



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



ESCUELA TÉCNICA
SUPERIOR INGENIEROS
INDUSTRIALES VALENCIA

TRABAJO FIN DE MASTER EN INGENIERÍA BIOMEDICA

DISEÑO Y DESARROLLO DE UN SISTEMA AUTOMÁTICO DE CLASIFICACIÓN DE ESTRUCTURAS GLANDULARES EN IMÁGENES HISTOLÓGICAS DE PRÓSTATA

AUTOR: José Gabriel García Pardo

TUTORA: Valery Naranjo Ornedo

COTUTOR: Adrián Colomer Granero

Curso Académico: 2017-18

AGRADECIMIENTOS

*A Valery, por darme la oportunidad de trabajar con ella,
por enseñarme y motivarme, por su amistad
y por su confianza depositada en mí.*

*A Adri, por ser el pilar fundamental de este proyecto,
por ser amigo antes que compañero
y por ser el referente a seguir.*

*A todo el grupo CVBLab, por su apoyo, por sus consejos
y por su disposición a ayudarme en todo momento.*

*A mis compañeros del fútbol, por encontrar en ellos
la ilusión y el sentimiento de un vestuario.*

*A mis amigos, por crecer conmigo
y por hacer grandes los pequeños momentos.*

*A Zaida, por creer siempre en mí, por cuidarme,
y por hacer que todos los lunes sean al sol.*

*A mi familia, por debérselo todo
y por ser siempre el lugar a donde quiero volver.*

RESUMEN

El presente TFM pretende abordar el desarrollo de un modelo de clasificación capaz de discriminar, de manera automática, las muestras histológicas de próstata de pacientes sanos con respecto a las de pacientes enfermos que presentan cáncer en una etapa temprana. El objetivo final es proporcionar un sistema de ayuda al diagnóstico que permita reducir la carga de trabajo, así como el nivel de subjetividad, de los médicos especialistas en anatomía patológica a la hora de analizar una determinada muestra biopsiada de la próstata.

En primer lugar, se aplican técnicas de *clustering* con la finalidad de encontrar elementos de interés a partir de los cuales identificar unidades de glándula. Para ello, se implementa un novedoso método de segmentación conocido como “*Locally Constrained Watershed Transform (LCWT)*”, el cual genera a su salida tres tipos de objetos: artefactos, glándulas benignas o glándulas patológicas. En este punto se aplican, por un lado, técnicas de *machine learning* clásico y, por otro lado, técnicas de *deep learning*, para comparar los resultados proporcionados por ambas aproximaciones. En cuanto a la primera, para determinar la clase de cada elemento segmentado, se lleva a cabo una profunda extracción de características haciendo uso de descriptores de forma y de textura, en combinación con características contextuales y un análisis fractal. Se obtienen 297 características que son sometidas a un exhaustivo análisis estadístico para seleccionar únicamente aquellas que presentan información relevante, en términos de correlación entre variables y capacidad discriminatoria con respecto a la clase. A partir de las muestras disponibles procedentes de 25 pacientes (3.200 artefactos, 3.195 glándulas benignas y 3.000 glándulas patológicas) se implementa una rigurosa estrategia de partición de datos para garantizar la robustez de los modelos construidos, así como la fiabilidad de los resultados obtenidos. Una vez llevada a cabo la partición de datos, se aborda el problema de clasificación desde una perspectiva multiclase aplicando, por un lado, clasificadores lineales y no lineales de *machine learning* clásico (*Support Vector Machine* y Perceptrón Multicapa) y, por otro lado, redes neuronales convolucionales, propias del *deep learning*, basadas en la arquitectura VGG19.

Los resultados obtenidos se comparan con los publicados por otros estudios del estado del arte con la intención de aportar nueva información de cara al cumplimiento del objetivo final, el cual reside en el desarrollo de un sistema de ayuda al diagnóstico precoz del cáncer de próstata.

Palabras clave: Cáncer de próstata, imagen histopatológica, *clustering*, segmentación, *constrained watershed*, extracción y selección de características, clasificación, SVM, redes neuronales, *machine learning*, *deep learning*.

RESUM

El present TFM pretén abordar el desenvolupament d'un model de classificació capaç de discriminar, de manera automàtica, les mostres histològiques de pròstata de pacients sans respecte de les de pacients malalts que presenten càncer en una etapa inicial. L'objectiu final és proporcionar un sistema d'ajuda al diagnòstic que permeti reduir la càrrega de treball, així com el nivell de subjectivitat, dels metges especialistes en anatomia patològica a l'hora d'analitzar una determinada mostra biopsiada de la pròstata.

En primer lloc, s'apliquen tècniques de *clustering* amb la finalitat de trobar elements d'interès a partir dels quals s'identifiquen unitats de glàndula. Per això, s'implementa un nou mètode de segmentació conegut com "*Locally Constrained Watershed Transform (LCWT)*", el qual genera a la seua eixida tres tipus d'objectes: artefactes, glàndules benignes o glàndules patològiques. En este punt s'apliquen, d'una banda, tècniques de *machine learning* clàssic i, d'altra banda, tècniques de *deep learning*, per a comparar els resultats proporcionats per ambdós aproximacions. Respecte a la primera, per a determinar la classe de cada objecte segmentat, es du a terme una profunda extracció de característiques fent ús de descriptors de forma i de textura, en combinació amb característiques contextuais i un anàlisi fractal. S'obtenen 297 característiques que són sotmeses a un exhaustiu anàlisi estadístic per a seleccionar únicament aquelles que presenten informació rellevant, en termes de correlació entre variables i capacitat discriminatòria respecte a la classe. A partir de les mostres disponibles procedents de 25 pacients (3.200 artefactes, 3.195 glàndules benignes i 3.000 glàndules patològiques) s'implementa una acurada estratègia de partició de dades per a garantir la robustesa dels models construïts, així com la fiabilitat dels resultats obtinguts. Una vegada realitzada la partició de dades, s'aborda el problema de classificació des d'una perspectiva multiclasse aplicant, d'una banda, tècniques lineals i no lineals de *machine learning* clàssic (*Support Vector Machine* i *Multilayer Perceptron*) i, d'altra banda, xarxes neuronals convolucionals, propi del *deep learning*, basades en l'arquitectura VGG19.

Els resultats obtinguts es comparen amb els publicats per altres estudis de l'estat de l'art amb la intenció d'aportar nova informació de cara al compliment de l'objectiu final, el qual resideix en el desenvolupament d'un sistema d'ajuda al diagnòstic precoç del càncer de pròstata.

Paraules clau: Càncer de pròstata, imatge histopatològica, *clustering*, segmentació, *constrained watershed*, extracció i selecció de característiques, classificació, SVM, xarxes neuronals, *machine learning*, *deep learning*.

ABSTRACT

This TFM pretends to address the development of a classification model capable of discriminating, automatically, the histological samples of prostate tissue from healthy patients regard to the samples of sick patients with cancer in an early stage. The final objective is to provide a computer aid-system that allows to reduce the workload, and the subjectivity level, when the pathologists analyse a certain sample of a prostate biopsy.

First, clustering techniques are applied with the aim of identifying elements of interest, from which to detect the gland unit. The strategy consists of a novel method known as “Locally Constrained Watershed Transform (LCWT)” that provides three types of objects: artefacts, healthy glands and pathological glands. In this point, classic machine learning and deep learning techniques are applied in order to compare the classification carried out by both strategies. Regarding the first one, a feature extraction stage is implemented in order to determine the class for each object. Several shape and textural descriptors are used in combination with contextual features and fractal analysis to achieve a total of 297 features. Later, a feature selection step is carried out to consider only the relevant information in terms of the correlation between variables and discriminatory capability concerning the class. Once the samples corresponding to 25 different patients are avoided (3.200 artefacts, 3.195 healthy glands and 3.000 pathological glands) a careful strategy of data partition is implemented to ensure the model’s robustness and the results reliability. Once de data is separated in different data sets, the classification problem is addressed by a multiclass approach applying, on the one hand, linear and non-linear machine learning techniques (Support Vector Machine and Multilayer Perceptron) and, on the other hand, convolutional neural networks based on the architecture known as “VGG19”, which corresponds to a deep learning methods.

The results obtained in this project are compared with other studies of the state of the art with the aim of providing new information in order to achieve the final objective, which consists of to develop a computer aid-system to early diagnosis of prostate cancer.

Keywords: Prostate cancer, histopathological image, clustering, segmentation, constrained watershed, extraction and selection of features, classification, SVM, neural networks, machine learning, deep learning.

ÍNDICE GENERAL

I. MEMORIA

II. PRESUPUESTO

I. MEMORIA

Índice de la Memoria

1. Introducción	1
1.1 Motivación y descripción del problema	2
1.1.1 Cáncer y cáncer de próstata.....	2
1.1.2 Imágenes histopatológicas.....	5
1.1.3 Sistema de clasificación de <i>Gleason</i>	8
1.1.4 Proyecto SICAP	9
1.2 Objetivos	10
1.3 Guía de la Memoria	11
2. Material y Métodos	13
2.1 Material	14
2.1.1 Base de datos de imágenes histopatológicas	14
2.1.2 Software utilizado	17
2.1.3 Hardware utilizado	17
2.2 Metodología	18
2.2.1 Preparación de la base de datos de imágenes.....	19
2.2.2 Pre-procesado (<i>Clustering</i>).....	21
2.2.3 Segmentación de candidatos a glándula.....	26
2.2.4 Extracción de características.....	31
2.2.5 Selección de características	43
2.2.6 Partición de datos	48
2.2.7 Clasificación supervisada.....	51
2.2.8 Creación del <i>groundtruth</i>	60
3. Resultados y discusión	67
3.1 Resultados de la segmentación	68
3.2 Resultados de la clasificación supervisada	71
3.2.1 Validación durante la etapa de Entrenamiento.....	71
3.2.2 Resultados durante la etapa de Test.....	75
3.3 Predicción y visualización	79
4. Conclusiones y líneas futuras	81
4.1 Conclusiones	82
4.2 Líneas futuras	84
Bibliografía	85

CAPÍTULO 1

1. Introducción

Índice de contenidos

1.1	Motivación y descripción del problema	2
1.1.1	Cáncer y cáncer de próstata.....	2
1.1.2	Imágenes histopatológicas.....	5
1.1.3	Sistema de clasificación de <i>Gleason</i>	8
1.1.4	Proyecto SICAP	9
1.2	Objetivos.....	10
1.3	Guía de la Memoria	11

1.1 Motivación y descripción del problema

1.1.1 Cáncer y cáncer de próstata

Cáncer

La Organización Mundial de la Salud (OMS) define el cáncer como un proceso de crecimiento de células incontrolado que puede afectar prácticamente a cualquier parte del organismo [1]. El cáncer aparece como resultado de una división anormal de células mutadas que no pueden repararse ni morir [2]. Por ello, la enfermedad del cáncer se constituye como una de las principales causas de *morbilidad*¹ y mortalidad en el mundo, y se prevé que el número de casos nuevos aumente aproximadamente en un 70% durante las próximas décadas [3].

Uno de los consejos más recomendables para prevenir el cáncer es evitar la exposición a factores de riesgo como pueden ser: la falta de actividad física, el consumo de tabaco y el consumo de alcohol. Lógicamente, esto no asegura el hecho de no padecer cáncer, ya que también deben tenerse en cuenta factores no controlables como la historia familiar, la raza y la edad. Cabe destacar que el cáncer es susceptible de tratamiento bien mediante técnicas de radioterapia o de quimioterapia, o bien empleando cirugía, principalmente si la enfermedad se detecta en un estado temprano. El problema subyace en que, a menudo, el cáncer se diagnostica en fases avanzadas, lo cual conlleva un tratamiento tardío que se asocia con un aumento de la mortalidad y un empeoramiento de la calidad de vida. Por ello, este trabajo pretende abordar la detección temprana de uno de los tipos de cáncer más frecuentes en la sociedad actual como es el cáncer de próstata. El principal objetivo es desarrollar un sistema de ayuda al diagnóstico con el que, a la larga, se haga factible la reducción de la mortalidad.

A continuación, se explica en mayor profundidad qué es el cáncer de próstata y la importancia de proporcionar un diagnóstico precoz. Para ello, se muestran diferentes estadísticas relativas a la *incidencia*², la *prevalencia*³ y la mortalidad, así como otros factores relacionados con el procedimiento diagnóstico y sus correspondientes tratamientos.

Cáncer de próstata

Hoy en día, el cáncer de próstata es uno de los tipos de cáncer más frecuentemente diagnosticados en el mundo. Además, la incidencia estimada de dicho cáncer en España fue de 30.076 nuevos casos en el año 2017, según reveló el estudio publicado por la Sociedad Española de Oncología Médica (SEOM) en enero de 2018 [3]. De este modo, el cáncer de próstata se sitúa como el segundo tipo de cáncer más frecuente, después del colorrectal, en lo que respecta a ambos sexos (véase la *Figura 1.1.a*). No obstante, se convierte en el más común cuando se analiza únicamente el sexo masculino, como se puede observar en la *Figura 1.1.b*.

¹ *Morbilidad*. Cantidad de personas que enferman en un lugar y en un período de tiempo determinados en relación con el total de la población.

² *Incidencia*. Medida de morbilidad relacionada con el número de casos nuevos de una enfermedad en una población en un periodo de tiempo determinado.

³ *Prevalencia*. Medida de frecuencia de una enfermedad que describe la proporción de la población que padece dicha enfermedad en un momento determinado.

1.1. Motivación y descripción del problema

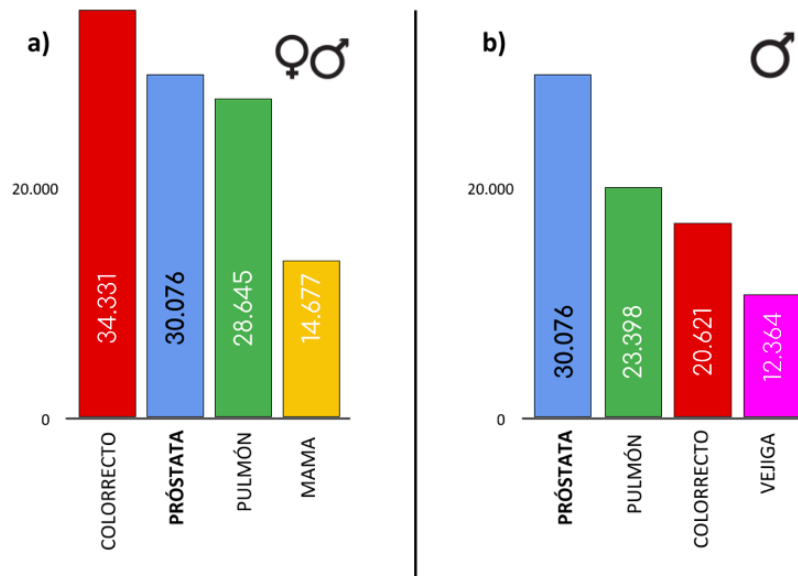


Figura 1.1. Incidencia estimada de los tumores más frecuentes en España en el año 2017. a) En ambos sexos. b) En varones [3].

Además de por la elevada incidencia que presenta el cáncer de próstata, la motivación de este proyecto reside fundamentalmente en la prevalencia de dicha enfermedad. Esta prevalencia viene determinada por la cantidad de años que un paciente puede sobrevivir con cáncer de próstata y, dado que la *tasa de supervivencia a 5 años*⁴ es del 99% [4], la prevalencia es muy alta (véase la *Figura 1.2.a*). De manera inversa, la tasa de mortalidad asociada al cáncer de próstata se presenta reducida con respecto a otros tipos como el cáncer de pulmón. No obstante, el cáncer de próstata seguía ocupando el tercer puesto en la tasa de mortalidad en varones españoles para el año 2016, como se muestra en la *Figura 1.2.b*.

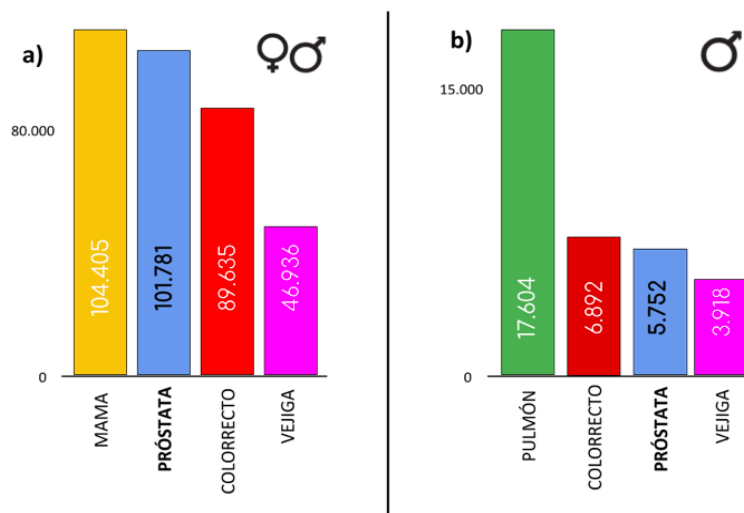


Figura 1.2. a) Estimación de la prevalencia a 5 años de tumores en España en ambos sexos para el año 2017. b) Tasa de mortalidad por tumores en varones en España para el año 2016.

⁴ *Tasa de supervivencia a 5 años.* Porcentaje de personas que viven al menos 5 años después de ser diagnosticadas con cáncer.

En base a estos gráficos, se hace evidente la necesidad de incorporar elementos preventivos y de diagnóstico precoz, a fin de afrontar una enfermedad con una incidencia y una prevalencia tan elevadas. Además, actualmente el 92% de los casos se detecta cuando la enfermedad está ubicada en la glándula prostática o en los órganos adyacentes, es decir, cuando la enfermedad se encuentra localmente avanzada [4].

Hoy en día, la prevención pasa principalmente por adquirir hábitos de vida saludables, aunque no eviten por completo el hecho de contraer la enfermedad. En cuanto al protocolo diagnóstico, que es en lo que realmente se basa este trabajo, existe una serie de pasos para determinar si el paciente presenta cáncer o no. En la *Figura 1.3* se puede observar el esquema del procedimiento común a implementar desde que se sospecha sobre un posible cáncer de próstata hasta que se emite el diagnóstico definitivo.

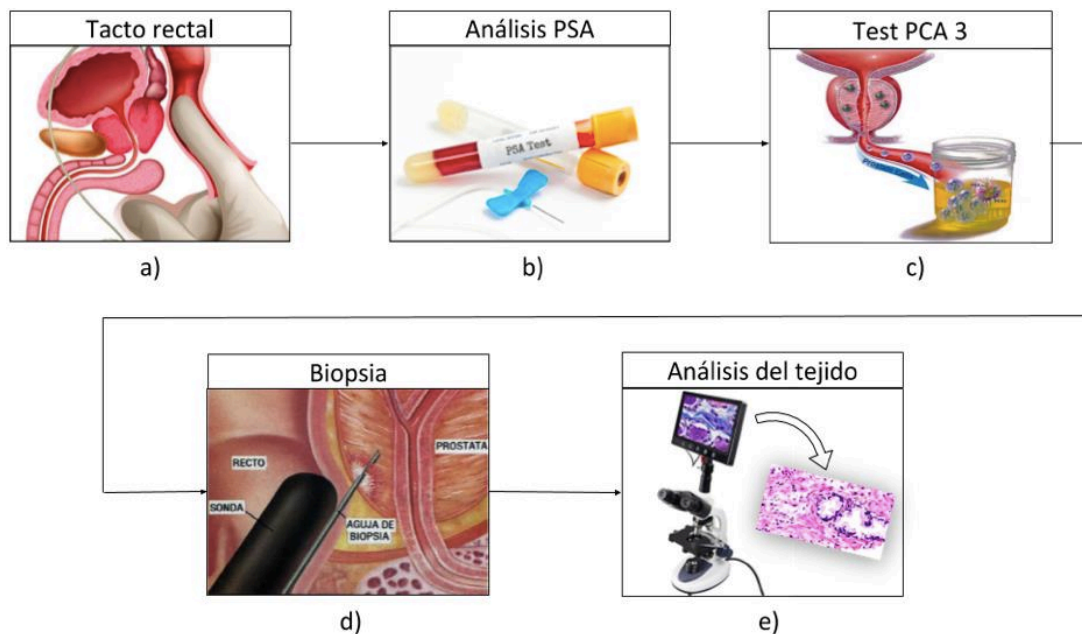


Figura 1.3. Procedimiento diagnóstico común para el cáncer de próstata [5]–[8].

El tacto rectal (*Figura 1.3.a*) consiste en un examen en el que el médico introduce un dedo en el recto para palpar la pared de la próstata en busca de nódulos o áreas anormales [9]. En caso de encontrar indicios de cáncer, el paciente es sometido a una prueba de laboratorio en la que se determinan sus niveles de PSA (*Prostate-Specific Antigen*) en sangre (*Figura 1.3.b*). El PSA es una sustancia producida específicamente por la próstata que actúa como marcador, ya que existe una relación directamente proporcional entre las concentraciones en sangre de dicha sustancia y los casos de pacientes con cáncer de próstata. No obstante, el análisis de PSA se caracteriza por ser una prueba altamente inespecífica, puesto que pueden darse anomalías no asociadas al cáncer de próstata que también generan niveles elevados de PSA en sangre, como por ejemplo un tamaño excesivamente grande de la glándula prostática. Por esta razón, surge también otra prueba conocida como PCA3 (*Figura 1.3.c*) cuyas siglas en inglés se refieren a “*Prostate-Cancer Antigen*”. Este test consiste en un análisis genético que, a diferencia del PSA, es específico del cáncer de próstata, por lo que discrimina mejor entre enfermedades prostáticas no cancerosas y el propio cáncer de próstata [10].

1.1. Motivación y descripción del problema

Las tres técnicas mencionadas anteriormente se emplean con la intención de emitir un diagnóstico que no conlleve la necesidad de realizar una *biopsia*⁵ (*Figura 1.3.d*), ya que es una técnica invasiva que puede derivar en diferentes complicaciones como reacciones alérgicas, infecciones, hemorragias, etc. [11]. Sin embargo, la biopsia se establece como el análisis definitivo y determinante para diagnosticar el cáncer de próstata, ya que el análisis bajo el microscopio (o por imagen digital) de los tejidos biopsiados (*Figura 1.3.e*) es lo que permite al *patólogo*⁶ determinar con seguridad si el paciente padece, o no, cáncer de próstata.

Por este motivo, el presente Trabajo Fin de Máster (TFM) está orientado a desarrollar un sistema de ayuda al diagnóstico, a partir de las muestras obtenidas tras realizar una biopsia. Con ello se pretende principalmente afrontar dos problemas: el nivel de subjetividad entre diferentes patólogos a la hora de evaluar una determinada muestra y la relación coste-efectividad en lo que se refiere al tiempo necesario para llevar a cabo el análisis.

1.1.2 Imágenes histopatológicas

Como se ha detallado en el párrafo anterior, el diagnóstico definitivo para el cáncer de próstata reside en el estudio de las muestras del tejido extraído en una biopsia. Dichas muestras, tras un proceso de preparación, se colocan en un portaobjetos (véase la *Figura 1.4*) para obtener las imágenes histopatológicas que posteriormente analizan los expertos en *histología*⁷, a fin de determinar la presencia o ausencia de cáncer en el paciente.

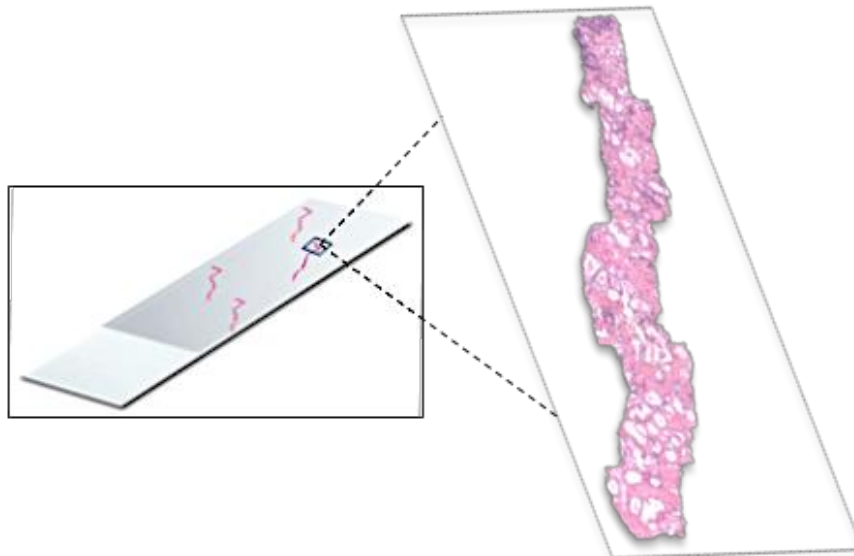


Figura 1.4. Portaobjetos con muestras de tejido prostático biopsiado con un zoom para observar la imagen histopatológica a mayor nivel de detalle.

⁵ *Biopsia*. Extirpación de una pequeña porción de tejido del cuerpo humano para su posterior análisis.

⁶ *Patólogo*. Médico especialista en anatomía patológica cuya función principal es detectar enfermedades mediante el estudio de células o tejidos bajo un microscopio.

⁷ *Histología*. Ciencia biológica que estudia la composición, la estructura y las características de los tejidos orgánicos de los seres vivos.

El *proceso histológico*⁸ más común desde que se realiza la biopsia de la próstata (*Figura 1.3.d*) hasta que se analizan en el microscopio las imágenes histopatológicas (*Figura 1.3.e*) es el que se presenta en la *Figura 1.5*. En primer lugar, se lleva a cabo la etapa de fijación mediante perfusión, cuyo objetivo es conservar las estructuras celulares y moleculares del tejido. A continuación, las muestras son embebidas en parafina completando así la fase de inclusión. Durante dicha fase, se endurecen las muestras con el fin de obtener secciones de tejido más delgadas. Una vez se han endurecido las muestras se lleva a cabo el corte de los tejidos, normalmente mediante un micrótopo, para conseguir secciones del orden de micrómetros. Posteriormente, las muestras son teñidas típicamente con un pigmento conocido como Hematoxilina y Eosina (H&E) que proporciona a las imágenes histológicas de la próstata su color característico [12]. De esta manera, es posible diferenciar los cuatro elementos básicos que se encuentran en este tipo de imágenes (véase la *Figura 1.7*).

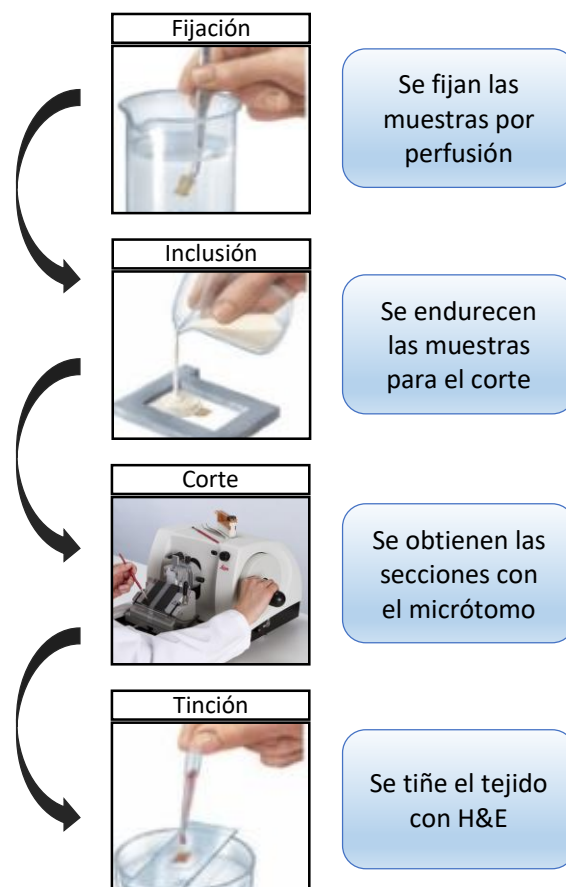


Figura 1.5. Proceso histológico que comprende los pasos inmediatamente posterior a la biopsia e inmediatamente anterior a la visualización del tejido bajo el microscopio [13].

Finalmente, las imágenes histopatológicas como las que se observan en la *Figura 1.6* pueden ser analizadas por los patólogos, bien con un microscopio óptico, o bien mediante un sistema digital. Este último es más cómodo, pero requiere el escaneado previo de las muestras.

⁸ *Proceso histológico.* Conjunto de técnicas utilizadas para la preparación de las muestras tisulares a fin de estudiar sus características morfológicas y estructurales.

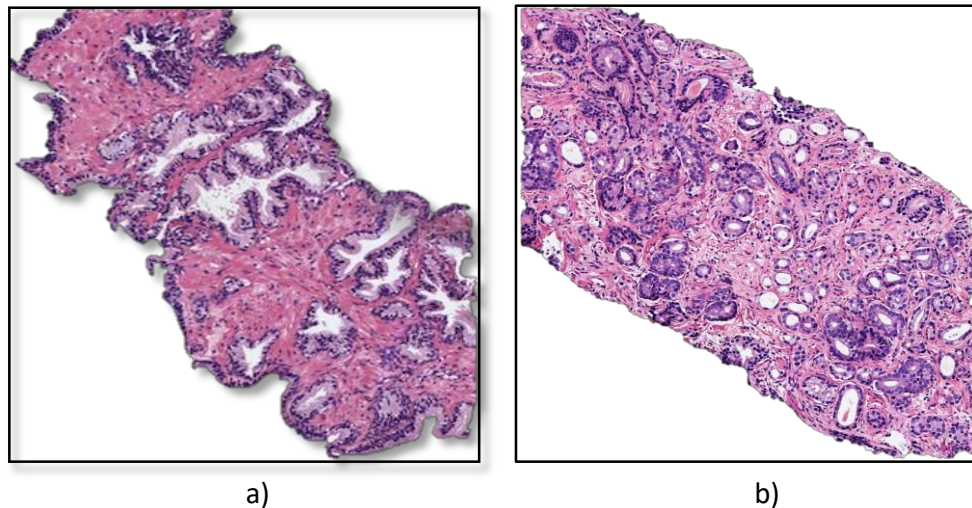


Figura 1.6. Ejemplo de dos imágenes histológicas de próstata. a) Muestra de tejido benigno. b) Muestra de tejido patológico (canceroso).

Como ya se ha comentado anteriormente, la tinción de hematoxilina y eosina proporciona, en las imágenes histológicas de la próstata, cuatro colores correspondientes a los cuatro elementos principales que constituyen la base para la detección de regiones de interés. Estos cuatro elementos son: los núcleos (elementos azules, más oscuros), el citoplasma (regiones púrpuras, más claras), el estroma (áreas rosas) y los lúmenes (zonas blancas rodeadas de citoplasma y núcleos) [14].

En este TFM se trabaja con imágenes cuyas regiones de interés se conocen como unidades glandulares o glándulas unitarias (*Figura 1.7*). La hipótesis de partida es que distintos patrones de unidades glandulares se asocian con distintos estadios de la enfermedad. Por esa razón, el objetivo de este proyecto se basa en: primero detectar y segmentar las regiones glandulares y, después, extraer características de dichas regiones para determinar el estadio de la enfermedad.

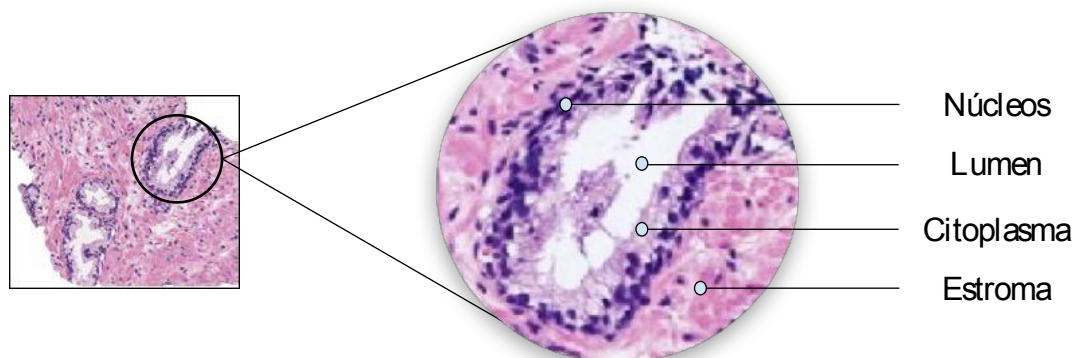


Figura 1.7. Región de interés correspondiente a una unidad glandular de una muestra de tejido benigno de próstata donde se resaltan sus cuatro elementos principales.

El diagnóstico definitivo se expresa de acuerdo al sistema de clasificación de *Gleason* detallado en el siguiente apartado. Dicho sistema, a grandes rasgos, permite agrupar las muestras en diferentes fases de la enfermedad según los patrones identificados en la imagen.

1.1.3 Sistema de clasificación de *Gleason*

El sistema de clasificación de *Gleason* constituye la herramienta fundamental para el diagnóstico en lo que se refiere al estudio de la próstata mediante los cortes histológicos obtenidos tras la biopsia. Dicho sistema permite llevar a cabo una comparación de las imágenes histopatológicas con una plantilla en la que se detallan distintos patrones tisulares (*Figura 1.8*) a fin de asignar una puntuación de *Gleason* que determine la gravedad de la enfermedad en una muestra bajo estudio.

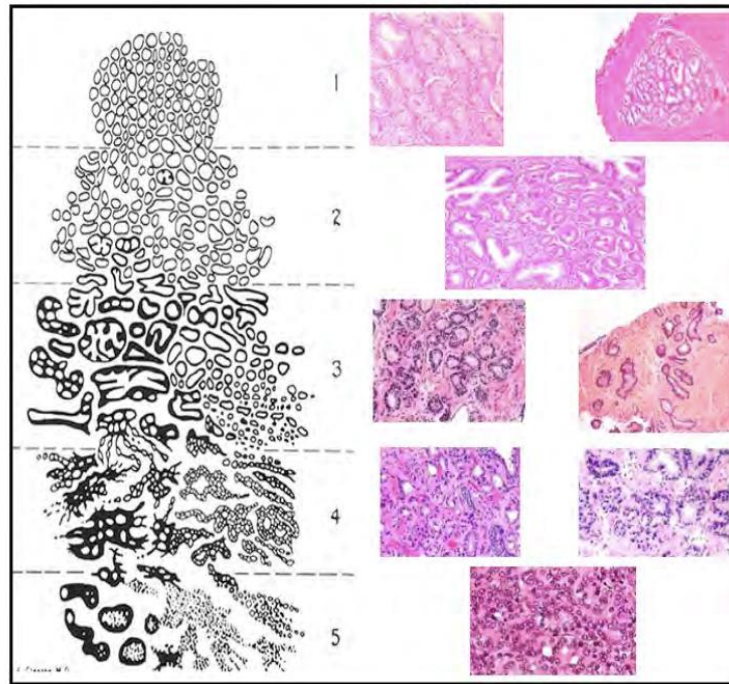


Figura 1.8. Modelo de clasificación de *Gleason* para el diagnóstico del cáncer de próstata [14].

La escala de puntuación de *Gleason* se hace imprescindible en el diagnóstico de una muestra biopsiada de la próstata. El procedimiento a seguir se compone de tres pasos fundamentales que se detallan a continuación:

- i) **Examen visual de la muestra bajo el microscopio óptico.** En primer lugar, los patólogos analizan las imágenes histopatológicas con un aumento pequeño (1x o 2x) con el objetivo de localizar regiones de interés. Después incrementan el zoom, a 5x o a 10x, para analizar la estructura de las glándulas y su distribución. Posteriormente, si con ese aumento no se distinguen evidencias claras de anomalías tisulares, se incrementa de nuevo el objetivo (a 20x o a 40x) para examinar la información citológica presente en el tejido.
- ii) **Asignación de una puntuación según la escala *Gleason*.** El método de clasificación de *Gleason* define cinco grados correspondientes a cinco patrones diferentes, como se muestra en la *Figura 1.8*. Los grados 1 y 2 hacen referencia a tejidos benignos, mientras que los grados 3, 4 y 5 corresponden a tejidos patológicos (ordenados de forma ascendente según la gravedad del cáncer). En base a ello, el patólogo asigna una puntuación a las imágenes histopatológicas según los patrones identificados.

- iii) **Suma de los dos grados mayoritarios.** Las muestras histológicas de la próstata suelen presentar simultáneamente grados diferentes, por lo que la puntuación final de *Gleason* se determina con la suma de los dos grados mayoritarios de la imagen [15]. Por ejemplo, si dichos grados son 3 y 4, el diagnóstico se expresa como 3+4 o 4+3. El orden de los factores es importante, ya que el primero indica el grado predominante de la imagen; y el segundo, el siguiente grado mayoritario. En función de ello, la severidad del cáncer varía y, por consiguiente, el tratamiento también.

Aunque con el método de la biopsia solo es posible analizar una parte del total de la próstata, el sistema de clasificación de Gleason se constituye como uno de los factores más importantes en el pronóstico, pues ayuda a los patólogos a determinar el tratamiento más apropiado. Además, se recomienda que las muestras sean examinadas por más de un patólogo para mejorar la precisión del diagnóstico. No obstante, esto no se lleva a cabo en la práctica, puesto que se trata de una tarea manual muy tediosa que supone un gran coste en términos de tiempo. Este hecho es uno de los principales motivos que dieron lugar al proyecto SICAP (Sistema de interpretación de Imágenes histopatológicas para la detección del Cáncer de Próstata). En la siguiente sección se detallan los objetivos de este proyecto, ya que es la base sobre la que se sustenta el presente trabajo final de máster.

1.1.4 Proyecto SICAP

SICAP es un proyecto nacional subvencionado por el Ministerio de Economía, Industria y Competitividad [DPI2016-77869-C2-1-R]. Actualmente, el grupo CVBLab (*Computer Vision and Behaviour Analysis Lab*), perteneciente al I3B (Instituto de Investigación e Innovación en Bioingeniería) de la Universitat Politècnica de València (UPV) colabora en este proyecto junto con la Universidad de Granada y el Servicio de Anatomía Patológica del Hospital Clínico Universitario de Valencia.

En el proyecto SICAP trabajan conjuntamente médicos, ingenieros y matemáticos como un equipo coordinado y multidisciplinar cuya finalidad es la de proporcionar un software (con vistas a la integración en la práctica clínica) para el diagnóstico automático del cáncer de próstata. Lo que se pretende con SICAP es trabajar en la segmentación automática y la extracción de características (ayudándose de la experiencia de los patólogos) para lograr clasificar automáticamente las muestras histopatológicas con la mayor precisión posible. Para ello, se implementan novedosas técnicas de segmentación, robustos métodos de extracción y selección de características y estrategias de *machine learning*⁹ y *deep learning*¹⁰ para la clasificación.

Con el proyecto SICAP se aspira a disminuir la carga de trabajo de los patólogos y a mejorar en términos de coste-efectividad, ya que el análisis manual bajo el microscopio es una tarea muy tediosa que conlleva mucho tiempo [14]. Además, también se pretende reducir el nivel de subjetividad existente entre diferentes expertos cuando intentan analizar una determinada muestra, pues es frecuente encontrar ciertas discordancias en su interpretación [16]–[18].

⁹ *Machine Learning*. Rama de la inteligencia artificial basada en el estudio y en la computación de modelos de aprendizaje automáticos en múltiples situaciones [79].

¹⁰ *Deep Learning*. Conjunto de algoritmos de aprendizaje automático que utiliza modelos computacionales compuestos por múltiples capas de procesamiento para aprender representaciones de datos con diferentes niveles de abstracción [80].

Por último, cabe reseñar que este TFM solo abarca una fase intermedia del proyecto general, ya que se centra únicamente en la distinción de glándulas benignas y patológicas de grado 3, sin tener en cuenta los grados 4 y 5. No obstante, SICAP solo ha alcanzado la mitad de su duración, por lo que las perspectivas de futuro son prometedoras, tanto para cumplir los próximos objetivos, como para mejorar los resultados de los ya conseguidos.

1.2 Objetivos

Como se ha explicado en el apartado anterior, el objetivo final del proyecto SICAP consiste en proporcionar un sistema de ayuda al diagnóstico que pueda ser integrado en hospitales, a fin de facilitar y mejorar el trabajo desempeñado por los patólogos a la hora de emitir un diagnóstico sobre el cáncer de próstata. Por lo que respecta al objetivo principal de este trabajo fin de máster, se pretende desarrollar un algoritmo que sea capaz de discriminar entre muestras sanas y patológicas de grado 3, atendiendo a las estructuras glandulares del tejido y de acuerdo a la escala *Gleason*. Este objetivo viene marcado porque, como se observa en la *Figura 1.8*, las glándulas benignas y las patológicas de grado 3 presentan características estructurales muy similares, pues ambas están bien definidas y se pueden observar de forma individual. Sin embargo, las imágenes correspondientes a los grados 4 y 5 se caracterizan, bien por presentar glándulas pobremente definidas y fusionadas, o bien por la ausencia de las mismas [14].

Para lograr la consecución del objetivo principal de este TFM se debe abordar una serie de objetivos específicos que se listan a continuación:

1. Revisar la literatura científica actual para conocer el estado del arte relativo a las técnicas de segmentación y clasificación implementadas en imágenes histopatológicas, principalmente de próstata y de colon.
2. Generar una base de datos de imágenes cuyas dimensiones sean adecuadas para poder trabajar sobre ellas, ya que las originales tienen un tamaño excesivamente grande. Para ello debe hacerse uso de la tecnología *Whole Slide Image (WSI)* que permite analizar imágenes digitales de alta resolución de muestras completas de biopsias.
3. Crear un *ground truth*¹¹ que pueda usarse tanto para entrenar los clasificadores durante la etapa del *training*, como para validar los modelos durante la fase de *test* en la clasificación. Previamente se debe generar también un *ground truth* para validar la técnica de segmentación a partir de la cual efectuar la clasificación.
4. Aplicar técnicas de clasificación no supervisadas para agrupar en diferentes clases las distintas estructuras que se revelan en las imágenes histológicas con la tinción de H&E.
5. Llevar a cabo una partición de datos que garantice la robustez del modelo al no testear con las mismas imágenes con las que se ha entrenado. Dicha partición no debe ser meramente aleatoria, sino que se ha de profundizar para que las muestras correspondientes a un paciente concreto pertenezcan siempre a un único conjunto de datos. Así, las imágenes relativas a un paciente solo se usarán como entrenamiento o como test, nunca como ambos, lo cual asegura la fiabilidad del método.

¹¹ *Ground truth*. Término utilizado en el ámbito del *machine learning* para hacer referencia a modelos contruidos manualmente que luego se comparan con los creados de forma automática para aceptar o rechazar una determinada hipótesis.

6. Realizar una exhaustiva extracción de características con la finalidad de encontrar patrones específicos de los diferentes tipos de tejido con los que se trabaja.
7. Implementar un minucioso análisis estadístico que permita la selección de las mejores características en términos de independencia entre clases y entre variables.
8. Abordar la tarea de clasificación tanto desde una perspectiva lineal como desde una no lineal. Asimismo, deben aplicarse diferentes metodologías basadas en aprendizaje automático y en aprendizaje profundo a fin de obtener los mejores resultados posibles.
9. Efectuar una comparación entre los diferentes métodos de clasificación propuestos en este trabajo y con respecto a los propuestos por otros autores del estado del arte actual.
10. Identificar y describir los problemas y limitaciones encontrados durante el estudio y proponer posibles mejoras para investigaciones futuras, teniendo en mente los próximos objetivos del proyecto SICAP.

1.3 Guía de la Memoria

En el capítulo 2 se hace referencia al material y a la metodología necesarios para llevar a cabo la clasificación de las muestras en sus respectivos grados, a partir de la segmentación automática de las regiones de interés constituidas por las glándulas unitarias. Además, se resume la información más importante relacionada con el estado del arte en los apartados relativos a métodos de algorítmica.

En el capítulo 3 se exponen, se discuten y se comparan con el estado del arte los resultados obtenidos con las diferentes técnicas propuestas para la segmentación, la extracción de características y la clasificación supervisada. También se presentan resultados tanto de la validación durante la fase de entrenamiento, como de test durante la fase de predicción. Además, se aportan resultados cualitativos y cuantitativos del análisis de nuevas imágenes totalmente desconocidas por el sistema, tal y como se llevaría a cabo para nuevas muestras en la práctica clínica.

Finalmente, en el capítulo 4, se recopilan las conclusiones más importantes que se han extraído de los estudios realizados y se detallan algunas propuestas como posibles líneas futuras de investigación dentro del marco del proyecto.

CAPÍTULO 2

2. Material y Métodos

Índice de contenidos

2.1	Material	14
2.1.1	Base de datos de imágenes histopatológicas	14
2.1.2	Software utilizado	17
2.1.3	Hardware utilizado	17
2.2	Metodología.....	18
2.2.1	Preparación de la base de datos de imágenes.....	19
2.2.2	Pre-procesado (<i>Clustering</i>).....	21
2.2.3	Segmentación de candidatos a glándula.....	26
2.2.4	Extracción de características	31
2.2.5	Selección de características	43
2.2.6	Partición de datos	48
2.2.7	Clasificación supervisada.....	51
2.2.8	Creación del <i>groundtruth</i>	61

2.1 Material

2.1.1 Base de datos de imágenes histopatológicas

El material de partida para este proyecto corresponde a los cortes histológicos obtenidos a partir de las biopsias realizadas a diversos pacientes en el Hospital Clínico de Valencia. Las muestras biopsiadas están disponibles, en formato digital, en una base de datos privada, donde únicamente tiene acceso el personal autorizado del Hospital Clínico de Valencia, el de la Universitat Politècnica de València y el de la Universidad de Granada, pues son las instituciones que participan en el proyecto SICAP detallado en el capítulo anterior. La privacidad de la base de datos tiene como finalidad asegurar en todo momento la confidencialidad de los pacientes, quienes firman un consentimiento informado para preservar su bienestar en términos de garantía pública, protección de los derechos y seguridad.

Una vez se tienen al alcance todas las imágenes de las biopsias, los patólogos utilizan una aplicación llamada “*MicroDraw*” con la que pueden anotar las muestras histológicas [19]. *MicroDraw* es una aplicación desarrollada sobre una librería de código abierto llamada *OpenSeadragon* que permite una visualización interactiva multi-resolución sin pérdida de calidad [20]. La plataforma *MicroDraw* está escrita en JavaScript, usando HTML5, CSS3 y jQuery, y ha sido modificada por los ingenieros del CVBLab, conjuntamente con los patólogos, con la finalidad de que puedan realizar anotaciones en las imágenes según los diferentes grados de la escala *Gleason*. Para poder utilizar la aplicación de *MicroDraw*, primero es necesario hacer uso de una librería de procesamiento de imagen llamada “*vips*” que permite convertir las imágenes completas, proporcionadas por el escáner en formato *.tiff, en otras con diferentes niveles de resolución en formato *.dzi. Esto permite que *MicroDraw* pueda ir cargando la imagen *.dzi que requiera cada nivel de aumento del que precisa el patólogo, manteniendo en todo momento la calidad original de la imagen.

Las imágenes histopatológicas con las que se trabaja en este proyecto pueden tener un tamaño del orden de [50000, 90000] x [90000, 190000] píxeles, lo que se traduce en un espacio de almacenamiento que puede ocupar desde los 300 hasta los 1.500 MB (1,5 GB). Por ello, se hace uso de una librería llamada “*OpenSlide*” [21] que proporciona una interfaz sencilla capaz de leer aquellas imágenes que se caracterizan por su alta *resolución espacial*¹² y sus elevadas dimensiones. Además, la librería *OpenSlide* permite resolver el inconveniente existente a la hora de cargar este tipo de datos con librerías estándares, las cuales están diseñadas para leer imágenes que puedan ser fácilmente descomprimidas en la RAM. A esta problemática hay que añadirle el hecho de que no existe ningún formato universal para trabajar con este tipo de imagen, puesto que cada vendedor maneja un formato propio distinto. Por este motivo, es necesario conocer tanto la resolución espacial como el tamaño de las imágenes, atendiendo a los diferentes formatos que pueden ser empleados y al *aumento óptico*¹³.

¹² *Resolución espacial*. Capacidad que tiene un sistema óptico de aislar dos puntos que se encuentran muy próximos entre sí, de manera que ambos se pueden observar por separado [81].

¹³ *Aumento óptico*. Ampliación que indica el número de veces que ha sido aumentada una determinada imagen [81].

Según lo expuesto, la base de datos empleada en el presente trabajo fin de máster está formada por una serie de muestras digitalizadas a un zoom de 40x con el objetivo de disponer de una resolución espacial lo más alta posible. Gracias a este nivel de detalle en la digitalización, en la fase de procesado de la imagen digital es posible captar información de los detalles más pequeños. Sin embargo, trabajar a máximo nivel de zoom también presenta inconvenientes relacionados con la ausencia de estructuras glandulares completas. (Véase la *Figura 2.1.b*). Por consiguiente, el aumento óptico utilizado para implementar los algoritmos es de 10x, ya que permite trabajar con una resolución que, aunque no es óptima, es la adecuada para poder distinguir todos los tipos de estructuras tisulares presentes en una imagen. Además, con una resolución de 10x, es posible encontrar un cierto número de glándulas en cada imagen de tejido, lo cual es necesario para realizar la segmentación de dichas glándulas y poder clasificar el tejido analizado en una clase u otra, atendiendo a la escala *Gleason*.

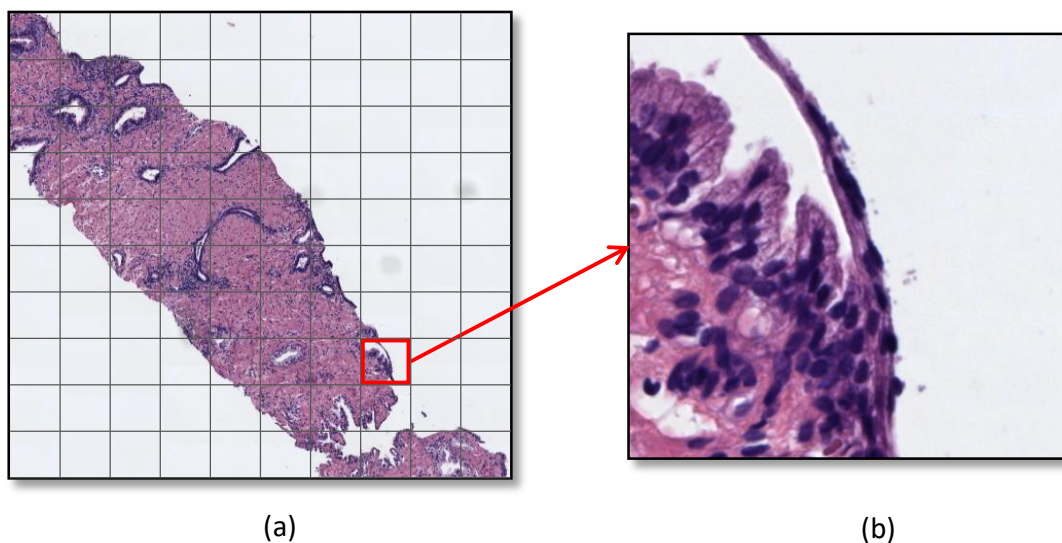


Figura 2.1. Representación gráfica de los conceptos *resolución espacial* y *aumento óptico* sobre imágenes histológicas. (a) Muestra con un aumento aceptable desde el punto de vista de la cantidad de criptas a segmentar, pero con una escasa resolución espacial que dificulta la extracción de características. (b) Fragmento de imagen adquirido de la misma muestra, donde el aumento óptico es demasiado elevado para poder segmentar varias glándulas, pero cuya resolución espacial es mucho más elevada, lo cual facilita el estudio de características que precisan mayor nivel de detalle, como por ejemplo los núcleos epiteliales.

En conclusión, las imágenes de partida para este proyecto corresponden a los cortes histológicos como el que se muestra en la *Figura 2.2.a* que, tras ser escaneados, se convierten en imágenes con un aumento de 40x. Esto conlleva la necesidad de implementar una preparación de la base de datos a fin de conseguir imágenes de 10x, como la mostrada en la *Figura 2.2.b*, con las que poder trabajar a una resolución adecuada y con un número de glándulas por imagen razonable.

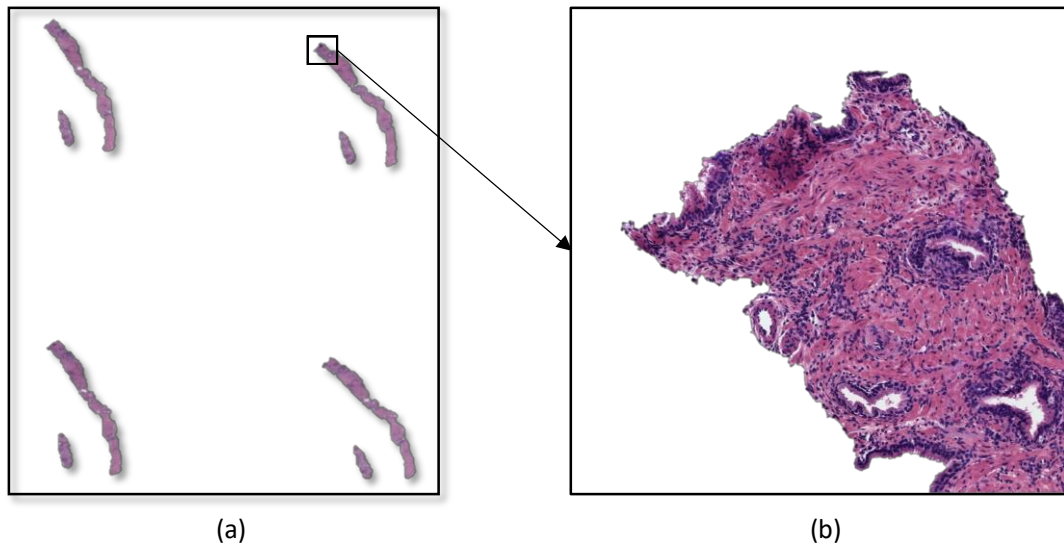


Figura 2.2. (a) Muestra histológica completa de tejido prostático a la mínima resolución. (b) Fragmento de imagen histopatológica de tejido prostático con un aumento de 10x.

Finalmente, cabe destacar que se han seleccionado 60 muestras como las de la *Figura 2.2.a*, de las cuales, 17 corresponden a tejido prostático benigno procedente de 8 pacientes diferentes, mientras que las restantes corresponden a tejido patológico de grado 3, 4 y 5. No obstante, de las 43 imágenes patológicas, solo se han usado en este trabajo aquellas regiones que han sido etiquetadas por los patólogos como grado 3, pues el objetivo final de este proyecto, como ya se ha expuesto en el capítulo 1, es poder discriminar entre glándulas sanas y glándulas patológicas de grado 3, de acuerdo a la escala *Gleason*. De esta forma, por lo que respecta a muestras patológicas, se hace uso finalmente de 18 imágenes de grado 3 procedentes de 17 pacientes diferentes. (Véase la *Tabla 1*).

	Tejido benigno	Tejido patológico de grado 3
Nº de pacientes	8	17
Nº de imágenes (originales)	17	18

Tabla 1. Número de pacientes y de imágenes originales por tipo de tejido analizado.

Por último, conviene señalar que se utilizaron todas las muestras disponibles hasta el momento para la creación de los modelos predictivos. Posteriormente, se recogieron nuevas imágenes de pacientes sin etiquetar para llevar a cabo una evaluación de dichos modelos, con el fin de corroborar la capacidad de generalización de los mismos simulando una situación real de la práctica clínica.

2.1.2 Software utilizado

Para la implementación de los distintos algoritmos utilizados y el cálculo de todos los resultados de este TFM se ha empleado el programa MATLAB® v.R2018a, de The MathWorks, Inc. (Natick, Massachusetts, Estados Unidos). El software MATLAB®, cuyo nombre proviene de *MATrix LABORatory*, es una plataforma optimizada para resolver problemas ingenieriles y científicos. Consiste en un lenguaje de alto nivel que combina tareas de programación, cálculo y visualización, expresando los resultados en notación matemática. Este software permite el desarrollo de algoritmos, la adquisición de datos y otras tareas relacionadas con el modelado, la simulación y la construcción de interfaces gráficas de usuario.

Por otra parte, para la etapa de entrenamiento de redes convolucionales basadas en *deep learning*, se hace uso de la librería *NVIDIA Cuda® Deep Neural Network* (cuDNN). Además, las librerías *OPENCV* con soporte *CUDA* también son necesarias en el ambiente de desarrollo. El lenguaje de programación para abordar esta fase ha sido Python3.5, utilizando PyCharm 2018.4 como IDE. Por último, en lo que a la parte de software respecta, cabe destacar que para la implementación de redes convolucionales profundas se empleó como *framework* de desarrollo Keras, con TensorFlow como *backend*.

2.1.3 Hardware utilizado

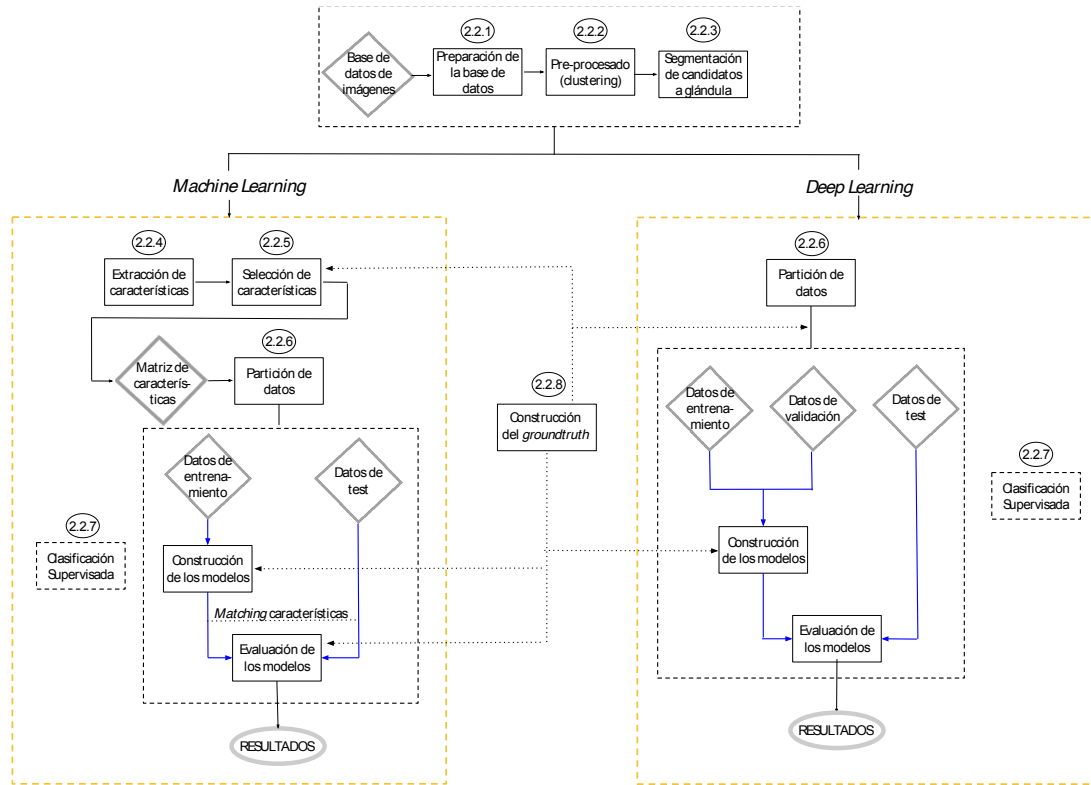
El proyecto ha sido desarrollado y ejecutado en un equipo compuesto por un procesador intel® Core™ i7-4790K @4 GHz. En cuanto a la memoria RAM, la capacidad es de 16 GB (dos DDR3 de 8 GB cada una) y, respecto al tipo de sistema operativo, se trata de un Windows 10 Pro de 64 bits. Además, el equipo cuenta con una tarjeta gráfica NVIDIA GeForce GTX760 de 2 GB.

Debido a la complejidad computacional de las técnicas que se desarrollan en el grupo de investigación, CVBLab cuenta con ordenadores de gran potencia de cálculo. Para llevar a cabo la etapa de entrenamiento de las redes neuronales convolucionales se ha hecho uso de un conjunto de cuatro servidores de computación de altas prestaciones compuestos por un procesador Intel i7 @4.20GHz, 32GB de RAM y tarjetas gráficas NVIDIA Titan XP. Además, el grupo también dispone de un servidor (Synology DS416) en el que se alojan los datos referentes a cada proyecto (señales, imágenes, vídeos, etc.) así como la documentación relativa a cada uno de ellos, con 16 TB de capacidad.

Cabe destacar finalmente que es importante conocer las características internas del hardware empleado para poder evaluar el funcionamiento y el rendimiento que se puede alcanzar a la hora de compilar y ejecutar los algoritmos en términos de tiempo. De esta manera, es posible establecer comparaciones futuras en caso de implementar los mismos programas en un hardware diferente.

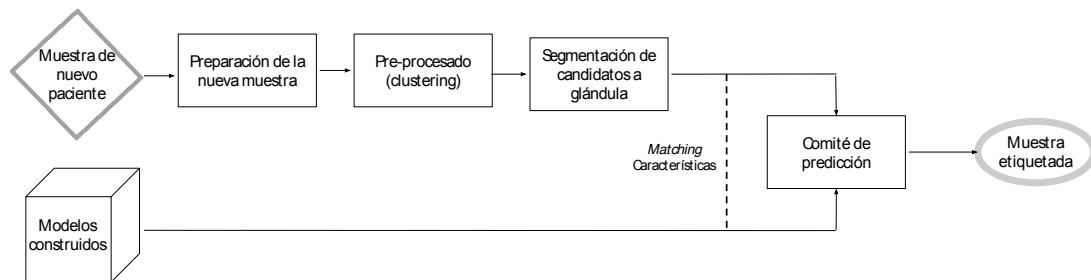
2.2 Metodología

Con el objetivo de abordar la presente sección de forma clara y ordenada, a continuación, se expone un diagrama de flujo donde se pueden observar las diferentes etapas correspondientes al proceso de creación de modelos para la detección automática de patrones anómalos en imagen histológica de próstata. En dicho diagrama de flujo es posible apreciar los distintos métodos implementados durante el presente trabajo fin de máster. A lo largo de la presente sección se detallará cada uno de ellos de manera exhaustiva.



Cabe destacar que normalmente se realiza la partición de datos antes de la selección de características. No obstante, en este proyecto conviene intercambiar el orden, ya que se implementan técnicas de validación cruzada externa (véase apartado 2.2.6).

Una vez detallados los métodos para la construcción de los modelos, al final de este trabajo se lleva a cabo una validación con imágenes de nuevos pacientes para simular la implementación de dichos modelos en una práctica clínica real. En este caso, el diagrama de flujo se establece de la siguiente forma:



2.2.1 Preparación de la base de datos de imágenes

El objetivo principal en esta primera etapa es conseguir lo expuesto en la *Figura 2.2*, es decir, obtener las imágenes histopatológicas con un aumento de 10x, a partir de las imágenes iniciales cuyo aumento es de 40x. El procedimiento para conseguirlo consiste, en primer lugar, en detectar en el corte histológico (*Figura 2.3.a*) las diferentes regiones de interés que albergan cada una de las muestras de tejido. Con ello, es posible eliminar la mayor parte de la información relativa a las regiones blancas (fondo) que aparecen en los cortes histológicos y que carecen de importancia (véase la *Figura 2.3.a*). En segundo lugar, es necesario implementar un algoritmo de ventana deslizante a resolución 10x que permita obtener sub-imágenes o *patches*¹⁴ de unas determinadas dimensiones, a partir de las imágenes correspondientes a las áreas de interés donde se encontraba la muestra de tejido (Véase la *Figura 2.3.b*). Concretamente, en el presente trabajo fin de máster, dichas dimensiones se fijan en 1024x1024x3, ya que se consideran adecuadas en lo que a resolución espacial y a número de glándulas presentes por imagen se refiere.

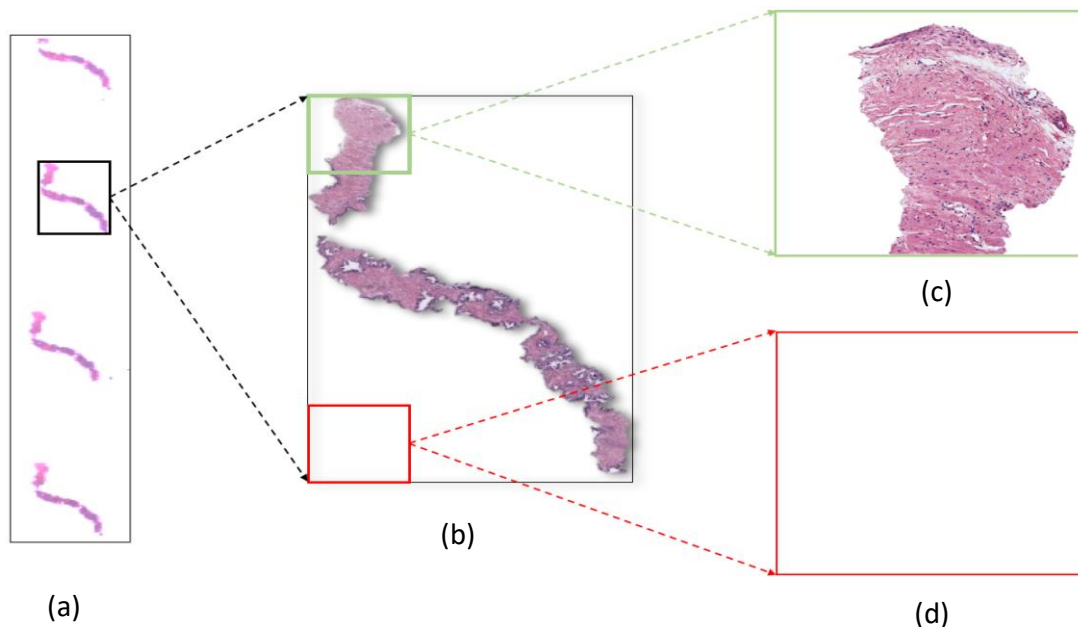


Figura 2.3. (a) Corte histológico de tejido prostático benigno. (b) Rectángulo mínimo que alberga una de las cuatro estructuras tisulares del corte histológico. (c) *Patch* de 1024x1024x3 con información tisular de interés. (d) *Patch* de 1024x1014x3 procedente de la misma imagen, pero sin información relevante.

Por otra parte, una vez se han convertido todas las imágenes originales en sub-imágenes de 10x, se requiere una segunda fase de acondicionamiento en el caso de las imágenes de tejido benigno. La diferencia con las de tejido patológico es que, en estas últimas los patólogos anotan regiones determinadas, por lo que únicamente se seleccionan para el estudio las zonas marcadas como patológicas por parte de los expertos. Sin embargo, en las imágenes de tejido

¹⁴ *Patches*. Fragmentos de una imagen que se obtienen al descomponerla en varias sub-imágenes con el objetivo de estudiar características de interés en regiones ampliadas.

benigno los patólogos no anotan regiones de interés, simplemente indican el carácter no maligno de la muestra en general. Esto conlleva que se genere una gran cantidad de *patches* correspondientes a regiones de fondo blanco sin información relevante. (Véase la *Figura 2.3.d*).

Finalmente, de las 17 imágenes originales benignas se obtienen 2.285 sub-imágenes de 1024x1024x3, mientras que, de las 18 originales patológicas, únicamente se obtienen 614 sub-imágenes, ya que solo se seleccionan las regiones de interés especificadas por los patólogos.

Para trabajar únicamente con la información relevante en las imágenes benignas, se aplica un filtro que permite descartar aquellas que contengan menos de un 5% de tejido. De esta forma, se garantiza que en cada sub-imagen haya la suficiente cantidad de información como para detectar al menos una unidad glandular como la mostrada en la *Figura 1.7*.

Por otra parte, puesto que se dispone de una cantidad de *patches* muy inferior en las muestras de tejido patológico, se realiza un ajuste implementando el algoritmo de ventana deslizante con un solapamiento del 50% en las sub-imágenes anotadas como patológicas de grado 3. Con esto se consigue añadir la información de glándulas prostáticas que quedaban seccionadas entre un *patch* y el siguiente, y que no se tenían en cuenta al descomponer la imagen inicialmente (Véase la *Figura 2.4*).

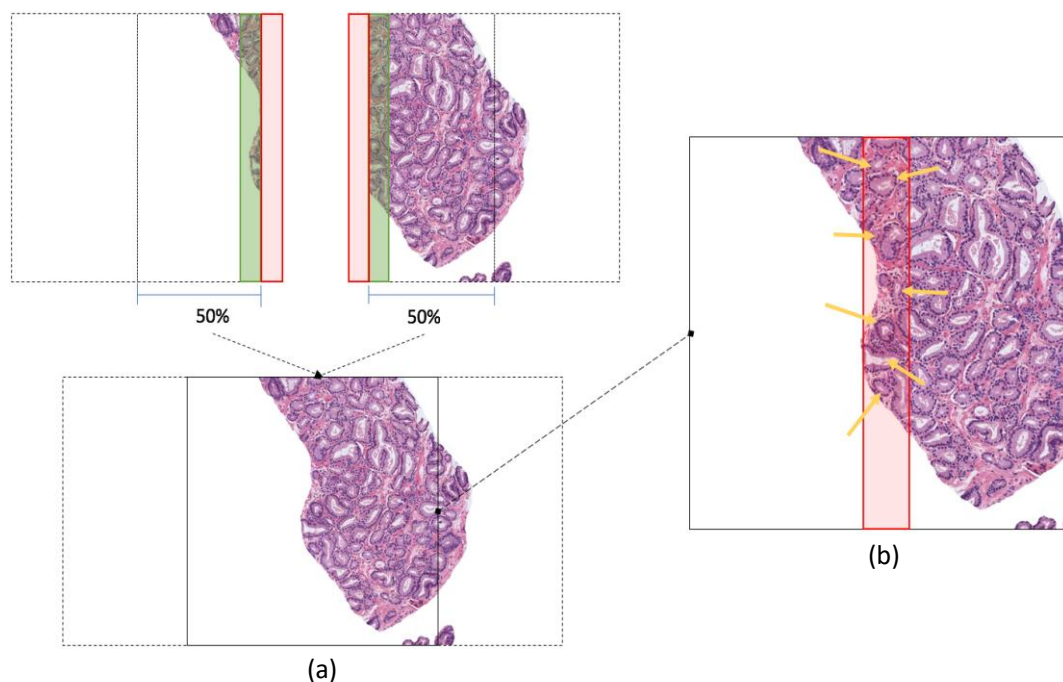


Figura 2.4. Representación gráfica de un solapamiento del 50%. (a) Nueva imagen generada a partir de “coser” el 50% de cada una de las anteriores. (b) Imagen ampliada donde se observan al menos 7 glándulas que sin el solapamiento no se habrían tenido en cuenta.

Todos estos métodos de procesamiento tienen el objetivo de conseguir un número de unidades glandulares balanceado de los dos tipos de tejido, ya que a partir de la segmentación de dichas glándulas se lleva a cabo su extracción de características y su posterior clasificación en la clase correspondiente. Por tanto, finalmente se obtienen 854 sub-imágenes benignas sin solapamiento y 614 sub-imágenes patológicas de grado 3 tras aplicar un porcentaje de

solapamiento del 50%. En la *Tabla 2* se recoge, a modo de resumen, toda la información expuesta para las muestras de tejido sano y para las de tejido canceroso.

	Tejido benigno	Tejido patológico de grado 3
Nº de sub-imágenes	854	614
% Tejido seleccionado	5 %	0 %
% Solapamiento	0 %	50 %

Tabla 2. Número de sub-imágenes de 1024x1024x3, porcentaje de tejido mínimo seleccionado en cada sub-imagen y porcentaje de solapamiento implementado para cada clase.

2.2.2 Pre-procesado (*Clustering*)

Una vez se dispone de las sub-imágenes con las dimensiones deseadas (1024x1024x3), lo que se pretende a continuación es llevar a cabo una clasificación que permita agrupar, en diferentes clases, los distintos componentes de interés que aparecen en las muestras histopatológicas de la próstata. Dichos elementos son los núcleos, el estroma, el citoplasma y los lúmenes, tal y como se indica en el capítulo 1, en el apartado de imágenes histopatológicas. (Véase la *Figura 1.7*).

Para llevar a cabo la agrupación de los datos en diferentes clases es necesario realizar una clasificación no supervisada, ya que solo se dispone de un conjunto de observaciones y no hay información sobre las etiquetas de cada muestra. El objetivo principal es deducir las relaciones intrínsecas existentes entre las variables y la categoría de la muestra u observación. De esta forma, partiendo de un grupo homogéneo de datos, se consigue llegar a clasificar dichos datos en diferentes grupos basándose en la información proporcionada por la tinción de hematoxilina y eosina de las muestras.

Con el fin de efectuar la clasificación no supervisada mencionada anteriormente, se implementan técnicas de *clustering*, las cuales consisten en clasificar los datos de una imagen en diferentes grupos (*clusters*) en base a unas características determinadas. Dentro de las técnicas de *clustering* se pueden distinguir los métodos jerárquicos y los no jerárquicos. Concretamente, en este TFM se ha utilizado una técnica conocida como "*kmeans*" que pertenece a la familia de los clasificadores no jerárquicos. El *kmeans* consiste en un método de agrupamiento que permite clasificar cada observación en el *cluster* que se encuentra más próximo en términos del centroide (media). Su objetivo, por tanto, es agrupar un conjunto de n observaciones en k grupos, de tal forma que cada observación pertenece al grupo cuyo centroide en valor medio es el más cercano.

En términos formales, la técnica del *kmeans* puede ser definida tal que: dado un conjunto de observaciones x_1, x_2, \dots, x_n donde cada una de ellas es un vector real de d dimensiones,

kmeans crea una partición de los datos en los *clusters* C_1, \dots, C_k , cuyos centroides vienen definidos por μ_j , con el objetivo de minimizar la siguiente expresión [22]:

$$\sum_{j=1}^k \sum_{i \in C_j} \|x_i - \mu_j\|^2 \quad (2.1)$$

El algoritmo de la técnica *kmeans* se basa en un procedimiento iterativo en el que se diferencian principalmente dos pasos (*Figura 2.5*) que se detallan a continuación:

1. **Paso de asignación.** En primer lugar, se define el número de k grupos iniciales de forma aleatoria y se calculan sus centroides. Después, para el conjunto de observaciones disponibles se calcula la distancia euclídea (ecuación 2.2) de cada elemento x a cada centroide μ de los k *clusters* y se asigna a cada elemento la etiqueta del *cluster* cuyo centroide es el más cercano, en términos de distancia euclídea.

$$d(x, \mu) = (x - \mu)(x - \mu)^T \quad (2.2)$$

2. **Paso de actualización.** Se calculan los nuevos centroides de los k *clusters* ($k=2$ en el ejemplo de la *Figura 2.5*) como los centroides de las observaciones en el grupo y se repite el procedimiento hasta que el algoritmo converge, es decir, hasta que ya no se produce ninguna reasignación al estabilizarse los elementos en alguno de los grupos.

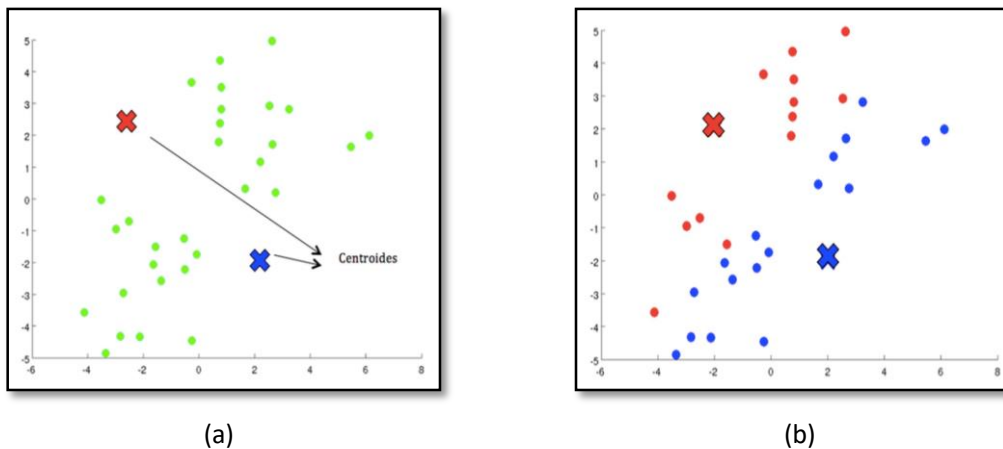


Figura 2.5. (a) Paso de asignación en el que se define el número *clusters* iniciales y se calculan los centroides a partir de un grupo homogéneo de datos. (b) Paso de actualización en el que se asigna a cada elemento una etiqueta en función del *cluster* cuyo centroide es el más cercano.

La etapa del pre-procesado es muy importante, pues es la que permite agrupar los píxeles de la imagen para obtener por separado los cuatro elementos de interés (núcleos, citoplasma, estroma y lúmenes) en función de los niveles de intensidad que presenta cada uno. De esta forma, tras aplicar las técnicas de *clustering*, es posible obtener una máscara binaria¹⁵ de cada uno de los elementos básicos de la imagen.

¹⁵ Máscara binaria. Matriz de datos formada únicamente por 0s y 1s, tal que se visualizan en blanco aquellos píxeles cuyo valor es 1, y en negro, los restantes.

Con la finalidad de obtener unas máscaras precisas, se utilizan diferentes espacios de color que generan imágenes con distintos niveles de intensidad. De esta forma, se resaltan más o menos los elementos básicos de un tejido histológico en función de los canales de color utilizados. Inicialmente, se utiliza la imagen original, es decir, la imagen en el espacio RGB (del inglés *Red, Green, Blue* – rojo, verde, azul) como punto de entrada al algoritmo *kmeans* para abordar la etapa del *clustering*. En consecuencia, puesto que los elementos básicos que se pretenden identificar son cuatro, el número de *clusters* se inicializa en $k = 4$. Se consigue así una imagen de etiquetas (*Figura 2.6.b*) caracterizada por presentar cuatro niveles de intensidad (visualmente, cuatro niveles de grises) que corresponden a los cuatro elementos mencionados. Dado que *kmeans* es una técnica que se basa en la intensidad de los píxeles para llevar a cabo la agrupación de los mismos, las etiquetas 1, 2, 3 y 4, corresponden a los núcleos, el citoplasma, el estroma y los lúmenes, respectivamente.

En la *Figura 2.6.c* se observa cómo se distribuyen los píxeles en el espacio de color RGB, en base al nivel de intensidad que presentan. Finalmente, se obtienen imágenes binarias de cada elemento, tal y como se expone en la *Figura 2.6 (d – g)*.

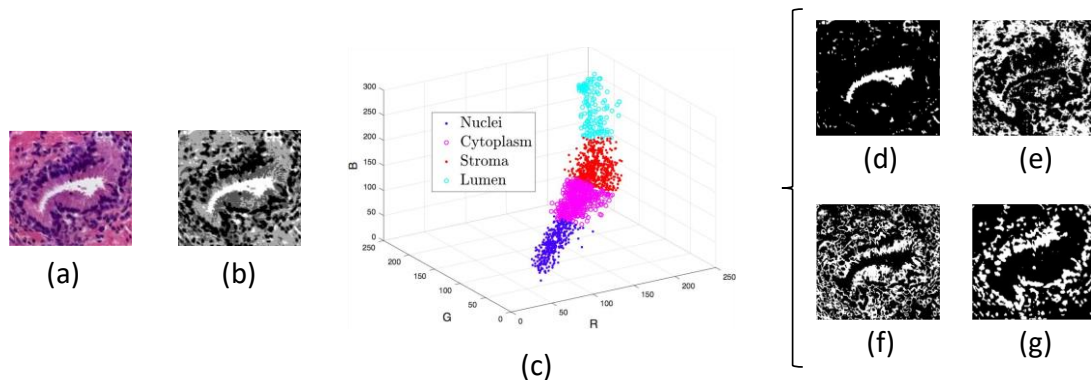


Figura 2.6. Proceso para llevar a cabo el *clustering*. (a) Imagen original en formato RGB. (b) Imagen de etiquetas con cuatro niveles de grises. (c) Diagrama tridimensional que muestra la distribución de los píxeles de cada *cluster* en el espacio. (d) Máscara de píxeles de los candidatos a ser lumen. (e) Máscara de píxeles de los candidatos a ser estroma. (f) Máscara de píxeles de los candidatos a ser citoplasma. (g) Máscara de píxeles de los candidatos a ser núcleos.

Como se puede apreciar en las máscaras de la figura anterior, únicamente la que corresponde a los candidatos a núcleos es aceptable, pues las restantes presentan mucho ruido, y además, una gran cantidad de píxeles no concuerda con la estructura que se observa en la imagen original. Por ese motivo, se realizan varias pruebas con diferentes espacios de color a fin de obtener unas máscaras precisas de los elementos tisulares. Una vez analizados, para el conjunto de entrenamiento, diferentes espacios de color, la metodología a seguir consiste en convertir la imagen original al espacio de color HSV (del inglés *Hue, Saturation, Value* – matiz, saturación, valor) para obtener la máscara de lúmenes, y al espacio de color CMYK (del inglés *Cyan, Magenta, Yellow, Key* – cian, magenta, amarillo, negro) para obtener la imagen binaria de candidatos a estroma y citoplasma. En la *Figura 2.7* se exponen las imágenes resultantes de aplicar los distintos espacios de color y las máscaras generadas a partir de cada uno de ellos.

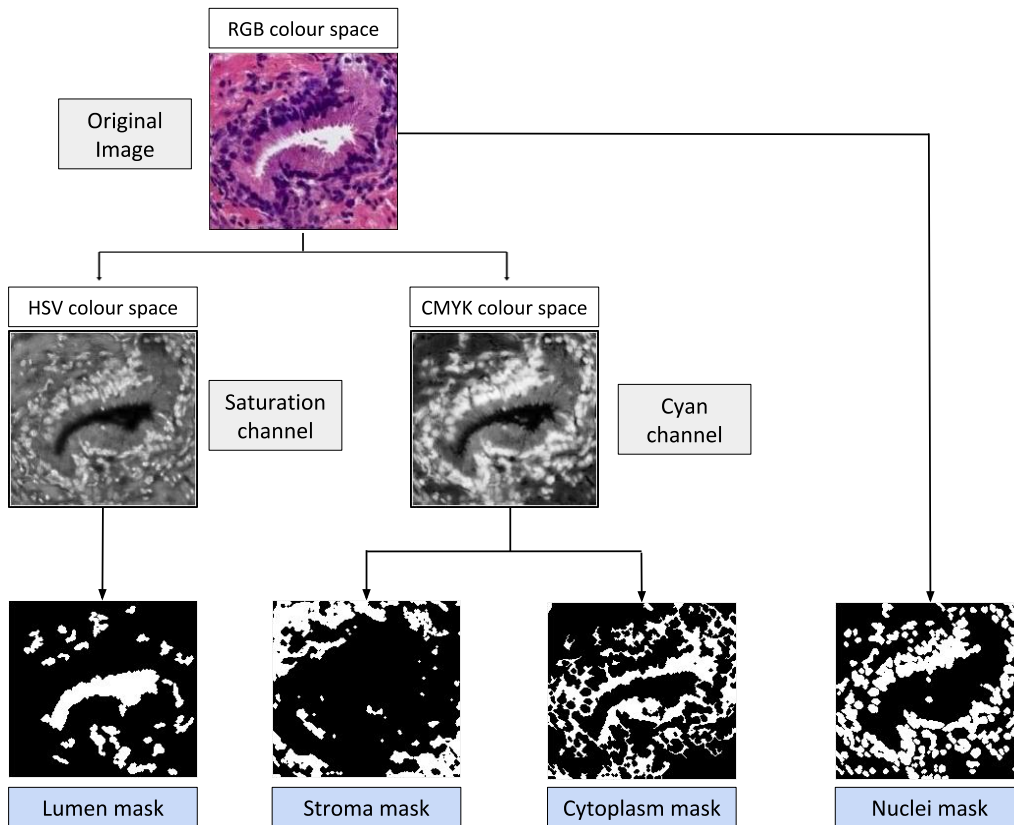


Figura 2.7. Diagrama explicativo del proceso de obtención de máscaras aplicando diferentes espacios de color.

Aunque las máscaras generadas en este punto también presentan ruido, se observa a simple vista la mejora con respecto a las máscaras de la *Figura 2.6*, ya que en este caso las estructuras en blanco coinciden en mayor medida con las de la imagen original. De este modo, es posible aplicar operaciones morfológicas, a modo de filtrado de ruido, tales como la dilatación (ecuación 2.3), la erosión (ecuación 2.4) y una técnica conocida como “*area opening*” (ecuación 2.5). Dichas técnicas se detallan a continuación:

- La **dilatación** es una operación morfológica extensiva que, a partir de una imagen binaria y dado uno o más píxeles en blanco, es capaz de fijar con el mismo nivel de intensidad los píxeles del vecindario. El número de píxeles vecinos que se convierten en blanco depende de la conectividad y del elemento estructurante (EE) elegidos. Siendo X la imagen binaria y B el EE, la dilatación $\delta_B(X)$ puede expresarse de la siguiente forma:

$$\delta_B(X) = X \oplus B \tag{2.3}$$

- La **erosión**, en cambio, es una operación morfológica anti-extensiva, pues dados una imagen binaria y un conjunto de píxeles en blanco, y siguiendo las mismas propiedades que la dilatación, es capaz de convertir en negro los píxeles más externos de los objetos blancos de la imagen. La erosión $\varepsilon_B(X)$ se expresa matemáticamente como:

$$\varepsilon_B(X) = X \ominus B \tag{2.4}$$

- La técnica **area opening** es una operación de filtrado que se aplica sobre una imagen binaria para eliminar todos aquellos objetos en blanco cuya área, en número de píxeles, es menor que un determinado umbral t [23]. Esta operación se expresa como sigue:

$$\gamma_t(X) = \bigvee_i \{\gamma_{B_i} \mid B_i \text{ es el objeto en blanco y la cardinalidad de } B_i = t\} \quad (2.5)$$

A lo largo de la etapa de pre-procesado, se hace uso de las tres técnicas para aumentar y disminuir el tamaño de determinados objetos (en términos de píxeles) como los núcleos y zonas del estroma, respectivamente, así como para eliminar algunos artefactos de lúmenes. El efecto de dichas técnicas de procesado puede apreciarse a modo de ejemplo en la *Figura 2.8*, aplicándose sobre la máscara de lúmenes con un EE cuadrático de tamaño 5 para la dilatación y la erosión, y un umbral $t = 250$ píxeles para el *area opening*.

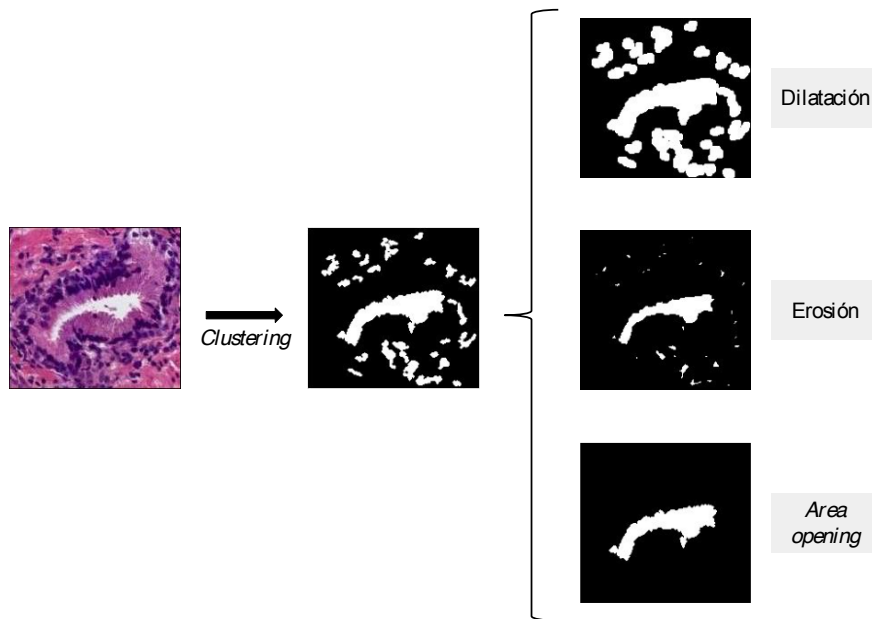


Figura 2.8. Aplicación de las operaciones morfológicas de dilatación, erosión y *area opening* sobre una imagen binaria correspondiente a la máscara de candidatos a lumen.

Cabe destacar que las máscaras obtenidas tras el proceso del *clustering* constituyen una pieza clave en la fase de la extracción de características. Con todo ello, la máscara de candidatos a lumen todavía adquiere una mayor importancia, ya que los lúmenes también son la base para llevar a cabo una buena segmentación de las glándulas (véase el siguiente apartado). No obstante, el pre-procesado no genera unas máscaras perfectas y, por tanto, en las imágenes binarias de los candidatos a lumen, algunos de ellos son lo que se conoce como *artefactos*¹⁶ en la literatura médica [14]. (Véase la *Figura 2.9*).

¹⁶ *Artefactos*. Áreas rotas de tejido que se caracterizan por presentar un color blanco en las imágenes teñidas con hematoxilina y eosina, lo cual puede inducir a una confusión con las regiones de lúmenes. No obstante, a diferencia de los lúmenes, los artefactos no suelen estar rodeados de estructuras de citoplasma ni de núcleos.

2.2.3 Segmentación de candidatos a glándula

El objetivo principal de este apartado reside en segmentar correctamente cada una de las glándulas individuales presentes en las sub-imágenes histopatológicas de 1024x1024x3, tanto de las benignas como de las patológicas de grado 3. Para ello, es necesario utilizar el lumen de una glándula como punto de partida para la segmentación. Sin embargo, como ya se ha comentado en el párrafo anterior, el *clustering* no es perfecto, por lo que la manera de proceder es segmentar todos los candidatos a lumen, sean artefactos o lúmenes, extraer y seleccionar las características oportunas y, después, aplicar clasificadores multiclase que permitan discriminar no solo las glándulas benignas y las patológicas de grado 3, sino también los artefactos.

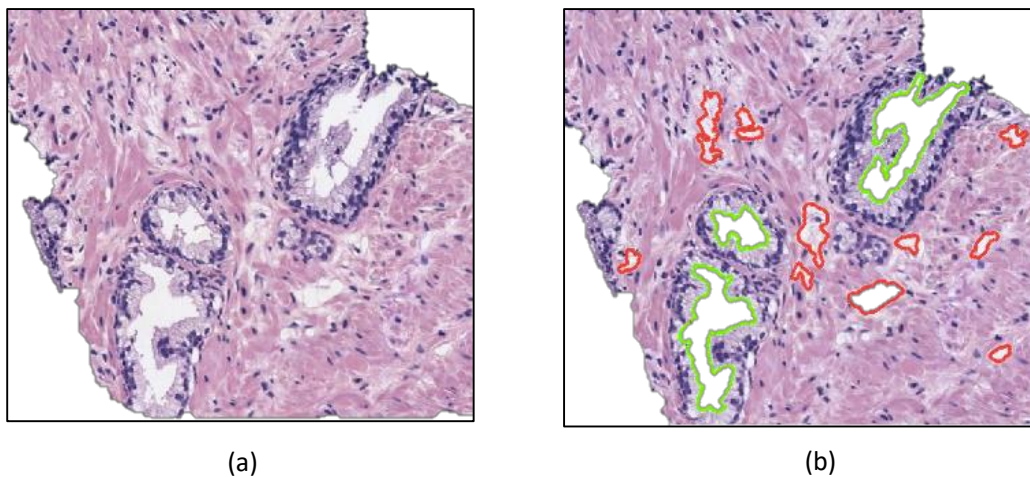


Figura 2.9. (a) Imagen original teñida con hematoxilina y eosina donde se observan tres glándulas individuales. (b) Misma imagen con los lúmenes de las glándulas resaltados en verde; y los artefactos, en rojo.

El estudio presentado en [24] hace hincapié en la necesidad de realizar una buena segmentación de glándulas con el objetivo de conseguir significantes mejoras a la hora de discriminar entre los diferentes grados propuestos según la escala *Gleason*. Por esa razón, en este trabajo se implementa una técnica conocida como *locally constrained watershed transform* (LCWT) [25]. Dicha técnica consiste en un novedoso algoritmo que ha sido diseñado e implementado durante este proyecto con el objetivo de mejorar los métodos existentes relativos al estado del arte, en cuanto segmentación de glándulas se refiere. No obstante, dicha segmentación se lleva a cabo únicamente como un paso intermedio para conseguir el verdadero objetivo que es la clasificación de los candidatos a glándula en sus respectivas clases.

Cabe destacar que, a diferencia del trabajo realizado en [26], donde los bordes de la glándula vienen definidos por los límites del citoplasma, en este proyecto dichos bordes quedan delimitados por la capa de núcleos que rodea a la glándula, de acuerdo con lo establecido en la literatura médica que propone el propio artículo. En trabajos posteriores, como en [14] y [27], la segmentación llevada a cabo es similar a la que se propone en este estudio, ya que también consideran los núcleos como los límites de las unidades glandulares. En la *Figura 2.10*, se expone la diferencia entre las segmentaciones realizadas en [26], en [27] y en este proyecto.

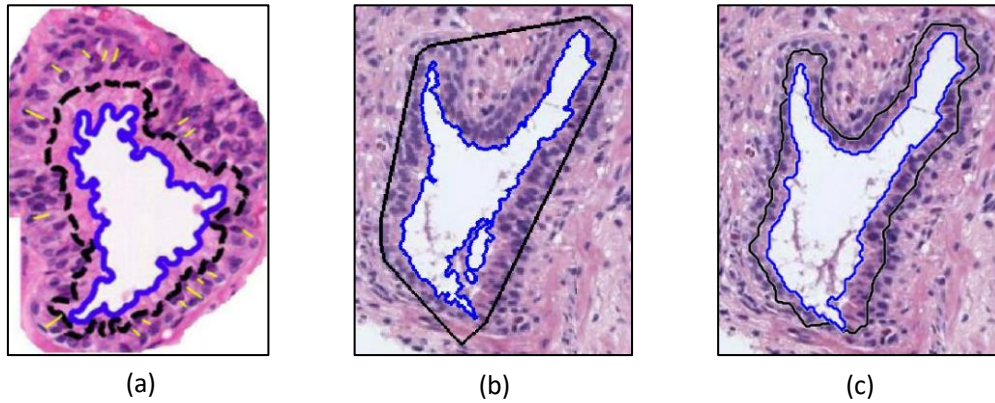


Figura 2.10. Comparación de la segmentación llevada a cabo por diferentes autores donde la línea azul delimita el lumen; y la negra, la glándula. (a) Ejemplo de glándula segmentada publicado en [26]. (b) Glándula segmentada con el código que propone el autor en [27]. (c) Misma glándula segmentada con el método propuesto en este TFM.

Como ya se ha comentado, la técnica utilizada en este estudio para la segmentación de las glándulas es la técnica LCWT, cuyo algoritmo se basa en una variante de los métodos tradicionales conocidos como “watershed” y “watershed con marcadores”. Por tanto, dichos métodos se explican primero con la finalidad de comprender correctamente el LCWT.

2.2.3.1 Watershed

La técnica *watershed* se basa en considerar la imagen inicial en escala de grises como un relieve topográfico donde se pueden distinguir dos tipos de elementos: las cimas, que serían los máximos locales; y los valles, que harían referencia a los mínimos locales [28]. De esta forma, se divide la imagen en las zonas de influencia de cada mínimo local simulando un proceso de inundación como el que se muestra en la *Figura 2.11*.

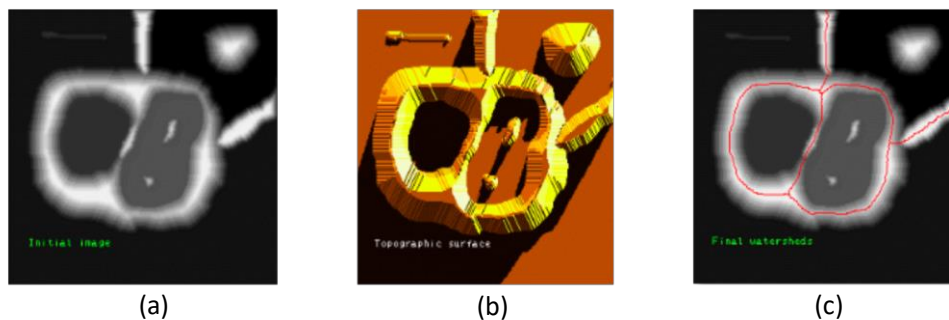


Figura 2.11. Procedimiento de segmentación mediante la técnica *watershed*. (a) Imagen de entrada en escala de grises. (b) Simulación de la imagen en un relieve topográfico. (c) Resultado de la segmentación marcado por las líneas rojas.

Típicamente, la técnica *watershed* se suele aplicar utilizando como entrada la imagen gradiente, ya que así, las fronteras de las regiones son máximos locales. Uno de los problemas del gradiente es que, si la imagen tiene mucho ruido, se crean demasiados mínimos y máximos locales, lo cual deriva en una sobsegmentación como la que se observa en la *Figura 2.12*.

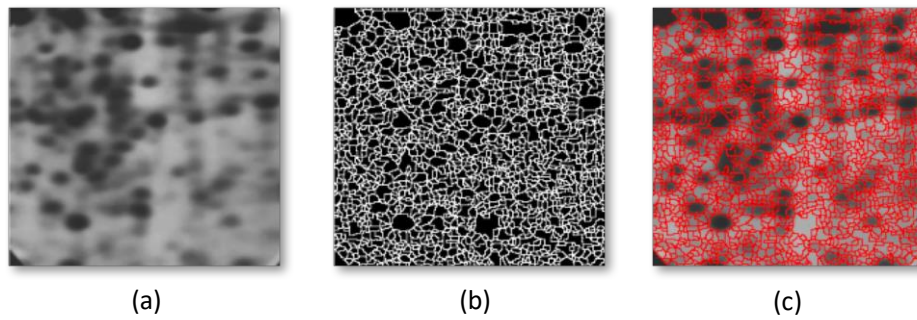


Figura 2.12. a) Ejemplo de imagen de entrada en escala de grises. b) Imagen donde se aprecia la sobresegmentación realizada. c) Resultado de sobresegmentación en la imagen de entrada.

2.2.3.2 Watershed con marcadores

Para solventar los problemas de la sobresegmentación se aplica una transformación sobre las imágenes a fin de obtener un único mínimo local por objeto. Para ello se utiliza el método *watershed* con marcadores, siendo un marcador un conjunto de píxeles conectados que actúan de referencia. Se hace uso de dos tipos de marcadores: los internos, que son aquellos que se colocan en el interior del objeto de interés en la imagen; y los externos, que se fijan fuera de la región de interés para delimitar la zona que debe ser segmentada.

El procedimiento se basa en simular la inundación característica del valle a partir del marcador interno localizado como mínimo local en el objeto de interés. Por otra parte, se utiliza el marcador externo para fijar la frontera en el máximo local (la cima) y evitar que el agua, con la que se inunda el valle, se extienda fuera de él. Es decir, el marcador externo limita la segmentación que comienza con el marcador interno. En la *Figura 2.13*, se demuestra cómo mejoran los resultados de la segmentación utilizando marcadores, con respecto a la técnica representada en la *Figura 2.12*.

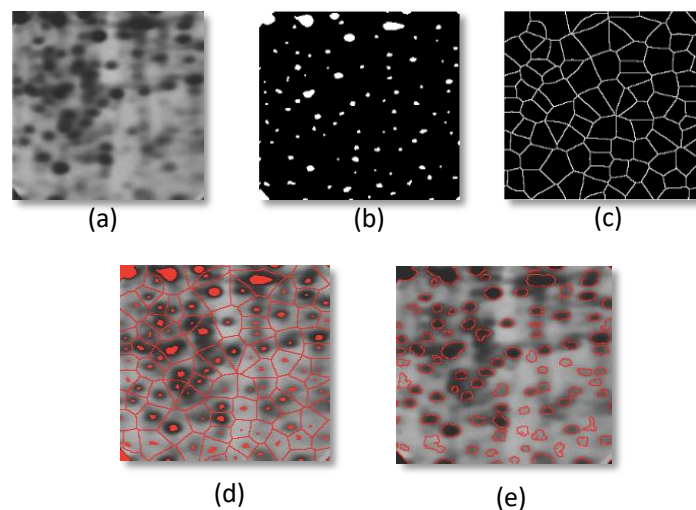


Figura 2.13. Segmentación mediante la técnica *watershed* con marcadores. (a) Imagen original en escala de grises. (b) Máscara de los marcadores internos. (c) Fronteras de segmentación. (d) Combinación de marcadores sobre la imagen original. (e) Resultado de la segmentación.

2.2.3.3 *Locally constrained watershed transform*

Como ya se ha comentado, esta técnica consiste en una variante del método explicado en el párrafo anterior, ya que también requiere el uso de marcadores internos y externos, aunque con ciertas condiciones. La necesidad de implementar esta variante reside en que los límites de una glándula están constituidos por los núcleos, los cuales se observan como píxeles separados unos de otros, alrededor de la glándula. Es decir, una estructura glandular no está delimitada por una línea continua, sino por una frontera abierta que hace inservible la técnica tradicional de marcadores a la hora de implementar la segmentación deseada. (Véase la *Figura 2.14*).



Figura 2.14. (a) Región de interés delimitada por una línea continua (frontera cerrada). (b) Región de interés delimitada por una línea discontinua (frontera abierta).

Siguiendo el símil del relieve topográfico, se puede entender el caso de las fronteras abiertas como una imagen donde los máximos locales son cimas individuales separadas tal que, al intentar inundar el valle con el método tradicional, el agua se escaparía por el espacio que hay entre las cimas y, por tanto, la segmentación de la región de interés no funcionaría en el caso de la *Figura 2.14.b*.

En definitiva, la técnica LCWT se aplica con el objetivo de solventar el problema de las fronteras abiertas y realizar una correcta segmentación de las unidades glandulares. Para su implementación es necesario especificar una serie de parámetros de entrada:

1. **Marcadores internos.** Definidos por la máscara de candidatos a lumen.
2. **Marcadores externos.** Definidos por los píxeles que conforman el fondo de la imagen y por muchos de los que componen la máscara del estroma.
3. **Imagen de entrada.** Definida por la máscara de núcleos.
4. **Elementos estructurantes para ambos marcadores.** Se refiere al tamaño de paso de los marcadores.

El funcionamiento del sistema consiste en que cada marcador analiza los píxeles de su vecindario intentando incorporarlos a su región. El avance en busca de integrar nuevos píxeles se lleva a cabo en función del EE definido para cada marcador. Por otra parte, los núcleos que componen la imagen de entrada actúan como restricciones al avance de los marcadores, ya que la expansión de un marcador puede verse detenida si el tamaño de su EE es mayor que la distancia entre las restricciones. En ese caso, el marcador no podría avanzar más y la frontera de segmentación quedaría definida entre las restricciones cuando dicho marcador entrase en contacto con el otro. (Véase la *Figura 2.15*).

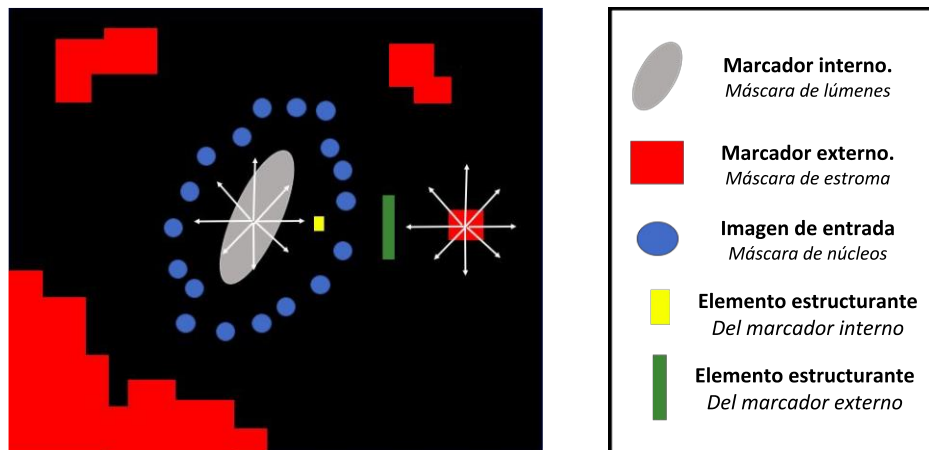


Figura 2.15. Ejemplo ilustrativo que explica visualmente el funcionamiento de la técnica *locally constrained watershed transform*.

Como puede apreciarse en la figura anterior, los dos marcadores intentan avanzar por la imagen, en todas direcciones, con tamaños de paso definidos por su respectivos EE. En el caso propuesto, el EE del marcador externo es de mayor tamaño que el del interno, ya que interesa que el marcador externo se quede estancado al intentar atravesar el espacio que hay entre los núcleos que delimitan la glándula. De este modo, el marcador interno, al crecer con un tamaño de paso mucho menor, puede entrar en contacto con el marcador externo obstaculizado entre las restricciones impuestas por los núcleos. Finalmente, se establece una frontera cerrada que viene definida por los píxeles de la imagen correspondientes a la unión de ambos marcadores, proporcionando así una segmentación de la glándula alrededor de los núcleos de la misma.

Trasladando la información esquemática de la *Figura 2.15* a imágenes reales del proyecto, se expone la *Figura 2.16*, donde se aprecian los marcadores internos, los marcadores externos y la imagen de entrada manteniendo la misma gama de colores que en el modelo ilustrativo.

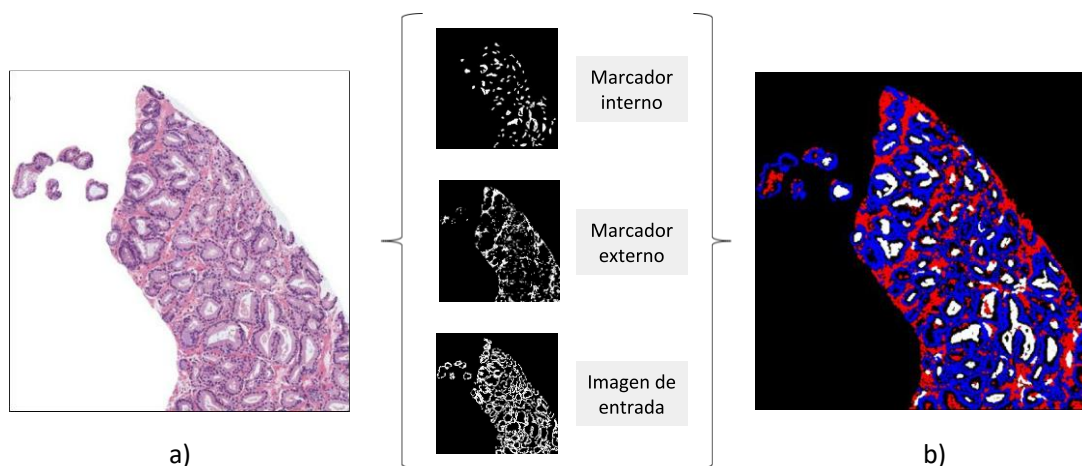


Figura 2.16. Esquema que muestra visualmente cómo se emplean los parámetros para realizar la segmentación mediante la técnica LCWT. a) Imagen original. b) Parámetros de entrada al algoritmo LCWT aplicados a partir de la imagen original. En blanco los lúmenes (marcadores internos), en rojo el estroma (marcador externo) y en azul los núcleos (imagen de entrada).

En la *Figura 2.17*, se expone la salida final de esta etapa para una imagen ejemplo de dimensiones 1024x1024x3. Se puede apreciar la segmentación automática llevada a cabo para todos los candidatos a glándula a partir de la máscara de posibles lúmenes. Cabe destacar que finalmente se obtienen imágenes más pequeñas correspondientes a cada uno de los candidatos a glándula, tal y como se muestra en la *Figura 2.17.b*. Estas imágenes más pequeñas se constituyen como el mínimo rectángulo capaz de albergar la glándula completa; y es en ellas donde se aplican los algoritmos de extracción y selección de características a fin de intentar discriminar si dichas imágenes corresponden a una glándula benigna, a una glándula patológica de grado 3 o a un artefacto.

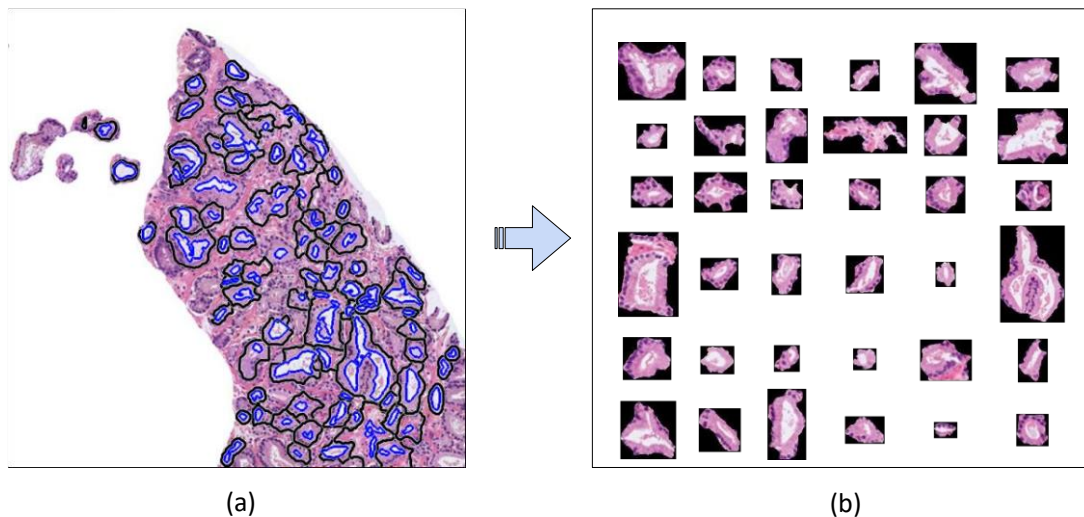


Figura 2.17. (a) Imagen de 1024x1024x3 donde se han segmentado las posibles glándulas (definidas por las líneas negras) a partir de los candidatos a lumen (delimitados por las líneas azules). (b) Candidatos a glándula segmentados automáticamente y almacenados de forma individual para su posterior estudio.

2.2.4 Extracción de características

La extracción de características constituye la primera etapa de la clasificación basada en *machine learning*. El objetivo es identificar patrones característicos en las imágenes de las glándulas individuales (*Figura 2.17.b*) que permitan entrenar modelos de clasificación capaces de aprender qué valores de qué características se asocian a una clase u otra.

Para abordar esta etapa, en primer lugar, se lleva a cabo una exhaustiva revisión bibliográfica en busca de las características más interesantes que se han implementado en el estado del arte a nivel de glándulas. La idea es computar inicialmente todas aquellas que puedan presentar información relevante y, posteriormente, llevar a cabo un análisis estadístico para seleccionar únicamente las que son independientes con respecto al resto de variables (las que no guardan correlación con otras) y dependientes con respecto a la clase (las que tienen una gran capacidad discriminativa).

Para abordar la tarea de extracción de características, en primer lugar, se realiza una revisión de la literatura médica para obtener información acerca de los patrones específicos de

las glándulas benignas y de las patológicas. A continuación, se detallan los rasgos distintivos de cada tipo de tejido, a fin de encontrar peculiaridades que ayuden a determinar su clase.

- **Tejido benigno.** Presenta grandes unidades de glándula, densamente compactas y separadas. Cada glándula está bien definida albergando grandes regiones de lumen y un límite grueso con núcleos prominentes, además de una estructura citoplasmática entre ambas capas. La forma de los lúmenes normalmente presenta un aspecto más fusiforme (ramificado) que los lúmenes del tejido patológico. No obstante, el lumen también puede ser ovalado o circular en este caso [14].
- **Tejido patológico de grado 3.** Se caracteriza por presentar glándulas de dimensiones menores y más circulares. Del mismo modo, los lúmenes que encierran dichas glándulas también son más pequeños y ovalados. Además, la densidad de núcleos en la glándula es menor, ya que solo hay una capa de células epiteliales que circunscribe la luz [14].

En la *Figura 2.18* se pueden apreciar las particularidades de cada tipo de glándula.

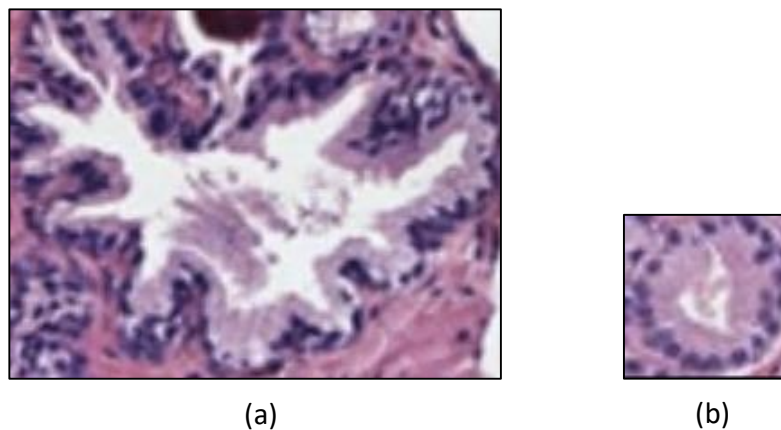


Figura 2.18. a) Unidad glandular benigna. b) Unidad glandular patológica de grado 3.

Con respecto al estado del arte, se observa que la gran mayoría de los estudios publicados sobre segmentación y clasificación de imágenes histopatológicas de próstata hacen uso principalmente de descriptores de forma y de textura durante la extracción de características. En [29] y [30] utilizan descriptores de textura para segmentar y descriptores de forma para llevar a cabo la tarea de clasificación entre muestras benignas y muestras patológicas. En [31], el autor extrae características morfológicas y de textura, mientras que en [32] solo se estudia la morfología de las glándulas. Por otra parte, el estudio llevado a cabo en [24] hace uso de descriptores de forma y de textura e incluye características del color de las glándulas para abordar la clasificación de las imágenes discriminando entre muestras patológicas de grado 3 y 4, de acuerdo a la escala *Gleason*. Concretamente, a diferencia de muchos estudios, como los llevados a cabo en [33] y en [27], donde únicamente se hace uso de 15 y 22 características morfológicas y contextuales, respectivamente, en este trabajo fin de máster se implementan un total de 297 características relacionadas con descriptores de forma, de textura, fractales y otras características de contexto. Cabe recordar que, a partir de las 297 características, se lleva a cabo un análisis estadístico para seleccionar únicamente aquellas que aportan información relevante al estudio en términos de independencia entre variables y dependencia con la clase.

A continuación, se detalla cada una de las métricas utilizadas en este proyecto según el tipo de descriptor al que pertenecen.

2.2.4.1 Descriptores de forma

Se caracterizan por definir la morfología y la estructura de los elementos de una imagen. Dentro de los descriptores de forma existen diferentes tipos. Los que se implementan en el presente trabajo fin de máster son, por un lado, los descriptores geométricos, que estudian la geometría y los contornos de las regiones de interés, por ejemplo, el área o el perímetro; y por otro lado, los descriptores topológicos, que estudian las configuraciones estructurales relacionadas con la *invariancia*¹⁷ de las transformaciones, por ejemplo, la compacidad o la suavidad.

Concretamente, en este proyecto, la aplicación de los descriptores de forma consiste en extraer la información directamente de las máscaras binarias correspondientes a las estructuras del tejido de los candidatos a glándula. En la *Tabla 3* se exponen las 34 métricas, de descriptores de forma, calculadas para los diferentes componentes de los candidatos a glándula segmentados en la etapa anterior.

	Características de los DESCRIPTORES DE FORMA (Total = 34)
GLÁNDULAS (11)	<i>area, compactness, convexArea, convexHull, eccentricity, equivDiameter, extent, orientation, perimeter, roundness y solidity.</i>
LÚMENES (11)	“
NÚCLEOS (8)	<i>numNucBB, numNucBBprop, numNucGland, numNucGlandprop, pixNucBB, pixNucBBprop, pixNucGland, pixNucGlandprop,</i>
CITOPLASMA (4)	<i>pixCytoBB, pixCytoBBprop, pixCytoGland, pixCytoGlandprop.</i>

Tabla 3. Descriptores de forma aplicados sobre las estructuras glandulares.

En base a la información recogida en la tabla anterior, se explica brevemente cada una de las métricas utilizadas en lo que a descriptores de forma se refiere.

- **area.** Cantidad de píxeles en blanco de la región de interés, que es la máscara de la glándula segmentada o el lumen de la misma.
- **compactness.** Compacidad del elemento (glándula o lumen). Mide su circularidad.
- **convexArea.** Número de píxeles en el área de la imagen convexa.
- **convexHull.** Polígono convexo más pequeño que es capaz de albergar la región de interés. Mide el carácter fusiforme del objeto.
- **eccentricity.** Longitud de la elipse formada por el objeto de interés (glándula o lumen), tal que un valor de 0 es un círculo perfecto y un valor de 1 es una línea.
- **equivDiameter.** Diámetro de un círculo con la misma área que la región de interés.

¹⁷ *Invariancia.* Término relativo a una propiedad de un objeto que mantiene su valor tras aplicar ciertas transformaciones como pueden ser un cambio de escala, una traslación, una rotación, etc.

- **extent.** Ratio de píxeles en la región de interés con respecto al área de la *BoundingBox*¹⁸.
- **orientation.** Ángulo formado entre el eje x y el eje mayor de la elipse que forma el elemento de interés.
- **perimeter.** Longitud definida por el número de píxeles que conforman el límite de la región de interés.
- **roundness.** Redondez de la región de interés.
- **solidity.** Proporción de píxeles en la *convexHull* que pertenece al área del objeto.
- **numNucBB.** Número de núcleos en la *boundingbox* (BB).
- **numNucBBprop.** Proporción entre el número de núcleos en la BB y el área de la misma.
- **numNucGland.** Número de núcleos en la región de interés.
- **numNucGlandprop.** Proporción entre el número de núcleos en la región de interés y el área de la misma.
- **pixNucBB.** Cantidad de píxeles de núcleos en la BB.
- **pixNucBBprop.** Proporción entre la cantidad de píxeles de núcleos en la BB y el área de la misma.
- **pixNucBBGland.** Cantidad de píxeles de núcleos en la región de interés.
- **pixNucBBGlandprop.** Proporción entre la cantidad de píxeles de núcleos en la región de interés y el área de la misma.
- **pixCytoBB, pixCytoBBprop, pixCytoGland y pixCytoBBGlandprop.** Hacen referencia a la misma información que las características de los núcleos con el mismo nombre, pero en este caso, se utiliza la máscara correspondiente al citoplasma.

2.2.4.2 Descriptores de textura

Se caracterizan por estudiar principalmente los cambios en la intensidad de los píxeles alrededor de una determinada región de interés. Concretamente, en este TFM se implementan dos tipos de descriptores de textura: la matriz de coocurrencia y los *Local Binary Pattern (LBP)*. Ambos se aplican sobre diferentes espacios de color, ya que cada uno presenta diferentes niveles de intensidad en la escala de grises y pueden ayudar a caracterizar los patrones de las imágenes de candidatos a glándula. Por ello, en primer lugar, se explican los tres tipos de espacios de color sobre los que se aplican los descriptores de textura y, posteriormente, se detallan las métricas implementadas para cada uno.

Los espacios de color utilizados son el canal cian (C), cuyo proceso de obtención ya se ha detallado en el apartado 2.2.2, y los canales relativos a la Hematoxilina (H) y a la Eosina (E), que se obtienen a partir de aplicar un método conocido como “*deconvolución de color*”. Esta técnica es de gran utilidad al trabajar con las imágenes histopatológicas teñidas con H&E, ya que permite separar las contribuciones de cada tinción mediante el parámetro de la Densidad Óptica (OD) [34], la cual viene definida por la siguiente ecuación:

$$OD_C = A * c_C \quad (2.6)$$

donde A es la absorbancia y c_C , la concentración de una tinción concreta. Además, cada tinción se caracteriza por presentar una densidad óptica específica en cada uno de los tres canales del espacio RGB. Por tanto, aplicando la *deconvolución de color* se pretende optimizar la aplicación

¹⁸ *BoundingBox*. Rectángulo más pequeño capaz de contener la región de interés.

de los descriptores de textura sobre imágenes cuyas diferencias de intensidad entre componentes estructurales es más evidente. En la *Figura 2.19*, se observan los tres espacios de color utilizados sobre los que se aplican las técnicas de análisis de textura.

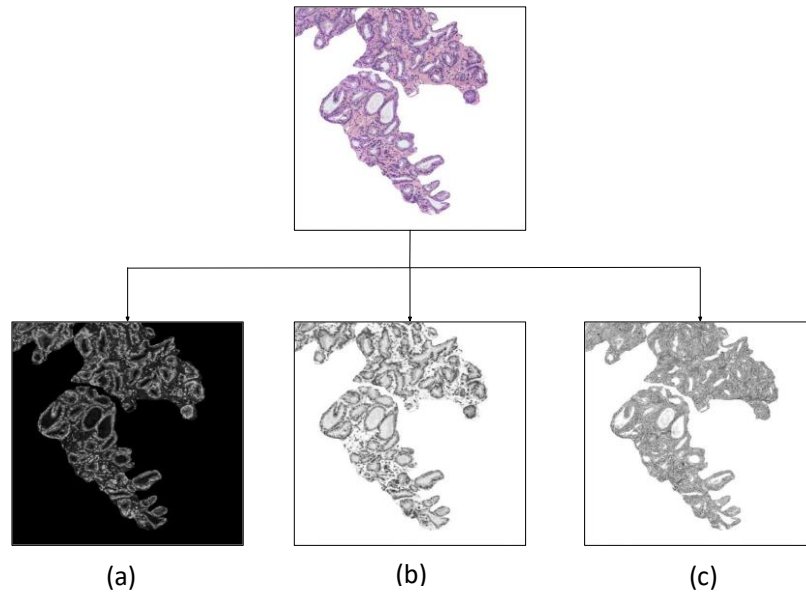


Figura 2.19. a) Canal Cian del espacio de color CMYK. b) Canal de Hematoxilina tras aplicar el método de *deconvolución de color*. c) Canal de Eosina de la *deconvolución de color*.

Una vez explicados los tres espacios de color sobre los que se implementan los descriptores de textura, se explican dichos descriptores y se detallan las métricas computadas.

➤ **Matriz de coocurrencia.** Consiste en una matriz de frecuencias en la que un píxel con un determinado nivel de gris i aparece en una relación de espacio específica con otro píxel de nivel de gris j . De esta forma, la matriz de coocurrencia es capaz de revelar determinadas propiedades sobre la distribución espacial de los grises en la textura de la imagen [35]. En [36] se propone un estudio para clasificar automáticamente muestras histológicas de próstata en los diferentes grados de la escala *Gleason*, a partir de la matriz de coocurrencia. En este proyecto, dicho descriptor se calcula para dos ángulos diferentes (0° y 45°) que corresponden a un 'Offset' de $[0,2]$ y $[-2,-2]$ respectivamente, al trabajar con una distancia $D = 2$ píxeles. Es decir, dado un píxel de interés, se analiza un número D de píxeles en la dirección definida por los ángulos. (Véase la *Figura 2.20*).

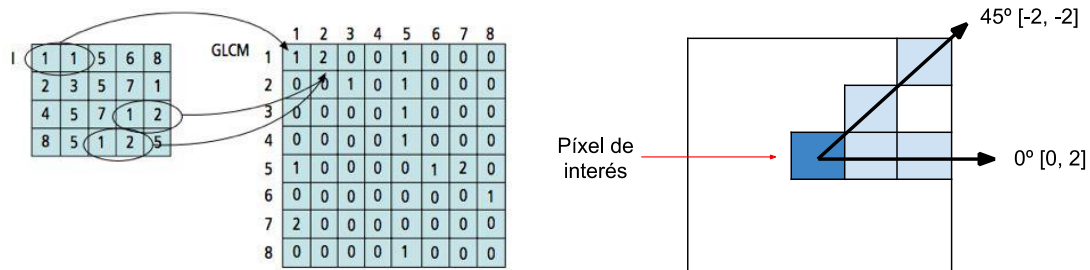


Figura 2.20. Proceso de construcción de una matriz de coocurrencia. (*Figura adaptada a partir de la publicada en [35]*).

Por defecto, para una imagen 2D, el número de niveles de grises utilizado es 8. En este TFM se multiplica por 5 ese número con el objetivo de tener una gama más amplia de intensidades que permita discriminar mejor las posibles glándulas en una de las tres clases que se consideran. (Véase la *Figura 2.21*).

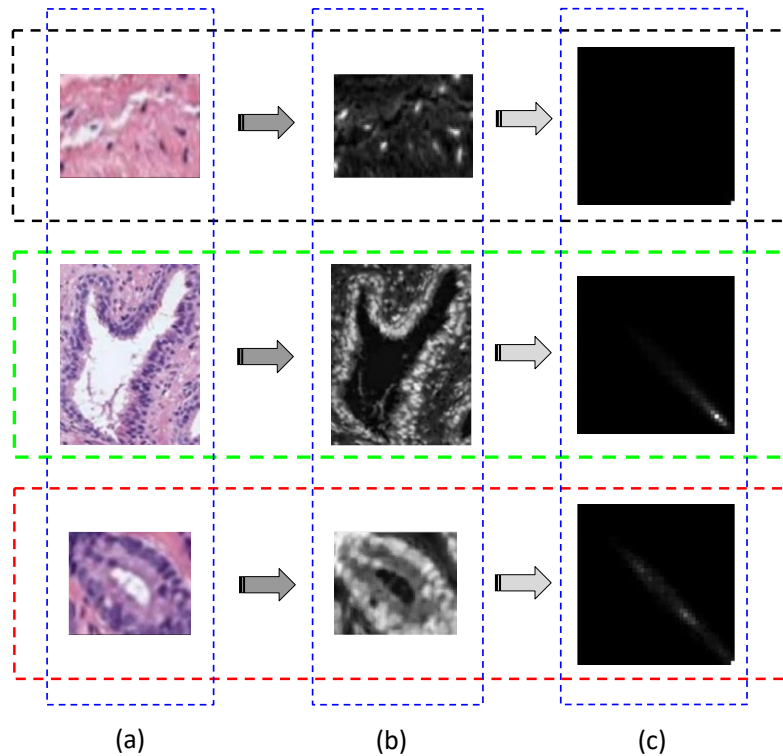


Figura 2.21. Procedimiento para la obtención de la matriz de coocurrencia. (a) Imágenes de un artefacto, una glándula benigna y una glándula patológica (en negro, verde y rojo, respectivamente). (b) Imágenes del canal cian tras aplicar una transformación al espacio de color CMYK de las imágenes originales. (c) Matrices de coocurrencia donde se aprecian las diferencias entre los tres tipos de elementos a clasificar.

A partir de la matriz de coocurrencia se obtienen las siguientes métricas, detalladas en [37]:

- **Contraste.** Parámetro que mide las variaciones de intensidad entre un píxel determinado y los de su vecindario.
- **Correlación.** Medida que determina lo parecido que es el valor de un píxel determinado con respecto a los de su alrededor.
- **Energía.** Se calcula como la raíz cuadrada del segundo momento angular (ASM), que es un parámetro que mide la similitud de todas las entradas de la matriz tal que, la energía es más baja cuanto mayor similitud existe entre las entradas.
- **Homogeneidad.** Valor que analiza la distribución de los elementos diagonales de la matriz, en términos de proximidad.
- **Media.** Parámetro calculado como la suma de todos los valores de cada columna entre el número total de filas.
- **Desviación Típica.** Medida de dispersión que refleja la representatividad de los promedios en función del margen de error. También se calcula por columnas.

➤ **Local Binary Pattern (LBP).** Consiste en un eficiente operador que describe patrones de la imagen a nivel local [38] proporcionando buenos resultados en aplicaciones de reconocimiento de texturas [39] y de localización de objetos y formas [40]. Por lo que respecta al estado del arte en imágenes histopatológicas de próstata, en [41] se hace uso de una variante de los LBP que, al igual que en este trabajo, se aplica sobre los espacios de color relativos a la Hematoxilina y a la Eosina. Por un lado, se utilizan los LBP uniformemente invariantes a la rotación ($LBP_{P,R}^{riu2}$) y, por otro lado, los invariantes a los cambios de contraste ($VAR_{P,R}$). A partir de ellos, se extraen métricas relacionadas con el histograma de la intensidad de los píxeles para representar las texturas de la imagen. A continuación, se explica en detalle la implementación de los LBP y, posteriormente, las métricas utilizadas basadas en el histograma.

Para entender las técnicas de LBP implementadas, en primer lugar, es necesario explicar el funcionamiento básico de los LBP originales propuesto en [42], como una segunda versión del método llevado a cabo en [43]. El algoritmo para obtener los LBP se basa en calcular la diferencia entre el valor del píxel de interés (central) y el valor de los píxeles vecinos fijando en cada celda el resultado correspondiente (Figura 2.22.b). Después, se lleva a cabo una umbralización tal que, si dicho resultado es positivo, se asigna la etiqueta 1; y si es negativo, la etiqueta 0 (Figura 2.22.c). Una vez generadas las etiquetas, se obtiene una cadena binaria que al multiplicarse por los valores de referencia (Figura 2.22.d) codifica un determinado número decimal mediante la suma de los valores resultantes (Figura 2.22.e). Para el ejemplo propuesto, el número binario es 00101111 que equivale al número decimal: $LBP = 4+16+32+64+128 = 254$.

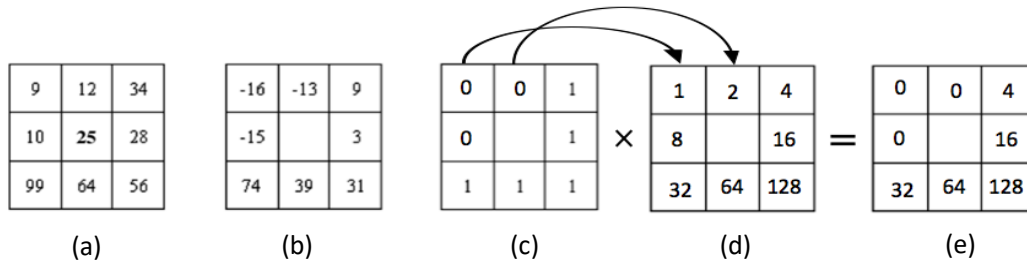


Figura 2.22. Proceso que muestra el funcionamiento interno del algoritmo basado en los LBP básicos. (Imagen modificada a partir de la publicada en [39]).

En [44] se expone una variante de los LBP básicos, cuyas diferencias radican principalmente en dos aspectos: 1) los píxeles vecinos alrededor del píxel de interés se calculan en forma de cadena circular y 2) los valores de grises de los píxeles de la diagonal se determinan por interpolación. De esta forma, se obtienen los $LBP_{P,R}$ que, por definición, son invariantes frente a cualquier transformación de la escala de grises y se expresa matemáticamente según la siguiente ecuación:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p, \quad s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (2.7)$$

donde P es el número de píxeles vecinos alrededor del píxel de interés que controla la cuantificación del espacio angular y R , el radio de dicha circunferencia que determina la resolución espacial con la que se trabaja (véase la Figura 2.23).

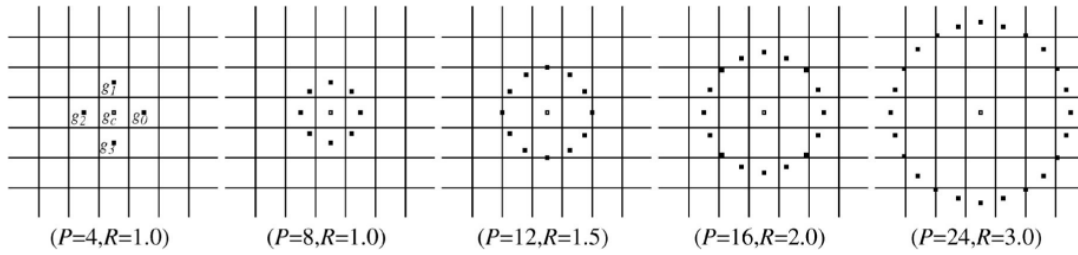


Figura 2.23. Ejemplos de ventanas circulares alrededor del píxel de interés con diferente número de puntos P y radios R para caracterizar a nivel local la textura de la imagen [44].

El inconveniente de este tipo de LBP es que son sensibles a las transformaciones de rotación, por lo que para eliminar el efecto de dichas transformaciones se asigna un único identificador para todas aquellas cadenas binarias que presenten el mismo número de *transiciones*¹⁹. Por ejemplo, dadas dos cadenas binarias: 11100011 y 01111000, cuyo número de transiciones de 0 a 1, o viceversa, es el mismo, se asignan a un único identificador normalizado tal que, los 0s se ponen a la izquierda y los 1s, a la derecha. El identificador para el caso propuesto sería 00001111, donde se observan 2 transiciones. Por tanto, para una $P = 8$, el operador $LBP_{P,R}$ produce $2^8 = 256$ valores de patrones diferentes, mientras que el nuevo operador invariante a la rotación $LBP_{P,R}^{ri}$ presenta únicamente 36 valores distintos, tal y como se muestra en la siguiente figura:

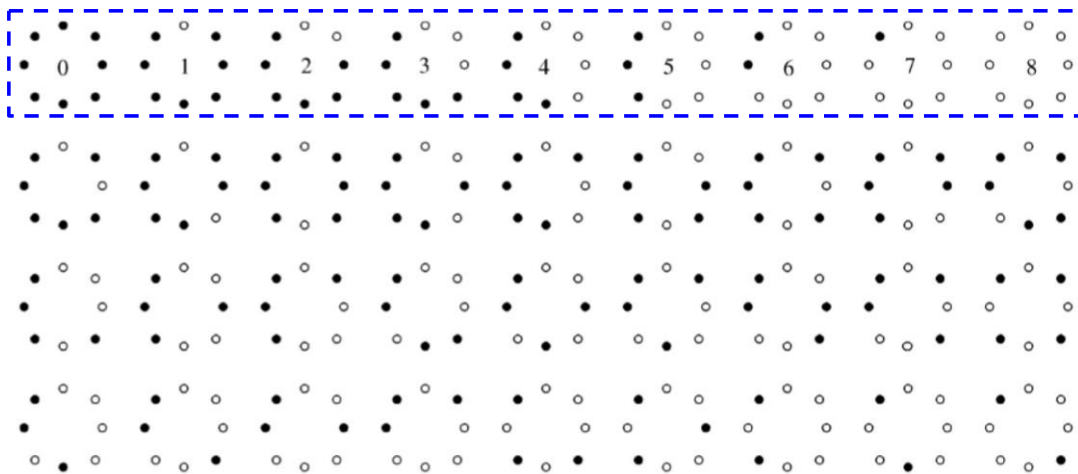


Figura 2.24. Ilustración de los únicos 36 *binary patterns* invariantes a la rotación que pueden ocurrir con un operador $LBP_{8,R}^{ri}$, donde los círculos negros y blancos corresponden a los valores de 0s y 1s, respectivamente. (Figura adaptada a partir de la publicada en [44]).

En [45] se descubrió que algunos de esos 36 patrones presentaban información redundante, por lo que en [44] se introdujo el concepto de “uniformidad” haciendo referencia al número de transiciones espaciales entre 0s y 1s. Se demostró en ese estudio que los patrones con un número de transiciones menor o igual que 2 tenían una capacidad discriminativa mucho mayor que el resto, por lo que definieron dichos patrones como LBP uniformes, proponiendo el siguiente operador invariante a transformaciones de rotación y de escala de grises:

¹⁹ Transiciones. Cambios de 0 a 1 o de 1 a 0 en una cadena binaria determinada.

$$LBP_{P,R}^{riu2} = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c) & \text{si } U(LBP_{P,R}) \leq 2 \\ P + 1 & \text{el resto} \end{cases} \quad (2.8)$$

donde

$$U(LBP_{P,R}) = |s(g_{P-1} - g_c) - s(g_0 - g_c)| + \sum_{p=1}^{P-1} |s(g_p - g_c) - s(g_{p-1} - g_c)| \quad (2.9)$$

De esta forma, utilizando una cadena binaria en forma circular de radio R con un número de píxeles vecinos P , el número exacto de patrones “uniformes” es $P+1$, cada uno de ellos etiquetados en el rango de 0 a P . Para el ejemplo propuesto de la *Figura 2.24*, los casos encuadrados en el rectángulo azul son los $P+1$ patrones uniformes, ya que el número de transiciones en ellos es menor que 2, mientras que los 27 restantes se agrupan en un conjunto ‘no uniforme’ llamado “ $P+1$ ”, tal y como se describe en la ecuación 2.8. De esta forma, $LBP_{P,R}^{riu2}$ tiene $P+2$ valores de salida distintos: de 0 a 8 para los patrones uniformes y la etiqueta $P+1$ para los no uniformes.

Por otra parte, cuando se utilizan los LBP para la descripción de texturas es común incluir también el operador conocido como *Rotational Invariant Local Variance* (VAR), ya que el $LBP_{P,R}^{riu2}$ es muy bueno para medir las estructuras espaciales de la textura de la imagen a nivel local, pero no es invariante a los cambios de contraste, ya que depende de la escala de grises. El operador $VAR_{P,R}$ (ecuación 2.10) por definición, es invariante frente a variaciones de contraste y, además, es una medida complementaria al $LBP_{P,R}^{riu2}$, cuya combinación se espera que proporcione interesantes resultados a la hora de discriminar entre las distintas clases presentes en las imágenes histopatológicas de próstata.

$$VAR_{P,R} = \frac{1}{P} \sum_{p=0}^{P-1} (g_p - \mu)^2, \quad \text{donde } \mu = \frac{1}{P} \sum_{p=0}^{P-1} g_p \quad (2.10)$$

De este modo, se obtienen como resultado las imágenes LBP y VAR de dimensiones $M \times N$ para cada uno de los tres canales (cian, hematoxilina y eosina) sobre los que se aplican los operadores (véase la *Figura 2.25*). El objetivo es obtener las métricas basadas en el histograma normalizado que proporciona la combinación de las imágenes LBP y VAR. Concretamente, en [38] se describen los *Local Binary Pattern Variance (LBPV) histogram*, que acumulan los valores de $VAR_{P,R}$ para cada etiqueta de los $LBP_{P,R}^{riu2}$ según la siguiente ecuación:

$$LBPV_{P,R}(k) = \sum_{i=1}^M \sum_{j=1}^N \omega(LBP_{P,R}(i,j), k), \quad k \in [0, K] \quad (2.11)$$

donde

$$\omega(LBP_{P,R}(i,j), k) = \begin{cases} VAR_{P,R}(i,j), & LBP_{P,R}(i,j) = k \\ 0, & \text{el resto} \end{cases} \quad (2.12)$$

siendo K la etiqueta máxima de los LBP.

En la *Figura 2.25*, se sigue el modelo de la *Figura 2.21* para representar las imágenes correspondientes a los $LBP_{P,R}^{riu2}$ y a los $VAR_{P,R}$. En el ejemplo se muestra también el histograma relativo a los LBPV de las tres clases de candidatos a glándula utilizando el canal cian como punto de partida para la extracción de este tipo de características.

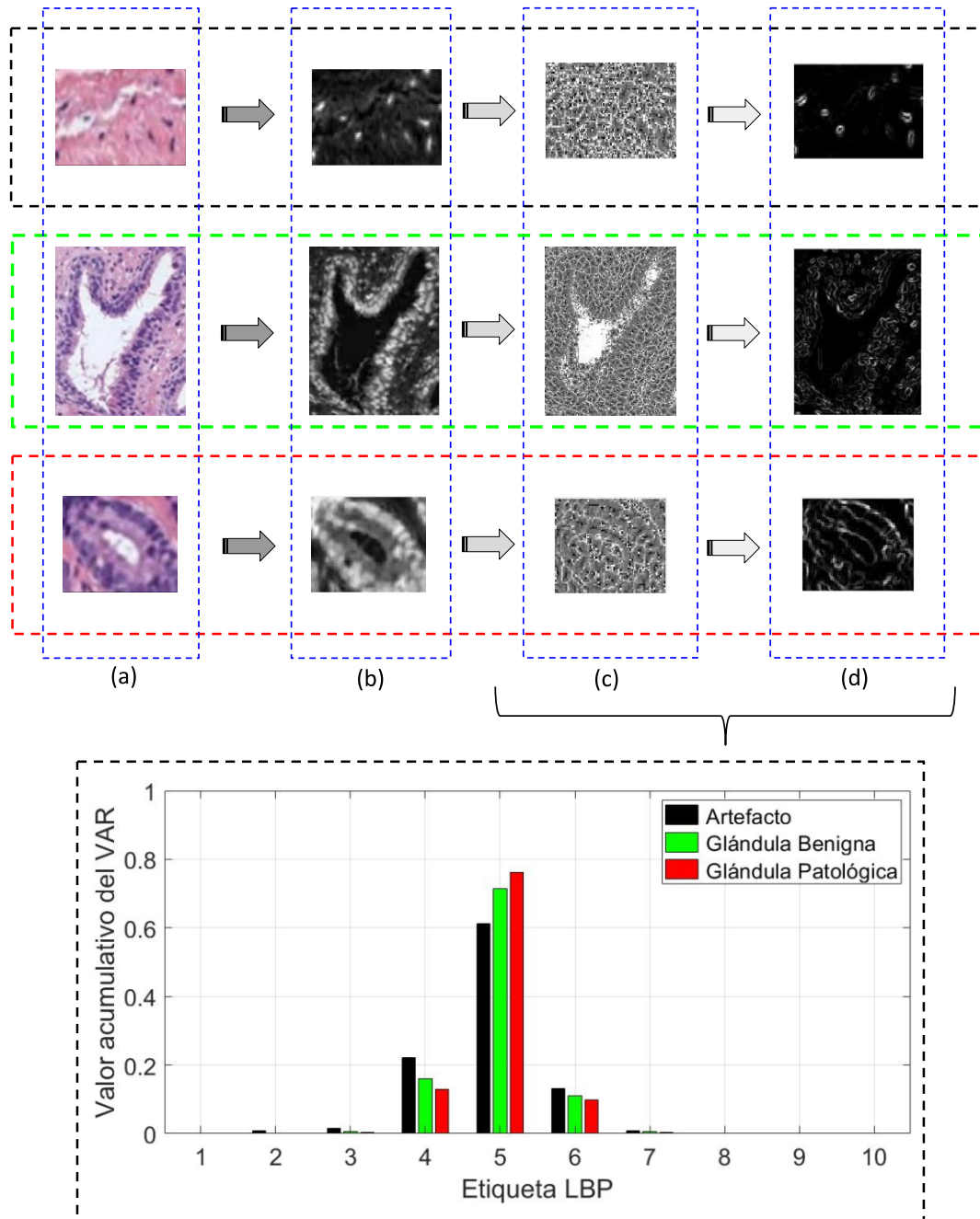


Figura 2.25. Procedimiento para la obtención del histograma LBPV normalizado. (a) Imágenes de un artefacto, una glándula benigna y una glándula patológica (en negro, verde y rojo, respectivamente). (b) Imágenes del canal cian tras aplicar una transformación al espacio de color CMYK de las imágenes originales. (c) Imágenes de los LBP uniformemente invariantes a la rotación. (d) Imágenes de los LBP uniformemente invariantes a los cambios de escala de grises.

Cabe destacar que además de los histogramas LBPV normalizados, también se obtienen los histogramas relativos a los $LBP_{p,R}^{riu^2}$ únicamente. Dichos histogramas se calculan sobre las imágenes con los candidatos a glándulas enmascarados y sin enmascarar.

A modo de resumen, se expone la *Tabla 4* en la que se recogen las métricas utilizadas en base a los descriptores de textura aplicados sobre los tres tipos de espacios de color de los candidatos susceptibles de ser glándulas.

	Matriz de coocurrencia (144)	LBP (96)
CIAN	<i>ContrastCyanF0, CorrelationCyanF0, EnergyCyanF0, HomogeneityCyanF0, meanGLCMcyanF0, stdGLCMcyanF0, ContrastCyanF45, CorrelationCyanF45, EnergyCyanF45, HomogeneityCyanF45, meanGLCMcyanF45, stdGLCMcyanF45</i>	<i>hist_LBPcyan, hist_LBP_VARcyan, hist_LBPcyanMask, hist_LBP_VARcyanMask</i>
HEMATOXILINA	“	“
EOSINA	“	“

Tabla 4. Listado de las 140 métricas de descriptores de textura aplicadas sobre los canales de color cian, hematoxilina y eosina.

2.2.4.3 Análisis fractal

Los fractales pueden definirse como objetos de geometría irregular que presentan patrones estructurales repetidos en una imagen y que son invariantes frente a cambios de escala. Permiten analizar la complejidad o la tortuosidad que presenta un determinado tejido u objeto de interés. En la literatura científica, el análisis fractal es ampliamente utilizado para caracterizar una gran variedad de tipos de imágenes biomédicas. Concretamente, en lo que a imágenes histopatológicas de próstata se refiere, en [30] se hace uso de la dimensión fractal de distancia y de la dimensión fractal de Higuchi para computar la distancia entre un núcleo y sus núcleos vecinos más cercanos. En [46] y en [47] también se hace uso de métodos basados en la dimensión fractal para analizar variaciones de intensidad y de textura en las regiones de interés, con el objetivo de abordar la clasificación del tejido en los diferentes grados según la escala *Gleason*. Concretamente, en este TFM se extrae localmente la dimensión fractal computando el exponente *Hurst* (H) a lo largo de 5 direcciones (0° , 30° , 45° , 60° y 90°) para los tres canales de color correspondientes al cian, al de Hematoxilina y al de Eosina. El exponente *Hurst* es una medida que se utiliza para determinar si los datos siguen patrones puramente aleatorios o mantienen tendencias subyacentes [48]. Consiste en un parámetro que gobierna la representación del movimiento fraccional browniano y que se relaciona con la dimensión fractal según la siguiente ecuación:

$$H = E + 1 - FD \quad (2.13)$$

donde E es la dimensión euclídea y FD , la dimensión fractal.

De esta forma, utilizando el exponente *Hurst* es posible estudiar, en diferentes direcciones, el comportamiento con “memoria a largo plazo” que existe entre las observaciones, es decir, se puede analizar la presencia de patrones repetidos en determinadas regiones de interés a lo largo de las imágenes.

2.2.4.4 Características contextuales

Este tipo de características hace referencia a la información que puede extraerse a partir de establecer determinadas relaciones de distancias, formas u otras similitudes alrededor de un elemento de interés. En este caso, dicho elemento es cada uno de los candidatos a glándula y las características de contexto que se implementan están relacionadas con las propiedades que presenta cada candidato en el entorno de la *BoundingBox* a la que pertenece. De esta forma, no se obtienen características directamente de los lúmenes, los núcleos o el citoplasma, sino de la interacción entre diferentes elementos. Con respecto al estado del arte en imágenes histopatológicas de próstata, se utilizan este tipo de características en [14] y en [27]. Siguiendo su ejemplo, en este proyecto se hace uso de 8 características contextuales que se han elegido a partir del conocimiento adquirido por parte del autor durante las observaciones de las muestras.

A continuación, se definen brevemente las 8 características contextuales implementadas en este proyecto:

- ***relNumNucGlandBB***. Proporción entre el número de núcleos en la región de interés y el número de núcleos totales en la *BoundingBox*, siendo la región de interés la máscara del candidato a glándula.
- ***relPixNucGlandBB***. Proporción entre la cantidad de píxeles de núcleos en la región de interés y la cantidad de píxeles de núcleos totales en la BB.
- ***relPixCytoGlandBB***. Misma medida que *relPixNucGlandBB*, pero aplicada sobre la máscara del citoplasma.
- ***LGrleation***. Relación entre el área del lumen y el área del candidato a glándula.
- ***meanDistLumen***. Media de la distancia de cada píxel del contorno del lumen a cada núcleo presente en la región de interés.
- ***varDistLumen***. Varianza de la distancia de cada píxel del contorno del lumen a cada núcleo perteneciente a la máscara del candidato a glándula.
- ***pixNCL***. Sumatorio de píxeles correspondientes a las máscaras proporcionadas por los núcleos, el citoplasma y el lumen de la región de interés.
- ***pixNCLprop***. Relación entre el sumatorio de píxeles de los núcleos, el citoplasma y el lumen, y la región de interés.

En resumen, durante la fase de la extracción de características, se obtienen 297 variables; 34 de ellas correspondientes a los descriptores de forma (teniendo en cuenta las estructuras de la glándula, del lumen, de los núcleos y del citoplasma); 140 correspondientes a los descriptores de textura (considerando la matriz de coocurrencia y dos tipos de *Local Binary Pattern*, aplicados sobre los canales de color cian, hematoxilina y eosina); 15 relativas a la dimensión fractal mediante el exponente *Hurst*; y 8 características relacionadas con la información contextual presente en las imágenes de los candidatos a glándula.

2.2.5 Selección de características

Del mismo modo que la etapa anterior, la fase de selección de características se lleva a cabo específicamente en la aproximación basada en *machine learning*. El objetivo es seleccionar, de las 297 características extraídas anteriormente, únicamente aquellas que presentan una información relevante de cara a resolver el problema de la clasificación. Para ello, se establece una serie de contrastes de hipótesis cuya finalidad es descartar, bien las variables que están correlacionadas con otras, o bien, las que no guardan una relación de dependencia con la clase.

En primer lugar, puesto que todas las características utilizadas en este proyecto son variables cuantitativas continuas, es necesario llevar a cabo una normalización. Esto es porque cada una puede tomar un rango de valores diferente, es decir, hay características que pueden presentar valores muy altos y otras que incluso pueden ser negativas. En ese caso, se le estaría otorgando más peso a aquellas variables cuyos valores son más altos, lo cual no tiene por qué implicar que esas características sean más importantes. La normalización de las variables se lleva a cabo mediante la siguiente ecuación que calcula el “*zscore*”:

$$z = \frac{(x - \bar{X})}{S} \quad (2.14)$$

donde x es un valor de una determinada característica, \bar{X} es la media de todos los valores de dicha característica y S , la desviación estándar.

Una vez las variables han sido normalizadas para que todas tengan la misma importancia dentro del estudio, el objetivo es determinar cuáles tienen una mayor capacidad de discriminación con respecto a las tres clases en cuestión: artefactos, glándulas benignas y glándulas patológicas. Por tanto, la finalidad es estudiar si una determinada característica toma valores muy diferentes según qué clase. Es decir, si la característica C adquiere siempre el valor A cuando se trata de la clase *Artefacto*, el valor B cuando se trata de la clase *Glándula Benigna* y el valor P cuando se trata de la clase *Glándula Patológica*, siendo $A \neq B \neq P$.

Para estudiar la capacidad discriminativa, en primer lugar, es necesario conocer la distribución de cada variable, ya que, si sigue una distribución normal, se compararán las medias de los valores de dicha variable asociadas a cada una de las tres clases. Por el contrario, si la variable no sigue una distribución normal, la comparación se llevará a cabo en base a la mediana, y no a la media. Por consiguiente, se implementa un contraste de hipótesis conocido como “test de *Kolmogorov-Smirnov*” y se realiza una prueba llamada “*one-sample K-S test*”, que permite comparar la distribución de la variable en cuestión con otra de referencia de media $\mu = 0$ y desviación típica $\sigma = 1$. Para implementar este contraste de hipótesis, primero se define un nivel de significancia α y se establecen una hipótesis nula H_0 y una hipótesis alternativa H_1 .

Concretamente en este TFM, se fija un nivel de significancia $\alpha = 0.0000001$ con el objetivo de que el nivel de confianza a la hora de aceptar o de rechazar una determinada hipótesis sea del 99,9999%. Por otra parte, la H_0 se formula tal que: la distribución de la variable V es normal. De esta forma, tras aplicar el *one-sample K-S test*, se obtiene un *p-value* que, si es menor que el nivel de significancia α , se determina que hay evidencias suficientes para rechazar la hipótesis nula, por lo que la variable no sigue una distribución $N(0,1)$ y, por tanto, se compararán las medianas de los valores asociados a cada una de las tres clases para decidir si la variable V es una característica interesante en términos de capacidad discriminativa.

En la *Figura 2.26* y en la *Figura 2.27* se exponen dos ejemplos donde se observa visualmente si las primeras 20 variables del estudio siguen una distribución $N(0,1)$, en función de si la nube de puntos se adapta a una línea recta especificada por los puntos rojos (en el caso de la *Figura 2.26*) o si el histograma de la variable concuerda con la campana de Gauss característica de una distribución normal (*Figura 2.27*).

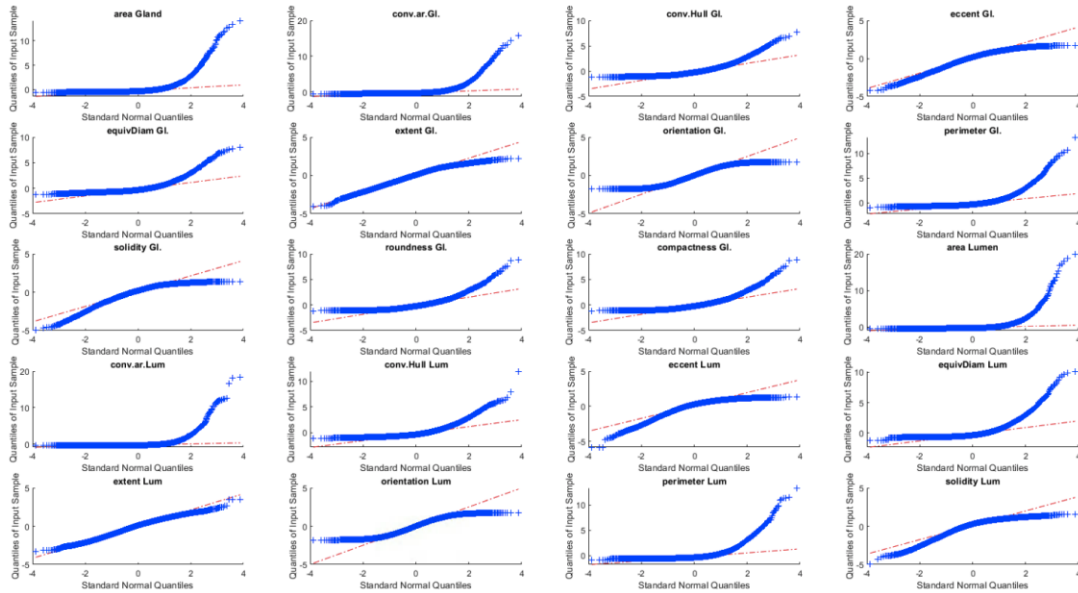


Figura 2.26. Gráfico que indica visualmente si las variables siguen una distribución $N(0,1)$ en base a la capacidad de adaptación de la nube de puntos azules sobre la línea roja.

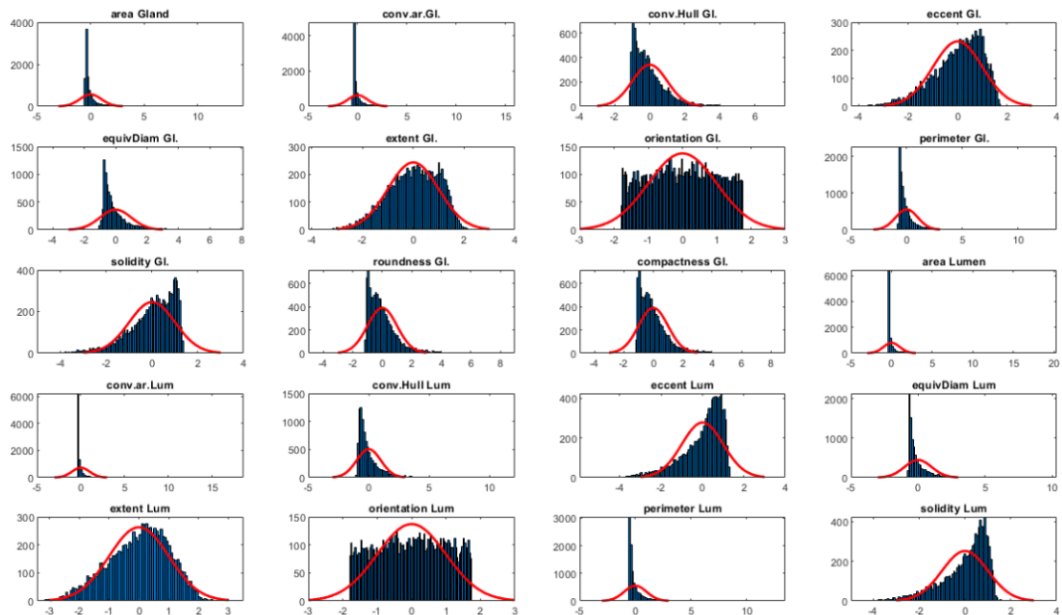


Figura 2.27. Gráfico que indica visualmente si las variables siguen una distribución $N(0,1)$ en base a la adaptación del histograma a la Campana de Gauss mostrada en rojo.

Según si se acepta o se rechaza la hipótesis nula, se lleva a cabo una comparación de medias o de medianas para determinar el nivel de independencia de cada variable con respecto a la clase. Concretamente, en este TFM se analiza el problema desde un punto de vista multiclase, pues hay que discriminar entre artefactos, glándulas benignas y glándulas patológicas, a partir de las mismas características. Para ello, se llevan a cabo las siguientes pruebas:

- **ANOVA** (del inglés, *ANalysis Of VAriance*). Se utiliza cuando la variable sigue una distribución normal. Esta técnica consiste en un tipo de prueba paramétrica que permite la comparación de las medias de los valores asociados a cada clase, para una determinada variable. De este modo, se establece un nuevo contraste de hipótesis, donde α sigue siendo igual a 0.0000001 y la H_0 defiende la igualdad de las medias. Lo que interesa es que se rechace la hipótesis nula, ya que, si las medias son iguales o parecidas, la variable tomará los mismos valores para las tres clases y, por tanto, dicha característica no será útil para discriminar. En ese caso, la variable en cuestión sería independiente de la clase y quedaría descartada del estudio.

- **Kruskal-Wallis**. Se utiliza cuando la variable no sigue una distribución normal. Este método es un tipo de prueba no paramétrica que consiste en una extensión de la prueba *U-Mann-Whitney*, con la diferencia de que, en este caso, se lleva a cabo una comparación de medianas para intentar discriminar entre tres clases, en lugar de dos. El contraste de hipótesis se realiza del mismo modo que en el método ANOVA.

Para analizar a nivel visual si una determinada variable presenta mayor o menor poder discriminativo, se exponen la *Figura 2.28* y la *Figura 2.29*, relativas a las primeras 10 variables.

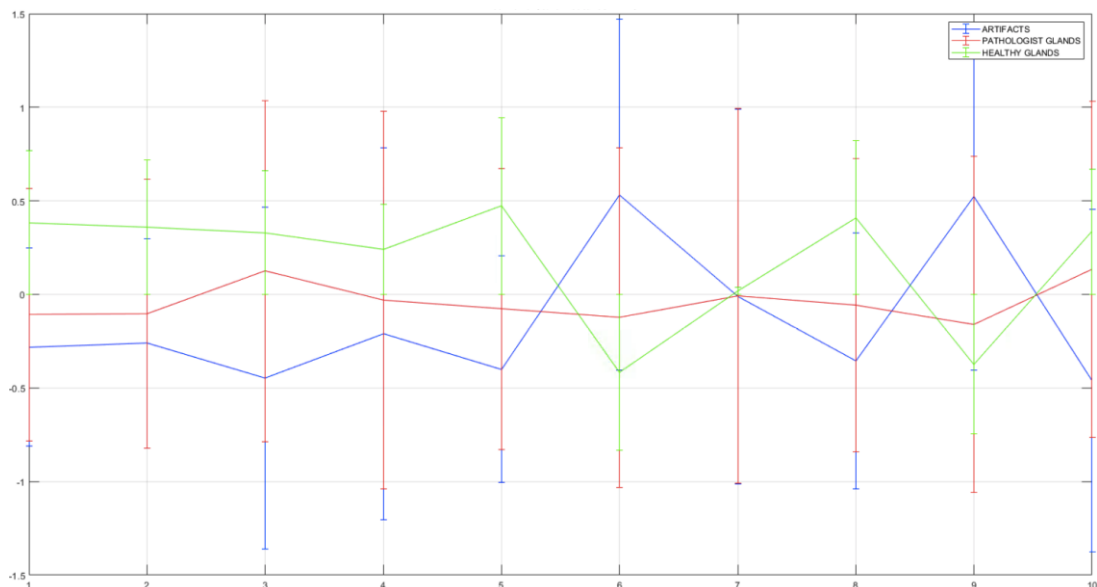


Figura 2.28. Gráfico de líneas que muestra visualmente la media y la desviación estándar de cada clase para cada una de las 10 características. En azul la clase *Artefacto*, en rojo la clase *Glándula Patológica* y en verde la clase *Glándula Benigna*.

En este gráfico se puede apreciar visualmente que la característica número 7 es independiente de la clase, ya que toma valores similares para las tres. Lógicamente, en ese caso, se acepta la hipótesis nula que sostiene la independencia de la variable con la clase y, por consiguiente, dicha variable es descartada del estudio por su escasa capacidad discriminativa.

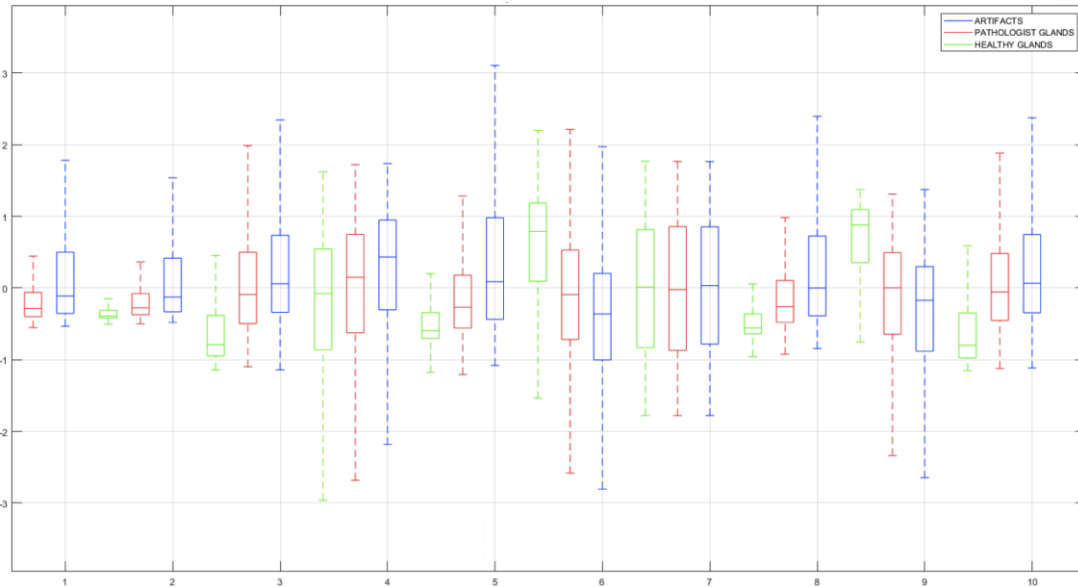


Figura 2.29. Diagrama de caja y bigotes que muestra visualmente el nivel de independencia de las diez primeras características del estudio con respecto a la clase. En azul la clase *Artefacto*, en rojo la clase *Glándula Patológica* y en verde la clase *Glándula Benigna*.

El diagrama de caja y bigotes es de gran utilidad para observar a nivel visual la capacidad discriminadora de las variables, ya que permite enfrentar para una misma característica los valores que toma en función de la clase. La interpretación de dicho diagrama se describe en la siguiente figura:

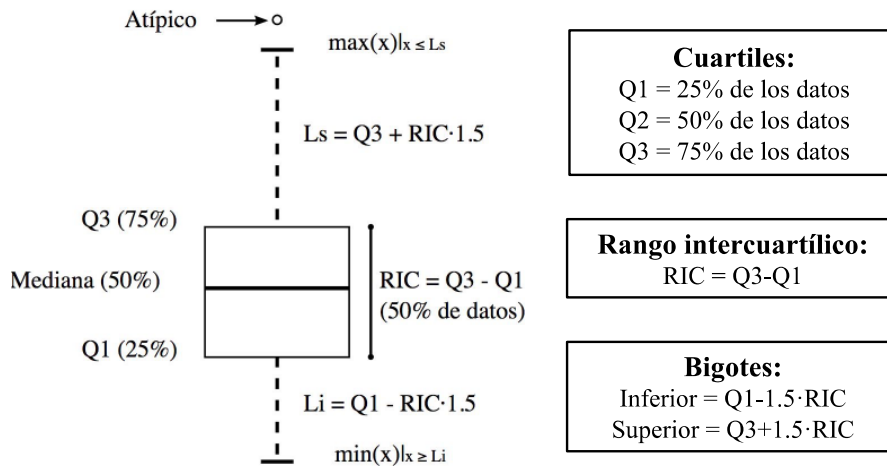


Figura 2.30. Esquema explicativo del diagrama de caja y bigotes [49].

Una vez estudiada la independencia de las variables con respecto a la clase, se analiza la independencia entre pares de variables con el objetivo de descartar aquellas que no aportan información nueva para el aprendizaje de los modelos de clasificación. La independencia entre pares de variables se analiza por medio de la correlación, la cual puede ser positiva o negativa. En este caso concreto, se trata la correlación en valor absoluto, pues lo que interesa es que las variables no estén correlacionadas sea cual sea el signo de la tendencia.

En la *Figura 2.31* se expone una matriz cuyas entradas son el coeficiente de correlación obtenido al enfrentar pares de variables. A nivel visual, los colores más rojos representan coeficientes de correlación más altos; y valores más azules, coeficientes de correlación más bajos. A nivel matemático, se lleva a cabo un contraste de hipótesis para determinar la independencia entre las variables. De nuevo, se fija el nivel de significancia como $\alpha = 0.0000001$ y la H_0 que mantiene la independencia entre variables. En este caso, a diferencia del anterior, interesa aceptar la hipótesis nula, ya que será indicativo de que las características extraídas aportan información nueva al estudio. El cálculo de la correlación proporciona dos valores: el *p-value* y un porcentaje que determina el grado de correlación. En función del *p-value* se acepta o se rechaza la hipótesis nula, es decir, se determina si dos variables están correlacionadas o no. No obstante, para este proyecto se define un umbral $t = 0.95$ para descartar no las variables que estén correlacionadas, sino las que estén correlacionadas por más de un 95%. Es decir, dos variables pueden estar correlacionadas en un 50% de los datos, pero el otro 50% es información que puede ser útil para el aprendizaje de los clasificadores. Por tanto, en lo que a independencia entre pares de variables se refiere, se descartarán únicamente aquellas características que aporten menos de un 5% de información nueva.

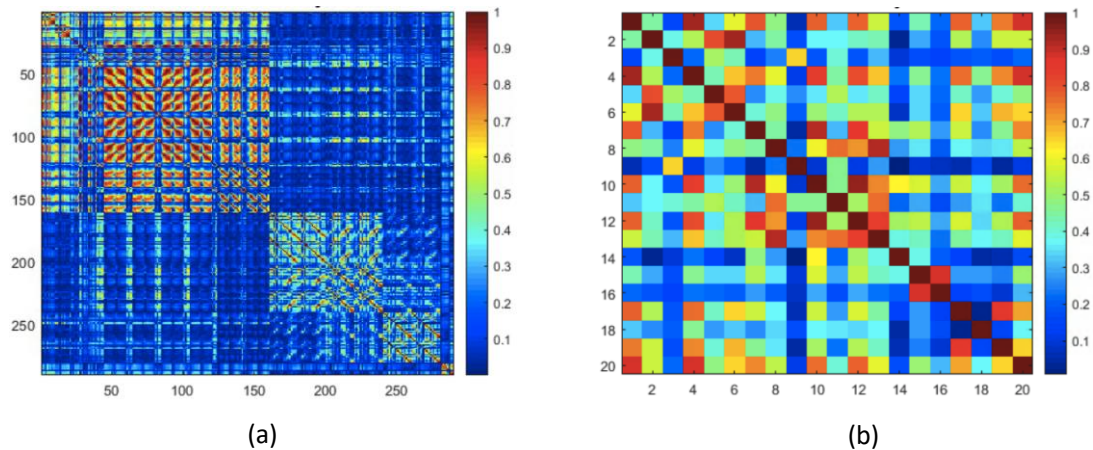


Figura 2.31. Matrices de coeficientes de correlación donde se analiza visualmente la independencia entre pares de variables. (a) Mapa de correlación relativo a las 297 características extraídas. (b) Mapa de correlación correspondiente a las 20 primeras características extraídas.

Al finalizar el proceso se obtienen 135 características que no han sido descartadas durante la etapa de cribado. De esta forma, se genera una matriz cuyo número de filas es el número de candidatos a glándula que se analizan y cuyo número de columnas es el número de características, más una columna adicional que viene determinada por la etiqueta de la clase a la que corresponde cada candidato. Con esta información, el algoritmo debe ser capaz de aprender qué valores de qué variables se asocian con qué clase. De esta forma, los clasificadores se entrenan para predecir en un futuro muestras de nuevos pacientes, a partir del cálculo de las mismas características.

2.2.6 Partición de datos

El objetivo de llevar a cabo una partición de datos consiste en garantizar la robustez de los modelos implementados. Es de vital importancia que las muestras usadas durante el entrenamiento de los modelos no sean también utilizadas para testarlos, ya que en ese caso se estarían proporcionando resultados muy buenos, pero erróneos. El algoritmo de aprendizaje debe ser capaz de predecir la clase de muestras nunca vistas anteriormente, a partir de lo aprendido con los datos de entrenamiento. Por tanto, lo que se hace es dividir los datos disponibles en diferentes conjuntos de datos tal que, unos se utilizan para entrenar y validar los modelos, y otros para testar su funcionamiento. A continuación, se presenta la metodología llevada a cabo para las diferentes aproximaciones de *machine learning* y *deep learning*.

2.2.6.1 Partición de datos adaptada para Machine Learning

En función de la cantidad de observaciones disponibles, las técnicas de partición de datos que se implementan son diferentes. Se listan a continuación las tres técnicas más populares:

- **Leave-one-out cross validation (LOOCV).** Consiste en una técnica que suele emplearse cuando la cantidad de datos disponibles es poca, ya que con esta aproximación se utilizan todas las observaciones para entrenar y solo se testea con una de ellas. Este proceso se repite tantas veces como observaciones hay, con el objetivo de que todas las muestras sean usadas para entrenar y para testear. Cabe destacar que en cada iteración la muestra del test no podrá ser utilizada como entrenamiento durante esa iteración.

- **Holdout method.** Se recomienda cuando existe una gran cantidad de datos. Esta técnica se caracteriza por llevar a cabo una única partición de datos aleatoria en dos conjuntos de datos (*training* y *test*). El tamaño de cada conjunto también es aleatorio, aunque normalmente el de entrenamiento suele ser mayor. La peculiaridad de este método es que solo se lleva a cabo una iteración, lo cual tiene el inconveniente de que los resultados serán muy dependientes de la aleatoriedad a la hora de dividir los datos en dos conjuntos de observaciones.

- **K-fold cross validation.** Este método por definición consiste en separar de forma aleatoria el conjunto de datos original en k conjuntos de las mismas dimensiones. De esta forma, un conjunto se utiliza para testear los modelos y $k-1$ conjuntos para entrenarlos y validarlos. El proceso se repite k veces para que todos los conjuntos pasen por la etapa de test y por la de entrenamiento, teniendo en cuenta que en cada iteración un conjunto solo puede pertenecer, bien al grupo de test, o bien al grupo de entrenamiento. Esta técnica también suele utilizarse cuando el conjunto original de datos del que se dispone es considerablemente grande.

En lo que se refiere al estudio de imágenes histopatológicas, la técnica más habitual que se observa en el estado del arte es *k-fold cross validation*, ya que permite crear una partición de los datos aleatoria que, al repetirla k veces, puede garantizar la robustez del modelo. Dentro de esta técnica el parámetro k puede tomar distintos valores, pero normalmente suele utilizarse $k=5$ o $k=10$. Cabe destacar que, normalmente, se implementan dos tipos de *cross validation*: uno externo (como se ha descrito) y otro interno. Este último consiste en llevar a cabo una nueva partición de datos, pero en este caso, con la finalidad de separar el conjunto de datos de entrenamiento (*training*) en otros dos subconjuntos: el de *train* y el de *validation*.

En la *Figura 2.32*, se expone un esquema a modo de explicación, donde se aprecia que para cada una de las 5 iteraciones del *cross validation* externo, se entrena con 4/5 de los datos y se prueba con 1/5. A su vez, los 4/5 de datos con los que se entrena se sub-dividen en 4/5 para *train* y 1/5 para *validation*, siendo este último el conjunto de datos utilizado para ajustar los hiperparámetros del clasificador. Cabe reseñar que, a diferencia del ejemplo de la figura, en este TFM se utilizan $k=5$ y $k=10$ *folds* para el *cross validation* externo e interno, respectivamente.

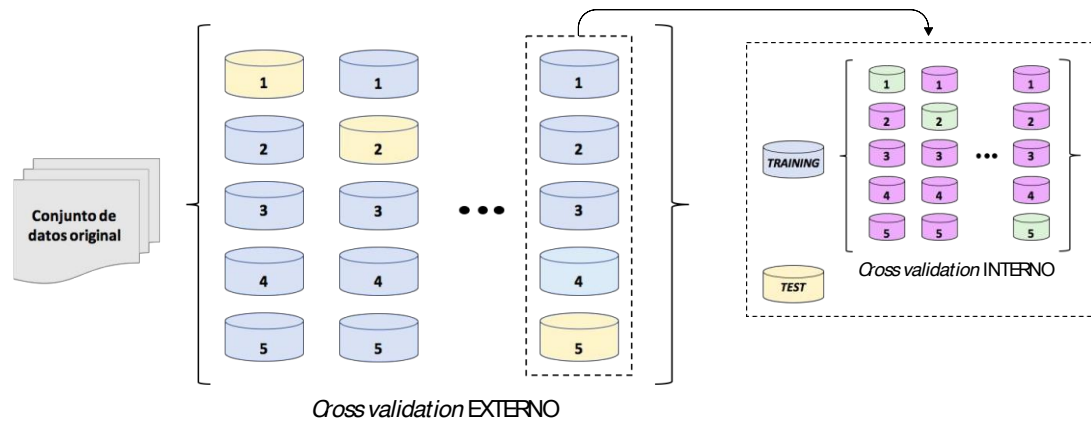


Figura 2.32. Esquema explicativo para una partición de datos basada en la técnica *k-fold cross validation* con $k = 5$ *folds*, tanto en la partición externa como en la partición interna. Con respecto al *cross validation* externo, las cajas azules representan los datos de *Training* y las naranjas, los de *Test*. En cuanto al *cross validation* interno, las cajas rosas representan los datos de *Train* y las verdes, los de *Validation*.

Por lo que respecta al conocimiento del autor, los métodos de partición de datos implementados en todos los estudios del estado del arte siguen un esquema similar al propuesto en la *Figura 2.32*. No obstante, en este proyecto, se ha pretendido ir más allá en dos aspectos:

1. **Clases balanceadas.** Para la realización de este TFM, la partición de datos no se ha llevado a cabo de forma totalmente aleatoria como establece por defecto la técnica *k-fold cross validation*. La peculiaridad en este caso es que se ha realizado una partición semialeatoria, asegurándose de que en cada conjunto de datos aleatorio (cajas amarillas y azules de la figura anterior) haya un número balanceado de candidatos correspondientes a cada clase. De esta forma, se garantiza que, a la hora de entrenar y testear un modelo, siempre habrá elementos de las tres clases, lo cual permite optimizar el aprendizaje de los algoritmos.

2. **Separación por historial clínico.** Este sub-apartado presenta uno de los puntos más novedosos implementados en todo el proyecto, con respecto al estado del arte. Se fundamenta en que, de cara a la robustez de los modelos, la partición de los datos no se realiza de forma meramente aleatoria, sino que se implementa un conjunto de algoritmos que permite separarlos en base a ciertas condiciones. Dichas condiciones residen en que todos los candidatos a glándula de un paciente con un historial clínico determinado deben pertenecer al mismo conjunto de datos. Es decir, una vez se dispone de todos los candidatos a glándula, no solo se dividen en diferentes conjuntos atendiendo a su clase, sino también atendiendo al paciente del que proceden. Esto conduce a una fiabilidad de los resultados mucho mayor, ya que, en todas las iteraciones, se testean las muestras de un paciente nuevo. La decisión de realizar esta partición de datos nace de la observación de que en cada paciente puede variar la nitidez a la

hora de teñir las imágenes histopatológicas, así como las estructuras glandulares de cada uno. Por tanto, aunque sería correcto entrenar y validar los candidatos a glándula de forma aleatoria (siempre y cuando no se valide con los mismos datos con los que se ha entrenado) la aproximación de separación por historial clínico proporciona una robustez de los modelos y una fiabilidad de los resultados mucho mayor.

Para entender en mayor profundidad lo expuesto en el punto 2, se puede abordar el problema desde una perspectiva como la que se muestra en la *Figura 2.33*. En dicha figura se realiza un símil con las conocidas muñecas rusas (*matrioskas*) para entender los niveles de organización que presenta el problema desde que se extrae el tejido biopsiado del paciente hasta conseguir las sub-imágenes de glándulas con las que se trabaja en este proyecto. De esta forma, se observa que cada paciente tiene un historial clínico (lo que sería la *matrioska* superior). Tras realizar las biopsias, se obtienen varios cortes histológicos como los que se representan en la segunda *matrioska*. Puesto que esos cortes proporcionan imágenes de dimensiones excesivamente grandes para su manipulación, se “trocean” y se da lugar a sub-imágenes de 1024x1024x3 como la que se observa en la tercera *matrioska*. Por último, en cada sub-imagen de este tipo, pueden encontrarse candidatos a glándula susceptibles de ser una glándula benigna, una glándula patológica o un artefacto. En la *matrioska* más pequeña se observa ese tipo de elementos de forma individual.



Figura 2.33. Modelo explicativo para entender la procedencia de los datos disponibles y para comprender el proceso de partición de datos implementado.

En base a lo expuesto, se entiende que los modelos son mucho más robustos si todos los candidatos a glándula (*matrioska* más pequeña) de un mismo historial clínico (*matrioska* más grande) se fijan en el mismo conjunto de datos. De este modo, nunca se testea ni se valida un candidato a glándula perteneciente al mismo paciente con cuyas muestras se ha entrenado.

En resumen, para la aproximación basada en *machine learning*, el presente trabajo fin de máster aborda la partición de datos de forma exhaustiva con el objetivo de garantizar la robustez de los modelos y la fiabilidad de los resultados proporcionados, mediante técnicas de *k-fold cross validation* externo ($k = 5$) e interno ($k = 10$). Además, dicha partición se lleva a cabo de forma semialeatoria asegurando que en cada subconjunto de datos hay una cantidad de muestras balanceadas de cada clase y que las muestras pertenecientes a un mismo paciente nunca pueden estar en dos subconjuntos de datos diferentes.

2.2.6.2. Partición de datos adaptada para Deep Learning

Para llevar a cabo la partición de los datos a fin de clasificar con técnicas de *deep learning*, se implementa un *k-fold cross validation* externo muy similar al realizado para la aproximación de *machine learning*. Es decir, se llevan a cabo 5 separaciones de datos tal que, en cada separación, 4/5 de los datos se utilizan como entrenamiento; y 1/5, como test. La diferencia con respecto al *machine learning* es que en este caso no se realiza un *cross validation* interno, sino que se lleva a cabo una única partición aleatoria del conjunto de entrenamiento (*training*) para separarlo en dos subconjuntos, donde el 90% de los datos (*train*) se utiliza para entrenar los modelos y el 10% (*validation*), para validarlos. (Véase la *Figura 2.34*).

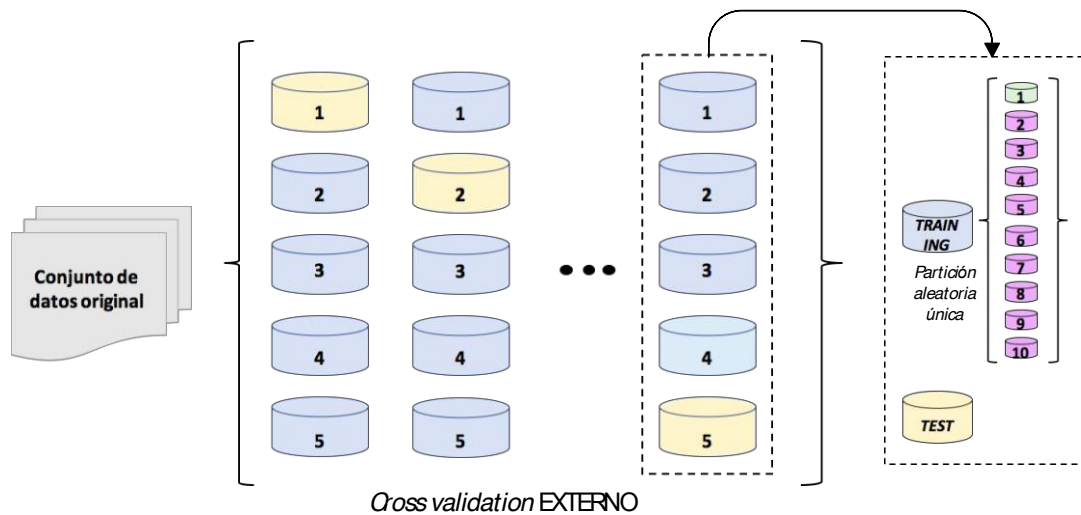


Figura 2.34. Partición de datos llevada a cabo para la aproximación basada en *deep learning*, donde las cajas azules representan los datos de entrenamiento (*training*) y las cajas naranjas los de test. El 90% (en morado) y el 10% (en verde) de los datos de *training* se utilizan para entrenar y validar los modelos, respectivamente.

2.2.7 Clasificación supervisada

La clasificación supervisada consiste en la construcción de unos modelos que son entrenados a partir de ejemplos con unas características determinadas que se asocian a una clase concreta. De esta forma, los modelos de clasificación pueden ser capaces de predecir la etiqueta de nuevas muestras a partir del cálculo de las mismas características seleccionadas durante el entrenamiento. Concretamente, este TFM aborda la tarea de clasificación desde un punto de vista multiclase, ya que el objetivo es discriminar entre artefactos, glándulas benignas y glándulas patológicas de grado 3, según la escala *Gleason*.

Para ello, y siguiendo el apartado anterior, se construyen cinco modelos de predicción correspondientes a los *k=5 folds* llevados a cabo durante la partición de datos. Por consiguiente, en cada iteración se elabora un modelo predictivo durante la fase de *Training* y se analiza el funcionamiento de dicho modelo durante la fase de *Test*.

Cabe recordar que en este proyecto se utilizan tanto estrategias de *machine learning* como de *deep learning* para la tarea de clasificación. El objetivo es, primero, analizar qué aproximación proporciona mejores resultados para las mismas imágenes y, segundo, compararse con el estado del arte actual en la medida de lo posible. A continuación, se explica en detalle cada una de las estrategias implementadas para llevar a cabo la clasificación supervisada.

2.2.7.1 Aproximación basada en Machine Learning

En esta etapa, se aborda la tarea de clasificación desde una perspectiva lineal y otra no lineal aplicando métodos ampliamente usados en la literatura [14]. Por una parte, se hace uso de técnicas de *Support Vector Machine* (SVM) con un *kernel* lineal, para abordar el enfoque lineal, y con un *kernel* de segundo orden para el enfoque no lineal. Asimismo, se emplean técnicas de *Feedforward Neural Network* (FNN), también conocidas como “perceptrón multicapa (MLP)”, con el fin de comparar el rendimiento de las dos aproximaciones no lineales.

La principal diferencia entre los clasificadores basados en SVM y los basados en MLP es que los primeros son de tipo no paramétrico y definen funciones bases centradas en determinados puntos para después seleccionar uno de ellos durante la fase de entrenamiento. En cambio, FNN utiliza formas paramétricas para definir funciones base donde los valores de los parámetros pueden adaptarse en el entrenamiento [50]. Además, SVM es un clasificador formulado originalmente para resolver problemas binarios, por lo que para abordar un enfoque multiclase es necesario dividir el problema inicial en sub-problemas binarios; mientras que FNN permite llevar a cabo la clasificación de todas las clases de forma simultánea [51].

Support Vector Machine (SVM)

Los modelos de SVM pretenden hallar un “hiperplano óptimo” que separe en dos regiones el espacio muestral, siendo el hiperplano óptimo aquel que se genera cuando la distancia entre los dos puntos más cercanos de cada clase al hiperplano es máxima [52]. (Véase la *Figura 2.35*). Por consiguiente, el procedimiento del clasificador SVM consiste, en primer lugar, en definir diferentes hiperplanos buscando aquellos que separen mejor los conjuntos de datos de cada clase (*Figura 2.35.a*); y, en segundo lugar, en buscar el hiperplano que presente el mayor margen, es decir, aquel cuya distancia a los dos puntos más próximos de cada clase sea máxima. Se consideran “vectores soporte” aquellos puntos que están en contacto con el límite del margen creado por el hiperplano óptimo (*Figura 2.35.b*).

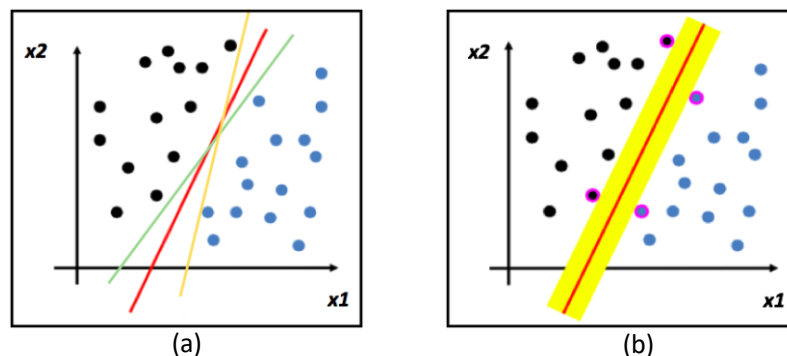


Figura 2.35. a) Definición de diferentes hiperplanos que separen los datos de cada clase. (b) Hiperplano óptimo con los vectores soporte resaltados en rosa.

El modelo explicado (LSVM) es el más sencillo, ya que los datos son linealmente separables. No obstante, en este proyecto también se hace uso del QSVM empleando un *kernel* cuadrático para abordar el problema en el caso de que los datos no se puedan separar de forma lineal. Los *kernels* permiten proyectar un espacio muestral D -dimensional a otro M -dimensional, donde $M > D$ tal que, la función de la proyección de los datos se define como $\phi: \mathbb{R}^D \rightarrow \mathbb{R}^M$. De esta forma, se buscan los vectores soporte en el espacio original y luego se hace una transformación a una dimensión mayor donde los datos sí son separables. (Véase la *Figura 2.36*).

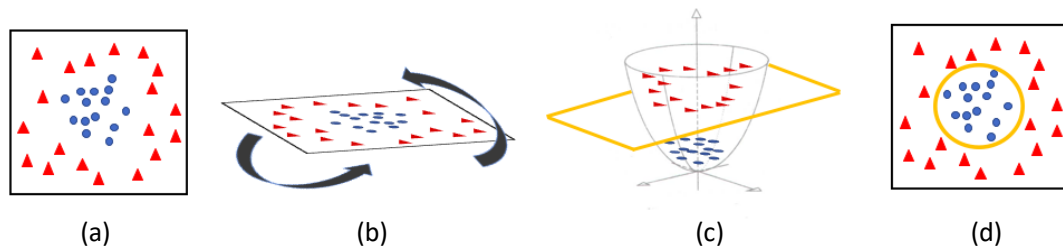


Figura 2.36. Ejemplo de clasificación no lineal utilizando una transformación al espacio tridimensional para poder separar linealmente los datos. (a) Observaciones originales de dos clases diferentes. (b) Rotación del plano para analizar los datos desde otra perspectiva. (c) Transformación al espacio tridimensional construyendo un paraboloide a partir del plano anterior, y fijando un hiperplano donde los datos en ese momento sí son linealmente separables. (d) Resultado final de la frontera de separación sobre la dimensión original bidimensional.

Para llevar a cabo la tarea de clasificación basada en SVM, se parte del modelo generado por una app disponible en el software de Matlab al instalar una *toolbox*²⁰ llamada *Statistics and Machine Learning*. Esta *toolbox* tiene diferentes funcionalidades, pero para este proyecto se utiliza una aplicación conocida como *App Classification Learner*, la cual incluye algoritmos para aprendizaje automático supervisado que permiten modelizar una variable de respuesta categórica en forma de función de uno o más predictores [53]. Dicha aplicación proporciona un código en la que ciertos parámetros son definidos por defecto, como por ejemplo los parámetros *BoxConstraint* que se fija en 1, y *KernelScale* que se define como ‘auto’. En este proyecto, se modifica el código generado por la aplicación de *Classification Learner* para cambiar los valores de algunos parámetros de entrada, tras aplicar un algoritmo de optimización de hiperparámetros. El funcionamiento de dicho algoritmo consiste en dividir los datos de entrenamiento en dos subconjuntos (*train* y *validation*) a partir de los cuales, intentar minimizar el error de una función objetivo, en base a unos determinados parámetros de entrada. En primer lugar, se fija un número de épocas para el aprendizaje (30 en este proyecto) y, después, en cada una de ellas, se modifican los parámetros intentando disminuir el error de la función.

Finalmente, se seleccionan los valores de los parámetros de la época que presentaba el menor error al intentar minimizar la función objetivo. Además, dado que en este TFM se aborda un problema de clasificación multiclase y que los modelos basados en SVM son por defecto binarios, es necesario especificar ciertas técnicas que permitan la comparación entre más de dos clases. En la literatura científica, una de las técnicas más populares es la “binarización”, que

²⁰ *Toolbox*. Entorno de desarrollo que permite introducir la programación en materias sin competencias informáticas.

consiste en encarar la tarea multiclase dividiendo la base de datos original en 2 subconjuntos e implementar un clasificador binario para cada uno [54]. Sin embargo, actualmente están muy en boga las técnicas basadas en *OneVsOne* (OvO) y *OneVsAll* (OvA) para abordar el problema de la clasificación multiclase [51]. (Véase la *Figura 2.37*).

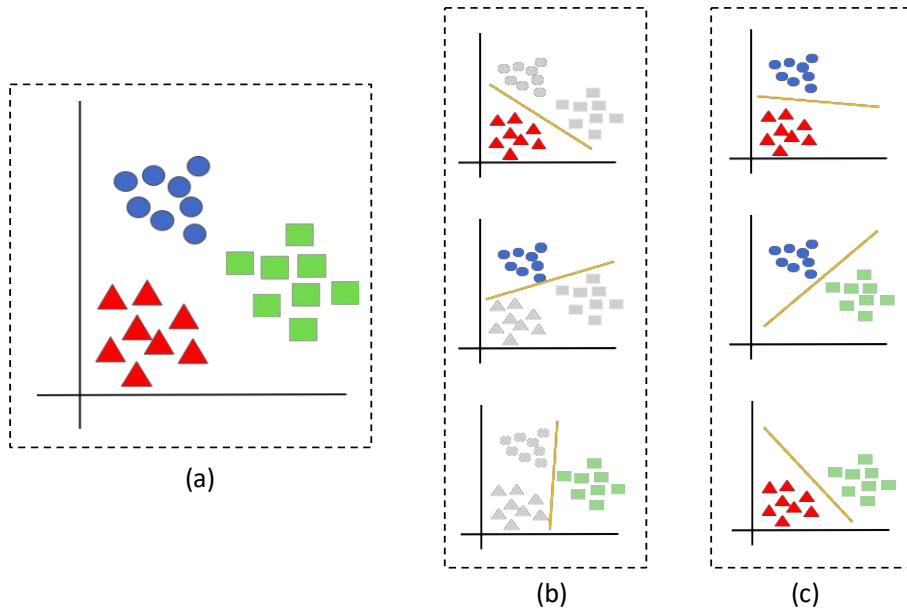


Figura 2.37. (a) Problema original de clasificación multiclase. (b) Enfoque *OneVsAll* (OvA) para discriminar una clase con respecto a las demás. (c) Enfoque *OneVsOne* (OvO) para considerar únicamente una clase contra otra.

En el estado del arte se encuentran estudios abordados desde ambas técnicas. Por ejemplo, en [55] se utilizan clasificadores SVM con la técnica OvO para distinguir entre muestras de grado 3, 4 y 5, según la escala *Gleason*, pero no aportan ninguna información sobre artefactos ni patrones benignos. Además, muchos trabajos como [14] y [56] no utilizan muestras de grado 5 en sus estudios debido a que presentan unas características fácilmente reconocibles con respecto al resto de grados. Por otra parte, en [57] y [58] utilizan la técnica OvA para la clasificación multiclase. En el primero se distinguen 7 clases de acuerdo a la escala *Gleason* y en el segundo, únicamente entre los grados 3, 4 y 5.

Concretamente, en este proyecto, la estrategia para seleccionar el tipo de clasificación binaria a partir del clasificador SVM multiclase, se lleva a cabo con el algoritmo de optimización de hiperparámetros mencionado anteriormente. En el capítulo 3.2.1 se expone una figura con los resultados que propone dicho algoritmo, siendo uno de ellos el “*coding*”, que permite decidir el enfoque óptimo de clasificación multiclase basado en OvO o en OvA.

Perceptrón multicapa (MLP)

Por otra parte, como ya se ha comentado, se emplean redes neuronales de tipo *feedforward* (conectadas totalmente hacia delante) que, según la literatura, constituyen los mejores modelos en cuanto a reconocimiento de patrones se refiere [50]. Los modelos de perceptrón multicapa (MLP) se caracterizan por ser adaptativos, distribuidos y paralelos. Se basan en un conjunto de regresiones logísticas donde la salida de una es la entrada de otra.

El elemento básico del perceptrón multicapa es lo que se conoce como “neurona”, la cual se compone de una función de agregación y de una función de activación no lineal que suele ser la tangente hiperbólica [50]. En la *Figura 2.38*, se expone un modelo típico de arquitectura MLP donde se observa el símil realizado con el comportamiento neuronal biológico.

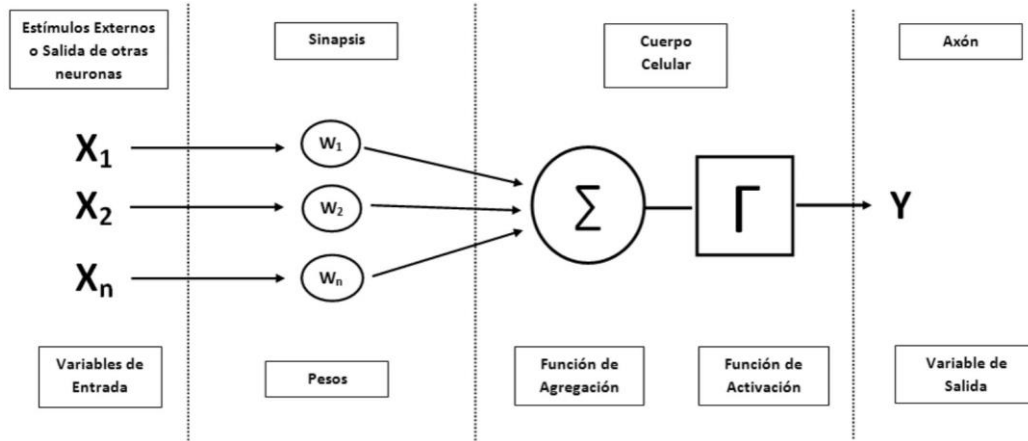


Figura 2.38. Arquitectura típica de perceptrón multicapa que expone la analogía con el funcionamiento de las neuronas biológicas [59].

Las neuronas se disponen en una serie de capas, siendo la primera la capa de entrada, la última la capa de salida, y las intermedias las capas ocultas. Según el número de neuronas y el número de capas, se obtiene una determinada red neuronal. (Véase la *Figura 2.39*).

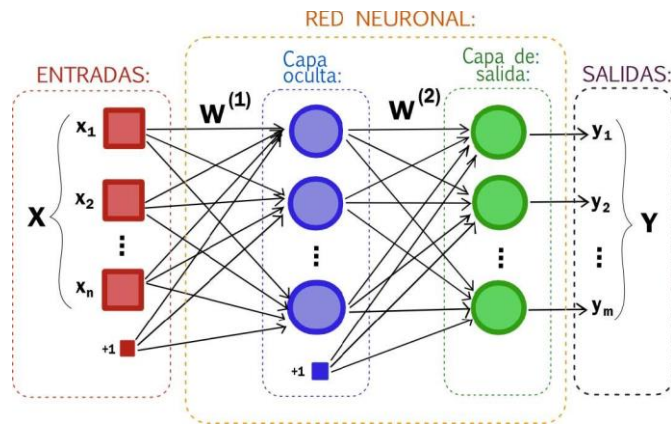


Figura 2.39. Modelo de red neuronal con n entradas x_i , una capa oculta y m salidas y_j donde se observan los pesos w_i y las neuronas (círculos azules) [60].

El funcionamiento del algoritmo de aprendizaje se basa en un modelo de retropropagación del error que sigue métodos de optimización de descenso por gradiente. Es decir, dado un conjunto de observaciones $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, donde x_i son las variables de entrada, e y_i el vector de salida (*outputs*), se debe minimizar la función de error:

$$E(w) = \frac{1}{N} \sum_{i=1}^N ||y_i - t_i||^2, \tag{2.15}$$

donde t_i es el valor (*target*, etiqueta verdadera de la observación) de la red neuronal con los pesos w para la entrada x_i .

El objetivo, por tanto, es obtener un vector de pesos w que minimice el error $E(w)$ que comete el perceptrón multicapa. (Véase la *Figura 2.40*).

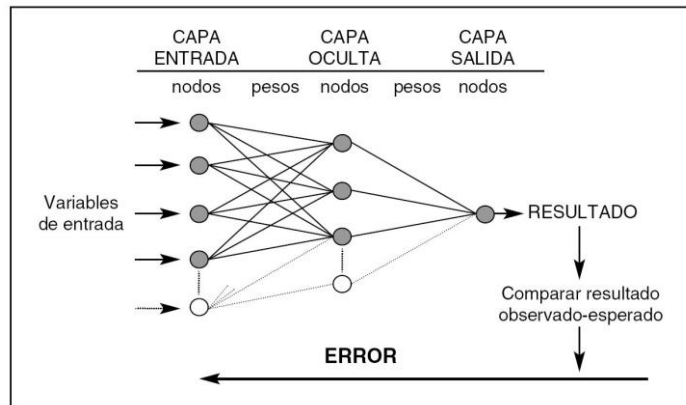


Figura 2.40. Esquema explicativo del modelo de retropropagación del error donde se pretende minimizar la función objetivo de la ecuación 2.15 [50].

Cabe destacar que las redes neuronales (MLP) con dos capas ocultas son capaces de establecer cualquier mapeo entre dos conjuntos de datos. Esto presenta una ventaja y a la vez un inconveniente, ya que, por un lado, se dispone de un modelizador universal, pero, por otro lado, dicho modelizador es demasiado flexible, lo cual puede dar lugar a *overfitting*. Es decir, puede darse una situación en la que el modelo se ajuste demasiado a los datos de entrenamiento, de tal forma que se empeore la capacidad predictiva y aumente el error en la validación. Por ejemplo, en la *Figura 2.41.a* la línea verde se adapta muy bien a los datos de entrenamiento separando las dos clases existentes, sin embargo, la línea negra sería capaz de predecir mejor la etiqueta de nuevos datos.

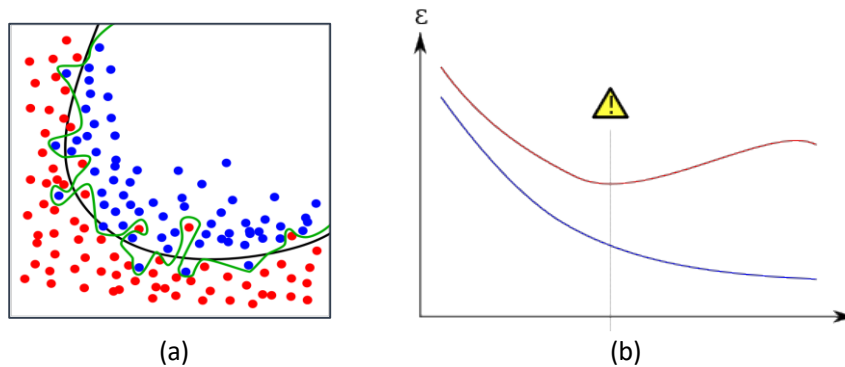


Figura 2.41. (a) *Overfitting* al implementar la clasificación propuesta por la línea verde. (b) Situación de *overfitting* que se produce cuando el error del conjunto de validación (curva roja) aumenta y el error del conjunto de entrenamiento (curva azul) disminuye [61].

En base a lo expuesto, en este trabajo fin de máster se desarrolla un MLP con una única capa oculta y 15 neuronas, con la finalidad de evitar el sobreajuste. Además, para llevar a cabo el aprendizaje de los modelos durante el entrenamiento, se implementa el algoritmo del gradiente conjugado escalado (*Scaled Conjugate Gradient*) que permite actualizar los pesos en cada época con el fin de minimizar el error de la función objetivo de la ecuación 2.15.

Cabe destacar que se han modificado determinados parámetros que vienen prefijados al implementar una red neuronal de este tipo. En primer lugar, es necesario cambiar la estrategia de separación de los datos de entrada, ya que, por defecto, la red utiliza una partición aleatoria en la que el 70% de los datos se usan para entrenar, el 15% para validar y el otro 15% para testear. En este caso, las entradas del modelo únicamente son los datos de entrenamiento, ya que son los que se utilizan para su aprendizaje. Por tanto, se modifican los parámetros para crear un *cross validation* interno que únicamente separe los datos de entrada en *train* y *validation*, en proporciones de 90% y 10%, respectivamente, de la misma forma que se hacía para los modelos de clasificación abordados con técnicas de SVM.

Por otra parte, se inicializa por defecto un número de épocas igual a 1.000 y un número máximo de fallos igual a 6. De esta forma, en cada época, la curva de validación (como la mostrada en la *Figura 2.41.b*) intenta disminuir el error aprendiendo del entrenamiento, pero si durante 6 épocas no es capaz de aprender y reducir el error, el algoritmo se detiene y se proporcionan los valores de la mejor época. En este trabajo, se modifican los parámetros tal que, el número de épocas se define en 200 y el número permitido de fallos se fija en 10.

Otro parámetro que se modifica es la función objetivo a minimizar, en lo relativo al algoritmo de retropropagación. Por defecto, se calcula el rendimiento de la red intentando minimizar la entropía cruzada (*cross-entropy*), cuya función se expone en la siguiente ecuación:

$$CE = \frac{1}{N} \sum_{i=1}^N -t_i * \log(y_i) \quad (2.16)$$

donde CE es la entropía cruzada, y_i es el vector de salida (*outputs*) y t_i es el valor (*target*) de la red neuronal para cada entrada. En este proyecto, se sustituye la entropía cruzada por el error cuadrático medio para evaluar el rendimiento de la red (ecuación 2.15).

En definitiva, se implementa una *feedforward neural network* con 135 variables de entrada correspondientes a las características seleccionadas y una capa oculta con 15 neuronas que proporciona una salida con tres tipos de etiquetas relativas a las tres clases que se pretende clasificar en este proyecto. En la *Figura 2.42* se muestra la arquitectura de la red neuronal implementada.

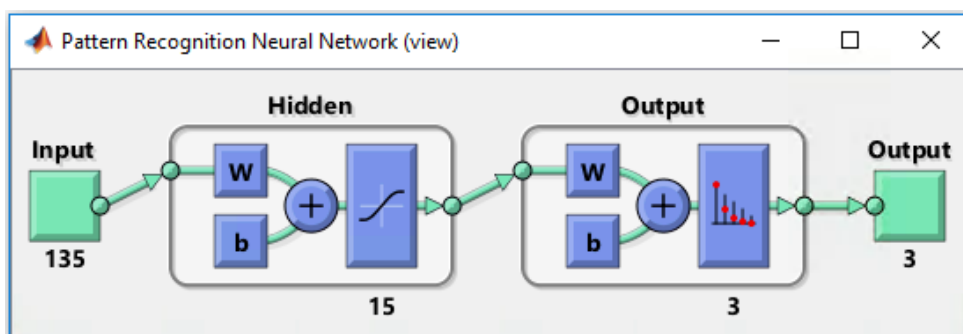


Figura 2.42. Modelo de red neuronal implementado para abordar la clasificación utilizando técnicas de *machine learning*.

2.2.7.2 Aproximación basada en Deep Learning

A diferencia de las estrategias basadas en *machine learning*, donde se ha hecho uso de diferentes descriptores para transformar la información visual percibida en las imágenes en datos numéricos, en esta fase se implementan técnicas de aprendizaje profundo (*deep learning*) para extraer automáticamente las características más relevantes de una imagen sin necesidad de seleccionar manualmente los descriptores óptimos, ya que el éxito de la clasificación está condicionado por el tipo de descriptores y características que se utilizan para extraer la información más relevante de las imágenes.

En lo que a imágenes histopatológicas de próstata se refiere, existen recientes trabajos en el estado del arte donde los autores aplican técnicas de *deep learning* para abordar diferentes tareas. En [62] se hace uso de dos redes neuronales convolucionales (CNN, del inglés – *Convolutional Neural Network*) una para detectar cada núcleo individual en las imágenes, y otra, para clasificarlos. Por otra parte, en los estudios llevados a cabo en [63] y en [64] se utilizan técnicas de *deep learning* para clasificar las imágenes en base a los diferentes grados de acuerdo al sistema de clasificación de *Gleason*. Concretamente en este TFM se hace uso de una arquitectura de CNN con el objetivo de abordar un problema de clasificación multiclase y distinguir de manera automática los tres tipos de candidatos a glándula con los que se trabaja: artefactos, glándulas benignas y glándulas patológicas; y poder comparar los resultados proporcionados con los obtenidos tras aplicar las técnicas de *machine learning*.

En los últimos años, han cobrado mucha importancia las redes neuronales convolucionales, ya que suponen una potente herramienta en los campos de *computer vision* a la hora de llevar a cabo tareas de caracterización y reconocimiento de patrones. La primera arquitectura de red neuronal convolucional que apareció fue el modelo conocido como “*LeNet*” [65] cuya aplicación se basaba en el reconocimiento automático de patrones. En 2012 se partió de la idea de utilizar una gran cantidad de convoluciones y de operaciones paralelas para dar lugar a la arquitectura conocida como “*AlexNet*” [66], donde se repetían las operaciones de convolución múltiples veces de forma paralela con el fin de aprender características cada vez más complejas.

El principal potencial de los modelos basados en CNN reside en su capacidad para extraer características altamente discriminantes utilizando varios niveles de abstracción. No obstante, este tipo de redes requiere previamente una gran cantidad de datos etiquetados para el aprendizaje de los modelos, además de equipos informáticos de alto rendimiento, tales como GPUs o *clusters* distribuidos a gran escala [67].

Por lo que respecta a las arquitecturas de red de tipo CNN, se requiere una imagen de entrada de dimensiones $M \times N \times 3$ para generar a la salida una cierta probabilidad de que la imagen de entrada pertenezca a una determinada clase. Internamente, las CNNs están compuestas por múltiples capas ocultas tanto lineales como no lineales que se pueden agrupar en cuatro: capas convolucionales, capas de *pooling*, capas totalmente conectadas (FC) (del inglés, *fully connected*) y capas de normalización, tal y como se observa en la *Figura 2.43*.

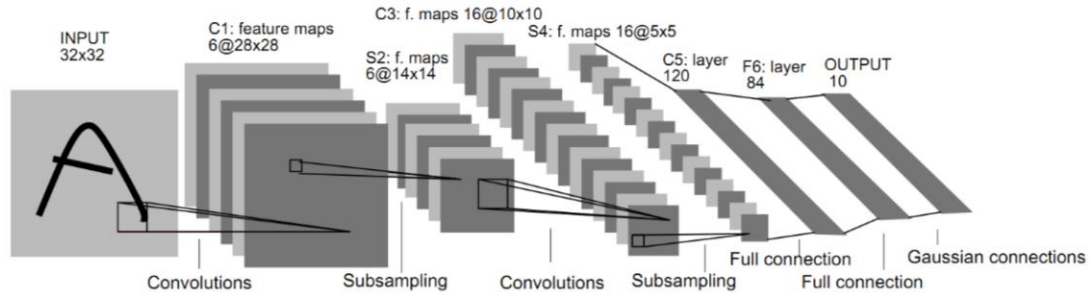


Figura 2.43. Típica arquitectura de red convolucional básica [68].

Cuando se diseña inicialmente una arquitectura de una CNN, la red no tiene ningún tipo de conocimiento, ya que los pesos establecidos son totalmente aleatorios. Dicho conocimiento lo adquiere aprendiendo a partir de los ejemplos que proporcionan las muestras etiquetadas, lo cual constituye la base de la etapa de entrenamiento de una red neuronal. El proceso de aprendizaje durante el entrenamiento puede dividirse en cuatro etapas: propagación hacia adelante, función de pérdidas, retropropagación y actualización de los pesos.

La idea es que, inicialmente la CNN toma unos pesos aleatorios en los filtros de las capas convolucionales, por lo que no pueden llevar a cabo una predicción correcta de la imagen de entrada en la primera iteración. No obstante, las salidas de cada iteración se comparan con los datos etiquetados del *groundtruth* y, a partir de dicha información, comienza el aprendizaje de la red mediante la actualización de los pesos de los filtros de acuerdo a una determinada función de pérdidas y al paso de retropropagación. De esta forma, el objetivo en cada iteración del entrenamiento es minimizar la función de pérdidas buscando los pesos óptimos para los filtros de las capas convolucionales. Cabe destacar que la función de pérdidas puede ser definida de varias maneras, pero en este caso se utiliza la medida del error cuadrático medio, igual que en la aproximación de *machine learning* basada en el perceptrón multicapa (MLP).

Por otra parte, los pesos varían en la dirección opuesta del gradiente y se actualizan según la siguiente ecuación:

$$\omega = \omega_i - \eta \frac{dL}{dW} E_{total} \quad (2.17)$$

donde ω_i son los pesos iniciales, ω son los pesos actualizados y η es la tasa de aprendizaje, que es un parámetro de diseño de la red relativo a los saltos del proceso de actualización tal que, un valor muy elevado de la tasa de aprendizaje, permite que el modelo converja en un tiempo menor. No obstante, si toma valores excesivamente grandes, puede que no sea posible alcanzar el mínimo global de la función de error, donde los pesos son óptimos.

Cabe destacar que el principal inconveniente que presenta el uso de las redes convolucionales es la cantidad de datos etiquetados disponible para llevar a cabo el aprendizaje de los modelos, ya que el procedimiento para conseguir muestras histológicas de próstata etiquetadas manualmente por parte de los patólogos es una tarea muy tediosa que consume mucho tiempo. Para solucionar este inconveniente, surgió una potente técnica llamada "*transfer learning*" [69] que consiste en un proceso en el que, a partir de un modelo de red pre-entrenado con una gran cantidad de datos genéricos, se soluciona un problema de clasificación específico utilizando una cantidad de datos reducida. Concretamente, en el presente trabajo fin

de máster se hace uso de una de las arquitecturas de redes convolucionales más populares en el estado del arte llamada “*Very Deep Convolutional Networks for Large-Scale Image Recognition*” (VGG19) [66] que fue entrenada utilizando el conjunto de datos *ImageNet* y los pesos de los filtros establecidos (Véase la *Figura 2.44*). De esta forma, utilizando la técnica de transferencia de conocimiento conocida como *fine-tuning*, todas las capas se inicializan con unos pesos pre-entrenados, excepto las últimas, sobre las que se lleva a cabo un re-entrenamiento con el fin de trasladar la información de las muestras específicas del problema bajo estudio.

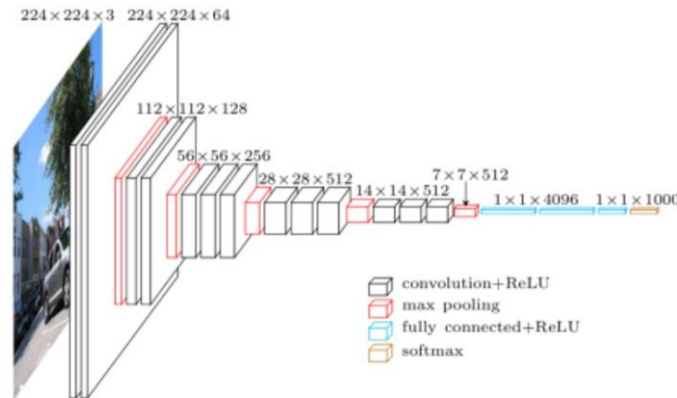


Figura 2.44. Arquitectura de la red neuronal convolucional basada en el modelo VGG19 [70].

Por último, en cuanto al valor de los parámetros introducidos para los algoritmos basados en *deep learning*, se han establecido: un número de épocas igual a 150, una tasa de aprendizaje de 1×10^{-4} , un criterio de parada de 15 épocas para evitar el *overfitting* en caso de que la precisión del modelo no mejore, y una arquitectura de CNN llamada VGG19, cuyo optimizador se basa en el gradiente estocástico descendente.

2.2.8 Creación del *groundtruth*

Para este trabajo fin de máster es necesario llevar a cabo dos tipos de *groundtruth*: uno para la parte de segmentación de las glándulas, y otro para la tarea de clasificación. El primero se utiliza para validar la técnica de segmentación *locally constrained watershed transform* propuesta en este proyecto, ya que consiste en una técnica novedosa que, por lo que respecta al conocimiento del autor, nunca antes había sido aplicada en imágenes histopatológicas de este tipo. Por otra parte, la construcción del *groundtruth* para la clasificación es determinante, ya que se aplica en diferentes puntos del proyecto, no solo para la validación de los modelos. En primer lugar, se requiere del *groundtruth* de clasificación en la parte de la selección de características, puesto que es necesario conocer la etiqueta para saber qué variables tienen una mayor capacidad de discriminación según qué clase. Por otra parte, el *groundtruth* de clasificación es de vital importancia durante la etapa de clasificación para el aprendizaje de los modelos tanto en *machine learning* como en *deep learning*, pues estos necesitan conocimiento previo basado en ejemplos etiquetados. Y, por último, el *groundtruth* se utiliza para validar los modelos durante la etapa de entrenamiento y evaluar los resultados finales durante la fase de test para cada uno de los cinco modelos implementados. Por tanto, a continuación, se explicará cada tipo de *groundtruth* por separado.

2.2.8.1 Groundtruth para la segmentación

Este procedimiento consiste en llevar a cabo una segmentación manual de las glándulas de una imagen histopatológica con el objetivo de evaluar el algoritmo de segmentación automática implementado en este proyecto. Lo ideal sería que las imágenes originales fueran segmentadas manualmente por expertos en anatomía patológica, ya que los resultados serían más fiables. No obstante, dicha tarea es una utopía, por lo que en este proyecto la segmentación manual ha sido realizada por miembros del grupo CVBLab en base al conocimiento aportado por los patólogos del Hospital Clínico de Valencia (HCV), colaboradores en el proyecto SICAP, y en base al conocimiento adquirido durante la carrera de Ingeniería Biomédica. Cabe destacar que dicha segmentación manual ha sido posteriormente validada por los patólogos del HCV.

Para llevar a cabo la segmentación manual, se ha hecho uso de una aplicación de Tablet que fue desarrollada por ingenieros del CVBLab como una de las fases del proyecto europeo "MINERVA" y cuya función era facilitar la tarea de marcado manual de las fronteras de las glándulas histológicas de colon. En este caso, es igualmente aplicable para las imágenes de la próstata. El procedimiento consiste en rodear con un lápiz las glándulas que aparecen en las imágenes de 1024x1024x3. A partir de ahí, se genera una máscara de las glándulas que se comparará con la máscara proporcionada por los algoritmos de segmentación automática para evaluar la bondad de la técnica LCWT.

La segmentación manual se lleva a cabo únicamente sobre 500 glándulas aleatorias, ya que es un trabajo muy tedioso que consume una gran cantidad de tiempo. Por consiguiente y, puesto que solo se requiere la segmentación manual como objetivo intermedio para validar el algoritmo automático LCWT, 500 es un número de glándulas suficientemente alto para determinar si la técnica es útil o si, por el contrario, es necesario implementar otra que no acarree un error considerable de cara a las siguientes etapas de extracción de características y de clasificación supervisada. En el apartado de 3.1.1 se exponen resultados cualitativos y cuantitativos de la segmentación automática implementada en este proyecto y se discuten con los proporcionados por otros autores del estado del arte.

A continuación, se definen y se explican las medidas utilizadas como indicadores para la evaluación de la segmentación automática implementada.

- **Coefficiente de Dice.** Estadístico utilizado para comparar, en términos de similitud, dos regiones de una imagen. Se expresa matemáticamente como el ratio de la intersección de las áreas de las regiones segmentadas automáticamente (A) y de las segmentadas manualmente (B) por la media del sumatorio de cada área individual [71], tal y como se expone a continuación:

$$Dice(A, B) = \frac{2|A \cap B|}{|A| + |B|} \quad (2.18)$$

donde *Dice* es el coeficiente de similitud que varía entre 0 y 1.

- **Coefficiente de Jaccard.** Parámetro que mide el índice de similitud entre dos conjuntos. Matemáticamente se define como la intersección de dos áreas dividida por la unión de las mismas [71], tal y como se observa en la ecuación 2.19, donde *Jaccard* toma valores entre 0 y 1.

$$Jaccard(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (2.19)$$

2.2.8.2 *Groundtruth para la clasificación*

La construcción del *groundtruth* para abordar la tarea de clasificación se basa en etiquetar manualmente cada uno de los candidatos a glándula según la clase a la que pertenezcan. Como ya se ha mencionado anteriormente, una vez se conoce qué tipo de elemento es cada candidato a glándula (artefacto, glándula benigna o glándula patológica de grado 3) es posible llevar a cabo la selección de características (apartado 2.2.5), el aprendizaje de los modelos durante la etapa de clasificación supervisada tanto en *machine learning* como en *deep learning* (apartado 2.2.7) y, en última instancia, la evaluación de dichos modelos (apartado 3.2).

Con el objetivo de construir el *groundtruth*, se desarrolla expresamente para este proyecto una interfaz gráfica de usuario (GUI, del inglés – *Graphic User Interface*) que permite ir etiquetando cada candidato a glándula en su clase correspondiente recorriendo todas las sub-imágenes de 1024x1024x3. En primer lugar, se define el directorio donde se encuentran las 854 imágenes benignas y se analiza cada una etiquetando cada posible candidato como artefacto o glándula benigna. Después se repite el mismo procedimiento para el directorio que contiene las imágenes patológicas catalogando cada candidato como artefacto o glándula patológica.

El algoritmo implementado en la GUI proporciona una interfaz muy sencilla para que cualquier persona, sin conocimiento de informática ni de programación, pueda llevar a cabo el etiquetado manual. De nuevo, al igual que para el *groundtruth* de la segmentación, sería conveniente que esta interfaz fuera usada por los patólogos, quienes tienen el conocimiento en anatomía patológica, aunque en este caso, las muestras han sido etiquetadas personalmente y después, validadas por los expertos.

Cuando se ejecuta el algoritmo, en la interfaz aparece la primera sub-imagen del directorio especificado, con el primer candidato resaltado en color cian. A partir de ahí, se determina la etiqueta para cada nuevo candidato, en función de si se considera glándula o no. Los candidatos etiquetados como artefactos se marcan en amarillo y los etiquetados como glándula, en azul; tal y como se muestra en la *Figura 2.45.a*. También se observan las distintas funcionalidades que presenta la interfaz, como la posibilidad de repetir el etiquetado del último candidato, la de cancelar y empezar desde el principio, la de modificar una imagen ya guardada, etc. Además, en la *Figura 2.45.b* se muestra cómo es posible hacer uso del *clustering* realizado al principio del proyecto para resaltar las zonas del citoplasma y de los núcleos que pueden ayudar a determinar si el candidato en cuestión es una glándula o un artefacto; pues una glándula es aquella que presenta un lumen rodeado de citoplasma y capas de núcleos.

La aplicación permite guardar de forma automática el etiquetado realizado, por lo que se puede abandonar la tarea a mitad y recuperar la información más adelante. Conforme se van etiquetando los candidatos, en función de la clase que se les asigna, se guardan en un directorio u otro. Finalmente, tras la construcción del *groundtruth* para las 1.468 sub-imágenes de 1024x1024x3, es posible recopilar toda la información original distribuida en diferentes niveles de organización, atendiendo al símil de las *matrioskas* explicado en el apartado 2.2.6. Dicha información se expone en la *Tabla 5*.

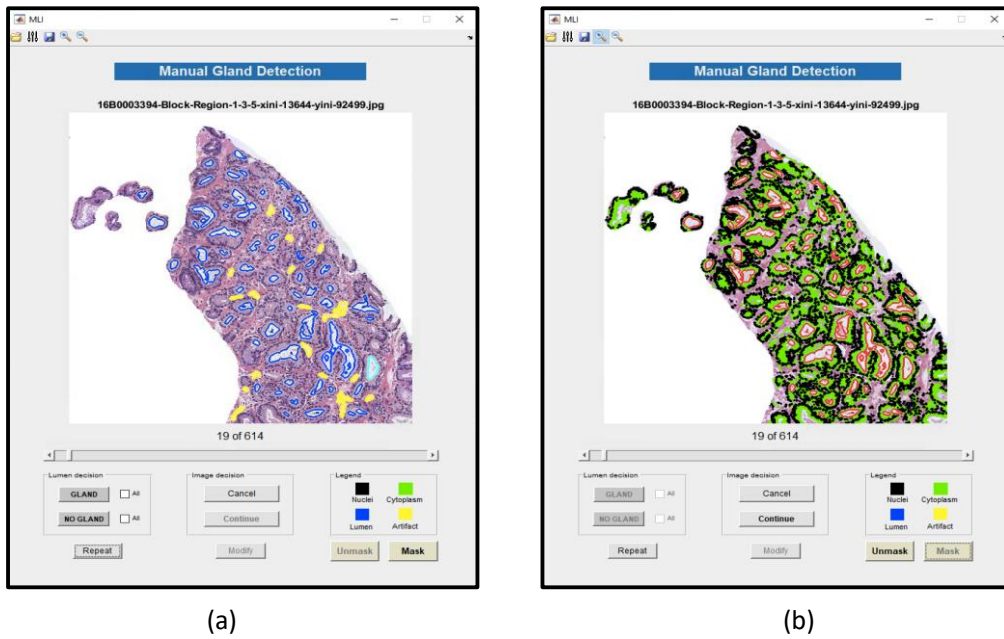


Figura 2.45. Interfaz de usuario para el etiquetado automático de candidatos a glándula donde se exponen diferentes funcionalidades. (a) Captura de imagen donde se observan los artefactos etiquetados en amarillo, los lúmenes de las glándulas en azul y el nuevo candidato bajo estudio, en cian. (b) Misma sub-imagen que presenta en rojo los candidatos etiquetados finalmente como glándulas, en verde la máscara del tejido citoplasmático y en negro, la de los núcleos; siendo el objetivo de las dos últimas ayudar visualmente a determinar la clase de los elementos para la construcción del *groundtruth*.





	Nº de pacientes	Nº de imágenes (originales)	Nº de sub-imágenes	Número de candidatos a glándula
Tejido benigno	8	17	854	25.052
Tejido patológico de grado 3	17	18	614	3.911
				

Tabla 5. Resumen de los datos originales distribuidos en diferentes niveles de organización.

Por otro lado, en la *Tabla 6* se detalla el número total de candidatos etiquetados en cada una de las clases. A partir de dichos candidatos se implementan las técnicas de pre-procesado para generar una base de datos en la que el número de elementos de cada clase sea balanceado.

	Artefactos	Glándulas	Total
Tejido benigno	21.857	3195	25.052
Tejido patológico de grado 3	911	3.000	6.195
Total	22.768	6.195	28.963

Tabla 6. Resumen de la información relativa a los candidatos a glándula.

A continuación, se explican los indicadores utilizados para evaluar la bondad de los modelos durante la etapa de la clasificación supervisada.

- **Sensibilidad.** Efectividad de un clasificador para identificar las etiquetas positivas. En el caso binario de *artefacto-glándula*, si consideramos la clase positiva *glándula* y la clase negativa *artefacto*, la sensibilidad mide la capacidad que tiene el clasificador para etiquetar un elemento con la clase *glándula* cuando realmente era una glándula [72].

$$Sensibilidad = \frac{VP}{VP + FN} \quad (2.20)$$

- **Especificidad.** Efectividad de un clasificador para identificar las etiquetas negativas. En el caso binario anterior, la especificidad hace referencia a la capacidad que presenta el clasificador para etiquetar como *artefacto* un elemento que realmente era un artefacto [72].

$$Specificity = \frac{VN}{VN + FP} \quad (2.21)$$

- **Valor Predictivo Positivo (VPP) o Precisión.** Eficacia real de una prueba diagnóstica cuyo resultado es positivo. Se trata de un valor post-test, ya que mide la probabilidad de que un candidato etiquetado con la clase *glándula* sea realmente una glándula [73].

$$Valor\ Predictivo\ Negativo\ (VPP) = \frac{VP}{VP + FP} \quad (2.22)$$

- **Valor Predictivo Negativo (VPN).** Eficacia real de una prueba diagnóstica cuyo resultado es negativo. Siguiendo el mismo caso anterior, el VPN haría referencia a la probabilidad de que un candidato etiquetado con la clase *artefacto* sea realmente un artefacto [73].

$$Valor\ Predictivo\ Negativo\ (VPN) = \frac{VN}{VN + FN} \quad (2.23)$$

- **Accuracy.** Exactitud con la que un clasificador es capaz de etiquetar correctamente todas las muestras. Es decir, la *accuracy* mide la tasa de acierto a la hora de etiquetar las glándulas con la clase *glándula* y los artefactos con la clase *artefacto*, en el caso binario [72].

$$Accuracy = \frac{VP + VN}{VP + VN + FP + FN} \quad (2.24)$$

- **F-Score.** Medida de exactitud de un test que es considerada como la media armónica de la sensibilidad y la precisión [74].

$$F - Score = \frac{2 * VPP * Sensibilidad}{VPP + Sensibilidad} \quad (2.25)$$

- **AUC** (del inglés, *Area Under Curve*). Mide el área bajo la Curva ROC (del inglés, *Receiver Operating Characteristic*). La Curva ROC es un gráfico que muestra la capacidad discriminatoria de un clasificador binario en base a un umbral determinado. Para ello, se representa gráficamente la tasa de verdaderos positivos (*True Positive Rate*) contra la tasa de falsos positivos (*False Positive Rate*). De esta forma, cuanto mayor sea el área bajo la curva, mayor será la capacidad discriminatoria del clasificador.

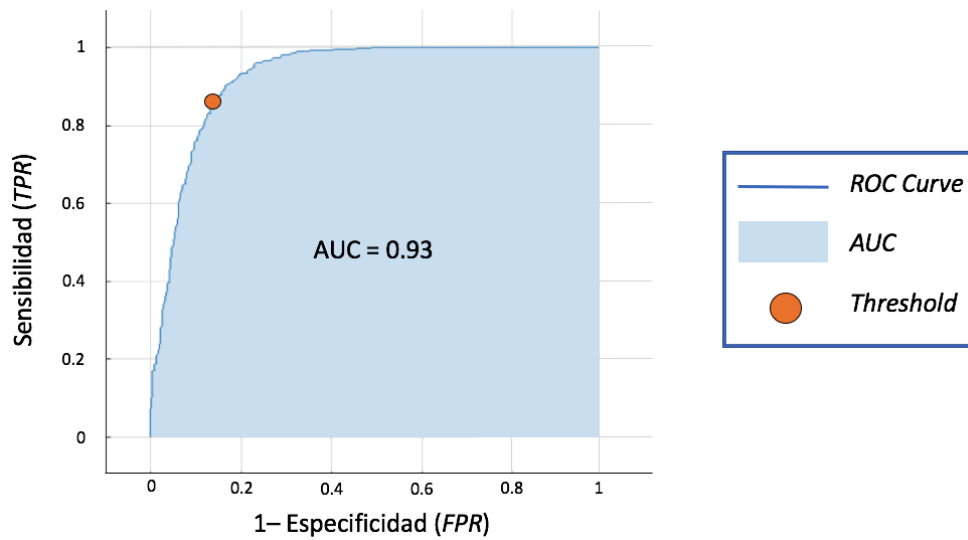


Figura 2.46. Ejemplo de Curva ROC y área bajo la curva (AUC).

CAPÍTULO 3

3. Resultados y discusión

Índice de contenidos

3.1	Resultados de la segmentación	68
3.2	Resultados de la clasificación supervisada	71
3.2.1	Validación durante la etapa de Entrenamiento	71
3.2.2	Resultados durante la etapa de Test.....	75
3.3	Predicción y visualización.....	79

A modo de introducción, cabe destacar que la comparación de resultados, tanto en la segmentación como en la clasificación, con respecto a los trabajos del estado del arte es muy complicada de llevar a cabo debido a diversos factores. Los principales consisten en:

- **Privacidad de la base de datos.** Las imágenes histopatológicas con las que se trabaja son de uso privado, ya que este TFM se enmarca dentro de un proyecto global cuya finalidad no es meramente de investigación, sino que tiene un objetivo superior basado en el desarrollo de un prototipo software que sirva de ayuda al diagnóstico para los patólogos del Hospital Clínico de Valencia. Además, no existe ninguna base pública de imágenes histopatológicas de próstata con la que evaluar los métodos y algoritmos desallorados.

- **Distintos objetivos en la clasificación.** Los autores de estudios relativos al estado del arte en imagen histológica abordan la tarea de clasificación con diferentes objetivos. Como ya se ha expuesto, algunos estudios tienen como fin último discriminar entre muestras benignas y patológicas, otros clasifican según los diferentes grados de la escala *Gleason*, otros tienen en cuenta los artefactos y otros centran sus esfuerzos en distinguir únicamente entre dos grados.

A su vez, existen también algunos inconvenientes de menor importancia como:

- *Dimensiones de las imágenes.* Cada autor trabaja, no solo con imágenes diferentes procedentes de distintos pacientes, sino también de diferentes dimensiones. Es decir, no hay un formato estándar en el que se trabaje siempre a 1024x1024 o 512x512, por ejemplo. El problema se aborda desde diferentes dimensiones y, por tanto, las estructuras glandulares varían en tamaño y resolución espacial. Por consiguiente, el método de otro autor no se podría aplicar sobre las imágenes de este proyecto porque habría que llevar a cabo una labor de adaptación, bien de las imágenes, o bien del método, lo cual desvirtuaría los resultados.

- *Subjetividad en la segmentación manual.* Para validar los resultados de la segmentación es necesario realizar un *groundtruth*, el cual depende en gran medida de la persona que lo lleve a cabo. De hecho, en algunos estudios como en [75], la segmentación manual es realizada por dos investigadores a fin de proporcionar unos resultados más fiables.

3.1 Resultados de la segmentación

Como se ha comentado a lo largo del proyecto, para llevar a cabo la validación de la segmentación automática basada en el método LCWT se ha realizado una segmentación manual con la aplicación del proyecto MINERVA. Una vez se dispone de las máscaras correspondientes a ambas segmentaciones, se proporciona una serie de resultados cualitativos y cuantitativos con el objetivo de analizar y discutir la bondad de la técnica implementada.

Cabe destacar que no es frecuente encontrar en la literatura médica estudios que expresen resultados cuantitativos sobre la segmentación de las unidades glandulares. Normalmente, en la mayoría de los estudios que pretenden clasificar las imágenes según la escala *Gleason*, primero se lleva a cabo una discriminación de las estructuras de las imágenes histopatológicas (núcleos, lumen, citoplasma y estroma) y, después, se aplican métodos de procesado para eliminar algunos de los objetos considerados como artefactos. Un ejemplo de ello se expone en [76]. En este TFM, se ha querido profundizar en la novedosa técnica implementada y, por ello, se presentan resultados de ambos tipos, a pesar de que el objetivo final es la clasificación.

3.1.1 Resultados cualitativos

En la *Figura 3.1* se exponen algunos ejemplos de la segmentación manual y automática superpuestas en imágenes RGB de glándulas de diferentes tamaños, con la finalidad de mostrar visualmente la precisión de la técnica LCWT.

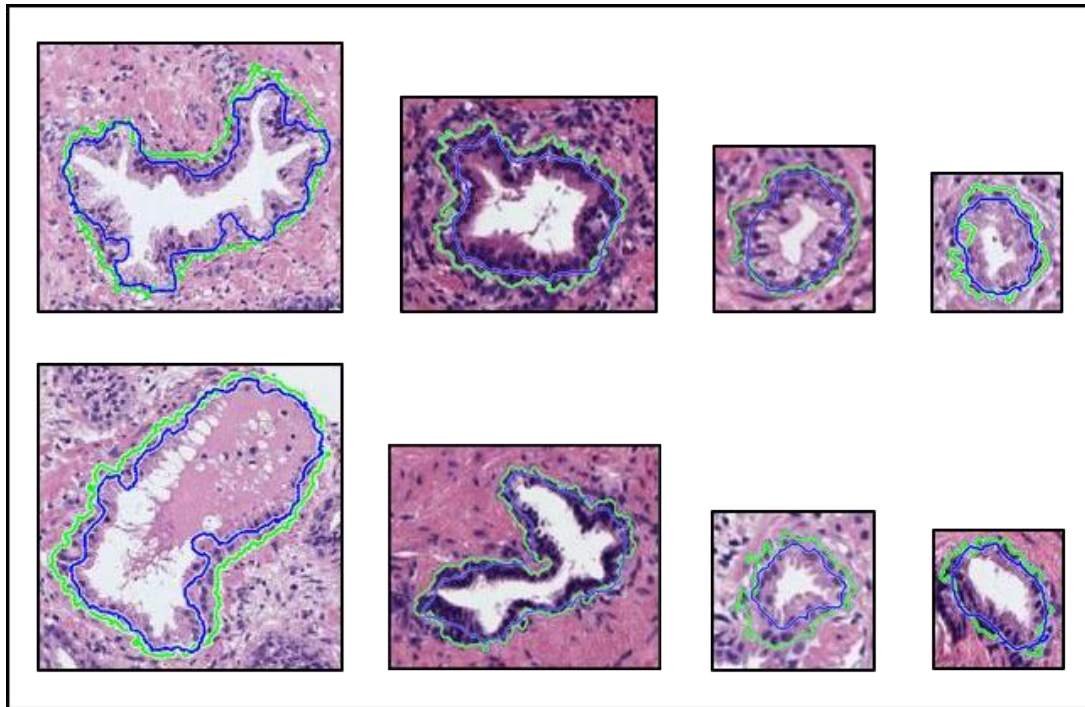


Figura 3.1. Ejemplos de imágenes de diferentes glándulas segmentadas manualmente (línea verde) y automáticamente (línea azul).

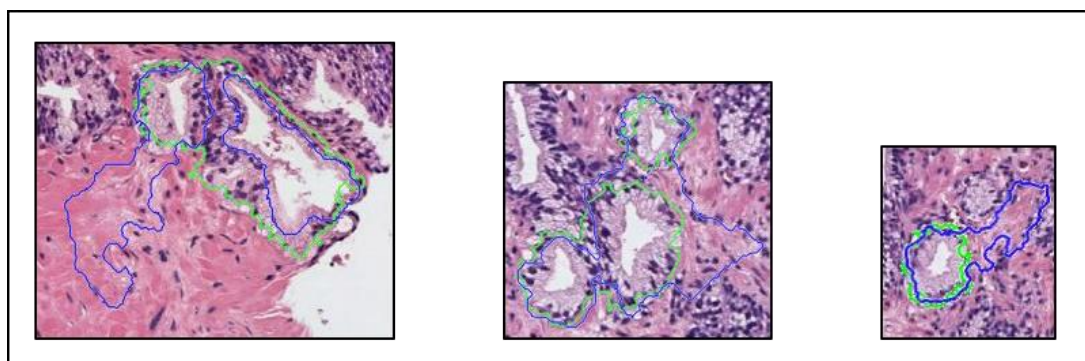


Figura 3.2. Ejemplos de glándulas donde la segmentación con la técnica LCWT proporciona resultados incorrectos.

En general, los resultados cualitativos de la segmentación de las unidades glandulares son considerablemente buenos, pues la segmentación se adapta perfectamente al contorno de la glándula, definiendo los núcleos como los bordes de la misma. No obstante, en raras ocasiones, la técnica LCWT proporciona una segmentación incorrecta cuando la glándula bajo estudio se presenta rodeada de una cantidad de núcleos muy pobre, tal y como se expone en la *Figura 3.2*.

3.1.1 Resultados cuantitativos

Como ya se ha comentado, en este TFM se presentan resultados cuantitativos atendiendo a los coeficientes de *Dice* y de *Jaccard* para evaluar la técnica LCWT. Cabe destacar que la segmentación no es el objetivo final del proyecto, por lo que únicamente se han segmentado a mano 500 glándulas aleatorias. La hipótesis de partida es que los resultados obtenidos para ese tamaño muestral pueden ser inferidos para toda la población, que en este caso se trata de 6.195 glándulas (3.195 benignas y 3.000 patológicas de grado 3).

En la *Tabla 7* se exponen los resultados cuantitativos de la segmentación obtenidos en este proyecto a nivel de glándula para los coeficientes especificados.

	$\mu \pm \sigma$
Coeficiente <i>Dice</i>	0.856 \pm 0.085
Coeficiente <i>Jaccard</i>	0.748 \pm 0.108

Tabla 7. Resultados cuantitativos proporcionados por la técnica *Locally Constrained Watershed Transform* para 500 glándulas seleccionadas de forma aleatoria.

En cuanto a la comparación de los resultados con respecto a otros autores del estado del arte, en primer lugar, cabe señalar que se lleva a cabo con algunos artículos cuya finalidad es expresamente la segmentación, ya que es poco frecuente encontrar resultados cuantitativos de segmentación en trabajos que buscan como último objetivo la clasificación. No obstante, existen algunos, como el propuesto en [27] que proporcionan resultados de ambas tareas. En segundo lugar, es importante recordar que la comparación de resultados es muy subjetiva, ya que las imágenes con las que se trabaja en cada estudio son muy diferentes, en cuanto a tinción, dimensiones y resolución se refiere. Se muestra la *Tabla 8* como evidencia de lo expuesto.

En [27], los autores seleccionan 309 glándulas cuyos bordes están bien definidos y generan un *groundtruth* para dichas glándulas. Utilizan el coeficiente *Jaccard* como indicador para la validación de su método de segmentación. Además, implementan sobre sus imágenes los métodos llevados a cabo por otros autores en [26] y en [77], cuyos resultados se recogen en la *Tabla 8*. Por otra parte, en [78] también se aborda como fin último la segmentación de glándulas de próstata, proponiendo varios métodos y proporcionando los resultados a partir de las imágenes completas (*whole-slide*).

	Naik et al. [26]	Nguyen et al. [27]	Mónaco et al. [58]	Xu et al. [59]	Método propuesto
Tipo de muestras	A nivel de glándula	A nivel de glándula	A nivel de glándula	<i>Whole-Slide image</i>	A nivel de glándula
Número de muestras	309	309	309	196	500
Coeficiente <i>Jaccard</i>	0.43	0.66	0.31	0.69	0.75

Tabla 8. Comparación de la segmentación con el estado del arte.

3.2 Resultados de la clasificación supervisada

En esta etapa, los objetivos principales residen, por un lado, en la comparación de los resultados de las dos estrategias que se abordan en este trabajo para la evaluación de los modelos, es decir, la comparación de resultados proporcionados por las aproximaciones de *machine learning* y de *deep learning*. Y, por otro lado, la comparación de resultados con los publicados por otros autores en el estado del arte, teniendo en cuenta las limitaciones detalladas en la introducción de este capítulo. A continuación, se presentan, en primer lugar, los resultados de validación proporcionados durante la etapa de entrenamiento para la construcción de los modelos y, en segundo lugar, los resultados de la evaluación de dichos modelos cuando se aplican sobre las imágenes de *Test*.

3.2.1 Validación durante la etapa de Entrenamiento

3.2.1.1 Aproximación basada en Machine Learning

Como se ha descrito en el apartado 2.2.7 relativo a la metodología, en este proyecto se aborda la clasificación supervisada basada en *machine learning* implementando clasificadores de tipo lineal (SVM con *kernel* lineal) y de tipo no lineal (SVM con *kernel* cuadrático y redes neuronales artificiales). A continuación, para ambos clasificadores, se presentan los resultados de los algoritmos de optimización implementados para la selección de los mejores parámetros.

Support Vector Machine (SVM)

En cuanto a los clasificadores de *Support Vector Machine* (SVM), se expone la *Tabla 9* donde se pueden apreciar los diferentes parámetros a optimizar a la hora de clasificar los candidatos a glándula en su clase correspondiente. En dicha tabla se recogen, para cada época, los errores obtenidos de la función objetivo, los tiempos de ejecución, el *coding* que permite enfrentar una clase contra otra (OvO) o una clase contra las demás (OvA), y otros parámetros de diseño como el *BoxConstraint* y el *KernelScale*. Además de los resultados cuantitativos, en la *Figura 3.3* también se expone un gráfico que muestra, en azul, el mínimo error conseguido hasta el momento; y en verde, el error estimado para cada época.

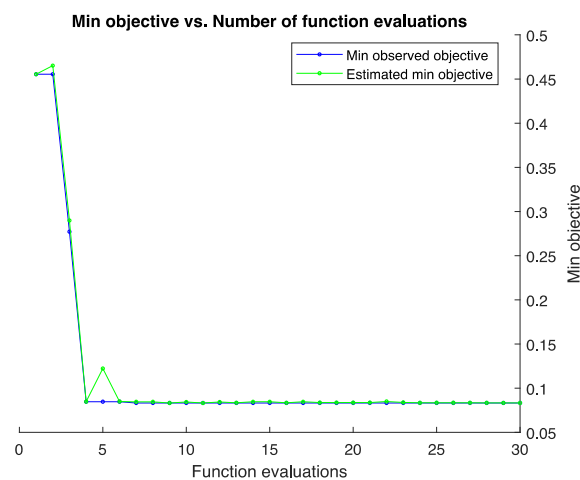


Figura 3.3. Representación gráfica del aprendizaje donde se muestra cómo se alcanza el mínimo de la función objetivo y el error estimado para cada etapa.

Iter	Eval	Objective	Objective	BestSoFar	BestSoFar	Coding	BoxConstraint	KernelScale
	result		runtime	(observed)	(estim.)		t	
1	Best	0.45546	814.59	0.45546	0.45546	onevsone	0.013818	0.0013012
2	Accept	0.65077	10.315	0.45546	0.46516	onevsone	0.0026695	218.35
3	Best	0.27718	1142.4	0.27718	0.28989	onevsone	61.291	0.0032659
4	Best	0.084752	175.66	0.084752	0.084779	onevsall	0.065696	0.37449
5	Accept	0.56432	1295.3	0.084752	0.12231	onevsall	19.513	0.0010221
6	Accept	0.085276	182.1	0.084752	0.084986	onevsall	0.063227	0.38726
7	Best	0.083312	1331.8	0.083312	0.084606	onevsall	2.7303	0.63761
8	Accept	0.25334	1179.5	0.083312	0.084616	onevsone	1.5418	0.029976
9	Accept	0.099555	14.356	0.083312	0.083305	onevsall	0.21203	4.335
10	Accept	0.11816	14.148	0.083312	0.084419	onevsall	0.0010241	1.2443
11	Accept	0.11278	12.99	0.083312	0.083252	onevsall	0.73795	24.727
12	Accept	0.083573	1501.6	0.083312	0.084388	onevsall	975.82	8.3366
13	Accept	0.28229	14.749	0.083312	0.083531	onevsall	0.0017676	893.21
14	Accept	0.28334	1686.1	0.083312	0.084663	onevsall	965.57	0.11639
15	Accept	0.089992	29.805	0.083312	0.084677	onevsall	962.62	116.6
16	Accept	0.085014	174.37	0.083312	0.08346	onevsall	315.31	26.571
17	Accept	0.084752	868.5	0.083312	0.084671	onevsall	14.906	2.6412
18	Accept	0.085276	129.91	0.083312	0.083719	onevsall	0.33856	0.94504
19	Accept	0.11475	8.9101	0.083312	0.083768	onevsall	964.31	989.85
20	Accept	0.26316	1022.8	0.083312	0.083748	onevsone	995.41	0.27446
21	Accept	0.089861	30.426	0.083312	0.083788	onevsall	0.0010208	0.11957
22	Accept	0.084883	1012.4	0.083312	0.084982	onevsall	952.17	18.75
23	Accept	0.084228	583.88	0.083312	0.083953	onevsall	107.42	8.6455
24	Accept	0.084883	866.49	0.083312	0.08347	onevsall	2.0724	0.99414
25	Accept	0.084359	662.57	0.083312	0.083569	onevsall	0.57678	0.58337
26	Accept	0.083442	1191.9	0.083312	0.083567	onevsall	178.27	5.1872
27	Accept	0.084228	1097.4	0.083312	0.083532	onevsall	277.58	8.8288
28	Accept	0.096411	11.875	0.083312	0.083426	onevsall	0.0010452	0.24648
29	Accept	0.084752	305.94	0.083312	0.083384	onevsall	0.018425	0.1496
30	Accept	0.08449	1011.4	0.083312	0.083325	onevsall	0.0010343	0.02368

Optimization completed.
MaxObjectiveEvaluations of 30 reached.
Total function evaluations: 30
Total elapsed time: 18402.941 seconds.
Total objective function evaluation time: 18384.1466

Best observed feasible point:

Coding	BoxConstraint	KernelScale
onevsall	2.7303	0.63761

Observed objective function value = 0.083312
Estimated objective function value = 0.083325
Function evaluation time = 1331.7628

Tabla 9. Salida del algoritmo de optimización implementado para el clasificador de tipo SVM durante la fase de entrenamiento de los modelos.

El proceso de optimización mostrado se lleva a cabo para cada una de las 5 iteraciones implementadas en la etapa de *cross validation* externo. La optimización en cada iteración se aplica sobre el conjunto de entrenamiento realizando 10 particiones en *train* y *validation*. En cada una de ellas se calculan los hiperparámetros óptimos (*coding*, *boxconstraint* y *kernelScale*) y se obtiene un valor de *accuracy*. Finalmente, se selecciona el modelo de entrenamiento que mejores resultados presenta para utilizarlo en la fase de *Test*.

Perceptrón Multicapa (MLP)

En cuanto a la validación de los modelos basados en *Feedforward Neural Network* (FNN), el funcionamiento es similar, ya que en cada una de las 5 iteraciones del *cross validation* externo se realiza una partición de los datos de entrenamiento en 10 subconjuntos para entrenar la red. En cada uno de esos 10 subconjuntos es posible extraer una determinada información acerca del rendimiento del modelo. Finalmente, se escoge aquel que presenta el menor error cuadrático medio. En la *Figura 3.4* se exponen algunas de las gráficas que se extraen para el mejor de los 10 subconjuntos del *cross validation* interno implementado en la primera iteración.

3.2 Resultados de la clasificación supervisada

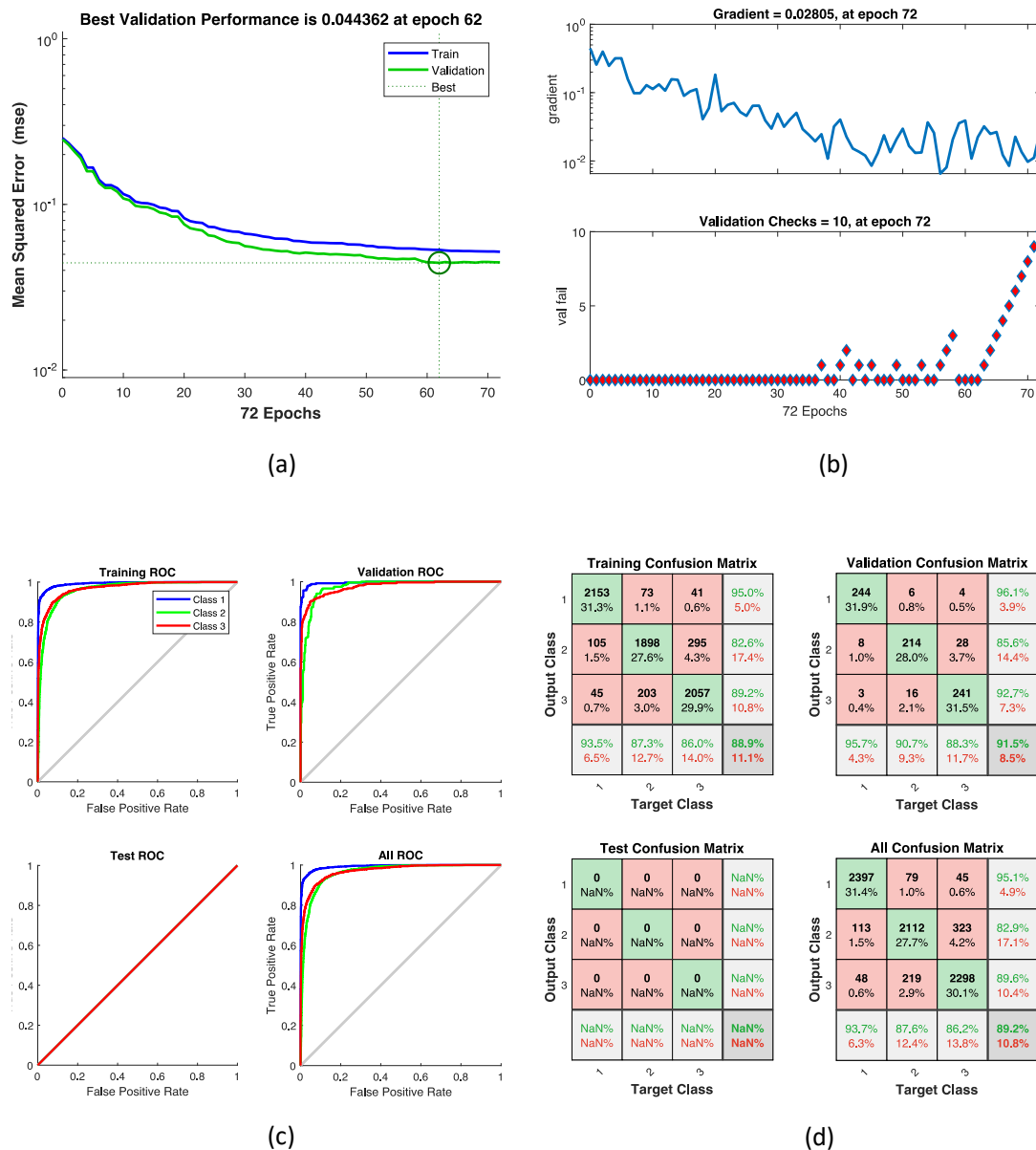


Figura 3.4. Resultados extraídos del mejor modelo de red neuronal llevado a cabo durante la etapa de validación en la primera iteración. (a) Curvas de aprendizaje del entrenamiento y de la validación, en azul y en verde, respectivamente. (b) Aprendizaje del modelo según el gradiente y el error cuadrático medio. (c) Curvas ROC relativas al conjunto de entrenamiento y al de validación. (d) Matrices de confusión para los datos de entrenamiento y de validación.

En la *Figura 3.4.a* se aprecia que el aprendizaje se detiene en la época 72 para no producir *overfitting*, ya que el mínimo error se obtiene en la época 62 y no mejora durante las 10 siguientes. Del mismo modo, en la *Figura 3.4.b*, también puede observarse que a partir de la época 62 no se disminuye el error cuadrático medio. Por otra parte, cabe destacar que en la *Figura 3.4(c-d)* se proporciona por defecto las curvas ROC y las matrices de confusión correspondientes al conjunto de datos de test. No obstante, dado que en esta fase solo se aborda el aprendizaje de los modelos de entrenamiento, los valores de test aparecen vacíos.

3.2.1.2 Aproximación basada en Deep Learning

A diferencia de los modelos de *machine learning* que son creados a partir de una extracción de características previa, los modelos basados en *deep learning* no tienen ningún conocimiento inicial. Por esta razón, dichos modelos necesitan una gran cantidad de datos etiquetados a partir de la cual poder aprender de forma automática. Para llevar a cabo la transferencia de conocimiento se hace uso de técnicas de *fine-tuning* que permiten incorporar la información de las redes pre-entrenadas a todas las capas de la CNN excepto a las últimas, ya que es en ellas donde traslada el conocimiento para resolver el problema bajo estudio mediante el conjunto de datos de los candidatos a glándula con su clase correspondiente.

Por otra parte, cabe destacar que, a diferencia de la aproximación basada en *machine learning*, en este caso no se implementan técnicas de *cross validation* interno, por lo que únicamente se obtienen una gráfica relacionada con la función de pérdidas y otra relacionada con el valor de *accuracy* que va tomando el modelo en cada época durante el aprendizaje, en cada una de las iteraciones del *cross validation* externo. (Véase la *Figura 3.5*).

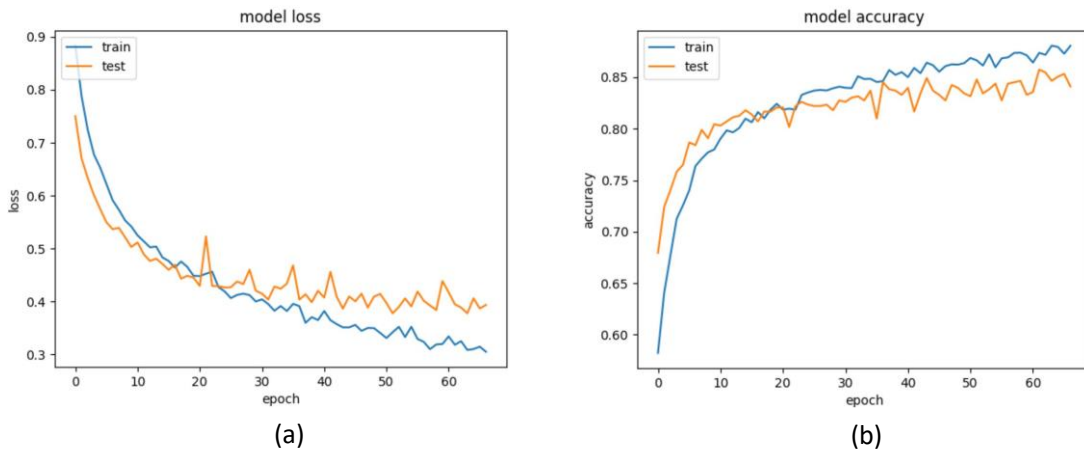


Figura 3.5. (a) Función de pérdidas donde la línea naranja representa la curva de validación y la línea azul, la de entrenamiento. (b) Resultados de *accuracy* que obtiene el conjunto de validación en cada una de las épocas.

En la *Figura 3.5.a*, se observa que los datos de entrenamiento siguen aprendiendo y reduciendo el error de la función objetivo en cada época, pero no ocurre lo mismo con la curva de validación. Por consiguiente, para no producir *overfitting* se establece un criterio de parada en 15 épocas. En la figura del ejemplo, el aprendizaje se detiene en la época 65 en lugar de continuar hasta la 150 (máximo permitido), puesto que a partir de la época 50 ya no se reduce la función del error cuadrático medio. Esta información se traslada a la *Figura 3.5.b*, donde se observa que, a partir de la misma época donde la curva de validación deja de aprender, los valores de *accuracy* empiezan a mantenerse constantes. Este procedimiento se lleva a cabo para cada uno de los 5 conjuntos de entrenamiento en los que se habían dividido inicialmente los datos, con el objetivo de construir unos modelos de clasificación a partir de los cuales predecir las etiquetas de los datos del conjunto de *Test*.

3.2.2 Resultados durante la etapa de Test

En esta etapa se exponen los resultados de la evaluación de los modelos de entrenamiento al aplicar los datos de los conjuntos de *Test*. Por tanto, en primer lugar, se presentan los resultados recogidos usando técnicas de *machine learning* y, después, los resultados obtenidos aplicando técnicas de *deep learning*, para cada uno de los 5 conjuntos en los que se habían dividido los datos en *Training* y *Test*. Finalmente, se lleva a cabo una comparación de los resultados de ambas técnicas para determinar cuál de ellos es el más fiable de cara a implementarlo en la simulación de una práctica clínica real. (Véase el apartado 3.3).

3.2.2.1 Aproximación basada en Machine Learning

Como ya se ha comentado, para abordar la fase de Test se utiliza el modelo del entrenamiento que mejores resultados había proporcionado durante el *cross validation* interno. De esta forma, para cada clasificador (LSVM, QSVM y FNN) se obtienen 5 resultados correspondientes a los 5 conjuntos de *Test* que se evalúan. Para medir la bondad de cada clasificador multiclase a la hora de discriminar entre las clases *Artefactos – Glándulas*, se expone la *Tabla 10* donde se recogen los parámetros de sensibilidad, especificidad, PPV, NPV, *accuracy*, *F-Score* y AUC, como la media de los 5 conjuntos de Test.

	LSVM	QSVM	FNN
Sensibilidad	0.927 ± 0.012	0.937 ± 0.001	0.936 ± 0.017
Especificidad	0.960 ± 0.021	0.954 ± 0.023	0.953 ± 0.031
VPP	0.922 ± 0.042	0.913 ± 0.052	0.912 ± 0.056
VPN	0.960 ± 0.012	0.964 ± 0.012	0.964 ± 0.012
Accuracy	0.863 ± 0.030	0.865 ± 0.024	0.843 ± 0.023
F-Score	0.924 ± 0.025	0.924 ± 0.027	0.923 ± 0.034
AUC	0.985 ± 0.006	0.986 ± 0.008	0.988 ± 0.009

Tabla 10. *Artefactos Vs Glándulas*. Resultados (en términos de media) de la clasificación supervisada para los tres clasificadores de *machine learning* implementados.

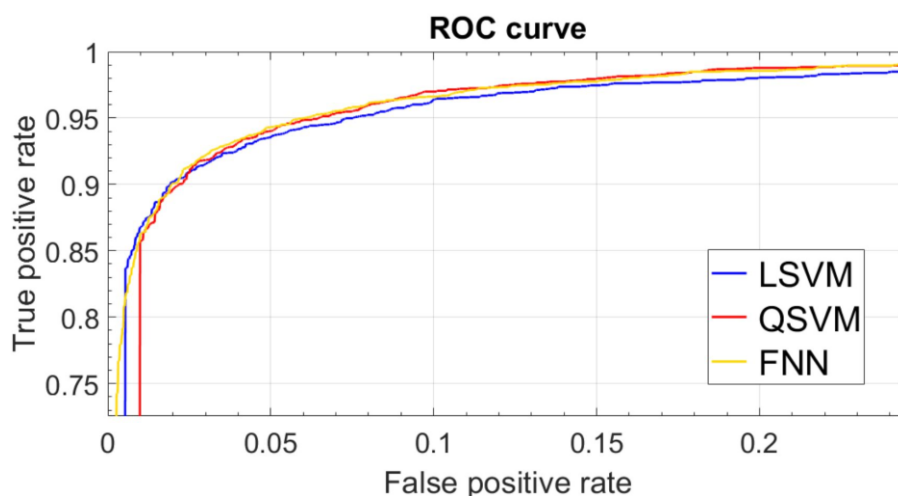


Figura 3.6. Curva ROC del estudio *Artefactos Vs Glándulas* usando técnicas de *machine learning*.

Se observa que los resultados proporcionados son muy similares para los tres clasificadores implementados. No obstante, el *Support Vector Machine* con *kernel* cuadrático (QSVM) presenta resultados ligeramente superiores frente al resto, a la hora de determinar si un determinado candidato a glándula es un artefacto o una glándula (de cualquier tipo).

A continuación, se presenta una tabla similar para evaluar qué clasificador presenta mejores resultados, en cuanto a la clasificación de glándulas benignas y patológicas se refiere.

	LSVM	QSVM	FNN
Sensibilidad	0.818 ± 0.051	0.794 ± 0.045	0.768 ± 0.100
Especificidad	0.848 ± 0.090	0.865 ± 0.083	0.817 ± 0.092
VPP	0.844 ± 0.093	0.853 ± 0.088	0.809 ± 0.084
VPN	0.825 ± 0.069	0.811 ± 0.053	0.790 ± 0.058
Accuracy	0.863 ± 0.030	0.865 ± 0.024	0.843 ± 0.023
F-Score	0.826 ± 0.027	0.818 ± 0.026	0.781 ± 0.045
AUC	0.925 ± 0.034	0.923 ± 0.037	0.902 ± 0.033

Tabla 11. Glándulas Benignas Vs. Glándulas Patológicas. Resultados (en términos de media) relativos a la clasificación supervisada para los tres tipos de clasificadores de *machine learning*.

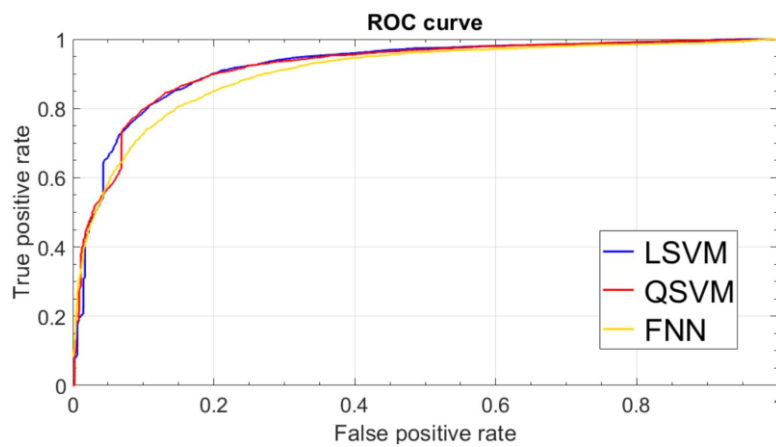


Figura 3.7. Curva ROC del estudio *Artefactos Vs Glándulas* usando técnicas de *machine learning*.

En este caso, se observa que las redes neuronales del perceptrón multicapa presentan peores resultados en comparación con las técnicas de SVM, tanto con un *kernel* lineal como con un *kernel* cuadrático. Entre estas dos últimas, LSVM presenta una ligera mejora frente al QSVM a la hora de clasificar entre glándulas benignas y glándulas patológicas, aunque los resultados son prácticamente iguales para ambas aproximaciones.

A continuación, se exponen los resultados conseguidos aplicando técnicas de *deep learning* para realizar una comparación con las estrategias basadas en *machine learning* a la hora de afrontar un problema de clasificación multiclase como el propuesto en este proyecto.

3.2.2.2 Aproximación basada en Deep Learning

Siguiendo el mismo procedimiento que para la aproximación basada en *machine learning*, se presentan en esta etapa los resultados correspondientes a la tarea de clasificación multiclase abordada mediante estrategias de *deep learning*. En la *Tabla 12* se expone la media de los valores obtenidos en cada una de las 5 iteraciones relativas a los 5 conjuntos de *Test*, tanto para la clasificación de *Artefactos – Glándulas*, como para la de *Glándulas Benignas – Glándulas Patológicas*.

	Artefactos Vs. Glándulas	Glándulas Benignas Vs. Glándulas Patológicas
Sensibilidad	0.900 ± 0.020	0.747 ± 0.109
Especificidad	0.938 ± 0.026	0.780 ± 0.123
VPP	0.884 ± 0.059	0.779 ± 0.101
VPN	0.945 ± 0.016	0.769 ± 0.047
Accuracy	0.812 ± 0.032	0.812 ± 0.033
F-Score	0.891 ± 0.033	0.753 ± 0.058
AUC	0.974 ± 0.010	0.889 ± 0.036

Tabla 12. Resultados proporcionados por los modelos de clasificación basados en la red VGG19 para abordar la etapa de *deep learning*.

3.2.2.3 Comparación de ambas aproximaciones

A continuación, se lleva a cabo una comparación de los resultados proporcionados por las dos metodologías de clasificación implementadas durante este proyecto. Para ello, se enfrentan los resultados obtenidos tras aplicar técnicas de *machine learning* con el clasificador QSVM y los obtenidos aplicando estrategias de *deep learning* basadas en la red VGG19.

	Artefactos Vs. Glándulas		Glándulas Benignas Vs. Patológicas	
	QSVM	VGG19	QSVM	VGG19
Sensibilidad	0.937 ± 0.001	0.900 ± 0.020	0.794 ± 0.045	0.747 ± 0.109
Especificidad	0.954 ± 0.023	0.938 ± 0.026	0.865 ± 0.083	0.780 ± 0.123
VPP	0.913 ± 0.052	0.884 ± 0.059	0.853 ± 0.088	0.779 ± 0.101
VPN	0.964 ± 0.012	0.945 ± 0.016	0.811 ± 0.053	0.769 ± 0.047
Accuracy	0.865 ± 0.024	0.812 ± 0.032	0.865 ± 0.024	0.812 ± 0.033
F-Score	0.924 ± 0.027	0.891 ± 0.033	0.818 ± 0.026	0.753 ± 0.058
AUC	0.986 ± 0.008	0.974 ± 0.010	0.923 ± 0.037	0.889 ± 0.036

Tabla 13. Resultados de la clasificación supervisada a partir de las aproximaciones basadas en *machine learning* y en *deep learning*.

Como se aprecia en la tabla anterior, los resultados alcanzados a partir de las estrategias basadas en *machine learning* superan los obtenidos con técnicas de *deep learning*. Este hecho es poco frecuente, ya que una de las ventajas del *deep learning* radica en la detección de características o patrones repetidos que están presentes en las imágenes de forma intrínseca y que son difíciles de identificar de forma manual durante la extracción de características habitual. No obstante, se parte de la hipótesis de que la estrategia de *deep learning* empleada presenta inconvenientes relacionados con las dimensiones de las imágenes de entrada, ya que estas son diezmadas o interpoladas para que todas tengan el mismo tamaño a la entrada de la red. Por el contrario, el hecho de que cada candidato a glándula presente unas dimensiones diferentes, supone una ventaja para las aproximaciones basadas en *machine learning*, ya que por definición el tamaño de los lúmenes y de las estructuras glandulares es una de las principales características que diferencian el tejido sano del tejido patológico. Por otra parte, es posible que los bajos resultados del *deep learning* sean debidos a que el número de muestras utilizadas de cada clase es demasiado pequeño, es decir, los modelos de clasificación no pueden aprender porque no disponen de la suficiente cantidad de datos etiquetados específicos de las imágenes histológicas de próstata.

3.2.2.4 Comparación con el estado del arte

Como se ha detallado al principio del capítulo 3, la comparación de resultados con respecto a otros estudios del estado del arte es muy subjetiva y no es posible llevarla a cabo de manera directa. A pesar de ello, en este trabajo se ponen en común los resultados obtenidos en este proyecto con los proporcionados por otros autores que también abordan la tarea de clasificación. Cabe destacar que cada autor tiene un objetivo de clasificación distinto y que expresa sus resultados con figuras de mérito diferentes. (Véase la *Tabla 14*).

	Naik et al. (2008)	Nguyen et al. (2012)	Kwak et al. (2017)	Ren et al. (2017)	Sahran et al. (2017)	Método propuesto	Método propuesto
OBJETIVO	Benigno, Patológico	Artefactos, Benigno, Patológico	Benigno, Patológico	Grados (3-4)	Benigno, Grados (3-4)	Artefactos, Glándulas	Benigno, Grado 3
Sensibilidad	–	–	–	0.70 ± 0.15	0.90	0.94	0.79 ± 0.05
Especificidad	–	–	–	0.89 ± 0.14	0.94	0.95 ± 0.02	0.87 ± 0.08
Accuracy	0.54 ± 0.12	0.77 ± 0.07	–	0.83 ± 0.03	0.91	0.87 ± 0.02	0.87 ± 0.02
F-Score	–	–	–	0.8460	–	0.92 ± 0.03	0.82 ± 0.03
AUC	–	–	0.98 ± 0.03	–	–	0.99 ± 0.01	0.92 ± 0.04

Tabla 14. Comparación de los resultados obtenidos en este proyecto con los publicados en otros estudios del estado del arte.

3.3 Predicción y visualización

Esta etapa consiste en hacer uso de los cinco modelos entrenados para crear un comité de evaluación que permita llevar a cabo la predicción de muestras de nuevos pacientes con las que no se ha trabajado durante ninguna fase del proyecto. Mediante esta aproximación se pretende simular una práctica clínica real en la que nuevas muestras de tejido prostático son biopsiadas y se asignan a una clase u otra en función de las predicciones de los modelos automáticos.

En la *Figura 3.8* se muestran dos ejemplos en los que la única información que se conoce es el carácter benigno o patológico de la muestra, pero no se dispone de datos etiquetados para obtener ningún tipo de métrica.

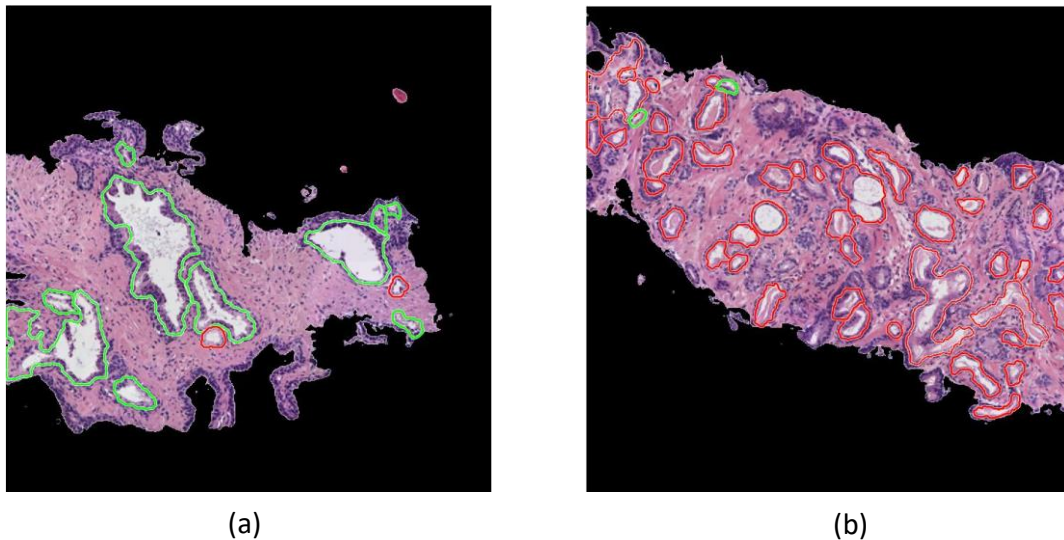


Figura 3.8. Predicción de etiquetas llevada a cabo mediante el comité de evaluación compuesto por los modelos *Support Vector Machine* con *kernel* cuadrático sobre las muestras histopatológicas. Se dejan fuera de la visualización los elementos etiquetados como artefactos y se marcan en verde y en rojo los etiquetados como glándulas benignas y glándulas patológicas, respectivamente. (a) Muestra de carácter benigno. (b) Muestra de carácter patológico.

Como se aprecia en la figura anterior, las etiquetas que genera el comité compuesto por los 5 modelos QSVM tienen un alto porcentaje de éxito. En este caso concreto, solo se observan las etiquetas de una pequeña parte del tejido biopsiado, pero los resultados son perfectamente extrapolables al resto del corte histológico. Por consiguiente, el resultado del análisis de la muestra se expresa como la proporción de glándulas de cada tipo presentes en la imagen.

En definitiva, en el presente trabajo fin de máster se ha expuesto una simulación de una práctica clínica, pero el cumplimiento del objetivo final reside en que una versión mejorada de este método sea utilizada en casos reales para ayudar a los patólogos a mejorar en términos de coste-efectividad a la hora de analizar muestras histopatológicas de próstata.

CAPÍTULO 4

4. Conclusiones y líneas futuras

Índice de contenidos

4.1	Conclusiones	81
4.2	Líneas futuras	81

4.1 Conclusiones

En primer lugar, cabe destacar que se ha llevado a cabo una exhaustiva revisión bibliográfica en cuanto a técnicas de segmentación y clasificación de imágenes histológicas de próstata se refiere. A partir de la información obtenida, se han desarrollado nuevas estrategias con el objetivo de mejorar algunos métodos y procedimientos del estado del arte actual.

En este proyecto, ha sido necesario llevar a cabo una preparación de la base de datos de imágenes, convirtiendo las muestras originales procedentes del Hospital Clínico de Valencia en sub-imágenes con unas dimensiones adecuadas en cuanto a resolución y número de glándulas por imagen se refiere. Una vez adaptadas las imágenes, se han implementado técnicas de *clustering* para distinguir los cuatro elementos del tejido que se revelan con la tinción de hematoxilina y eosina: lúmenes, núcleos, citoplasma y estroma. Posteriormente, se han aplicado técnicas de pre-procesado para eliminar parte del ruido que presentan dichos elementos.

A partir de los lúmenes obtenidos tras el proceso de *clustering*, se pretende determinar cuáles son realmente lúmenes de glándula y cuáles son artefactos. Para ello, en primer lugar, se ha desarrollado un novedoso método de segmentación, nunca antes aplicado en imágenes de este tipo, conocido como *Locally Constrained Watershed Transform* (LCWT). Con esta técnica es posible llevar a cabo, a partir del lumen, la segmentación de todos los candidatos a glándula, que constituyen la entrada al algoritmo de la clasificación supervisada. Una de las novedades que incluye la técnica LCWT es la segmentación de la estructura glandular alrededor de la capa de núcleos epiteliales que presenta, que coincide con lo establecido según la literatura médica.

Tras realizar la segmentación de todos los candidatos a glándula, se aplican técnicas de *machine learning* clásico y de *deep learning* para clasificar cada candidato en su clase correspondiente y para establecer una comparativa entre ambas aproximaciones. Con respecto a las estrategias de *machine learning*, se ha llevado a cabo una robusta extracción de características combinando descriptores de forma y de textura con características texturales y un análisis fractal, con el objetivo de encontrar patrones en las imágenes que permitan discriminar tres tipos de elementos: artefactos, glándulas benignas y glándulas patológicas. En total, se computan 297 características como punto de partida al aprendizaje de los clasificadores. No obstante, primero se ha llevado a cabo un exhaustivo análisis estadístico cuya finalidad es seleccionar, de las 297 variables, únicamente aquellas que presentan información relevante en términos de independencia entre variables y dependencia con la clase.

A partir de la matriz de características con todos los candidatos a glándula, se han implementado técnicas de *k-fold cross validation* tanto de manera externa como de manera interna. En la partición externa se han dividido las imágenes en 5 conjuntos de datos, de tal forma que con 4 de ellos se entrena un modelo de clasificación, y con el otro, se evalúa dicho modelo. Esta operación se repite 5 veces para que todas las muestras sean utilizadas como parte del conjunto de entrenamiento y como parte del conjunto de test. Para la construcción de cada modelo, también se ha implementado un *cross validation* interno con $k=10$ particiones para hallar el valor de los hiperparámetros óptimos que permiten definir el mejor modelo posible. Cabe destacar que la partición de los datos basada en las técnicas de validación cruzada no se ha llevado a cabo de forma meramente aleatoria, sino que se ha implementado, como novedad

con respecto al estado del arte, una rigurosa estrategia en la que todas las muestras relativas a un mismo paciente se han establecido en un conjunto de datos único. De esta forma, no solo se particionan los datos en diferentes conjuntos atendiendo a la clase de cada uno, sino también atendiendo al historial clínico del que proceden. Por consiguiente, dado que en cada iteración se testean las muestras de un paciente nuevo, se consiguen una robustez de los modelos y una fiabilidad de los resultados mucho mayores.

Tras realizar la partición de los datos en los diferentes subconjuntos, se ha llevado a cabo la tarea de clasificación utilizando, por un lado, clasificadores de *machine learning* clásico, como SVM y perceptrón multicapa y, por otro lado, redes neuronales convolucionales basadas en una de las arquitecturas más populares de los métodos de *deep learning*, conocida como VGG19. Cabe destacar que se han definido un *kernel* lineal y otro cuadrático en los clasificadores de SVM, con el objetivo de afrontar la clasificación desde diferentes perspectivas.

Por otra parte, a fin de evaluar la técnica de segmentación desarrollada, se ha construido un *groundtruth* de 500 glándulas seleccionadas aleatoriamente. La bondad del método LCWT se ha determinado en base a los coeficientes *Jaccard* y *Dice*; y los resultados obtenidos se han comparado con los publicados en otros estudios del estado del arte, proporcionando relativas mejoras frente a los métodos existentes en este tipo de imágenes.

Cabe destacar que también se ha elaborado un *groundtruth* relativo a la clasificación, el cual es necesario en tres etapas fundamentales del proyecto. Por un lado, en el proceso de selección de características, para determinar qué variables tienen una mayor capacidad discriminativa. Por otro lado, en la tarea de la partición de datos, para llevar a cabo una separación balanceada de los mismos en función de su clase. Y, por último, en el proceso de clasificación supervisada, para evaluar el rendimiento de los clasificadores entrenados y para realizar una comparación entre las dos aproximaciones implementadas. Las métricas utilizadas para determinar el buen funcionamiento de los clasificadores han sido: sensibilidad, especificidad, valor predictivo positivo, valor predictivo negativo, *F-Score*, *accuracy* y área bajo la curva ROC (AUC). Conviene mencionar que, sorprendentemente, las estrategias basadas en *machine learning* clásico, sobre todo QSVM, proporcionan mejores resultados que las correspondientes al *deep learning*. También cabe destacar que se ha efectuado una comparación con los resultados proporcionados por otros autores en el estado del arte. No obstante, ha sido inviable realizar dicha comparación de forma directa, ya que no existen bases de datos de imágenes públicas sobre las que aplicar los diferentes métodos.

Finalmente, se ha realizado una simulación de una práctica clínica real mediante la creación de un comité de predicción, compuesto por los 5 modelos que se habían entrenado tras aplicar la partición de datos. Para ello se han seleccionado imágenes con las que no se ha trabajado en todo el proyecto y se ha intentado predecir la etiqueta de las glándulas identificadas en ellas. Los resultados conseguidos son realmente prometedores y se pretende adaptar la versión desarrollada en el presente TFM con vistas a la integración en el Hospital.

En definitiva, se puede concluir que se ha logrado diseñar y desarrollar un sistema automático de clasificación de estructuras glandulares en imágenes histológicas de próstata, con el objetivo de discriminar entre artefactos, glándulas benignas y glándulas patológicas de grado 3, de acuerdo al sistema de clasificación de *Gleason*.

4.2 Líneas futuras

Como ya se ha comentado, este trabajo aborda una de las etapas intermedias que forman parte de un proyecto nacional de mayor envergadura. Por tanto, la principal línea de futuro a contemplar se correspondería con los objetivos finales del proyecto SICAP, es decir, con el desarrollo de un prototipo completo de software que permita no solo distinguir artefactos, glándulas benignas y glándulas patológicas de grado 3, como en este proyecto, sino también estudiar la clasificación de los grados 4 y 5, a fin de proporcionar un sistema de ayuda al diagnóstico precoz del cáncer de próstata.

Basándose exclusivamente en lo que respecta a este proyecto de fin de máster, la línea principal de investigación que debería contemplarse reside en la implementación de nuevas arquitecturas de redes neuronales convolucionales, en lo que a técnicas de *deep learning* se refiere. En este proyecto, por motivos de tiempo, únicamente se ha utilizado la arquitectura VGG19, pero en líneas venideras sería conveniente hacer uso de otras redes como *Microsoft ResNet*, *GoogLeNet* o *Xception*. No obstante, para conseguir los buenos resultados esperados será necesario hacer uso de una cantidad de imágenes muy superior a la actual. Por este motivo, existe una continua comunicación entre los ingenieros del grupo CVB Lab y los patólogos del Hospital Clínico de Valencia.

Entrando más a nivel de detalle, es posible encontrar aspectos que mejorar en la mayoría de los métodos implementados, ya que ninguno es completamente perfecto. Conviene recordar que este proyecto se basa en un enfoque *end-to-end*, ya que únicamente se le pasa como entrada una imagen histológica de próstata y el conjunto de algoritmos devuelve dicha imagen con las glándulas marcadas en un color específico según el tipo de glándula detectado. Por tanto, la mejora de los algoritmos intermedios de cualquiera de las etapas contribuiría a una clasificación de mayor éxito.

Los márgenes de mejora más amplios radican en los métodos correspondientes al *clustering* y a la segmentación de las unidades de glándulas. Para el primero, en este TFM se han utilizado diferentes espacios de color en busca de aquel que fuese más eficaz a la hora de separar los distintos componentes de las imágenes. No obstante, todavía existe un amplio margen de mejora a la hora de erradicar por completo los artefactos de lúmenes, sin eliminar las regiones de interés. Con respecto a la segmentación, en este TFM se ha propuesto un método totalmente nuevo para intentar mejorar el problema de la clasificación a partir de estudiar glándulas unitarias. Sin embargo, es necesario que la segmentación sea precisa para llevar a cabo un enfoque como el propuesto en este trabajo. A pesar de que los resultados de la segmentación son considerablemente buenos, se espera que la optimización de los parámetros de entrada a la segmentación conlleve una gran mejora de los resultados de la posterior clasificación.

Bibliografía

- [1] “Cáncer,” *Organización Mundial de la Salud*, 2017. [Online]. Available: <http://www.who.int/topics/cancer/es/>. [Accessed: 24-Jun-2018].
- [2] “¿Qué es el cáncer y cómo se desarrolla?,” *Sociedad Española de Oncología*. [Online]. Available: <https://www.seom.org/es/informacion-sobre-el-cancer/que-es-el-cancer-y-como-se-desarrolla>. [Accessed: 24-Jun-2018].
- [3] “Las cifras del cáncer en España,” *Sociedad Española de Oncología*, 2018. [Online]. Available: <https://www.seom.org/es/noticias/106525-las-cifras-del-cancer-en-espana-2018>. [Accessed: 24-Jun-2018].
- [4] “Estadísticas del cáncer de próstata,” *American Society and Clinical Oncology*,. [Online]. Available: <https://www.cancer.net/es/tipos-de-cancer/cancer-de-prostata/estadisticas>. [Accessed: 24-Jun-2018].
- [5] “Tacto rectal,” *Galened Internacional*. [Online]. Available: <http://galened.com/tacto-rectal/>. [Accessed: 25-Jun-2018].
- [6] “Prueba del PSA,” *dreamstime*. [Online]. Available: <https://es.dreamstime.com/foto-de-archivo-prueba-del-psa-image66865075>. [Accessed: 25-Jun-2018].
- [7] “PCA3 test,” *Harley Street Group*. [Online]. Available: <http://www.harleystreetinternational.com/article.php?id=28>. [Accessed: 25-Jun-2018].
- [8] A. Pérez, “Biopsia de próstata.” [Online]. Available: <https://alejandroperezvazquez.wordpress.com/ultrasonido-y-biopsia-de-prostata/>. [Accessed: 25-Jun-2018].
- [9] “Cáncer de próstata,” *Sociedad Española de Oncología*. [Online]. Available: <https://www.seom.org/es/info-sobre-el-cancer/prostata?showall=1>. [Accessed: 25-Jun-2018].
- [10] “PCA3,” *Laborcentro*. [Online]. Available: <http://www.laborcentro.es/Test-PCA3-Prostata/>. [Accessed: 25-Jun-2018].
- [11] “Biopsia de próstata,” *Society of Interventional Radiology*. [Online]. Available: <https://www.radiologyinfo.org/sp/info.cfm?pg=prostate-biopsy>. [Accessed: 25-Jun-2018].
- [12] “Proceso histológico,” *Atlas de Histología Vegetal y Animal*, 2017. [Online]. Available: <https://mmegias.webs.uvigo.es/6-tecnicas/1-proceso.php>. [Accessed: 25-Jun-2018].
- [13] A. Hernández, “La célula, unidad estructural y funcional,” *Slideshare*, 2013. [Online]. Available: <https://www.slideshare.net/Alberkar/la-clula-unidad-estructural-y-funcional-el-ncleo-2013-81584011>. [Accessed: 25-Jun-2018].
- [14] K. Nguyen, B. Sabata, and A. K. Jain, “Prostate cancer grading: Gland segmentation and structural features,” *Pattern Recognit. Lett.*, vol. 33, no. 7, pp. 951–961, May 2012.
- [15] J. Gordetsky and J. Epstein, “Grading of prostatic adenocarcinoma: current state and prognostic implications,” *Diagn. Pathol.*, vol. 11, no. 1, p. 25, Dec. 2016.
- [16] A. Andrion *et al.*, “Malignant mesothelioma of the pleura: interobserver variability.,” *J. Clin. Pathol.*, vol. 48, no. 9, pp. 856–60, Sep. 1995.
- [17] G. D. Thomas, M. F. Dixon, N. C. Smeeton, and N. S. Williams, “Observer variation in the histological grading of rectal carcinoma.,” *J. Clin. Pathol.*, vol. 36, no. 4, pp. 385–91, Apr. 1983.
- [18] C. Gunduz-Demir, M. Kandemir, A. B. Tosun, and C. Sokmensuer, “Automatic segmentation of colon glands using object-graphs,” *Med. Image Anal.*, vol. 14, no. 1, pp. 1–12, Feb. 2010.
- [19] “MicroDraw.” [Online]. Available: <http://microdraw.pasteur.fr/>. [Accessed: 10-Jul-2018].
- [20] “OpenSeadragon.” [Online]. Available: <http://openseadragon.github.io/>. [Accessed: 10-Jul-2018].

- [21] M. S. Adam Goode, Benjamin Gilbert, Jan Harkes, Drazen Jukic, "OpenSlide." [Online]. Available: <https://openslide.org/>. [Accessed: 02-Jul-2018].
- [22] S. Dasgupta and Y. Freund, "Random projection trees for vector quantization," 2008.
- [23] P. Soille, *Morphological Image Analysis: Principles and Applications*. Springer Berlin Heidelberg, 2004.
- [24] J. Ren, E. Sadimin, D. J. Foran, and X. Qi, "Computer aided analysis of prostate histopathology images to support a refined Gleason grading system," 2017, vol. 10133, p. 101331V.
- [25] R. Beare, "A locally constrained watershed transform," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 7, pp. 1063–1074, Jul. 2006.
- [26] S. Naik, S. Doyle, M. Feldman, J. Tomaszewski, and A. Madabhushi, "Gland Segmentation and Computerized Gleason Grading of Prostate Histology by Integrating Low-, High-level and Domain Specific Information."
- [27] K. Nguyen, A. Sarkar, and A. K. Jain, "Structure and Context in Prostatic Gland Segmentation and Classification," Springer, Berlin, Heidelberg, 2012, pp. 115–123.
- [28] "The Watershed Transform," *Mathworks*. [Online]. Available: <https://es.mathworks.com/company/newsletters/articles/the-watershed-transform-strategies-for-image-segmentation.html>. [Accessed: 04-Jul-2018].
- [29] R. Farjam, H. Soltanian-Zadeh, K. Jafari-Khouzani, and R. A. Zoroofi, "An image analysis approach for automatic malignancy determination of prostate pathological images," *Cytom. Part B Clin. Cytom.*, vol. 72B, no. 4, pp. 227–240, Jul. 2007.
- [30] J. T. Kwak and S. M. Hewitt, "Multiview boosting digital pathology analysis of prostate cancer," *Comput. Methods Programs Biomed.*, vol. 142, pp. 91–99, Apr. 2017.
- [31] J. Diamond, N. H. Anderson, P. H. Bartels, R. Montironi, and P. W. Hamilton, "The use of morphological characteristics and texture analysis in the identification of tissue composition in prostatic neoplasia," *Hum. Pathol.*, vol. 35, no. 9, pp. 1121–1131, Sep. 2004.
- [32] S. Naik, S. Doyle, M. Feldman, J. Tomaszewski, and A. Madabhushi, "Gland Segmentation and Computerized Gleason Grading of Prostate Histology by Integrating Low-, High-level and Domain Specific Information.," *Proc. 2nd Work. Microscopic Image Anal. with Appl. Biol.*, no. September 2015, pp. 1–8, 2007.
- [33] K. Nguyen, B. Sabata, and A. K. Jain, "Prostate cancer grading: Gland segmentation and structural features," in *Pattern Recognition Letters*, 2012, vol. 33, no. 7, pp. 951–961.
- [34] A. C. Ruifrok and D. A. Johnston, "Quantification of histochemical staining by color deconvolution," *Anal Quant Cytol Histol*, vol. 23, pp. 291–299, 2001.
- [35] J. Ríos-Díaz, J. Javier Martínez-Payá, and E. Del Baño Aledo, "El análisis textural mediante las matrices de co-ocurrencia (GLCM) sobre imagen ecográfica del tendón rotuliano es de utilidad para la detección cambios histológicos tras un entrenamiento con plataforma de vibración Textural Analysis by Means of a Grey Level Co-Occurrence Matrix Method on Patellar Tendon Ultrasonography is Useful for the Detection of Histological Changes after Whole-Body Vibration Training."
- [36] M. D. Difranco, G. O. Hurley, E. W. Kay, and R. W. G. Watson, "Automated Gleason Scoring of Prostatic Histopathology Slides Using Multi-channel Co-occurrence Texture Features," *Proc. 3rd Work. Microscopic Image Anal. with Appl. Biol. New York, NY, USA.*, 2002.
- [37] M. Presutti, "Tema do Trabalho: Recursos Didáticos para o Ensino de Sensoriamento Remoto."
- [38] Zhenhua Guo, Lei Zhang, and D. Zhang, "A Completed Modeling of Local Binary Pattern Operator for Texture Classification," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1657–1663, Jun. 2010.
- [39] G. Zhao and M. Pietikainen, "Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29,

- no. 6, pp. 915–928, Jun. 2007.
- [40] Xiangsheng Huang, S. Z. Li, and Yangsheng Wang, “Shape Localization Based on Statistical Method Using Extended Local Binary Pattern,” in *Third International Conference on Image and Graphics (ICIG’04)*, pp. 184–187.
- [41] N. Ing *et al.*, “Machine learning approaches to analyze histological images of tissues from radical prostatectomies,” *Comput. Med. Imaging Graph.*, vol. 46, pp. 197–208, Dec. 2015.
- [42] T. Ojala, M. Pietikäinen, and D. Harwood, “A comparative study of texture measures with classification based on featured distributions,” *Pattern Recognit.*, vol. 29, no. 1, pp. 51–59, Jan. 1996.
- [43] L. Wang and D.-C. He, “TEXTURE CLASSIFICATION USING TEXTURE SPECTRUM,” *Pattern Recognit.*, vol. 2390, no. 8, pp. 5–910, 1990.
- [44] T. Ojala, M. Pietikainen, and T. Maenpaa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
- [45] M. Pietikäinen, T. Ojala, and Z. Xu, “Rotation-invariant texture classification using feature distributions,” *Pattern Recognit.*, vol. 33, no. 1, pp. 43–52, Jan. 2000.
- [46] A. Tabesh *et al.*, “Multifeature prostate cancer diagnosis and gleason grading of histological images,” *IEEE Trans. Med. Imaging*, vol. 26, no. 10, pp. 1366–1378, 2007.
- [47] Po-Whei Huang and Cheng-Hsiung Lee, “Automatic Classification for Pathological Prostate Images Based on Fractal Analysis,” *IEEE Trans. Med. Imaging*, vol. 28, no. 7, pp. 1037–1050, Jul. 2009.
- [48] HURST and H. E., “Long Term Storage,” *An Exp. Study*, 1965.
- [49] “Diagrama de caja y bigotes,” *wikipedia*. [Online]. Available: https://es.wikipedia.org/wiki/Diagrama_de_caja#/media/File:Boxplot.svg. [Accessed: 06-Jul-2018].
- [50] M. Jordan, J. Kleinberg, and B. Schölkopf, “Pattern Recognition and Machine Learning.”
- [51] S. Sahran, A. Abdullah, D. Albashish, and S. J. Abu-Taleb, “Multi-level learning approach for prostate histopathology images classification,” in *2017 8th International Conference on Information Technology (ICIT)*, 2017, pp. 947–955.
- [52] C.-C. Chang and C.-J. Lin, “LIBSVM: A Library for Support Vector Machines,” 2001.
- [53] “Statistics and machine learning toolbox,” *Mathworks*. [Online]. Available: <https://es.mathworks.com/products/statistics/features.html#aprendizaje-automático>. [Accessed: 11-Jul-2018].
- [54] J. Fürnkranz, “Round Robin Classification,” *J. Mach. Learn. Res.*, vol. 2, pp. 721–747, 2002.
- [55] A. Shaukat *et al.*, “Automatic Cancerous Tissue Classification using Discrete Wavelet Transformation and Support Vector Machine,” *J. Basic. Appl. Sci. Res*, vol. 6, no. 7, pp. 1–1, 2016.
- [56] K. Nguyen, A. Sarkar, and A. K. Jain, “Prostate Cancer Grading: Use of Graph Cut and Spatial Arrangement of Nuclei,” *IEEE Trans. Med. Imaging*, vol. 33, no. 12, pp. 2254–2270, Dec. 2014.
- [57] S. Doyle, M. D. Feldman, N. Shih, J. Tomaszewski, and A. Madabhushi, “Cascaded discrimination of normal, abnormal, and confounder classes in histopathology: Gleason grading of prostate cancer,” *BMC Bioinformatics*, vol. 13, no. 1, p. 282, Oct. 2012.
- [58] A. Almuntashri, S. Agaian, I. Thompson, D. Rabah, O. Zin Al-Abdin, and M. Nicolas, “Gleason grade-based automatic classification of prostate cancer pathological images,” in *2011 IEEE International Conference on Systems, Man, and Cybernetics*, 2011, pp. 2696–2701.
- [59] C. Rico, M. Paredes, N. Fernández, and B. Marino, “MODELACIÓN DE LA ESTRUCTURA JERÁRQUICA DE MACROINVERTEBRADOS BENTÓNICOS A TRAVÉS DE REDES NEURONALES ARTIFICIALES Modeling of the Hierarchical Structure of Freshwater Macroinvertebrates Using Artificial Neural Networks,” *Acta biol. Colomb*, vol. 14, pp. 71–96, 2009.

- [60] V. G. Vílchez, “Estimación y clasificación de daños en materiales utilizando modelos AR y redes neuronales para la evaluación no destructiva con ultrasonidos,” 2010.
- [61] “Sobreajuste,” *wikipedia*. [Online]. Available: <https://es.wikipedia.org/wiki/Sobreajuste>. [Accessed: 08-Jul-2018].
- [62] N. Kumar *et al.*, “Convolutional neural networks for prostate cancer recurrence prediction,” 2017, vol. 10140, p. 101400H.
- [63] N. Zhou, A. Fedorov, F. Fennessy, R. Kikinis, and Y. Gao, “Large scale digital prostate pathology image analysis combining feature extraction and deep neural network.”
- [64] E. Arvaniti *et al.*, “Automated Gleason grading of prostate cancer tissue microarrays via deep learning,” *bioRxiv*, p. 280024, Mar. 2018.
- [65] Y. Lecun *et al.*, “LEARNING ALGORITHMS FOR CLASSIFICATION: A COMPARISON ON HANDWRITTEN DIGIT RECOGNITION.”
- [66] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks.” pp. 1097–1105, 2012.
- [67] J. Dean *et al.*, “Large Scale Distributed Deep Networks.”
- [68] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998.
- [69] Y. Bengio, “Deep Learning of Representations for Unsupervised and Transfer Learning,” vol. 27, pp. 17–37, 2012.
- [70] J. Utrera, “VGG19 - Deep learning.” [Online]. Available: <http://jesuutrer.com/articles/article06.html>. [Accessed: 16-Jul-2018].
- [71] H.-H. Chang, A. H. Zhuang, D. J. Valentino, and W.-C. Chu, “Performance measure characterization for evaluating neuroimage segmentation algorithms,” *Neuroimage*, vol. 47, no. 1, pp. 122–135, Aug. 2009.
- [72] M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, Jul. 2009.
- [73] D. G. Altman and J. M. Bland, “Statistics Notes: Diagnostic tests 2: predictive values,” *BMJ*, vol. 309, no. 6947, pp. 102–102, Jul. 1994.
- [74] Y. Sasaki and R. Fellow, “The truth of the F-measure,” 2007.
- [75] Y. Peng, Y. Jiang, L. Eisengart, M. A. Healy, F. H. Straus, and X. J. Yang, “Segmentation of prostatic glands in histology images,” in *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2011, pp. 2091–2094.
- [76] C. M. Lopez *et al.*, “Exploration of efficacy of gland morphology and architectural features in prostate cancer gleason grading,” in *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2012, pp. 2849–2854.
- [77] J. P. Monaco *et al.*, “High-throughput detection of prostate cancer in histological sections using probabilistic pairwise Markov models,” *Med. Image Anal.*, vol. 14, no. 4, pp. 617–629, Aug. 2010.
- [78] J. Xu, A. Janowczyk, S. Chandran, and A. Madabhushi, “A high-throughput active contour scheme for segmentation of histopathological imagery,” *Med. Image Anal.*, vol. 15, no. 6, pp. 851–862, Dec. 2011.
- [79] R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, *Machine Learning : an Artificial Intelligence Approach*. Springer Berlin Heidelberg, 1983.
- [80] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
- [81] D. Narváez, “Microscopía.” [Online]. Available: http://www.medic.ula.ve/histologia/anexos/microscopweb/MONOWEB/capitulo3_4.htm. [Accessed: 02-Jul-2018].

II. PRESUPUESTO

Índice del presupuesto

1.	Presupuestos parciales	II
1.1	Coste de personal.....	II
1.2	Coste de los materiales	II
1.3	Coste de las herramientas hardware y software	III
2.	Presupuesto total.....	IV

1. Presupuestos parciales

En este apartado del informe se detallan las diferentes partes que conforman los presupuestos parciales de este proyecto basado en el diseño y desarrollo de un sistema automático de clasificación de estructuras glandulares en imágenes histológicas de próstata. Para cuantificar los presupuestos parciales se han desglosado los costes en tres grupos bien diferenciados: (i) coste de personal, (ii) coste de los materiales y (iii) coste de las herramientas hardware y software, utilizadas para llevar a cabo el TFM.

1.1 Coste de personal

En este punto se tienen en cuenta los recursos humanos necesarios para el desarrollo del presente trabajo fin de máster. De esta manera, se detalla la remuneración de cada participante atendiendo a una estimación de los costes que viene determinada por el tiempo dedicado al proyecto, en unidades de horas. Por tanto, debe considerarse la contribución al trabajo de: D^a Valery Naranjo Ornedo (como tutora del proyecto), de D. Adrián Colomer Granero (como cotutor) y de D. José Gabriel García Pardo (como alumno y autor del proyecto). Teniendo en cuenta que el presente TFM se ha llevado a cabo a lo largo de 9 meses, manteniendo una hora de reuniones semanales con los dos supervisores, se recogen los costes del presupuesto de personal en la siguiente tabla:

Tabla 1. Presupuesto para el coste de personal.

Denominación	Uds.	Cantidad	Precio unitario (€)	Total (€)
Directora del TFM (Catedrática)	h	36	42,00	1.512,00
Cotutor del TFM (Doctor)	h	36	20,50	738,00
Autor del TFM (Alumno de MUIB)	h	600	13,80	8.280,00
TOTAL				10.530,00

1.2 Coste de los materiales

Para llevar a cabo el presupuesto de este apartado se ha tenido en cuenta, por una parte, el coste de la realización de una prueba de biopsia, a partir de la cual se obtienen las imágenes con las que se trabaja en este proyecto. El número de biopsias se toma como el número de pacientes individuales que han sido analizados en este estudio. Por otra parte, se considera el coste que supone la adquisición de cada muestra relativa al corte histológico donde se presentan las imágenes de tejido prostático teñidas con hematoxilina y eosina (H&E). En este caso, las muestras corresponden a cada uno de los *slides* donde se fijan las secciones de tejido extraídas al realizar la biopsia.

Tabla 2. Presupuesto para el coste de los materiales.

Denominación	Uds.	Cantidad	Precio unitario (€)	Total (€)
Biopsia	u	25	600,00	15.000,00
Muestras	u	35	10,00	350,00
TOTAL				15.350,00

1.3 Coste de las herramientas hardware y software

En este apartado se detalla el presupuesto parcial correspondiente a los recursos hardware y software necesarios para llevar a cabo la totalidad del proyecto. Dado que estas herramientas no se han obtenido específicamente para la elaboración del trabajo, es necesario tener en cuenta el periodo de amortización para cada uno de ellos. Ese periodo se relaciona con la vida útil del material y el intervalo de tiempo amortizado para cada herramienta. Cabe destacar, en este último punto, que toda la parte correspondiente al software implementado para la realización del proyecto ha tenido un coste cero, puesto que, por una parte, las licencias de Microsoft Word y de MATLAB 2018a[®] son proporcionadas gratuitamente por la Universitat Politècnica de València a sus estudiantes. Y, por otra parte, el IDE PyCharm 2018.4 y las librerías cuDNN y OPENCV son de código abierto para cualquier persona.

Tabla 3. Presupuesto de las herramientas hardware y software.

Denominación	Uds.	Cantidad	Precio unitario (€)	Periodo de amortización (años)	Intervalo amortizado (meses)	Total (€)
<i>Software para machine learning y deep learning</i>						
Licencia de Office 365	u	1	0,00	1	9	0,00
Licencia de MATLAB 2018a [®]	u	1	0,00	1	9	0,00
Licencia de Python 3.5	u	1	0,00	1	9	0,00
Librerías cuDNN y OPENCV	u	1	0,00	1	9	0,00

Hardware relativo al equipo de <i>Machine Learning</i>						
Procesador Intel® Core™ i7-4790K	u	1	387,30	4	9	72,62
Memoria DDR3 de 8 GB	u	2	68,90	4	9	25,84
Tarjeta gráfica GeForce GTX760 2GB	u	1	210,00	4	9	39,28
Almacenamiento HDD de 2TB	u	1	62,00	4	9	11,63
Almacenamiento SSD de 256 GB	u	1	45,00	4	9	8,44
Hardware relativo al equipo de <i>Deep Learning</i>						
Procesador Intel i7 @4.20GHz	u	1	344,00	4	9	64,50
Memoria DDRAM4 16 GB	u	2	165,00	4	9	61,88
Tarjeta Gráfica NVIDIA Titan XP	u	1	1299,00	4	9	243,56
Disco SSD de 250 GB	u	1	77,00	4	9	14,44
TOTAL						542,19

2. Presupuesto total

Para el presupuesto total del proyecto es necesario tener en cuenta la suma de los presupuestos parciales definidos en el capítulo anterior. Además, deben añadirse al cálculo dos porcentajes establecidos: el de gastos generales (13%) y el asociado al beneficio industrial (6%). Por último, es preceptivo repercutirle al resultado final el 21% de IVA, obteniendo de esta forma el presupuesto total que supondría la realización de este trabajo fin de máster.

Tabla 4. Presupuesto total.

CAPÍTULOS	IMPORTE (€)
1. Coste de personal	10.530,00
2. Coste de los materiales	15.350,00
3. Coste de herramientas hardware y software	542,19
PRESUPUESTO DE EJECUCIÓN DE MATERIAL	26.422,19
13% de gastos generales	3.434,88
6% de beneficio general	1.585,33
SUMA	31.442,40
21% de IVA	6.602,90
PRESUPUESTO TOTAL	38.045,30

