joachims. *Making large-Scale SVM Learning Practical. Advances in Kernel Methods - Support Vector Learning*. MIT-Press, 1999.
- [Kall 92] M. Kallergi, K. Woods, L. Clarke, *et al.* “Image segmentation in digital mammography: comparison of local thresholding and region growing algorithms”. *Computer med. Imaging Graph.*, Vol. 16, No. 5, pp. 323–331, 1992.
- [Kars 96] N. Karssemeijer and G. Brake. “Detection of stellate distortions in mammograms”. *IEEE Trans. Med. Imaging*, Vol. 15, No. 5, pp. 611–619, 1996.
- [Kim 97] J. Kim, J. Park, K. Song, and H. Park. “Adaptive mammographic image enhancement using first derivative and local statistics”. *IEEE Transactions on Medical Imaging*, Vol. 16, No. 5, pp. 495–502, 1997.
- [Koho 95] T. Kohonen. “Self-organizing maps”. In: *Springer Series in Information Sciences*, Berlin, Heidelberg, 1995.

- [Krat 88] C. Kratzik, A. Hainz, *et al.* “Texture Analysis: A New Method of Differentiating Prostatic Carcinoma from Prostatic Hypertrophie”. *Urological Research*, Vol. 16, No. 5, pp. 395–397, 1988.
- [Krup 96] E. Krupinski. “Visual scanning patterns of radiologists searching mammograms”. *Academic Radiology*, Vol. 3, No. 2, pp. 137–144, 1996.
- [Kuo 02] W. Kuo, R. Chang, , W. Moon, *et al.* “Computer-aided Diagnosis of Breast Tumors with Different US Systems”. *Academic Radiology*, Vol. 9, No. 7, pp. 793–799, 2002.
- [Kupi 97] M. Kupinski and M. Giger. “Feature selection and classifiers for the computerized detection of mass lesions in digital mammography”. In: *IEEE International Conference on Neural Networks*, pp. 2460–2463, 1997.
- [Lain 94] A. Laine, S. Schuler, J. Fan, and W. Huda. “Mammographic feature enhancement by multiscale analysis”. *IEEE Trans. Med. Imaging*, Vol. 13, No. 4, pp. 725 –774, 1994.
- [Lain 95] A. Laine, J. Fan, and W. Yang. “Wavelets for contrast enhancement of digital mammography”. *IEEE Eng. Med. Biol.*, Vol. 14, No. 5, pp. 536–550, 1995.
- [Laye 90] G. Layer, I. Zuna, *et al.* “Computerized ultrasound B-scan texture analysis of experimental fatty liver disease: influence of total lipid content and fat deposit distribution”. *Ultrasonic Imaging*, Vol. 12, No. 3, pp. 171–188, 1990.
- [Laye 91] G. Layer, I. Zuna, *et al.* “Computerized ultrasound B-scan texture analysis of experimental diffuse parenchymal liver disease: correlation with histopathology and tissue composition”. *Journal of Clinical Ultrasound*, Vol. 19, No. 4, pp. 193–201, 1991.
- [Lee 89] K. Lee. *Automatic Speech Recognition: The Development of the SP-HINX System*. Kluwer Academic Publishers, 1989.
- [Li 99] L. Li, W. Qian, L. Clarke, *et al.* “Improving mass detection by adaptive and multi-scale processing in digitized mammograms”. In: *Proc. of SPIE - The International Society for Optical Engineering*, pp. 490–498, 1999.
- [Lizz 97] F. Lizzi, M. Astor, E. Feleppa, *et al.* “Statistical framework for ultrasonic spectral parameter imaging”. *Ultrasound Med Biol.*, Vol. 23, No. 9, pp. 1371–1382, 1997.

- [Llob 03] R. Llobet, A. Toselli, J. Perez-Cortes, and A. Juan. “Computer-aided Prostate Cancer Detection in Ultrasonographic Images”. In: *Proceedings of the 1st Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA)*, pp. 411–419, Puerto de Andratx (Mallorca, Spain), 2003.
- [Llob 04] R. Llobet and J. C. Perez-Cortes. “Breast Cancer Detection in Digitized Mammograms Using Non-Parametric Methods”. In: *Proceedings of the 2nd International Conference on Advances in Biomedical Signal and Information Processing (MEDSIP)*, pp. 281–287, Sliema, Malta, 2004.
- [Llob 05] R. Llobet, R. Paredes, and J. Perez-Cortes. “Comparison of Feature Extraction Methods for Breast Cancer Detection”. In: *Iberian Conference on Pattern Recognition and Image Analysis*, pp. 495–502, Springer-Verlag, Estoril (Portugal), June 2005.
- [Llob 06] R. Llobet, J. Perez-Cortes, A. Toselli, and A. Juan. “Computer-aided Detection of Prostate Cancer”. *International Journal of Medical Informatics*, p. in press., 2006.
- [Loch 00] T. Loch, I. Leuschner, *et al.* “Improvement of transrectal ultrasound. Artificial neural network analysis (ANNA) in detection and staging of prostatic carcinoma”. *Der Urologe. Ausg.*, Vol. 39, No. 4, pp. 341–347, 2000.
- [Loch 99] T. Loch, I. Leuschner, *et al.* “Artificial neural network analysis (ANNA) of prostatic transrectal ultrasound”. *The Prostate*, Vol. 39, No. 3, pp. 198–204, 1999.
- [Lowe 04] D. Lowe. “Distinctive Image Features from Scale-Invariant Keypoints”. *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91–110, 2004.
- [Lu 94] J. Lu, D. Healy, and J. Weaver. “Contrast enhancement of medical images using multiscale edge representation”. *Opt. Eng.*, Vol. 33, No. 7, pp. 2151–2161, 1994.
- [Mall 92] S. Mallat and S. Zhong. “Characterization of signals from multiscale edges”. *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 14, No. 7, pp. 710–732, 1992.
- [Mart 79] J. Martin, M. Moskowitz, *et al.* “Breast cancer missed by mammography”. *AJR*, Vol. 132, pp. 737–739, 1979.
- [Metz 84] C. Metz, P. Wang, *et al.* “A new approach for testing the significance of differences between ROC curves measured from correlated data”. In:

- Proc. of the Information Processing in Medical Imaging*, pp. 432–445, 1984.
- [Metz 96] C. Metz. “Evaluation of digital mammography by roc analysis”. *Excerpta Medica*, Vol. 1119, pp. 61–68, 1996.
- [Moor 91] A. Moore. *Efficient memory based learning for robot control*. PhD thesis, University of Cambridge, 1991. PhD thesis.
- [Mori 99] K. Morik, P. Brockhausen, and T. Joachims. “Combining statistical learning with a knowledge-based approach - A case study in intensive care monitoring”. In: *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 283–285, 1999.
- [Nam 98] S.Ñam and J.Y.Choi. “A method of image enhancement and fractal dimension for detection of microcalcifications in mammogram”. In: *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1009–1012, 1998.
- [Papp 92] T. Pappas. “An adaptive clustering algorithm for image segmentation”. *IEEE Trans. Signal Processing*, Vol. 40, No. 2, pp. 901–914, 1992.
- [Parz 62] E. Parzen. “On the estimation of a probability density function and mode”. *Annals of Mathematical Statistics*, Vol. 33, pp. 1065–1076, 1962.
- [Pere 00] J. Perez-Cortes, R. Llobet, and J. Arlandis. “Fast and Accurate Handwritten Character Recognition using Approximate Nearest Neighbours Search on Large Databases”. In: *Workshop on Statistical Pattern Recognition SPR-2000*, pp. 767–776, Alicante (Spain), 2000.
- [Pere 02] J. C. Perez-Cortes, A. Juan, and E. Vallada. “Textural Analysis Of Prostate Cancer In Transrectal Ultrasound Images”. In: *Proc. of Biosignal 2002*, pp. 283–285, Brno (Czech Republic), June 2002.
- [Perr 96] M. Perrotti, R. Kaufman, *et al.* “Endo-rectal coil magnetic resonance imaging in clinically localized prostate cancer: Is it accurate?”. *Journal of Urology*, Vol. 156, pp. 106–109, 1996.
- [Pisa 98] E. Pisano, S. Zong, B. Hemminger, *et al.* “Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms”. *J. Digital Imaging*, Vol. 11, No. 4, pp. 193–200, 1998.
- [Pres 96] J. Presti, H. Hricak, *et al.* “Local staging of prostatic carcinoma: comparison of transrectal sonography and endorectal MR imaging”. *American Journal of Roentgenology*, Vol. 166, pp. 103–108, 1996.

- [Rabi 89] L. Rabiner. "A Tutorial of Hidden Markov Models and Selected Applications in Speech Recognition". *Proc. IEEE*, Vol. 77, pp. 257–286, 1989.
- [Rabi 93] L. R. Rabiner and B. H. Juang. *Fundamentals of Speech Recognition*. Prentice-Hall, Englewood Cliffs, New Jersey, USA, 1993.
- [Raud 91] S. Raudys and A. Jain. "Small sample size effects in statistical pattern recognition: recommendations for practitioners". *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 13, No. 3, pp. 252–264, March 1991.
- [Raud 97] S. Raudys. "On Dimensionality, Sample Size, and Classification Error of Nonparametric Linear Classification Algorithms". *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 6, pp. 667–671, 1997.
- [Riss 95] T. Rissanen, H. Makarainen, *et al.* "Mammography and ultrasound in the diagnosis of contralateral breast cancer". *Acta Radiol.*, Vol. 36, pp. 358–366, 1995.
- [Robb 02] R. Robb. "Three-dimensional visualization and analysis in prostate cancer". *Drugs of Today*, Vol. 38, No. 3, pp. 153–165, 2002.
- [Roeh 98] J. Roehrig, T. Doi, *et al.* "Clinical results with R2 ImageChecker system". In: N. Karssemeijer, M. Thijssen, *et al.*, Eds., *Digital Mammography*, pp. 395–400, Kluwer Academic Publisher, Dordrecht, 1998.
- [Rose 95a] J. de la Rosette, R. Giesen, *et al.* "Automated analysis and interpretation of transrectal ultrasonography images in patients with prostatitis". *European Urology*, Vol. 27, No. 1, pp. 47–53, 1995.
- [Rose 95b] J. de la Rosette, A. Huynen, *et al.* "Computerized Analysis of Transrectal Ultrasonography Images in the Detection of Prostate Carcinoma". *British Journal of Urology*, Vol. 75, pp. 485–491, 1995.
- [Rowl 98] H. Rowley, S. Baluja, *et al.* "Neural Network-Based Face Detection". *IEEE Transactions on PAMI*, Vol. 20, No. 1, pp. 23–38, 1998.
- [Sahi 01] B. Sahiner, N. Petrick, H. Chang, *et al.* "Computer-Aided Characterization of Mammographic Masses: Accuracy of Mass Segmentation and Its Effects on Characterization". *IEEE Trans. on Medical Imaging*, Vol. 20, No. 12, pp. 1275–1284, 2001.
- [Sahi 98] B. Sahiner, H. Chan, N. Petrick, *et al.* "Design of a high-sensitivity classifier based on a genetic algorithm: application to computer-aided diagnosis". *Phys. Med. Biol.*, Vol. 43, No. 10, pp. 2853–2871, 1998.

- [Sall 99] M. Y. Sallam and K. W. Bowyer. “Registration and difference analysis of corresponding mammogram images”. *Medical Image Analysis*, Vol. 3, No. 5, pp. 103–118, 1999.
- [Scha 00] R. Schapire and Y. Singer. “BoosTexter: A boosting-based system for text categorization”. *Machine Learning*, Vol. 39, No. 2/3, pp. 135–168, 2000.
- [Scha 02] R. Schapire. “The Boosting Approach to Machine Learning. An Overview”. In: *MSRI Workshop on Nonlinear Estimation and Classification*, pp. 411–419, 2002.
- [Scha 90] R. Schapire. “The strength of weak learnability”. *Machine Learning*, Vol. 5, No. 2, pp. 197–227, 1990.
- [Skaa 97] P. Skaane, K. Engedal, and A. Skjennald. “Interobserver variation in the interpretation of breast imaging. Comparison of mammography, ultrasonography, and both combined in the interpretation of palpable noncalcified breast masses”. *Acta Radiol.*, Vol. 38, pp. 497–502, July 1997.
- [Soll 02] P. Sollich. “Bayesian Methods for Support Vector Machines: Evidence and Predictive Class Probabilities”. *Machine Learning*, Vol. 46, pp. 21–52, 2002.
- [Spro 91] R. Sproull. “Refinements to Nearest-Neighbor Searching in k-Dimensional Trees”. *Algorithmica*, Vol. 6, pp. 579–589, 1991.
- [Stew 01] C. Stewart, B. Leibovich, *et al.* “Prostate cancer diagnosis using a saturation needle biopsy technique after previous negative sextant biopsies”. *The Journal of Urology*, Vol. 166, pp. 86–91, 2001.
- [Ston 77] M. Stone. “Asymptotics for and against cross-validation”. *Biometrika*, Vol. 64, pp. 29–35, 1977.
- [Suck 94] J. Suckling, J. Parker, *et al.* “The mammographic images analysis society digital mammogram database”. In: *Excerpta Medica. International Congress Series*, pp. 375–378, 1994.
- [Swet 79] J. Swets. “ROC analysis applied to the evaluation of medical imaging techniques”. *Invest. Radiol.*, Vol. 14, No. 2, pp. 109–121, 1979.
- [Tao 05] Q. Tao, G.-W. Wu, *et al.* “Posterior Probability Support Vector Machines for Unbalanced Data”. *IEEE Trans. on Neural Networks*, Vol. 16, No. 6, pp. 1561–1573, 2005.
- [Trei 85] A. Treisman. “Pre-attentive processing in vision”. *Computer Vision, Graphics and Image Processing*, Vol. 31, pp. 156–177, 1985.

- [Vapn 95] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, New York, USA, 1995.
- [Vero 99] K. Veropoulos, N. Cristianini, and C. Campbell. “The Application of Support Vector Machines to Medical Decision Support: A Case Study”. In: *ACAI99*, 1999.
- [Vito 96] J. Viton, M. Rasigni, G. Rasigni, *et al.* “Method for characterizing masses in digital mammograms”. *Opt. Eng.*, Vol. 35, No. 12, pp. 3453–3459, 1996.
- [Wall 91] M. Wallis, M. Walsh, *et al.* “A review of false negative mammography in a symptomatic population”. *Clin Radiol*, Vol. 44, pp. 13–15, 1991.
- [Wata 85] S. Watanabe. *Pattern recognition: human and mechanical*. John Wiley, New York, USA, 1985.
- [Wesz 76] J. Weszka, C. Dyer, and A. Rosenfield. “A Comparative Study of Texture Measures for Terrain Classification”. *IEEE Trans. on Systems, Man and Cybernetics*, Vol. 6, No. 4, pp. 269–285, 1976.
- [Will 98] L. Williams, R. Prescott, *et al.* “Computer-aided cancer detection in the UK breast screening programme”. In: N. Karssemeijer, M. Thijssen, *et al.*, Eds., *Digital Mammography*, pp. 359–362, Kluwer Academic Publisher, Dordrecht, 1998.
- [Wils 98] M. Wilson, R. Hargrave, S. Mitra, *et al.* “Automated Detection of Microcalcifications in Mammograms through Application of Image Pixel Remapping and Statistical Filter”. In: *Proceedings of the IEEE Symposium on Computer-Based Medical Systems*, pp. 270–274, 1998.
- [Wong 98] K. Wongsritong, K. Kittayaruasiriwat, F. Cheevasuvit, *et al.* “Contrast enhancement using multipeak histogram equalization with brightness preserving”. In: *IEEE Asia-Pacific Conference on Circuits and Systems*, pp. 455–458, 1998.
- [Wood 91] K. Woods, L. Clarke, and R. Welthuizen. “Enhancement of digitized mammograms using a local thresholding technique”. In: *International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 114–115, 1991.
- [Wu 92] Y. Wu *et al.* “Computerized Detection of Clustered Microcalcifications in Digital Mammograms: Applications of Artificial Neural Networks”. *Medical Physics*, Vol. 19, pp. 555–560, 1992.
- [Yin 94] F. Yin, M. Giger, K. Doi, *et al.* “Computerized detection of masses in digital mammograms: investigation of feature-analysis techniques”. *J. Digital Imaging*, Vol. 7, pp. 18–26, 1994.

-
- [Youn 97] S. Young, J. Odell, D. Ollason, V. Valtchev, and P. Woodland. *The HTK Book: Hidden Markov Models Toolkit V2.1*. Cambridge Research Laboratory Ltd, March 1997.
- [Zaka 96] P. Zakarauskas and J. Ozard. “Refinements to Nearest-Neighbor Searching in k-Dimensional Trees”. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 18, No. 6, pp. 663–668, 1996.
- [Zhen 01] L. Zhen and A. Chan. “An artificial intelligent algorithm for tumor detection in screening mammogram”. *IEEE Trans. Med. Imaging*, Vol. 20, No. 7, pp. 559–567, 2001.

