The logo of the Universidad Nacional de Educación a Distancia (UNED) is displayed in white, bold, sans-serif capital letters against a dark green rectangular background.

UNED

Máster Universitario en Inteligencia Artificial
2016/2017

Trabajo Fin de Máster:

**Segmentación y Extracción de
Biomarcadores en Pulmones
infectados por *Mycobacterium
Tuberculosis***

Pedro Macías Gordaliza

Tutor/es

Mariano Rincón Zamorano

Arrate Muñoz Barrutia

Madrid, 17 de Septiembre de 2017

Palabras clave: Tuberculosis, pulmón, segmentación, clasificación, lesión, infección, imagen médica, Tomografía Axial Computarizada, implementación de *software*.

Resumen: En este trabajo se introducen dos nuevas metodologías esenciales en la evaluación de nuevos fármacos destinados a la erradicación de la Tuberculosis utilizando imágenes de tomografía axial computarizada de alta resolución. Ambas técnicas hacen uso de múltiples herramientas del campo de la inteligencia artificial, con el fin de automatizar los procedimientos y extraer información relevante dentro de un contexto altamente multidisciplinar. Así la primera de las herramientas presentadas, presenta un método capaz de segmentar automáticamente pulmones infectados. Mientras que la segunda metodología expone un nuevo flujo de trabajo para la clasificación de regiones anómalas en pulmones infectados, mediante el uso de la textura en imágenes afrontando un problema clásico en los datos procedente de aplicaciones biomédicas, como es el desequilibrio entre las clases a clasificar.

Índice

1	Introducción	6
2	Motivación y objetivos	8
3	Revisión de la bibliografía	11
3.1	Técnicas para el diagnóstico de la tuberculosis	11
3.2	Caracterización de la tuberculosis en imágenes de tomografía computarizada	12
3.3	Segmentación de pulmones en imágenes radiológicas	14
3.4	Clasificación de regiones anómalas en la imagen radiológica	17
3.4.1	Clasificación basada en la textura de la imagen	18
3.4.2	Aprendizaje utilizando datos no balanceados	18
3.4.3	Selección de características	20
4	Materiales	22
5	Métodos	23
5.1	Métodos: Segmentación del pulmón	23
5.1.1	Preprocesado	23
5.1.2	Segmentación inicial del conjunto de pulmones y vías aéreas	24
5.1.3	Umbralización	24
5.1.4	Análisis de conectividad y topológico	27
5.1.5	Extracción de las vías aérea	27
5.1.6	Inclusión de las lesiones en el interior del pulmón	30
5.1.7	Refinamiento basado en el patrón de las lesiones de tuberculosis	31
5.1.8	Métodos para la evaluación de los resultados de la segmentación	32
5.2	Métodos: Clasificación de lesiones en pulmones infectados	33
5.2.1	Extracción de características de textura	33
5.2.2	Selección de características relevantes	34
5.2.3	Clasificación de características: MI-mRMR	35
5.2.4	Evaluación de características	36
5.2.5	Balanceo de datos	37
5.2.6	Clasificación de regiones utilizando clasificadores optimizados	37
6	Desarrollo de software	40
6.1	Software para la segmentación de pulmón	40
6.1.1	Aplicación e instalación	42
6.2	Software para la extracción de características de textura	45
7	Resultados Experimentales	47
7.1	Resultados para la segmentación del pulmón	47
7.1.1	Resultados cuantitativos	48
7.1.2	Resultados cualitativos	49
7.2	Resultados de la clasificación de lesiones	52
7.2.1	Selección de características	52
7.2.2	Resultados del proceso de optimización	56
8	Discusión	58
8.1	Discusión de los resultados de segmentación	58
8.2	Discusión de los resultados de clasificación	59

9 Conclusiones	61
Apéndice A: Tablas de resultados para la selección de características	62
Apéndice B: Definición de características de textura y expresiones utilizadas para el cálculo	64

Lista de figuras

1	Proyección descendente de la incidencia de la tuberculosis	7
2	Espectro continuo de la infección por Tuberculosis	8
3	Pulmones de una imagen <i>Computed Tomography (HRCT)</i>	9
4	Imagen de CT para TB	12
5	Imagen de granulomas y conglomeraciones causados por la TB	13
6	Imagen de árbol en brote o <i>tree in bud</i> causada por la tuberculosis.	13
7	Imagen de consolidación causada por la tuberculosis.	13
8	Imagen de caverna causada por la tuberculosis.	14
9	Imagen de <i>ground glass opacity</i> causado por la tuberculosis.	14
10	Estado del arte de los algoritmos de segmentación de pulmón	16
11	Estado del arte de los algoritmos de segmentación de pulmón-2	17
12	Esquema general del flujo de trabajo implementado para realizar la segmentación de pulmones infectados por TB	24
13	Ejemplo de distribución bimodal de una imagen de CT	26
14	Extracción de las vías aéreas	30
15	Inclusión de lesiones de tuberculosis	31
16	Esquema general para la clasificación de lesiones	39
17	Esquema de validación cruzada introduciendo balanceo de los datos de entrenamiento	39
18	Diagrama de clases para la herramienta de segmentación	42
19	Instalación de herramienta: Configuración de CMake	44
20	Instalación de herramienta: Compilar con <i>Unix Makefiles</i>	44
21	Instalación de herramienta: Make	45
22	Instalación de herramienta: Compilar filtros	45
23	Dispersión de la similitud 1	49
24	Dispersión de la similitud 2	49
25	Comparación de segmentación automática y manual	51
26	Segmentación automática en modelo animal 2	52
27	Mapa de calor para el ranking de características	54
28	Selección de características con mRMR	54
29	Diagrama de cajas de los resultados de clasificación utilizando 10 características mRMR	56
30	Diagramas de cajas de los resultados de la validación cruzada.	57

Índice de tablas

1	Lesiones etiquetadas	22
2	Desplazamientos correspondientes a las 13 direcciones entre vóxeles adyacentes en el espacio tridimensional.	34
3	Resultados de similitud para la segmentación automática	48
4	Resultado de un ANOVA (<i>ANalysis Of VAriance</i>) de tres factores: el nivel de cuantificación utilizado, el número de características utilizadas y la partición de datos. Se muestra el valor para el estadístico <i>F</i>	55
5	Tukey Test para diferencia de medias entre niveles de cuantización	55
6	Resultados de clasificación utilizando 10 características de MI-mRMR	56
7	Resultados optimización para 8,16 y 32	57
8	Resultados optimización para 8,16 y 32	57
9	Resultado de un ANOVA (<i>ANalysis Of VAriance</i>) de tres factores: el nivel de cuantificación utilizado, el número de características utilizadas y el balanceo. Se muestra el valor para el estadístico <i>F</i>	58
10	Selección de características: Precisión empleando entre 1 y 5	62
11	Selección de características: Precisión empleando entre 6 y 10	62
12	Selección de características: Precisión empleando entre 11 y 15	62
13	Selección de características: Precisión empleando entre 16 y 20	62
14	Selección de características: Precisión empleando entre 21 y 25	63
15	Ranking de características	63

1 Introducción

La *Tuberculosis* (TB) es una enfermedad infecciosa que se contagia por vía aérea y que principalmente afecta al sistema respiratorio, especialmente a los pulmones, aunque en fases muy avanzadas de la enfermedad esta puede presentarse extra-pulmonarmente [1, 2].

El agente causante de esta enfermedad, la bacteria *Mycobacterium tuberculosis* (*Mtb*) o *bacilo de Koch*, fue descubierto hace más de un siglo, en 1882, por el premio Nobel de Medicina que le da nombre, *Robert Koch*. Esto ha supuesto que esta patología haya podido ser ampliamente estudiada por la comunidad científica. Desde un primer momento se establecieron relaciones consecuentes entre el desarrollo de la enfermedad y el estado del sistema inmune de los pacientes. De esta forma, aquellos sujetos cuyo sistema inmunológico se encuentra en mejores condiciones, generalmente favorecido por unas condiciones de vida adecuadas, son capaces de contener la infección, mientras que aquellos que se encuentran inmunodeprimidos (generalmente debido a la desnutrición o a la presencia de enfermedades que atacan el sistema inmune) se ven afectados por una progresión severa de la tuberculosis. Este hecho ha provocado que bajo la visión individualista de occidente, donde las causas de la inmunodeficiencia se suelen atajar con éxito entre la mayoría de la población, la TB sea vista tradicionalmente, como una enfermedad marginal. En concreto se piensa que afecta sólo a los más desfavorecidos (países en vías de desarrollo o países del tercer mundo) o a pacientes crónicos afectados por enfermedades como el sida/VIH (infección por el virus de la inmunodeficiencia humana y el síndrome de inmunodeficiencia adquirida), provocando un desinterés social y colocando a la TB entre las denominadas enfermedades olvidadas (*neglected diseases*) [3].

Sin embargo, esta postura choca dramáticamente con la realidad. En 2017, la TB continua siendo la segunda enfermedad que más muertes causa en el mundo sólo superada por el VIH. Además, aunque clásicamente es una enfermedad asociada a países con índices de pobreza altos, la enfermedad en su modo asintomático y no contagioso, pero no curado, denominado *tuberculosis latente* (*LTB*) alcanza cifras peligrosamente altas también en occidente.

Estos datos han provocado la reacción apremiante de varios gobiernos, instituciones y asociaciones en occidente. Principalmente, la *World Health Organization* (WHO) recoge en su último informe global sobre tuberculosis, conocido en inglés como *Global Tuberculosis Report* [4] que en 2016 se detectaron 10.4 millones de nuevos casos de TB y 1.4 millones de muertes. Además, los test inmunológicos muestran evidencias de que un tercio de la población mundial, dos billones de personas, padece tuberculosis latente [5]. Estos números se ven agravados al tener en cuenta los datos del mismo informe referentes a la resistencia a los fármacos utilizados (*Multidrug-Resistant Tuberculosis*, *MDR-TB*), que ascendieron a 480.000 personas en 2015.

Por ello, esta misma organización se encargó de la reciente puesta en marcha de un plan de actuación conocido como *End TB Strategy* [6]¹, con el fin de erradicar la TB antes del 2035 (ver figura 1), aunque los últimos informes señalan el 2050.

¹Actualmente la mayoría de proyectos que habían sido promovidos por otras organizaciones, como la iniciativa 6 de *United Nations Millennium Development Goals* están incluidas en el plan dirigido por la WHO

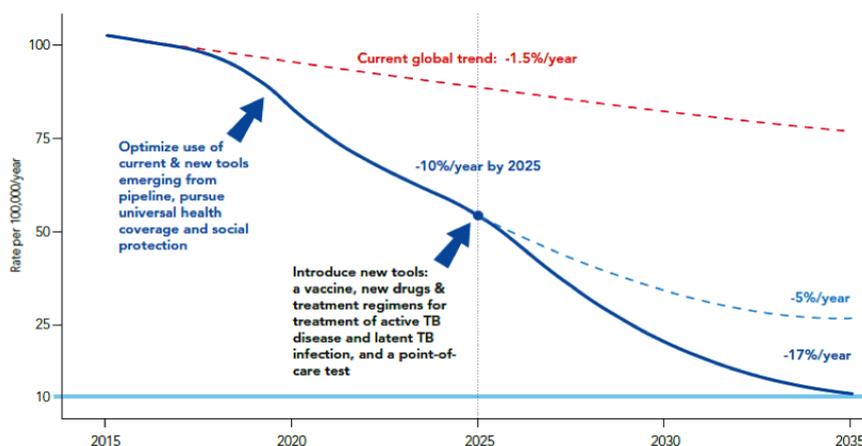


Figura 1: Decrecimiento proyectado de la incidencia de la *tuberculosis*(TB) con la optimización de las herramientas actuales. Fuente([7])

El plan elaborado por la WHO manifiesta que, para alcanzar estos objetivos ambiciosos es necesario mejorar las herramientas existentes e introducir técnicas novedosas. Así es necesario el desarrollo de nuevas vacunas más efectivas y medicamentos más eficientes y seguros para lidiar con la MDR-TB. Se hace necesaria también la implementación de nuevos protocolos para el diagnóstico basados en medidas cuantitativas.

Este trabajo se enmarca dentro de este ámbito, el desarrollo de nuevas técnicas que ayuden a la erradicación de la TB, concretamente, en el desarrollo de métodos que permitan evaluar de manera cuantitativa la efectividad de los nuevos medicamentos contra la TB, a partir de una imagen radiológica de los pulmones de un sujeto infectado.

Los motivos sobre la idoneidad del uso de imagen radiológica, frente a otros métodos tradicionales para evaluar el estado de la enfermedad. Así como la necesidad de utilizar, adaptar y desarrollar técnicas de visión artificial y clasificación, se esclarecen en la siguiente sección de este texto, Motivación. En ella, se presentan las limitaciones presentes en las técnicas utilizadas en la actualidad cuando se desea realizar un diagnóstico certero de la infección causada por el *Mtb*.

2 Motivación y objetivos

Como se comentó en la introducción, los fármacos que se utilizan tradicionalmente en el tratamiento de la TB, resultan ineficaces en muchos de los tratamientos al estar diseñados para combatir una familia determinada del *Mtb*. Mediante la acción de mecanismos evolutivos naturales, una familia del *Mtb* origina nuevas cepas inmunes a dichos fármacos (MDR-TB). La aparición de nuevas variedades se debe generalmente a un abandono temprano de los fármacos, debido a la desaparición de los síntomas de la enfermedad pero no necesariamente del causante de la misma, el *Mycobacterium tuberculosis*.

Este análisis delata el principal escollo cuando se afronta el problema de la TB. Esto es, no se puede entender la patología de forma binaria (se tiene TB o no se tiene), sino como un espectro continuo [5,8], como se muestra en la figura 2. Por ello, se debería ser capaz de caracterizar la cantidad de carga bacteriana en un individuo infectado, asintomático o no.

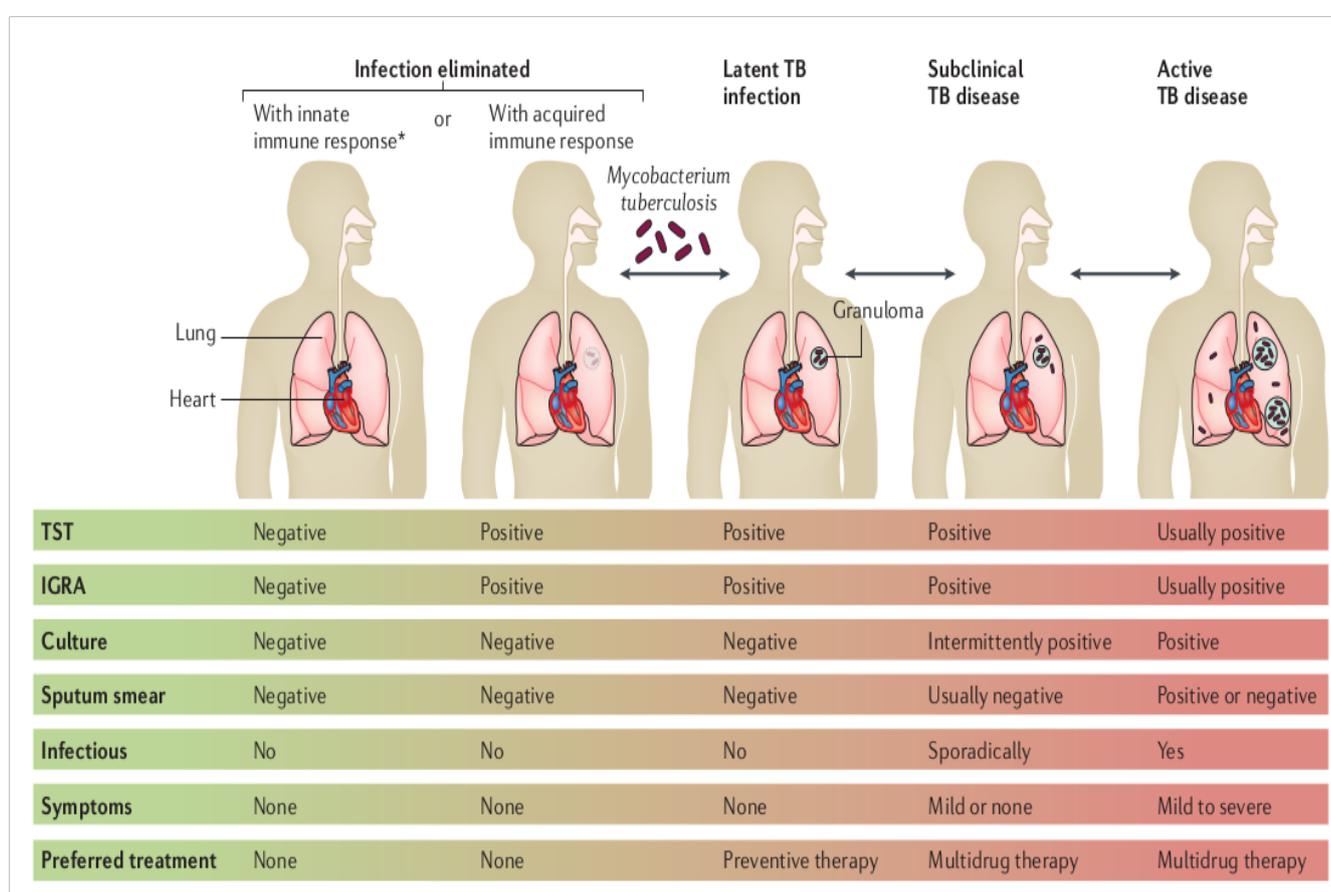


Figura 2: Espectro continuo de infección para *Tuberculosis* (TB). De la infección por *Mycobacterium tuberculosis* (*Mtb*) a TB activa. La figura especifica la precisión de los test habituales en cada uno de las etapas de la enfermedad, así como los síntomas y el tratamiento. NOTA: TST, Tuberculin Skin Test; Interferon Gamma Release Assay, IGRA. Imagen Extraída de [8]

Desafortunadamente, la aproximación binaria responde a una limitación técnica. Las pruebas realizadas tradicionalmente (Interferon gamma release assay, IFN- γ , IGRA) [9], los

cuales se presentan brevemente dentro la revisión de la bibliografía en la sección Técnicas para la detección de la tuberculosis, no tienen la suficiente sensibilidad para medir cuantitativamente el nivel de infección.

Por lo tanto, para evaluar la eficacia de los nuevos fármacos desarrollados, se necesitan nuevas herramientas capaces de medir de forma precisa la evolución de la enfermedad y con capacidad para medir cambios de la carga bacteriana. En este contexto, la imagen radiológica y concretamente, la tomografía axial computarizada de alta resolución o *High Resolution Computed Tomography (HRCT)*, o de forma abreviada CT, dispone de gran sensibilidad y especificidad [9,10] (ejemplo de una de imagen de HRCT fig.3).

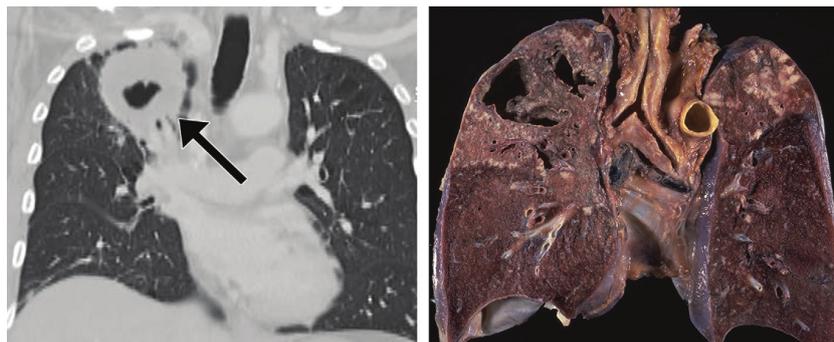


Figura 3: Vista coronal de los pulmones de una imagen *Computed Tomography (HRCT)* mostrando una cavidad debida a la TB (a la izquierda). Fotografía de un paciente diferente para una cavidad en la misma zona del pulmón (a la derecha). Imagen Extraída de [9]

Existen varios estudios, realizados tanto con humanos [11] como con modelos animales [12], donde utilizando bio-marcadores identificados por radiólogos expertos en imágenes de CT correspondientes a los pulmones (ver sección 3.2), se consigue cuantificar con éxito la evolución de la enfermedad. Sin embargo, esta tarea supone una cantidad enorme de recursos (tiempo, dinero, etc.) y está sujeta a la subjetividad de cada evaluador [13]. Estas limitaciones suponen un problema de enorme calado: **Es inviable utilizar interpretaciones visuales de las imágenes realizadas por expertos para estudios a gran escala.**

Los ensayos realizados por las empresas farmacéuticas contienen cientos o miles de imágenes volumétricas, correspondientes a varios sujetos, expuestos a diferentes dosis de infección y tratados con diferentes combinaciones de medicamentos. Es por ello que **la automatización de esta tarea resulta esencial.**

Resulta evidente que al tratarse de un problema complejo compuesto por múltiples partes, este puede abordarse desde muy distintas perspectivas y aproximaciones metodológicas. En este trabajo se explotan las virtudes que proporcionan diversas técnicas del campo de la inteligencia artificial con la pretensión de poder replicar la tarea realizada por un humano, en este caso un experto radiólogo.

Para ello, el trabajo se ha dividido en dos objetivos principales, que definen dos de los hitos fundamentales a completar de forma exitosa en un análisis de imágenes radiológicas para modelos de infección por *Mycobacterim tuberculosis*. Estos objetivos se definen de forma concisa como sigue:

1. **Segmentación automática de pulmones infectados:** Implementación de una herramienta para la extracción automática de la zona afectada por la TB, el parénquima pulmonar, a partir de la imagen volumétrica completa del CT.
2. **Clasificación automática de regiones anómalas en pulmones infectados:** Desarrollo de un flujo de trabajo que permita clasificar las diferentes irregularidades que aparecen en el parénquima pulmonar a raíz de la infección (ver sec.3.2).

Evidentemente, estos problemas, a su vez, se dividen en diferentes tareas que se exponen en el desarrollo del texto.

3 Revisión de la bibliografía

Actualmente, los equipos de HRCT permiten la adquisición isotrópica con resolución submilimétrica del tórax completo. La calidad de imagen que ofrece es muy superior a la de otras modalidades, como la resonancia magnética (*Magnetic Resonance Imaging*, MRI) o la tomografía por emisión de positrones (*Positron Emission Tomography*, PET)². Sin embargo, esta gran ventaja conlleva la capacidad de examinar grandes cantidades de datos de muy diversas formas [14], aunque parece claro que actualmente, las técnicas de visión artificial y aprendizaje resultan esenciales para facilitar la interpretación de imágenes de CT. Este hecho ha supuesto que las mejoras en las técnicas de adquisición se acompañen de numerosos estudios en torno al análisis de imágenes torácicas de CT.

En la literatura existen gran cantidad de aproximaciones que recogen diferentes metodologías referentes a los principales problemas que se plantean en este trabajo: la segmentación automática del pulmón y la clasificación de lesiones en el parénquima pulmonar. Sin embargo, la inmensa mayoría de estas, están centradas en el estudio de segmentación y clasificación para pulmones afectados por el cáncer de pulmón o el enfisema en humanos y no en el efecto de enfermedades infecciosas como la TB en modelos animales como los que se tratan en este trabajo.

El estudio de una patología diferente, en modelos animales específicos para esta, implica la adaptación de técnicas utilizadas con éxito en problemas similares y el desarrollo de nuevas, a partir del conocimiento del modelo a tratar. Por esta razón, los dos primeros apartados de esta sección se ocupan de problemas específicos a tener en cuenta para el desarrollo del trabajo. Primeramente, se definen las técnicas existentes en la detección de la TB (Técnicas para el diagnóstico), con el fin de poner de manifiesto la problemática que estas implican y la necesidad de utilizar imágenes radiológicas. En la segunda parte (Caracterización de la TB en imágenes de CT), se describen las diferentes lesiones y anormalidades que se pueden encontrar en pulmones infectados por TB y que son propias de esta patología. Conocer la caracterización de las lesiones es fundamental tanto para la etapa de segmentación, dado que esta fase se puede refinar incorporando el modelo de crecimiento e información sobre la ubicación de las lesiones, como para la clasificación de las mismas por razones evidentes.

Los últimos apartados tratan sobre las diferentes técnicas utilizadas en la segmentación de pulmón y las técnicas más utilizadas para problemas de clasificación en el entorno médico.

3.1 Técnicas para el diagnóstico de la tuberculosis

Existen diferentes técnicas, para diagnosticar la infección por TB. Aunque, como ya se mencionara en la motivación de este trabajo (sec.2), algunas únicamente muestran sensibilidad suficiente para determinar si el individuo está infectado pero no para evaluar la progresión de la enfermedad. Basándonos en este criterio podemos encontrar:

- Técnicas para la detección de la infección:
 - *Tuberculin skin test* (TST): Se trata del test más simple y usado. Consiste en una inyección subcutánea de un líquido denominado *tuberculina*. Si en el periodo de 48 a 72 horas aparece un bulto rojo, la probabilidad de estar infectado por TB es alta.

²La MRI obtiene su contraste de tejidos con alta densidad, sin embargo los pulmones están fundamentalmente repletos de aire. La imagen de PET no proporciona información espacial precisa.

- Análisis de sangre para TB: Conocido como *interferon-gamma release assays* (IFN- γ , IGRA). Miden la intensidad de la reacción del sistema inmune a una infección.

- Técnicas para la evaluación de la progresión:

- Test de esputo: El esputo es producido en los pulmones y vías aéreas. En este caso, la muestra de esputo es analizada en el laboratorio, donde se mide la cantidad de carga bacteriana de la muestra. Sin embargo, esta técnica presenta algunos inconvenientes, por ejemplo, los niños más pequeños no son capaces de producir esputos. Además, este tipo de análisis conllevan de 4 a 8 semanas, tiempo que suele resultar crucial en el desarrollo de la TB.

Por esta razón, los análisis basados en imagen cada vez reciben más atención.

- Análisis de imagen: La evaluación de imágenes moleculares (imágenes que permiten visualizar y caracterizar procesos biológicos [15]) presentan como grandes ventajas, que la técnica es no invasiva (no es necesario sacrificar a un sujeto para realizar la evaluación, como si sucede en un histología) y la posibilidad de identificar espacialmente las alteraciones que produce la infección, lo que permite realizar estudios longitudinales de gran precisión.

Existen múltiples técnicas de imagen molecular que permiten evaluar diferentes características de la TB (ver figura) [1, 9, 16]. El análisis específico y concienzudo de cada una de ellas va más allá del ámbito de este trabajo, por lo que, únicamente se realiza un estudio pormenorizado del tipo de imágenes empleadas en la realización del mismo, las imágenes de tomografía axial computarizada (HRCT). Dada la importancia de esta modalidad de imagen en el proyecto, la descripción referente a la evaluación de imágenes de HRCT para la cuantificación de TB se realiza de forma individual en la siguiente sección Caracterización de la tuberculosis mediante HRCT.

3.2 Caracterización de la tuberculosis en imágenes de tomografía computarizada

Las imágenes adquiridas con los equipos de tomografía axial computarizada (*Computer Tomography*, CT) [17], resultan especialmente útiles en el diagnóstico de la infección provocada por la TB dada su alta sensibilidad para detectar cambios en el parénquima pulmonar (fig.4). Por otro lado, se trata de imágenes donde los valores presentados por los vóxeles que identifican diferentes tejidos se encuentran acotados dentro de la denominada escala de *Hounsfield* [18] mediante un valor en unidades de *Hounsfield* o *Hounsfield Units*, (HU)

Además, es posible diferenciar diferentes clases de lesiones, lo que ayuda a los radiólogos a estudiar la evolución de la enfermedad e incluso distinguir entre diferentes estadios de la misma.



Figura 4: a) Corte coronal de una imagen de abdomen para un sujeto infectado por TB. b) Corte axial. c) Corte sagital

A continuación, se introducen brevemente las principales clases de lesiones detectadas actualmente por los radiólogos expertos [19]:

- Granuloma: Se trata de las lesiones principales producidas por la TB. Tienen forma esférica y un valor alto en la escala de *Hounsfield*. Debido a su parecido en esta escala con los vasos sanguíneos y el tejido correspondiente al mediastino, resulta sencillo clasificarlos de manera incorrecta (fig.5).
- Conglomeración: Se trata de una adhesión de granulomas, se suele producir cuando la enfermedad está avanzada y no tratada. Su estructura puede observarse en la figura 5

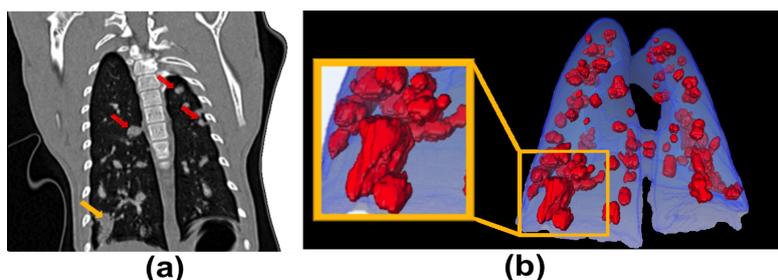


Figura 5: a) (a) Imagen de CT de un sujeto 13 semanas después de haber sido infectado. Granulomas (flechas rojas). Conglomeración (flecha amarilla) en la parte baja del lóbulo derecho. (b) Imagen 3D del pulmón infectado

- Árbol en brote o *tree in bud*: Aparece cuando la infección comienza a ocupar estructuras que normalmente se encuentran repletas de aire (alvéolos, vías aéreas, etc.). Por esta razón, se identifican mediante la observación de opacidades (zonas muy brillantes) en estas estructuras como se puede ver en la figura 6

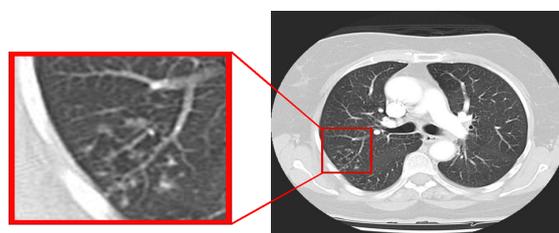


Figura 6: *Tree in bud*

- Infiltrado o consolidación: Opacificación de espacios aéreos dentro del parénquima pulmonar. La consolidación o el infiltrado pueden ser densos o irregulares y pueden tener bordes irregulares, mal definidos o nebulosos.

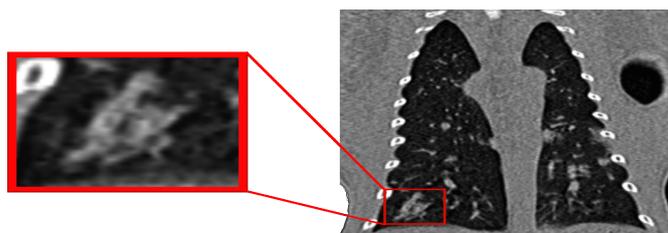


Figura 7: Consolidación

- Cavernización: Las cavernas son zonas repletas de aire en el interior de los granulomas. Se origina al inicio de la enfermedad en sujetos inmunodeprimidos, en los cuales, los mecanismos de defensa no son capaces de contener la infección. Primero, se constituye una lesión proliferativa (con un componente inflamatorio reactivo alrededor del foco infeccioso) que tiene tendencia a producir necrosis en la parte central. La necrosis recibe el nombre de caseosis por el aspecto blanquecino que recuerda al queso. Las cavernas se forman cuándo los focos de caseosis se vacían.

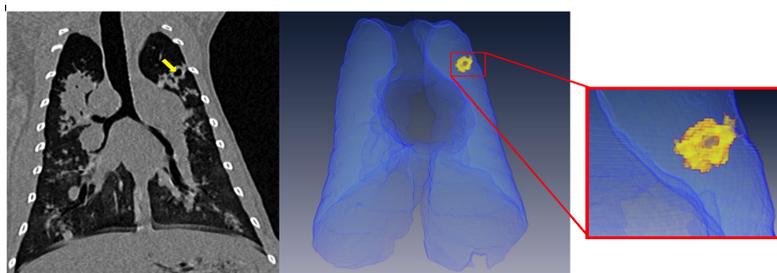


Figura 8: Caverna

- Vidrio deslustrado o *Ground Glass Opacity*(GGO): Se refiere a un área de mayor atenuación en el pulmón en la tomografía computarizada (TC) con marcas bronquiales y vasculares preservadas.

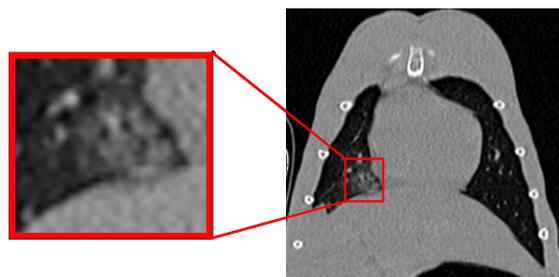


Figura 9: Ground glass opacity

3.3 Segmentación de pulmones en imágenes radiológicas

La segmentación de estructuras en imagen radiológica es en si mismo un campo muy amplio dentro de la disciplina del reconocimiento de patrones en imágenes. Existen numerosas aproximaciones [20, 21] que explotan y adaptan las metodologías aplicadas con éxito en problemas de visión artificial genéricos, como por ejemplo el uso de modelos deformables (contornos activos, captura de movimiento, etc.) [22] o modelos estadísticos [23]. Además, los diferentes algoritmos se ajustan para cada modalidad de imagen médica, lo que supone que aparecen familias completas de nuevos algoritmos, por ejemplo, para las imágenes adquiridas mediante resonancia magnética [24] o ultrasonido [25].

En particular, la segmentación de pulmones utilizando imagen de HRCT es un tema recurrente en la literatura actual dado el creciente interés en el estudio del cáncer de pulmón y las lesiones que en este se producen [26–30]. Sin embargo, la literatura respecto a la segmentación de pulmones infectados por TB en modelos animales es significativamente menor [31, 32]. La tendencia es utilizar los métodos empleados con éxito en modelos con cáncer y adaptarlos a los problemas en los que se trata con la segmentación de pulmones anómalos. De esta manera, en la literatura actual podemos distinguir las siguientes aproximaciones dedicadas a la segmentación de pulmón:

1. Métodos basados en umbralización: Se trata de los métodos más simples y tradicionalmente más utilizados [33, 34](Figura.10.A). Estos métodos extraen objetos (*blobs*) a partir de la distribución de los valores de grises en la imagen. Para ello, se realiza un proceso que intenta encontrar el valor umbral que mejor separa los objetos de interés (los pulmones, en este caso). Este tipo de procesado funciona bien con las imágenes de CT dado que, como se mencionó, estas presentan niveles de gris asociados a los diferentes tejidos buscados. Sin embargo, este tipo de técnicas son muy sensibles al ruido y los patrones anormales que presenta el tejido infectado, por lo que es necesario aplicar varias operaciones morfológicas para obtener una segmentación sub-óptima.
2. Métodos basados en el parecido de los vóxeles en una región (Figura.10.B): Estos métodos explotan el parecido de los vóxeles en ciertas regiones espaciales de la imagen [35, 36]. El procesado más común dentro de esta familia de métodos es el algoritmo de crecimiento de semilla, el cual, consiste en evaluar los vóxeles cercanos a un vóxel establecido anteriormente de forma determinista, la semilla, y fijar criterios para decidir si los vóxeles evaluados pertenecen a la misma región o tejido que la semilla. La precisión de esta clase de métodos depende en gran medida de: a) la identificación correcta de vóxeles significativos que puedan ser considerados semillas, b) la definición de la vecindad, es decir, que vóxeles se consideran suficientemente cercanos a esta para ser evaluados y c) la definición de los criterios de parecido en la vecindad. Por esta razón, la intervención de un usuario experto que controle todas las variables se hace indispensable, en muchas ocasiones, para lograr segmentaciones suficientemente aceptables que permitan extraer resultados cuantitativos.
3. Métodos basados en información anatómica conocida *a priori* (Figura. 10.C): Estos métodos utilizan información sobre la estructura a segmentar obtenida a partir de un modelo o un atlas que ha sido definido o construido por expertos [37]. Un atlas consiste en una plantilla para imágenes de CT correspondientes a pulmón en el caso que nos ocupa, que contiene etiquetas de las estructuras anatómicas de las regiones de interés. Esta plantilla se alinea o en el argot del procesado de imagen, se registra, con la imagen a segmentar optimizando un indicador de similitud entre ambas, por ejemplo, la información mutua entre las imágenes. Una vez que ambas imágenes están alineadas, se segmenta la región de interés. Estos métodos funcionan suficientemente bien para casos donde las lesiones no han modificado de forma abrupta la estructura esperada. Sin embargo, para el caso de la Tuberculosis, suelen resultar ineficaces dado que los pulmones de los sujetos dañados difieren de forma significativa de unos pulmones sanos. Por su parte, los métodos basados en el parecido a un modelo, optimizan el ajuste de la imagen que se quiere segmentar a un modelo preconcebido con el fin de evitar en mayor medida las variaciones en la forma del tejido.
4. Métodos basados en la localización de la anatomía adyacente (Figura.11.A): Esta clase de métodos utilizan información sobre las estructuras que se espera encontrar alrededor del órgano de interés. Por ejemplo, la caja torácica, el corazón o el diafragma [38], en el caso de los pulmones. Este tipo de procedimiento es muy útil cuando la zona de interés se encuentra tan dañada que es imposible reconocerla. Su principal problema radica en que requiere para funcionar de forma correcta que los órganos vecinos no se encuentren dañados o afectados por artefactos de adquisición de la imagen.
5. Métodos basados en técnicas de aprendizaje máquina o *Machine Learning* (Figura.11.B): Existen múltiples aproximaciones recientes dado el crecimiento que ha experimentado el campo del aprendizaje automático o aprendizaje máquina, ejemplarizado en la explosión de las técnicas de aprendizaje profundo (*Deep*

Learning) [39] en los últimos años [40, 41]. El proceso general suele consistir en la extracción de características de la imagen a partir de pequeñas regiones de la misma (*patches*) con el fin de clasificar todos los vóxeles dentro de cada región como un tejido específico. Estas características son de naturaleza muy variada y su efectividad cambia con el problema en concreto a resolver. La principal desventaja de estos métodos es la enorme cantidad de datos que deben ser clasificados previamente por un experto para que la función de clasificación (que se obtiene tras el proceso de entrenamiento) vierta resultados válidos y generales. Es por ello que las metodologías más recientes basadas en aprendizaje máquina intentan realizar la segmentación a partir de etiquetas no clasificadas [42, 43].

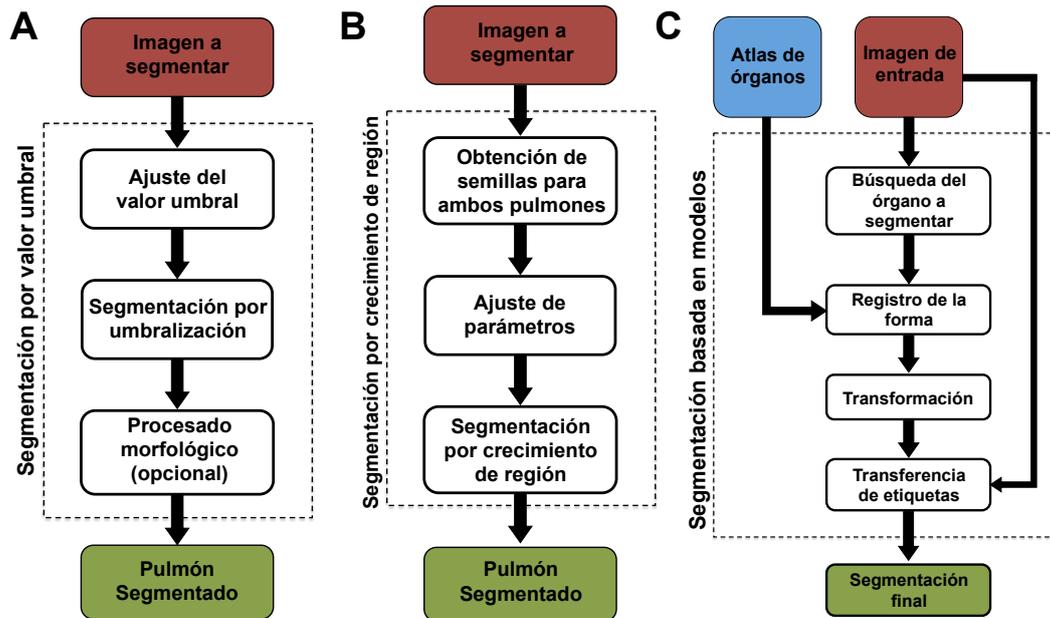


Figura 10: Flujo de trabajo de las diferentes aproximaciones para la segmentación de pulmón. A) Umbralización. B) Basada en regiones. C) Basada en información anatómica.

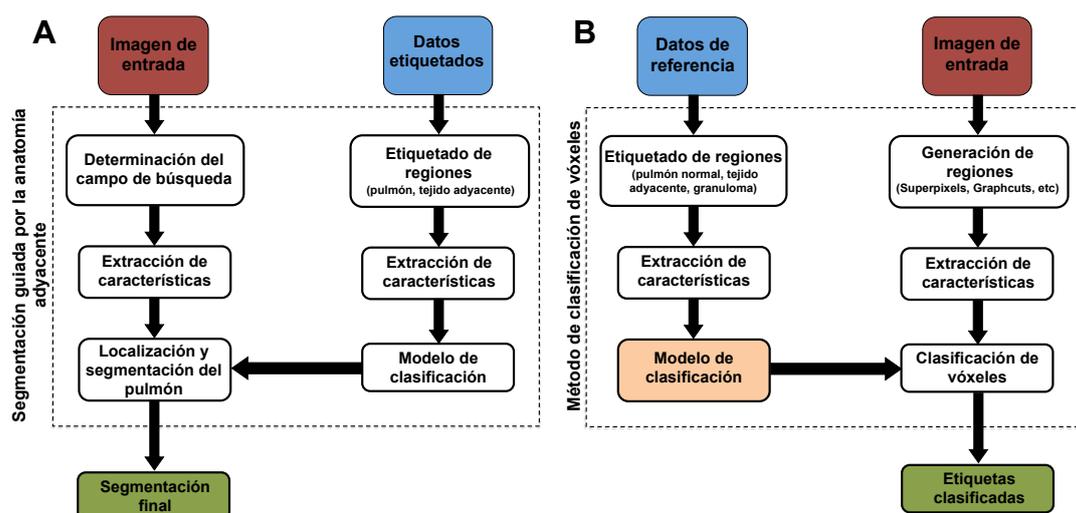


Figura 11: Flujo de trabajo de las diferentes aproximaciones para la segmentación de pulmón. A) Basados en el guiado por la localización de la anatomía adyacente. B) Basados en aprendizaje máquina, *Machine Learning* o clasificación por vóxel.

3.4 Clasificación de regiones anómalas en la imagen radiológica

Existe un considerable número de técnicas usadas en la clasificación de regiones irregulares en el parénquima pulmonar a partir de imágenes de CT, fundamentalmente en el entorno de la detección de nódulos pulmonares propios del cáncer de pulmón. Actualmente, estas técnicas ocupan desde las recurrentes y populares redes neuronales artificiales con multitud de capas ocultas [44, 45], hasta técnicas fundamentadas en el modelo anatómico [46–48], basándose el éxito de estas en la detección de la forma esférica característica de los nódulos. Sin embargo, este patrón no es tan reconocible en las lesiones de TB, ni tan siquiera en lesiones como los granulomas que únicamente mantienen la forma esférica durante fases muy tempranas de la enfermedad.

Sin embargo, aparecen lesiones con mayor similitud en el proceso de destrucción de parénquima que se produce con el enfisema o en las enfermedades pulmonares intersticiales (*Interstitial Lung Disease, ILD*) [49, 50]. En estos casos, la literatura, no demasiado numerosa en comparación con la referida al cáncer, muestra sus mejores resultados cuando se utilizan análisis de textura en imágenes, tanto para enfisema [51, 52], como para ILD [53–57].

Estos resultados, motivan el uso de características de textura, las cuales se pueden definir de múltiples maneras [53, 56, 58]. En la aproximación que se realiza en este trabajo, se emplearán las características de co-ocurrencia de nivel de gris, las cuales se introducen en la próxima sección.

Dado que la dependencia entre las múltiples características de textura es conocida [59, 60], así como los problemas que puede presentar utilizar clasificadores con características irrelevantes o redundantes, también se repasan en esta sección los métodos fundamentales para la selección de características (sec.5.2.2).

Finalmente, se introducen las metodologías que se utilizan con mayor frecuencia para tratar los problemas de aprendizaje con datos muy desbalanceados. Dicho fenómeno es muy común

en los problemas biomédicos y no es una excepción para la clasificación en TB.

3.4.1 Clasificación basada en la textura de la imagen

Dentro de la medicina, como en la inmensa mayoría de campos en nuestros días, existe un interés creciente en el análisis de la inmensa cantidad de datos que la tecnología actual es capaz de producir, espoleado por el avance exponencial de las técnicas de aprendizaje estadístico. En el campo de la investigación de la imagen radiológica, este interés se ha materializado en una disciplina propia, generalmente denominada por su termino en inglés, *Radiomics* [61, 62].

Entre las múltiples técnicas de análisis de datos de imagen médica, existe especial interés en la caracterización (automática o no) de regiones de la imagen en función de la cuantificación de la textura que está presente.

Aunque existen diferentes definiciones para la textura en una región de un imagen, de forma simple podemos entender esta como, la suavidad o la rugosidad y tosquedad que esta presenta. Esta definición resulta bastante abierta, lo que supone que dependiendo de la aplicación se definan diversos métodos e indicadores, que intentan cuantificar algún tipo de relación espacial entre los píxeles (o vóxeles) de una imagen, con el fin de representar numéricamente, lo que un humano simboliza cualitativamente.

La aproximación más tradicional para intentar cuantificar la textura de una región es utilizar los estadísticos clásicos (media, desviación típica, etc.) sobre el histograma de la imagen. Estos indicadores son sencillos de obtener y proporcionan información global sobre la región que resulta de gran utilidad, sin embargo, se encuentran fuertemente limitados dado que no son capaces de sintetizar la información que ofrece la distribución espacial de los vóxeles en una región.

En los últimos años, han aparecido diversos métodos que permiten cuantificar algunas de las relaciones espaciales entre vóxeles que caracterizan la textura visualizada. Estas métricas son generalmente denominadas *características de textura locales* o *características de textura de segundo orden*. Entre los diferentes descriptores son especialmente reconocidos, debido a su exitosa aplicación en el campo de la imagen radiológica [63, 64], aquellos derivados de la denominada Matriz de Co-ocurrencia de nivel de Gris (*Gray-Level Co-occurrence Matrix*, GLCM). Por esta razón, en este estudio, se utilizan las características globales en conjunción con las características derivadas a partir de la GLCM para caracterizar las regiones a clasificar dentro del parénquima pulmonar.

3.4.2 Aprendizaje utilizando datos no balanceados

En la enorme mayoría de aplicaciones que enlazan las técnicas de aprendizaje estadístico con datos médicos (radiológicos o de otro tipo) se presenta la siguiente situación: El porcentaje de instancias del problema que caracterizan alguna patología o trastorno suele ser significativamente menor que aquellas que caracterizan un estado basal. Este problema es frecuentemente referido por su termino en inglés *class imbalance* [65] y surge en gran variedad de disciplinas. Sin embargo, resulta muy común en medicina donde los desequilibrios entre clases son del orden de 100:1, 1000:1 e incluso 10000:1 [66, 67].

El caso de clasificación abordado en este trabajo, es un ejemplo claro de datos no balanceados, donde como se presenta en la sección de 4, algunas clases cuentan con número de instancias relativamente alto en relación al resto. Por ello, la aplicación de técnicas de

balanceo debe de ser un motivo de estudio.

Los datos no balanceados perjudican notablemente el proceso de aprendizaje dado que la mayoría de algoritmos de aprendizaje máquina están diseñados para trabajar con datos que presentan un número de instancias similares entre las clases [68]. Por esta razón, han surgido nuevas técnicas capaces de hacer frente a conjuntos de datos con clases no pareadas. A continuación, se introducen las tres variantes principales: a) **técnicas de submuestreo**, consistente en eliminar instancias de las clases mayoritarias; b) **técnicas de sobremuestreo**, que genera nuevas instancias de las clases menos representadas y c) **técnicas combinadas**, en las cuales se combinan las anteriores.

a) Técnicas de submuestreo:

- 1 Submuestreo aleatorio (*Random Under-Sampling*, RUS): Se eliminan aleatoriamente un conjunto de muestras de las clases con más muestras hasta igualar el número de muestras en las clases minoritarias o una parte proporcional de estas.
- 2 Enlace Tomek (*Tomek Link*, TL): De manera general consisten en eliminar aquellas muestras de las clases mayoritarias que se encuentran más próximas a muestras de otras clases, la relación entre una muestra de la clase mayoritaria con su variable más cercana de otra clase se denomina *Tomek Link* [69].
- 3 Submuestreo mediante agrupamiento (*Clustering Under-Sampling*, CUS): Emplea el algoritmo de *k-means* [70] para agrupar las instancias de las clases mayoritarias en k grupos, donde k es el número de muestras de la clase que cuenta con menos instancias. Finalmente, se seleccionan únicamente los k centroides obtenidos.
- 4 Regla de limpieza del vecindario (*Neighborhood Cleaning Rule*, NCR) [71]: En este caso, cada muestra del conjunto de datos se examina en función de sus k vecinos más cercanos. Si la muestra pertenece a la clase con más instancias pero la mayoría de sus vecinos no, la muestra se elimina. Si por el contrario, la muestra no pertenece a la clase mayoritaria pero más de la mitad sus vecinos sí, son estos últimos los eliminados.

b) Técnicas de sobremuestro:

- 1 Sobremuestreo aleatorio: Se replican aleatoriamente muestras de las clases con menor número de instancias.
- 2 Técnica de sobremuestreo minoritario sintético (*Synthetic Minority Oversampling Technique*, SMOTE) [72]: Consiste en generar muestras sistemáticamente, situando estas entre una muestra de la clase minoritaria y sus k vecinos más cercanos.

c) Técnicas combinadas

- 1 SMOTE + TL: En este caso se combinan ambas técnicas, de manera que tras aplicar SMOTE se realiza un proceso de submuestreo para eliminar datos espurios que se hayan podido generar, de forma que se prevenga el sobre-entrenamiento [65].
- 2 SMOTE + NCR: Se procede de manera similar al caso anterior pero en este caso aplicando NCR.

3.4.3 Selección de características

De forma simplificada, se puede definir una característica como una variable medible de una entidad durante un proceso que es objeto de estudio.

Utilizando varias características que definen un problema concreto, los algoritmos de aprendizaje son capaces de establecer una clasificación.

Como se comentó con anterioridad, las posibilidades que ofrece actualmente la técnica ha supuesto un crecimiento masivo del número de características que se pueden recoger de un proceso. El hecho es que en unos pocos años, los algoritmos para el reconocimiento de patrones han pasado de trabajar con decenas de características a ordenes de magnitud mucho mayores. Este hecho, que en un principio puede parecer una ventaja dado que se cuenta con más información para resolver el problema, puede significar una bajada considerable en el rendimiento de los clasificadores. Debido sobre todo, a los problemas para trabajar con conjuntos de datos de alta dimensionalidad [73] y al desconocimiento sobre la información real aportada por las características.

Para contrarrestar estas limitaciones, se han desarrollado diferentes técnicas para reducir el número de variables de un problema en función de su relevancia y su redundancia.

El proceso de selección ayuda a entender mejor el problema, lo que resulta transcendental en el contexto de la investigación biomédica donde se ubica este trabajo, además permite reducir los tiempos de computación y mejora el rendimiento dado por los clasificadores.

Aunque la literatura al respecto de la selección de variables es amplísima [74–76], las técnicas se pueden resumir principalmente en dos tipos:

- a) Métodos de filtrado: Este tipo de métodos realizan una clasificación jerarquizada, *ranking*, de las variables utilizando un criterio para ordenarlas. A cada variable se le asigna un valor, referido también como figura de mérito para a continuación, elegir aquellas por encima de un valor umbral predeterminado.

Las figuras de mérito se basan en una medida que establezca cuanta información útil contiene una variable en relación a las diferentes clases del problema, es decir, cómo de relevante es.

El elemento más importante de este tipo de métodos es, por tanto, definir la manera más adecuada de medir la relevancia. Lo que ha supuesto la aparición de numerosas técnicas [77], más o menos exitosas en función de la aplicación, que tienen como objetivo medir como de grande es la dependencia entre una variable y las clases del problema con independencia del resto de variables. Son las técnicas basadas en información mutua (*Mutual Information, MI*), las que suelen ofrecer mejores resultados [78]. En particular han demostrado una alta capacidad de diferenciación en problemas de visión artificial en los que se emplean texturas [79].

Este tipo de técnicas resultan computacionalmente eficientes al no incluir procesos de aprendizaje. Por contrapartida, presentan algunos inconvenientes: Las características que resultan irrelevantes cuando se analizan de forma individual son descartadas, sin tener en cuenta cómo de relevantes resultan al ser combinadas con otras variables.

- b) Métodos empaquetadores o *Wrappers*: Esta clase de técnicas utiliza los clasificadores como una caja negra que facilitan una predicción utilizada para medir la bondad de un conjunto de características. Dado que la evaluación de todos los subconjuntos de características entre las N posibles requeriría de 2^N procesos de aprendizaje y la evaluación de un clasificador, se utilizan métodos heurísticos que permitan acelerar el proceso de búsqueda [75, 78].

De manera general, se puede hablar de métodos basados en *wrapper* de selección

secuencial de características [80] y de búsqueda heurística. Los primeros comienzan con un subconjunto vacío al que añaden de forma secuencial nuevas características hasta alcanzar un umbral. Por otro lado, los algoritmos de búsqueda heurística evalúan diferentes subconjuntos configurados de forma aleatoria o mediante algún algoritmo de optimización (por ejemplo, algoritmos genéticos, descenso de gradiente) [81].

4 Materiales

Para evaluar las segmentaciones de pulmón que se obtienen utilizando la propuesta descrita en la metodología (sec.5.1) e implementada como se apunta en la sección dedicada específicamente al software (sec.6), se utilizarán imágenes de CT correspondientes a modelos animales de infección por *Mtb*. Dichas imágenes fueron cedidas por el *Laboratorio de Imagen Médica (LIM)* del hospital *Gregorio Marañón*, el departamento de imagen médica (BiiG) de la universidad *Carlos III de Madrid* y la compañía farmacéutica *GlaxoSmithKline (GSK)*.

Por motivos de confidencialidad expresados por la compañía farmacéutica, no se recogen datos específicos de los modelos animales empleados, pero si los detalles de las imágenes adquiridas por los equipos de HRCT.

Concretamente, para la validación de la segmentación se emplean imágenes de dos modelos animales. Un primer modelo correspondiente a mamíferos de tamaño medio, al que en esta memoria se hace referencia como modelo animal A, y un segundo modelo, para mamíferos de pequeño tamaño, referido como modelo animal B. Las características de las imágenes adquiridas para ambos modelos son las siguientes:

1. **Modelo animal A** de infección por *Mtb*: Nueve imágenes volumétricas de HRCT ($0.26mm \times 0.26mm \times 0.63mm$), correspondientes a tres sujetos, que en este trabajo se nombran como subj-1, subj-2 y subj-3, en el momento de ser infectados (semana 0) y 16 y 28 semanas tras el insulto.
2. **Modelo animal B** de infección por *Mtb*: Cuatro imágenes de CT ($0.087mm \times 0.087mm \times 0.087mm$), correspondientes a dos sujetos, en el momento de inocular la infección por *Mycobacterium tuberculosis* (semana 0) y 4 semanas tras el insulto.

Además, se cuenta con la segmentación manual de los pulmones de los sujetos del modelo A. Esta segmentación no fue realizada por radiólogos expertos, pero si por expertos en procesado de imagen médica. Por lo cual, aunque se pueden esperar incorrecciones, se cuenta con una referencia que permite establecer una evaluación cuantitativa (ver sección 7) de la calidad de la segmentación automática.

De estos sujetos segmentados también se cuenta con las etiquetas que se utilizan en la segunda parte del estudio correspondiente a la clasificación de lesiones de TB. Para esta tarea se tiene el siguiente número de lesiones etiquetadas:

Lesión	Num. Etiquetas
Granuloma	2140
Conglomeración	350
<i>Tree in Bud</i>	82
Consolidación	80
Ground Glass Opac.	53

Tabla 1: Lesiones etiquetadas

5 Métodos

Por motivos de claridad, esta sección se ha dividido en dos grandes bloques correspondientes el primero, a los algoritmos propuestos para la segmentación de pulmones infectados por TB (sec.5.1) y un segundo gran bloque, dedicado al proceso de clasificación de las lesiones (sec.5.2).

5.1 Métodos: Segmentación del pulmón

Existen múltiples ejemplos [20, 21, 82], que muestran la necesidad de aislar la región anatómica que es objeto de estudio de forma automática en una imagen radiológica [31, 32, 82], del resto de estructuras que pueden aparecer, con el fin de proporcionar análisis posteriores de mayor calidad.

En el caso de este trabajo, la extracción automática correcta del parénquima pulmonar es primordial para los procesos de análisis posteriores, como puede ser la clasificación de lesiones del tejido pulmonar causadas por la TB.

La segmentación de pulmones infectados es un proceso complejo. Los pulmones se encuentran en continuo movimiento durante la fase de adquisición de la imagen de CT debido a la respiración, formando zonas difusas en las que resulta complicado diferenciar entre tejido extra-pulmonar y el parénquima. Además, la TB se manifiesta de manera muy distinta en los pulmones durante las diferentes fases de la enfermedad como se comentó en la sección 3.2, esto supone la inclusión de una fuente más de variación que facilita el cambio morfológico de la estructura y la composición del tejido a segmentar.

Por estas razones, este trabajo tiene como uno de sus objetivos principales, la implementación de una herramienta versátil, para la segmentación automática de pulmones infectados con TB. Dicha herramienta es completamente automática y capaz de trabajar con imágenes procedentes de diferentes modelos animales, simplemente cambiando parámetros básicos de configuración que dan lugar a flujos de trabajo específicos para cada modelo.

A continuación, se describen las diferentes etapas específicas de procesado de imagen, diseñadas para resolver el problema de la segmentación por partes independientes, de manera que cada una de estas puede ser utilizada, eliminada o mejorada con facilidad sin tener efectos en el resto del proceso.

El esquema completo de procesado se muestra en la Figura 12.

5.1.1 Preprocesado

Existen diferentes fuentes de ruido derivadas del proceso de adquisición de imágenes de CT que fundamentalmente se manifiestan en forma de artefactos de movimiento [17, 83], y ruido impulsivo o *salt and pepper noise*. Los primeros producen regiones difusas en la imagen y existen múltiples algoritmos que intentan paliar los efectos, como por ejemplo, los algoritmos de *gating* respiratorio retrospectivo [84]. En cuanto a los segundos, son bien conocidas las bondades del filtro de mediana y sus variaciones [85]. Dichos filtros ayudan a mejorar la imagen, lo cual tiene un efecto directo sobre la forma presentada por el histograma, que como se verá en la siguiente sección, influye sobre los resultados logrados por la aplicación. En concreto, para esta aplicación se realiza un filtrado anisotrópico de la imagen, aunque como se expone en la sección dedicada al software implementado, 6Desarrollo de software, se posible sustituirlo por cualquiera de los filtros genéricos de reducción del ruido (media, mediana, gaussiano, etc).

Es bien sabido que el filtrado anisotrópico contrarresta el efecto de bordes de sierra o *aliasing* [86, 87]. Aplicarlo en las imágenes radiológicas se traduce en una mejor definición de las superficies que delimitan las estructuras. Esto supone, una mejora en la calidad de la textura de la imagen que resulta adecuada para la cuantificación basada en estas características, la mejora en la definición de las zonas difusas aparecidas debidas al artefacto de movimiento que introduce la respiración y la reducción del ruido impulsivo.

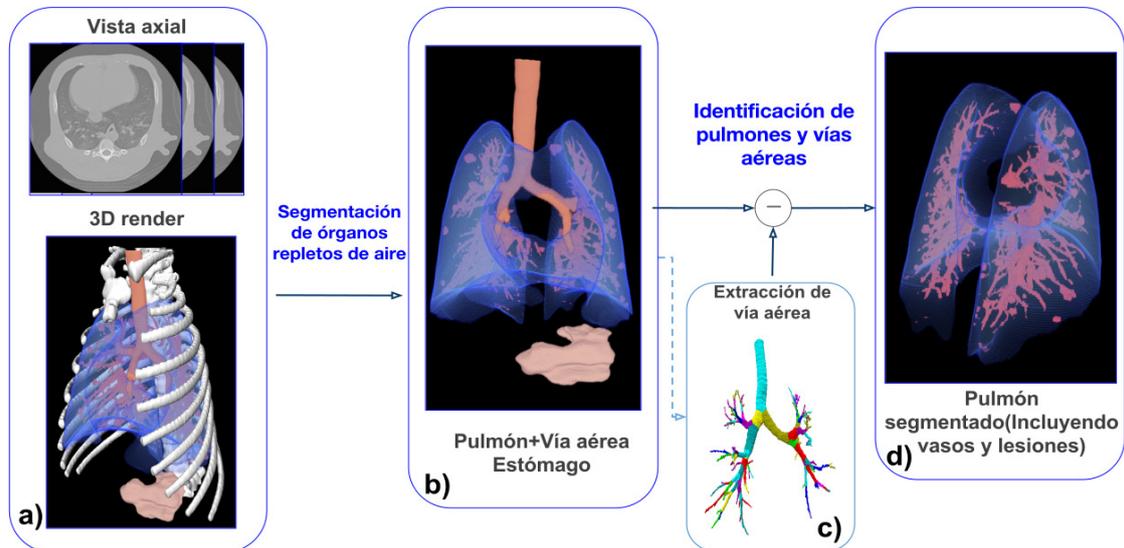


Figura 12: Esquema general del flujo de trabajo implementado para a segmentación de los pulmones: **a)** Imagen inicial adquirida con un equipo de alta resolución para tomografía axial computarizada (*High Resolution Computed Tomography, HRCT*) junto con una representación 3D de la región de interés (*Region of Interest, ROI*); **b)** Estructuras segmentadas tras aplicar alguno de los métodos de umbralización 5.1.3 y antes de utilizar las metodologías para seleccionar únicamente el parénquima pulmonar 5.1.4; **c)** Vía aérea extraída tras utilizar el algoritmo de *Fast Marching Level Set* 5.1.5; **d)** Pulmón segmentado tras sustraer la traquea al tejido previamente identificado como parénquima pulmonar y tras realizar las operaciones correspondientes de postprocesado 5.1.6 para incluir las lesiones generadas por la infección por *Mycobacterium Tuberculosis*.

5.1.2 Segmentación inicial del conjunto de pulmones y vías aéreas

5.1.3 Umbralización

Este proceso aprovecha la información conocida *a priori* sobre la región a segmentar, los pulmones, y su representación en imágenes de CT.

De manera muy simplificada podemos decir que, el contraste en las imágenes de CT se obtiene midiendo la pérdida de energía que se produce cuando una fuente de rayos-X controlada *ad-hoc* atraviesa un tejido. Esta pérdida es mayor cuanto más denso es el tejido que se atraviesa. De forma que cuando los rayos-X proporcionados por la fuente traspasan el parénquima pulmonar, o las vías aéreas o el estómago, en general estructuras fundamentalmente repletas de aire, la energía perdida por los rayos-X es mucho menor que cuando atraviesa el hígado, el corazón o un hueso. Esto se traduce en la aparición de estructuras muy oscuras en la imagen en el primer caso y en estructuras más brillantes en el

segundo caso ³. Esta propiedad permite que las imágenes de CT puedan calibrarse dentro de una escala común de unidades, la escala Hounsfield [17, 88], que establece que el aire debe estar por debajo de -1000 unidades Hounsfield o *Hounsfield Units* [HU], el pulmón entre -700 y -600 o el hueso por encima de 700. Aunque evidentemente, estas son unidades de referencia para tejidos sanos en humanos adquiridas en inspiración, ayudan a tener cierta información sobre lo que se espera observar en una imagen correspondiente a un tórax realizada con un equipo de CT, aún cuando estas adquisiciones provengan de otro tipo de mamíferos. En otras palabras, se espera encontrar un histograma bimodal de la Figura 13.

El conocimiento de esta información *a priori*, se puede explotar para comenzar el proceso de aislamiento de estructuras dentro de una imagen volumétrica de CT, aplicando una de las operaciones clásicas de visión artificial a bajo nivel, la umbralización [89].

La idea es separar automáticamente dos grandes bloques de objetos en las imágenes. Por un lado, todos aquellos más oscuros, correspondientes a las estructuras que contiene aire, como los pulmones el estomago o el propio aire. Por otro lado, todas aquella estructuras que se pueden identificar como tejido blando, tejido duro (hueso) e incluso piezas que forman el equipo y que son inevitables⁴. Concretamente, para la herramienta implementada se han tenido en cuenta las siguientes técnicas:

- Filtro de Hu: Este filtro toma su nombre a partir del artículo publicado por *Hu et al.* [33]. Se trata de un filtro sobradamente conocido en el entorno de la imagen médica, siendo ampliamente utilizado en la segmentación de pulmones sanos, aunque no es más que una pequeña adaptación de un método de filtrado básico.

La implementación del mismo supone hallar un umbral, T , para el histograma de forma iterativa. Dicho umbral debe separar las regiones del histograma en dos distribuciones asociadas a los valores de intensidad ligados a las estructuras más aireadas y a las regiones más densas de forma como sigue:

$$T_{i+1} = \frac{\mu_a + \mu_{na}}{2}, \quad (1)$$

donde μ_a y μ_{na} representan la intensidad media de los vóxeles por debajo y encima del umbral T respectivamente, en la iteración i . El algoritmo se detiene cuando se llega a una situación de convergencia, $T_{i+1} = T_i$ ⁵. Es decir, no hay vóxeles cambiantes entre iteraciones consecutivas.

- Método de Otsu [90]: Este método realiza automáticamente la umbralización de una imagen basada en el agrupamiento (clustering) o, la reducción de una imagen en niveles de gris a una imagen binaria. El algoritmo supone que la imagen tiene dos clases de píxeles que siguen un histograma bimodal (píxeles de primer plano y píxeles de fondo). El umbral óptimo para separar las dos clases se calcula de modo que su dispersión combinada (variación intra-clase) sea mínima o equivalentemente (debido a que la suma de las distancias cuadradas pares es constante), de modo que su varianza inter-clase sea máxima. Como consecuencia, el método de Otsu es aproximadamente un análogo discreto unidimensional al análisis discriminante de Fisher.

³El lector debe tener en cuenta que la intensidad de la señal proporcionada por los rayos-X que atraviesan el tejido relleno de aire es mayor que la señal cuando se atraviesa tejido, sin embargo, se realiza una inversión de la magnitud del contraste para obtener las imágenes clásicas de CT o radiografía.

⁴Los sujetos dentro de un equipo HRCT se encuentran tendidos en una camilla que aparece en cualquier imagen de tomografía axial

⁵Generalmente se rebaja esta restricción por cuestiones de estabilidad numérica, de manera que suele ser suficiente cuando la diferencia entre umbrales en dos iteraciones sucesivas es menor a un número pequeño, por ejemplo, $\epsilon = 0.01$

La búsqueda dispersión conjunta (o variación intra-clase) habitualmente usada en el campo de la estadística se define como:

$$\sigma_w^2(T) = q_1(T)\sigma_1^2(T) + q_2(T)\sigma_2^2(T), \quad (2)$$

donde q_1 y q_2 son las probabilidades asociadas a cada una de las distribuciones dado el umbral T :

$$q_1(T) = \sum_{i=1}^T P(i) \quad q_2(T) = \sum_{i=T+1}^N P(i), \quad (3)$$

siendo N el número total de vóxeles en la imagen y donde las medias y las varianzas asociadas a cada clase vienen dadas por su definición habitual, esto es:

$$\mu_1 = \frac{\sum_{i=1}^T iP(i)}{q_1(T)} \quad \mu_2 = \frac{\sum_{i=T+1}^N iP(i)}{q_2(T)} \quad (4)$$

$$\sigma_1^2(T) = \sum_{i=1}^T [i - \mu_1(T)]^2 \frac{P(i)}{q_1(T)}; \quad \sigma_2^2(T) = \sum_{i=T+1}^N [i - \mu_2(T)]^2 \frac{P(i)}{q_2(T)} \quad (5)$$

Por lo tanto, el método de *Otsu* consiste en un problema de minimización de (2), en función del parámetro T .

Esta minimización se puede realizar mediante búsqueda exhaustiva. Sin embargo, esta operación puede resultar demasiado costosa computacionalmente cuando se utilizan resoluciones de imagen altas y/o imágenes de tamaño considerable⁶. Por lo que es común replantear el problema teniendo en consideración que la varianza de la imagen, σ^2 , tiene un valor determinado que se suele expresar como la suma de la varianza *within-class* definida anteriormente y la varianza entre las clases o *between-class*, σ_b^2 , esto es:

$$\sigma^2 = \sigma_w^2 + \underbrace{q_1(T)[1 - q_1(T)][\mu_1(T) - \mu_2(T)]^2}_{\sigma_b^2} \quad (6)$$

Por lo tanto la minimización de 2 propuesta, como se indicó anteriormente, es equivalente a la maximización de esta varianza, lo que computacionalmente implica un ahorro de recursos.

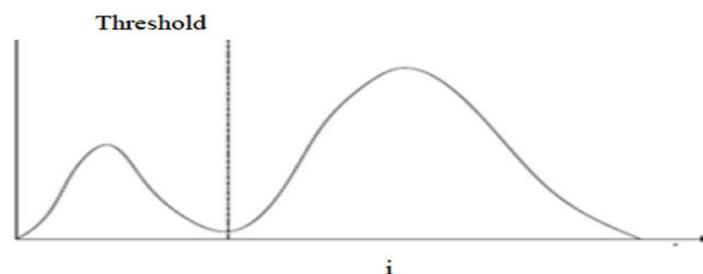


Figura 13: Ejemplo de distribución bimodal formada por dos distribuciones quasi-gaussianas separadas por un umbral óptimo en términos de minimización de la varianza asociada a cada distribución.

⁶Cada una de las imágenes de HRCT utilizadas en este trabajo contiene alrededor de 0.5 GB de información representados por vóxeles que utilizan 16 bits para codificar su valor

5.1.4 Análisis de conectividad y topológico

Dado que el proceso anterior proporciona una segmentación de todas aquellas estructuras que contienen mayoritariamente aire (ver figura 12.b) es necesario desarrollar e implementar métodos capaces de discernir entre todos los órganos resultantes del proceso de umbralización, con el fin de seleccionar la región de interés (*Region Of Interest, ROI*), los pulmones.

El resultado de aplicar el proceso de umbralización a una imagen de CT de tórax, es una imagen binaria que contiene los pulmones unidos a las vías aéreas, el estomago, el diafragma y por supuesto, el aire entre el sujeto y los límites del campo de visión (*Field of View, FOV*) proporcionado por las características del equipo de adquisición. Para diferenciar entre las estructuras, se realiza un análisis de conectividad entre los vóxeles segmentados para obtener máscaras asociadas a cada una de ellas, seguido por un estudio de la topología de cada objeto segmentado.

Así, para evaluar la conectividad entre todos los vóxeles segmentados en el proceso de umbralización se utiliza el método de división y unión de regiones desarrollado por *Lehmann* [91], que proporciona una implementación computacionalmente más eficiente que las utilizadas clásicamente.

Tras obtener una serie de estructuras con la división de regiones, se analiza el tamaño y la posición de éstas. Utilizando el tamaño, se descartan regiones muy pequeñas procedentes de pequeños cavidades de aire, como por ejemplo, las formadas en el tracto intestinal o regiones muy grandes, como el aire en el fondo de la imagen. La posición de la estructura segmentada viene dada por el centroide de la misma, de manera que se pueda obtener la distancia entre cada centroide y el centro de la imagen. Teniendo en cuenta que se espera que el conjunto formado por pulmones y vías aéreas⁷ se encuentre más centrado que cualquier otro objeto segmentado, se establece el siguiente criterio para localizarlo:

$$\arg \min_i d(c, c_i) \quad i = 1, 2, \dots, K, \quad (7)$$

donde i representa el i -ésimo objeto segmentado, K el número de objetos totales, c y c_i el centro y el centroide de la imagen y el objeto i , respectivamente y d la distancia euclidiana entre estos.

5.1.5 Extracción de las vías aérea

La distribución natural de las vías aéreas intrincadas en el parénquima pulmonar, hace imposible la extracción limpia de este órgano empleando métodos tradicionales de segmentación.

Con el objetivo de obtener una segmentación precisa del pulmón, evitando que otros órganos sean incluidos, aparece la necesidad de implementar un algoritmo dedicado de forma exclusiva a la extracción de las vías aéreas [30], capaz de segmentar la mayor parte posible del tracto respiratorio enmarañado en el parénquima pulmonar.

Existen en la literatura diversas aproximaciones dedicadas a la segmentación de las vías aéreas en humanos. Dependiendo del objetivo concreto de la aplicación, se pueden encontrar modelos basados en la morfología [92], en la orientación de las vías aéreas [93] o en

⁷Apreciar que mediante estas técnicas de filtrado básicas es imposible separar pulmones y vías aéreas.

crecimientos de semilla guiados por patrones anatómicos [94], entre otros. Las principales aproximaciones al respecto, se pueden encontrar en el concienzudo trabajo de *Lo et al.* [30].

Para el trabajo presentado en este documento, se ha optado por adaptar las propuesta utilizadas por *Schlathoelter et al.* [95] y *Bulöw et al.* [96] dado que suelen asegurar una tasa de éxito alta en problemas de extracción de objetos cilíndricos distribuidos en forma de árbol. Estas metodologías, utilizan contornos activos [89] para la segmentación de estructuras tubulares. Concretamente, el contorno evoluciona simulando un frente de ondas esférico que avanza desde un punto inicial localizado en la tráquea y recorre las vías aéreas sin invadir los pulmones. Se trata de un proceso extenso en el que intervienen distintos parámetros, que puede descomponerse en las siguientes partes (ver figura 14):

- Detección de la tráquea e inicialización de las vías (Fig.14.A): El frente de ondas se inicializa a partir de una semilla hallada al comienzo de la traquea en la imagen binaria obtenida tras la umbralización. El método encargado de detectar la tráquea utiliza el primer corte axial de la imagen volumétrica de CT (ver imagen de corte axial 4). Sobre dicho corte, se aíslan las manchas ('blobs') correspondientes a las estructuras que más aire presentan en su interior. Se calcula el centro de masas de las mismas y finalmente, se selecciona como semilla para el frente de ondas el centroide de la estructura situada más próxima del centro de la imagen. A partir de esta semilla se configura un frente de ondas esférico tomando los vóxeles vecinos a la semilla hallada.
- Propagación (Fig.14.B): El frente de ondas avanza por la traquea añadiendo vóxeles que conforman nuevos segmentos⁸ (en gris en la figura 14) de la vía aérea. En la implementación proporcionada junto a este trabajo (sec.6), los vóxeles se van añadiendo guiados por el algoritmo basado en contornos activos [97], *3D Fast Marching Level Set*⁹ descrito por *Deschamps et Cohen* [98]. El éxito de este tipo de algoritmos depende en gran medida de la definición de las restricciones de energía que debe cumplir el contorno y la zona segmentada de interés. Así como de la restricción en la *velocidad de propagación* que se define con el termino paso temporal o *time step*. En el caso del algoritmo aditivo implementado, se añaden nuevos vóxeles si cumplen con dos criterios que permiten mantener la energía baja o en otras palabras, una velocidad de propagación controlada. Dichos criterios están relacionados entre sí y son los siguientes:
 1. La intensidad del vóxel evaluado debe ser similar a los vóxeles del segmento de vía aérea (es decir, su vecindario delimitado por el frente de ondas esférico) en el que se ubicaría.

Se considera que este es el caso, cuando su intensidad no supera un umbral variable que se define como:

$$T_i = \mu_s + \alpha \cdot \max(\sigma_{s-1}, \sigma_{s-2}), \quad (8)$$

donde μ_s es la intensidad media de los vóxeles en el segmento donde se estudia incluir el nuevo vóxel, σ_{s-1} y σ_{s-2} corresponden con las desviaciones estándar de los segmentos delimitados en una y dos iteraciones anteriores mientras que α ¹⁰ es un parámetro a definir en función de la aplicación.

⁸Por segmento se entiende cada una de las partes que forman las distintas estructuras tubulares que a su vez configuran las vías aéreas

⁹Deschamps y Cohen adaptan los trabajos "clásicos" sobre segmentación utilizando contornos activos para el entorno de la imagen médica

¹⁰el valor de α se define para cada experimento en los Resultados

2. El gradiente de intensidad, entre el vóxel evaluado y sus vecinos, debe ser bajo. Se utiliza el operador tridimensional de *Sobel* [89,99], de manera que el vóxel se rechaza si el valor de este es mayor que un umbral T_s .

- Detección de bifurcaciones (Fig.14.C): Aparecen bifurcaciones en la vía aérea cuando el frente de ondas se divide en dos. Para evaluar cuando sucede esto, se utiliza el siguiente criterio:

$$\begin{cases} r_s \leq \beta \cdot r_e, & \text{No hay bifurcación} \\ r_s > \beta \cdot r_e, & \text{Hay Bifurcación,} \end{cases} \quad (9)$$

siendo r_s , el radio del actual segmento delimitado por el frente de ondas, r_e el radio esperado y β un factor para controlar la variación del radio esperado permitida.

En la implementación, cuando se detecta una bifurcación, se establece el segmento que está ocupado por el frente de onda como segmentado. Los nuevos segmentos encontrados se añaden a una cola que se encarga de repetir el proceso hasta que no se pueden encontrar nuevos segmentos sin introducir fugas en la segmentación.

- Detección de fugas (Fig.14.D): Debido a varios factores, como ruido en la imagen, artefactos de movimiento o efectos de volumen parcial [100] entre otros; el frente de ondas pueden alcanzar los pulmones provocando las denominadas fugas [101]. Para evitar esto, se restringe el número de nuevas vías que se pueden crear tras detectar una bifurcación a dos. Además, se implementan las siguientes restricciones que los segmentos deben cumplir para poder ser considerados válidos:

- Tasa de crecimiento

$$TC = \frac{1}{N} \sum_{i=1}^N \frac{|NV_i|}{|NV_{i-1}|} < U_{TC}, \quad (10)$$

siendo NV el número de vóxeles en un segmento en la i -ésima iteración del algoritmo, N el número de iteraciones y U_{TC} un umbral fijado.

- Compactibilidad: [102]

$$C = \frac{n - \frac{A}{6}}{n - n^{\frac{2}{3}}}, \quad (11)$$

donde n es el número de vóxeles que conforman el frente de ondas y A el área.

- Diferencia de longitud entre los frentes de onda: Empíricamente se ha demostrado que diferencias mayores al 10% entre los tamaños de los frentes de ondas que conforman cada segmento son un síntoma evidente de fugas, por esta razón se define:

$$|W_{ultimo} - W_{primero}| < T_W, \quad (12)$$

siendo W_{ultimo} y $W_{primero}$, el tamaño del frente de ondas al final y al principio del segmento evaluado, respectivamente, y T_W un umbral definido por el usuario.

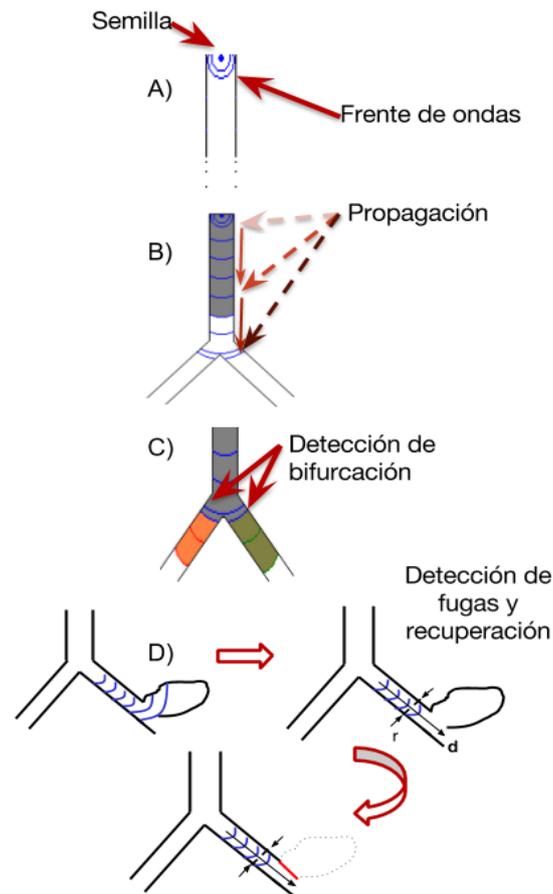


Figura 14: Extracción de las vías aéreas: A) Establecimiento de la semilla y el frente de ondas en la vecindad; B) Propagación del frente de ondas; C) Detección de una bifurcación y división del frente de ondas en dos nuevos frentes; D) Corrección de las fugas hacia el parénquima pulmonar

5.1.6 Inclusión de las lesiones en el interior del pulmón

La aplicación sucesiva de los métodos detallados en las secciones anteriores a una imagen de CT, permiten obtener una segmentación del tejido pulmonar sano. Sin embargo, no se realiza la inclusión de las lesiones causadas por el *Mycobacterium tuberculosis* como parte del parénquima pulmonar (ver sec.3.2). Esto supone la aparición de nichos en la imagen, como se puede observar en el ejemplo de la figura 15.

Dado que el objetivo es obtener todo el tejido perteneciente al pulmón, tanto sano como enfermo, para realizar una posterior evaluación del estado de la enfermedad, se hace uso de una técnica de post-procesado dedicada a rellenar de forma automática, siguiendo un determinado criterio, las cavidades internas.

Para este fin, se utiliza de manera iterativa una operación morfológica de cerrado, frecuentemente utilizada en aplicaciones de imagen médica [103].

Mediante este proceso, se evalúa que vóxeles de una imagen binaria, como la mostrada a la izquierda de la figura 15, son incluidos dentro de la segmentación final¹¹ del pulmón

¹¹En el caso de la figura 15 se muestra la máscara binaria correspondiente al tejido identificado como pulmón y no la segmentación de la imagen propiamente

evaluando la conveniencia de incluir cada uno de ellos a partir de la intensidad de los vóxeles adyacentes. La decisión final sobre cada vóxel dependerá del valor mayoritario de los vóxeles en el vecindario seleccionado.

En el caso particular de esta aplicación, se utiliza un filtro iterativo para rellenar cavidades basado en la votación por mayoría con un radio para definir las regiones de 1 mm, de manera similar a como se describe en el trabajo de *Janaszewski et al.* [104].



Figura 15: Post-procesado para incluir partes del parénquima pulmonar, pertenecientes a las lesiones de tuberculosis, excluidas de la segmentación inicial. A la izquierda, se muestra la máscara de segmentación para un corte transversal de un CT de tórax. Aparecen múltiples cavidades, correspondientes a tejido pulmonar, debido a las limitaciones de los procesos de umbralización. A la derecha, la máscara de la misma rodaja tras aplicar el post-procesado, únicamente aparecen cavidades correspondientes a los vasos de gran tamaño (las arterias pulmonares) y las vías aéreas (los bronquios, en este caso).

5.1.7 Refinamiento basado en el patrón de las lesiones de tuberculosis

Como se mostró en el proceso anterior (5.1.6), la extracción de las lesiones resulta compleja, especialmente en el caso de aquellas que se encuentran pegadas a la pleura [44], debido a que presentan niveles de gris muy similares a los del tejido que las rodea. El proceso de inclusión descrito en la sección anterior puede resultar insuficiente cuando un sujeto muestra diversas lesiones en este área específica del pulmón. Debido a esto, es necesario incluir un último paso para refinar la segmentación, que sea capaz de detectar lesiones en los bordes del tejido extraído en el paso anterior del flujo de trabajo. Para este fin, se utilizan métodos basados en curvas de nivel dado que han demostrado eficacia en tareas de segmentación similares [47, 105].

Este proceso se realiza extrayendo la superficie del volumen obtenido tras el paso descrito en la sección anterior, a partir del cual se obtienen una serie de semillas situadas en las zonas dañadas de los límites del pulmón. Dichas semillas se extraen utilizando la propiedad de que estos vóxeles muestran una intensidad muy por encima de las de sus vecinos. Una vez extraídas se utilizan para establecer los contornos iniciales utilizados por un algoritmo de *fast marching level sets* [106]. A continuación, el contorno se expande pudiéndose llegar a unirse con contornos cercanos si la condición del gradiente impuesta lo permite. De esta manera, se obtendrán varias curvas de nivel que se pueden ajustar mejor a las lesiones utilizando el algoritmo de contornos activos geodésicos (*Geodesic Active Contours*, GAC) descrito por *Caselles et al.* [97]. De esta forma un contorno se ajusta a la región siguiendo la siguiente ecuación diferencial:

$$\frac{\partial \Psi}{\partial t} = -\alpha \mathbf{A}(\mathbf{x}) \cdot \nabla \Psi - \beta P(\mathbf{x}) |\nabla \Psi| + \gamma Z(\mathbf{x}) \kappa |\nabla \Psi|, \quad (13)$$

siendo Ψ la curva de nivel inicial, \mathbf{x} un punto en el contorno, $\mathbf{A}(\mathbf{x})$ control el proceso de advención, $P(\mathbf{x})$ la propagación y $Z(\mathbf{x})$ la modificación espacial de la curvatura media; κ ; α , β y γ son escalares para la modulación de la evolución del contorno.

5.1.8 Métodos para la evaluación de los resultados de la segmentación

Con el propósito de evaluar la herramienta implementada, en la sección de resultados para la segmentación (7.1), se comparan las máscaras obtenidas automáticamente con el software, con las máscaras extraídas manualmente por personal experto con conocimientos en imagen médica.

Existen varios métodos para medir el parecido entre segmentaciones [107], normalmente encargados de extraer diferencias, bien entre los volúmenes, o bien, entre las superficies de estos. Para la evaluación de la herramienta, se utilizan los métodos más comunes, dado que permiten extraer con facilidad conclusiones generales sobre la calidad de la segmentación, además de establecer comparaciones con otros métodos existentes en la literatura [108]. Estos son:

- Coeficiente de Sorensen-Dice (*Dice-Sorensen Similarity Coefficient, DSC*) [109, 110]: El DSC muestra cual es el grado de solapamiento entre una imagen volumétrica segmentada A y una imagen de referencia B .

$$DSC = \frac{2(A \cap B)}{A + B} \quad (14)$$

- Falsos positivos y falsos negativos (*False Positive Error, FPE* y *False Negative Error, FNE*) [111]: Se define el error por falsos positivos como, el volumen de una imagen B no solapado con la región definida por una imagen A , normalizado por el volumen de B . Mientras que el error por falsos negativos, se achaca a el volumen de A no coincidente con el de la imagen B . Matemáticamente:

$$FNE = \frac{\sum_{i=1}^N B_i \setminus A_i}{\sum_{i=1}^N B_i} \quad FPE = \frac{\sum_{i=1}^N A_i \setminus B_i}{\sum_{i=1}^N A_i}, \quad (15)$$

donde i representa el i -ésimo vóxel de la máscara y N el número total de vóxeles.

- Distancia de Hausdorff [112] (*Hausdorff Distance, HD*) y distancia media de Hausdorff (*Hausdorff Distance Average, HDA*): La HD representa la máxima distancia posible entre las dos superficies segmentadas.

$$HD = \max \left(\max_{s_a \in S(A)} d(s_a, S(B)), \max_{s_b \in S(B)} d(s_b, S(A)) \right), \quad (16)$$

donde s_a y s_b representan puntos en las superficies de A y B , $S(A)$ y $S(B)$ el conjunto de todos los puntos en las superficies de A y B , respectivamente y d la distancia euclidiana.

Evidentemente, la distancia media de Hausdorff se obtiene promediando todas las distancias entre las superficies y proporciona información poco sensible a deformaciones anómalas que se hayan podido producir en la segmentación de la estructura. Se define como se muestra a continuación:

$$HDA = \frac{1}{N_a + N_b} \left[\sum_{i=1}^{N_a} d(s_a, S(B)) + \sum_{j=1}^{N_b} d(s_b, S(A)) \right], \quad (17)$$

siendo N_a y N_b el número de vóxeles segmentados en cada volumen.

5.2 Métodos: Clasificación de lesiones en pulmones infectados

El modelo de clasificación empleado se muestra en la figura 16. Este diseño está especialmente orientado a:

1. La extracción de un conjunto pequeño de características de textura informativas.
2. Análisis del efecto del desequilibrio entre el número de instancias de cada clase.

Las técnicas empleadas para este análisis se describen en las secciones subsiguientes.

5.2.1 Extracción de características de textura

Como se comentó en la revisión de la bibliografía (sec.3.4.1), existen diferentes variantes para caracterizar una textura numéricamente. Sin embargo, las basadas en la matriz de co-ocurrencia de nivel de gris resultan especialmente útiles en aplicaciones radiológicas, por lo que se emplean en este estudio.

Dada la relevancia que estas características tienen en la consecución de los objetivos del trabajo, se ha implementado una herramienta para el cálculo de las mismas que permite evitar los inconvenientes de otras existentes. Los detalles de la implementación se recogen en la sección 6.2.

La matriz de co-ocurrencia del nivel de gris (*Gray-Level Co-occurrence Matrix*, GLCM), proporciona un método para generar características de textura de segundo orden en una imagen. Esta matriz representa mediante cada uno de sus elementos, la frecuencia conjunta de todas las combinaciones de niveles de gris entre parejas de vóxeles que se encuentran separados una distancia d en una dirección Θ en una imagen. Dicha definición se puede generalizar para tres dimensiones (en este trabajo se utilizan volúmenes) introduciendo el término desplazamiento u *offset* para referirse a la posición de los vóxeles a comparar de forma relativa. De esta manera, cada vóxel $v_{x,y,z}$ en una imagen o una Región de Interés (ROI)¹², se compara con un vóxel pareado, $v_{x+\Delta_x,y+\Delta_y,z+\Delta_z}$, definido por su posición relativa al vóxel que sirve como referencia, dado por un desplazamiento $(\Delta_x, \Delta_y, \Delta_z)$.

Utilizando la definición original presentada por *Haralick et al.* [113, 114] y ampliándola para el caso tridimensional, la matriz de co-ocurrencia, C , se establece formalmente como:

$$C_{\Delta_x, \Delta_y, \Delta_z}(i, j) = \sum_{x=1}^n \sum_{y=1}^m \sum_{z=1}^p \begin{cases} 1, & \text{si } I(x, y, z) = i \text{ y } I(x + \Delta_x, y + \Delta_y, z + \Delta_z) = j \\ 0, & \text{en el resto} \end{cases}, \quad (18)$$

donde $I(x, y, z)$ es la intensidad del vóxel en la posición (x, y, z) de la imagen.

Para este trabajo, se obtuvieron las frecuencias conjuntas entre vóxeles adyacentes. Es decir, se consideró un desplazamiento unitario, en las 13 direcciones posibles en una imagen tridimensional (ver tabla 2). Esto supone trabajar con trece GLCMs por región, sin embargo, es sabido que la textura en imágenes radiológicas de tomografía axial no presenta grandes variaciones en función de la orientación [115], por lo que las trece matrices son promediadas para obtener una matriz final que represente cada una de nuestras regiones segmentadas previamente. A partir de los elementos de la GLCM, se obtienen 22 indicadores de textura expuestos en el apéndice de definición de características de textura obtenidas a partir de la matriz GLCM (sec.9), que se utilizaron para caracterizar cada una

¹²Carece de interés calcular matrices de co-ocurrencia para una matriz completa dado que se utilizan para caracterizar zonas específicas con características particulares. En este trabajo en particular, interesa caracterizar las texturas de las lesiones causadas por la tuberculosis.

de las lesiones etiquetadas.

Comúnmente, la GLCM no se calcula sobre las imágenes con $2^{16} - 1$ niveles de gris¹³ que se pueden encontrar en una imagen de CT, sino que se realiza una cuantización, es decir, una reducción del número de niveles de gris, utilizando técnicas de redondeo y truncado. De esta manera, se consigue reducir el coste computacional del cálculo de la GLCM, a la par que se consigue minimizar el ruido [116], aunque supone como es de esperar, una pérdida de detalles. Esto puede ser fundamental en la textura, dado que si ésta se encuentra influenciada por los artefactos presentes en la imagen (por ejemplo, el movimiento de los pulmones) no representará fielmente el tejido que *a priori* caracteriza.

Debido a ello, este trabajo estudia el efecto que tiene la cuantización, con el objetivo de determinar si es preferible trabajar con imágenes más suavizadas debido al ruido del proceso de adquisición o si por el contrario es mejor utilizar toda la información disponible.

Con este objetivo se realiza el estudio de clasificación utilizando 8, 16, 32, 64, 128 y 256 niveles de gris de forma independiente.

Dirección	offset
(0°, 45°)	(1,0,1)
(0°, 90°)	(1,0,0)
(0°, 135°)	(1,0,-1)
(45°, 45°)	(1,1,1)
(45°, 90°)	(1,1,0)
(45°, 135°)	(1,1,-1)
(90°, 45°)	(0,1,1)
(90°, 90°)	(0,1,0)
(90°, 135°)	(0,1,-1)
(135°, 45°)	(-1,1,1)
(135°, 90°)	(-1,1,0)
(135°, 135°)	(-1,1,-1)
(-, 0°)	(0,0,1)

Tabla 2: Desplazamientos correspondientes a las 13 direcciones entre vóxeles adyacentes en el espacio tridimensional.

5.2.2 Selección de características relevantes

Como se ha mencionado anteriormente, los conjuntos de datos, especialmente aquellos que cuentan con cientos o miles de características, suelen contener un alto número de componentes irrelevantes y redundantes que afectan negativamente al proceso de aprendizaje automático [73].

Las características irrelevantes son aquellas que no contribuyen a la distinción entre diferentes clases. Las características redundantes son aquellas que dan la misma información que otras, siendo ambas relevantes. El hecho es que el desconocimiento de una de ellas no afecta negativamente al rendimiento del proceso de clasificación. Por ello, es recomendable eliminar tanto las características irrelevantes como las redundantes.

El proceso de suprimir características irrelevantes y redundantes sin afectar a la bondad de la clasificación [117], se denomina selección de características (*feature selection*).

¹³apreciar que, normalmente no se utilizan todos los niveles procurados al utilizar 16 bits, dado que las HU suelen establecerse entre -1024 y 3024

Para el problema que ocupa esta parte del trabajo, referente a la clasificación automática de las distintas lesiones que causa la tuberculosis en el parénquima pulmonar, optamos por calcular un número reducido de características (concretamente 26 definidas en 5.2.1 y el Apéndice B). Sin embargo, estas características se encuentran fuertemente relacionadas, lo que implica una alta dependencia entre ellas [118]. Por lo tanto, existe el riesgo de incluir características redundantes que pueden introducir errores en la clasificación. Además, para nuestra aplicación, resulta especialmente útil conocer que características resultan más informativas, dado que la interpretación a nivel biológico de los valores que éstas presentan, pueden ser fundamentales para el diagnóstico y el tratamiento de la enfermedad [9, 11].

Por esta razón, el flujo de trabajo introducido en la figura 16 realiza un análisis de las características compuesto de dos partes. En primer lugar, se realiza una clasificación de las características extraídas de cada región siguiendo el criterio MI-mRMR (*Mutual Information minimal Redundancy Maximal Relevance*). Seguidamente, se emplea un clasificador simple, para evaluar el rendimiento ofrecido por diferentes conjuntos de características formados por grupos de estas de magnitud creciente. Es decir, en cada iteración se evalúa un conjunto de características de textura que cuenta con un indicador más que el grupo utilizado en la iteración anterior (ver, m , número de características utilizadas en clasificación en la figura 16). Hay que añadir que el clasificador realiza la operación de aprendizaje utilizando datos balanceados (ver figura 17).

De esta manera se pretenden aprovechar las ventajas de los métodos de filtrado, mediante el criterio de mRMR y de *wrapper* de forma conjunta.

5.2.3 Clasificación de características: MI-mRMR

Como se indicó anteriormente, las técnicas basadas en criterios de información mutua se utilizan ámpliamente para la clasificación jerarquizada (*ranking*) de características, especialmente en conjunción con el criterio de mínima-redundancia-máxima-relevancia (mRMR) [76, 117, 119]. Por ello, se ha utilizado como técnica de filtrado en este trabajo.

La MI de dos características (variables aleatorias) viene dada por la función de densidad de probabilidad marginal de cada una de ellas y por su función de distribución conjunta como sigue:

$$I(x; y) = \sum_{i,j} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \quad (19)$$

El objetivo de esta técnica consiste en ordenar el conjunto de características, Ω , en función del mRMR.

Siguiendo este criterio, la primera variable del conjunto de características seleccionadas, S , es la que presenta mayor dependencia con una de las clases que aparecen en el problema de clasificación, c .

El mRMR utiliza la MI en dos operaciones para generar la clasificación [117]:

- Emplea la condición de redundancia mínima para seleccionar aquellas características que son mutuamente exclusivas, esto es:

$$\min R(S), R = \frac{1}{m^2} \sum_{x_i, x_j \in S} I(x_i; x_j), \quad (20)$$

siendo m el número de características incluidas en S .

- Emplea la condición de máxima relevancia para medir la dependencia entre una característica x_i y una clase c , matemáticamente:

$$\max D(S, c), D = \frac{1}{m} \sum_{x_i \in S} I(x_i; c). \quad (21)$$

Las condiciones 20 y 21 se pueden combinar en única expresión capaz de determinar la figura de mérito de cada característica, siguiendo el criterio mRMR:

$$\max_{x_i \in \Omega_S} \left[I(x_i; c) - \frac{1}{m} \sum_{x_j \in S} I(x_i; x_j) \right], \quad (22)$$

donde Ω_S representa el conjunto de características ya seleccionado, $\Omega_S = \Omega - S$.

5.2.4 Evaluación de características

La clasificación aportada por el algoritmo de MI-mRMR, proporciona información sobre la relevancia de cada una de las características. Sin embargo, no lo hace sobre el impacto que supone emplear sólo el conjunto de las características más relevantes, o alternatively, cuantas de estas características se deben usar para no perder precisión al clasificar.

Por esta razón, se evalúan conjuntos de características empleando un clasificador simple, muy diferente a los clasificadores optimizados que se utilizan para lograr la máxima precisión con un conjunto de características predefinido, como se realiza en apartado de optimización (sec.5.2.6). Esta aproximación permite:

- Procesos de entrenamientos rápidos, al trabajar con clasificadores no paramétricos, que resultan mucho más adecuados dado que se realizarán de forma múltiple.
- Establecer relaciones potenciales de causalidad entre las características y los resultados de la clasificación. De esta manera, se facilita su interpretación en procedimientos médicos como el diagnóstico, con respecto a los resultados ofrecidos por clasificadores más complejos, destinados a ofrecer la máxima precisión.

En esta aproximación se utiliza Análisis Discriminante Lineal (*Linear Discriminant Analysis*, LDA) [120], que aparece de forma recurrente en la literatura médica que emplea el análisis de texturas [121, 122].

El clasificador basado en LDA, utiliza el teorema de Bayes para asignar a un vector de características, \mathbf{x} (las texturas que representan una región en el caso de este trabajo), la clase c que maximice la probabilidad *a posteriori* $P(y = c|\mathbf{x})$, esto es:

$$P(y = c|\mathbf{x}) = \frac{P(\mathbf{x}|y = c)P(y = c)}{P(\mathbf{x})} = \frac{P(\mathbf{x}|y = c)P(y = c)}{\sum_k P(\mathbf{x}|y = k) \cdot P(y = k)}. \quad (23)$$

El cálculo del denominador de esta expresión resulta complejo y es objeto de una amplia discusión en el campo de la estadística y el aprendizaje estadístico [123, 124]. En el caso de LDA se asume que las distribuciones *a priori* y condicionales, $P(y = c)$ y $P(\mathbf{x}|y = c)$, son *gaussianas*, por lo que la ecuación 23 quedaría como sigue:

$$p(\mathbf{x}|y = c) = \frac{1}{(2\pi)^n |\Sigma_c|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu_c)^T \Sigma_c^{-1}(\mathbf{x}-\mu_c)} \quad (24)$$

Expresión que se simplifica asumiendo que la matriz de covarianza Σ_c es idéntica para todas las clases, lo que supone que se puede establecer el siguiente criterio para clasificar dos clases (c y k):

$$\log \left(\frac{P(y=c|\mathbf{x})}{P(y=k|\mathbf{x})} \right) = 0 \Leftrightarrow (\mu_c - \mu_k) \Sigma^{-1} \mathbf{x} = \frac{1}{2} (\mu_c^T \Sigma^{-1} \mu_c - \mu_k^T \Sigma^{-1} \mu_k). \quad (25)$$

Lo que impone que se establezcan superficies lineales de separación en el espacio hiperdimensional del problema, que no proporcionan resultados óptimos pero que facilitan la interpretación de los resultados.

Así, el LDA se utilizará para clasificar conjuntos compuestos por $m = \{1, 2, 3, \dots, M\}$ características de forma iterativa (ver fig.16) siendo M el número total de características extraídas. Para reducir los efectos de la tendencia o *bias* y la varianza, se realiza validación cruzada.

5.2.5 Balanceo de datos

Para lidiar con el problema que supone tener conjuntos de datos donde existen muchas más instancias de unas clases que de otras, esto es, la tendencia del aprendizaje a predecir siempre la clase más numerosa, se utilizan dos técnicas de muestreo encadenadas, SMOTE + TL.

En primer lugar, se realiza un sobre-muestreo controlado de las clases minoritarias utilizando SMOTE. Este procedimiento crea cada muestra nueva, \mathbf{x}_n , a partir de una instancia de la clase minoritaria a sobre-muestrear, \mathbf{x}_i , y sus k vecinos más cercanos, donde k es normalmente establecida de forma determinista. Específicamente esta nueva muestra se obtiene como:

$$\mathbf{x}_n = \mathbf{x}_i (\mathbf{x}_i - \hat{\mathbf{x}}_i) \times \alpha, \quad (26)$$

donde $\hat{\mathbf{x}}_i$ representa a uno de los k vecinos más cercanos escogido aleatoriamente y α es un número aleatorio entre 0 y 1.

Dado que este proceso tiene problemas bien conocidos y estudiados respecto a su generalización y la varianza de las muestras creadas [68], se limita su número. Además, se incluye un segundo paso consistente en aplicar sub-muestreo mediante enlaces Tomek o *Tomek Links*. Este proceso de sub-muestreo se realiza para todas las clases y consigue eliminar las instancias más cercanas de clases diferentes. De esta forma, se puede ampliar la separación en las regiones que forman las soluciones de manera que el clasificador pueda operar de manera más óptima.

5.2.6 Clasificación de regiones utilizando clasificadores optimizados

La última parte de la metodología evalúa el rendimiento obtenido por un modelo de clasificación optimizado. En un caso, cuando se utilizan todas las características de textura disponibles. En otro caso, cuando únicamente se utilizan el subconjunto de características más relevantes obtenido en los pasos anteriores. En ambos casos, con y sin balanceo de los datos. Esto resulta en un total de $6 \times 2 \times 2$ clasificadores optimizados, correspondientes a los 6 niveles de cuantización, los 2 conjuntos de características y el análisis del balanceo.

Para esta evaluación se optimizan clasificadores basados en máquinas de soporte de vectores (*Support Vector Machines, SVM*) [125] con función de kernel de base radial gaussiana. La función de clasificación de una SVM, se puede extraer directamente del trabajo de *Vapnik*,

dada por $sign(f_{\mathbf{x}})$, donde $f_{\mathbf{x}}$ se define como:

$$f(\mathbf{x}) = \mathbf{w}\phi(\mathbf{x}) + b = \sum_{i \in S} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b, \quad (27)$$

siendo K la función de base o *kernel* y S el subconjunto de multiplicadores de Lagrange, los cuales sopesan cada instancia en función de su importancia para la minimización de:

$$\Psi(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i, \quad (28)$$

restringido a:

$$y_i(\mathbf{w}\phi(\mathbf{x}) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \quad i = 1, 2, \dots, N, \quad (29)$$

empleando el *kernel* gaussiano dado por:

$$K(\mathbf{x}_i, \mathbf{x}) = e^{(-\gamma \|\mathbf{x}_i - \mathbf{x}\|^2)}, \gamma \geq 0, \quad (30)$$

Siendo γ y C los hiperparámetros a optimizar mediante el método de búsqueda exhaustiva (*grid search*) y validación cruzada, en el que de nuevo se realiza el balanceo de los datos de entrenamiento (ver fig.17).

Para evaluar entre los distintos clasificadores en el proceso de optimización, se utiliza la medida F1 por cada clase ponderada por el número de muestras de la misma. Esto es:

$$\frac{1}{\sum_k |y_k|} \sum_{k \in C} |y_k| F_1(y_k, \hat{y}_k), \quad (31)$$

siendo y_k e \hat{y}_k , las etiquetas verdaderas y el valor estimado por el predictor para la clase k , perteneciente al conjunto de clases del problema C .

Es importante apreciar que para evaluar la calidad de un clasificador para un problema multiclase como el que concierne a este trabajo, las medidas habituales de los modelos binarios (exactitud, Precision, Recall, F-1) resultan inaplicables o ineficientes.

Los modelos pueden obtener precisiones altas con facilidad, dado que suelen presentar sesgos que provocan la clasificación de las instancias como casos de las clases más representadas. Por esta razón resulta fundamental utilizar medidas de evaluación que combinen los resultados obtenidos para cada clase de forma de forma individual.

Diagrama del flujo de trabajo empleado en la clasificación

Las siguientes figuras sintetizan el flujo de trabajo seguido para la clasificación de lesiones utilizando seis niveles de de cuantización diferentes, realizando el proceso de selección de características y teniendo en cuenta el balanceo de los datos.

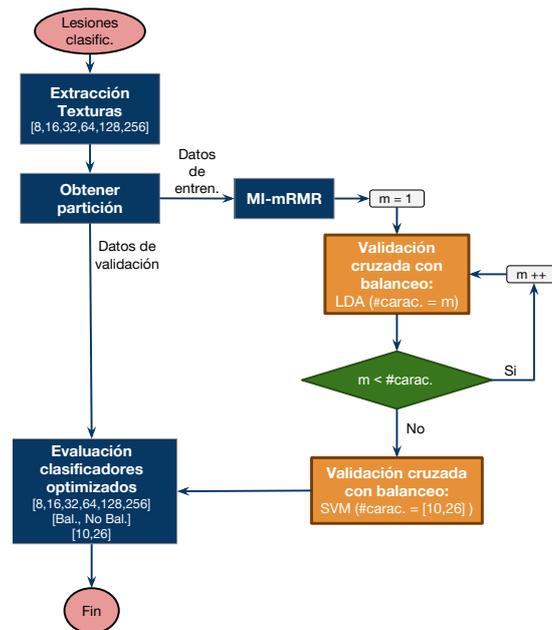


Figura 16: Diseño seguido para la extracción de características de textura relevantes para cada nivel de cuantización utilizado (8, 16, 32, 64, 128, 256), según el criterio MI-mRMR y validación cruzada de los clasificadores con y sin balanceo. Apreciar que los clasificadores con balanceo (representados por las cajas en tonos ocre) siguen el esquema presentado en la figura 17

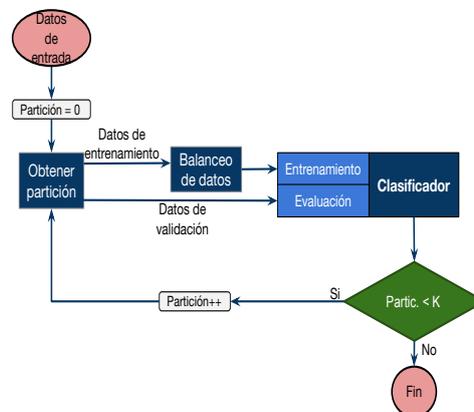


Figura 17: Esquema genérico para cualquier clasificador utilizando balanceo de datos. Se puede apreciar que únicamente los datos de entrenamiento son balanceados.

6 Desarrollo de software

Esta sección recoge un resumen detallado del software implementado para alcanzar los objetivos del proyecto.

Así, en la primera sección (sec.6.1) se pormenorizan las especificaciones junto con los detalles de implementación de la herramienta desarrollada para la segmentación de pulmones infectados con tuberculosis en imágenes de CT de modelos animales. Se trata del apartado más extenso en lo que refiere a desarrollo de código, principalmente debido a que al tratarse del problema más específico a resolver dentro del proyecto, las librerías ya implementadas y testadas que se pueden reutilizar (en este caso librerías para el desarrollo de aplicaciones de visión artificial), contienen utilidades de propósito mucho más general.

En el segundo apartado, se recogen los detalles del software desarrollado desde cero para la extracción de características de textura a partir del computo de las matrices GLCM.

El código implementado se encuentra disponible en el siguiente repositorio:

https://github.com/PeterMcGor/TFM_Code

6.1 Software para la segmentación de pulmón

Como se recoge en la sección de la metodología referente a la segmentación de pulmón, las diferentes técnicas planteadas se basan en nuevas propuestas que se combinan con filtros y funciones habituales en las aplicaciones de procesado de imagen.

Existen varias plataformas que facilitan estas funciones, sin embargo, una de las más conocidas y utilizadas (si no la más) para la construcción de aplicaciones en el contexto de la imagen médica, debido a sus altas prestaciones computacionales y flexibilidad es: *The Insight Toolkit*, ITK [126], bien en su versión original para *C++* o en su versión *wrapeada* para *Python*, *SimpleITK*.

ITK es un ejemplo paradigmático de la programación orientada a objetos, presentando una estructura muy modular. Cada técnica de procesamiento se encapsula en una clase, que utilizando la terminología habitual en el campo del procesamiento de señales, se denomina filtro. Evidentemente, cada uno de estos filtros se conforma por la encapsulación de otros que realizan operaciones requeridas previamente. Esta arquitectura, basada en la conexión de varios objetos de tipo filtro, se completa definiendo interfaces, que resultan transparentes para los usuarios e incluso en la mayoría de ocasiones, para los desarrolladores. Con esta estructura, los filtros que representan diferentes operaciones son ensamblados de forma inmediata, teniendo como único requisito, establecer para cada entrada de una operación de filtrado, la salida del filtro que le precede, sin interactuar con el resto del algoritmo.

Tanto para el desarrollo de la metodología, como para la implantación de futuras técnicas, no resulta especialmente óptimo utilizar únicamente software desarrollado con *C++*¹⁴, si no que, puede resultar muy conveniente utilizar las funciones *wrapeadas* de *SimpleITK*, los filtros desarrollados puramente en *Python*, las funciones de *ImageJ* [127] o *Matlab*, entre otros muchos ejemplos. Por esta razón, la arquitectura implementada, siguiendo la filosofía de ITK, proporciona interfaces entre distintos entornos de procesamiento, de manera que se pueden encadenar funciones entre estos.

¹⁴*C++*, aunque es un lenguaje muy potente que permite minimizar los tiempos de computación, puede resultar tedioso para realizar prototipos sobre todo cuando no se maneja con fluidez el lenguaje, algo crítico en un entorno tan abierto y multi-disciplinar como la investigación.

En la última versión del código facilitado en el repositorio, se proporciona la arquitectura general, implementada en *Python*, con las interfaces desarrolladas hasta el momento. Dichas interfaces permiten unir filtros de *ITK(C++)*, compilados y proporcionados en forma de ejecutables individuales, filtros de *SimpleITK* y los métodos desarrollados puramente en *Python* para el tratamiento de las imágenes.

La arquitectura se recoge principalmente en el archivo *Segmentation.py*. Dentro de este se encuentra la codificación correspondiente a la estructura de clases (fig.18) que permite ejecutar de manera encadenada la sucesión de operaciones descrita para la tarea de segmentación propuesta.

La clase principal, *A_Filter*, se define como una clase abstracta, obligando a las clases derivadas a implementar el método *execute*, siendo en este método donde se recogen las particularidades de cada tipo de filtro a utilizar.

Para poder realizar filtrados encadenados, *A_Filter* cuenta con los atributos *Input_Path_image* y *Output_Path_Image* que indican las direcciones donde se almacenan las imágenes, de entrada y de salida del filtro respectivamente. Siempre que proceda, dado que no siempre se cuenta con un objeto imagen (filtros externos) o no siempre se opta por guardar la imagen resultante de un filtro (la implementación permite decidir que resultados guardar).

Los filtros desarrollados con *Python* derivan directamente de *A_Filter*. En esta implementación sólo se ha incluido *Hu_Threshold*, utilizado en la umbralización (ver sec.5.1.2).

Por otro lado, los filtros desarrollados en *C++* con *ITK* se encadenan utilizando la clase abstracta derivada de *A_Filter*, *External_Filter*. Esta clase impone la inclusión de un método *set_params* en las clases que heredan (filtros de *ITK*). Este último método, necesita que se establezcan los parámetros a utilizar en las funciones (definidas en la variable *path*) que se llaman de forma externa en el método *execute* de *External_Filter*.

Se incluyen los siguientes filtros de tipo *External_Filter*:

- *Apply_Anisotropic*: Encargado de realizar el pre-procesado de la imagen para eliminar ruido. La implementación en *C++* se adjunta en el archivo *apply_anisotropic.cxx* incluido en *utilities*.
- *Otsu_Threshold*: Utilizado en la umbralización. También incluido en *utilities*
- *Labelize*: Filtro para realizar la operación de conectividad. El código fuente se encuentra en la carpeta *lungs* con el nombre *labelize_main.cxx*
- *Choose_Lungs*: Este filtro busca el objeto más centrado en la segmentación de los diferentes *blobs* (ver sec.5.1.4). También está incluido en la carpeta *lungs* como *choose_lungs_main.cxx*.
- *Extract_Trachea*: Este filtro aúna todo el código necesario para el proceso complejo de segmentación de la vía aérea. El código fuente se recoge dentro de la carpeta *airways* en los archivos *label_air_main.cxx*, *SegmentAirwayTree_FM.cxx* y *itkTreeSegmentationImageFilter.txx*,

En el caso de la última interfaz implementada hasta el momento, los filtros *wrapeados* para *Python* de la librería de *SimpleITK*, se deriva de *A_Filter* la clase abstracta *SimpleITK_Filter*, que en este caso impone la definición de un objeto filtro de *SimpleITK* y de sus parámetros.

Dentro del fichero *Filters.py* se encuentran cuatro clases derivadas de *SimpleITK_Filter*:

Dilation y *Binarize* encargadas de realizar operaciones morfológicas y binarizaciones en varios procesos, *Mask* utilizada para extraer la vía aérea de la imagen original y obtener el resultado final de los pulmones segmentados mediante enmascarado y *File_Holes* para el postprocesado que permite incluir las lesiones causadas por la *tuberculosis*.

Finalmente, para facilitar la creación de flujos de trabajo específicos (por ejemplo, diferentes modelos animales, tipos de adquisiciones de imágenes de HRCT), se introduce la clase *Pipe*. Dicha clase contiene una lista de filtros que se derivan de la clase *A_Filter*, y que se encadenan tras la llamada al método *execute*.

El resto de archivos que componen el código fuente incluyen funciones de ayuda para realizar diferentes tareas básicas a lo largo del procesado, como las que aparecen en el archivo *UsefullFunctions.py*. El código utilizado para obtener los coeficientes mostrados en los Resultados se puede encontrar en el fichero *metrics.py*, mientras que la llamada principal a las tareas de segmentación están incluidas, como aparece en la sección de ejemplos de uso, en el archivo *Application.py*.

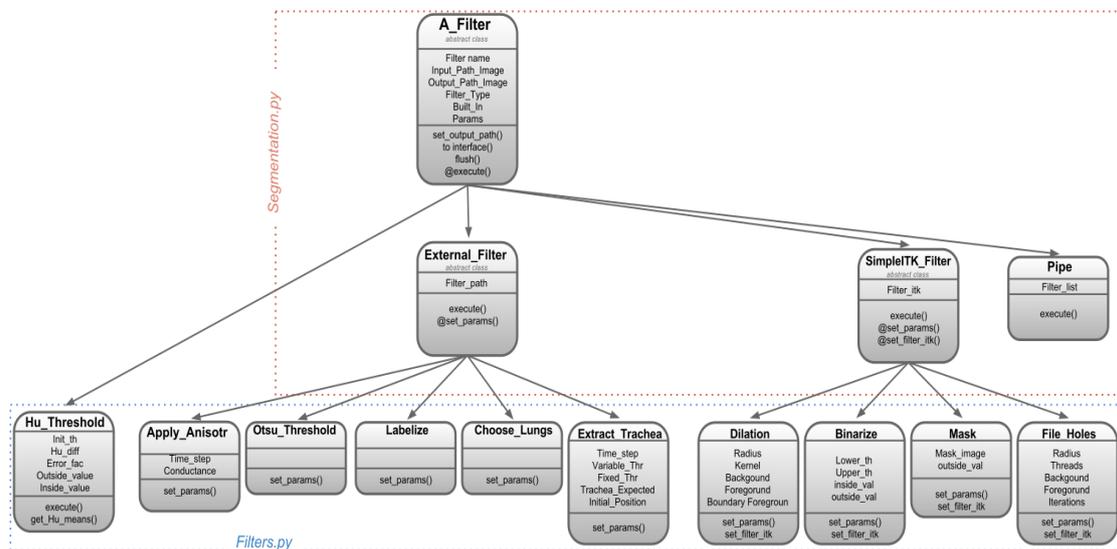


Figura 18: Representación de la arquitectura principal recogida en el archivo *Segmentation.py* y los filtros que realizan las distintas operaciones de procesado definidas en el fichero *Filters.py*. La clase de base *A_Filter* contiene las reglas a definir para incluir filtros de diferentes aplicaciones. La clase *External_Filter* define lo necesario para utilizar los filtros que se dan como ejecutables, es decir, los filtros de ITK desarrollados en *C++* para esta aplicación. La clase *Simple_ITK_Filter* proporciona la interfaz para todos los filtros *wrapeados* en *SimpleITK*. La clase *Pipe* permite establece la técnica para ejecutar en cadena varios filtros.

6.1.1 Aplicación e instalación

Se facilita el código desarrollado y testado para *Linux*, *C++11* y *Python 2.7*¹⁵. El código facilitado (carpeta *TLS*) se divide en dos partes: La carpeta *Segmentation_Modules* que contiene el código en *C++* correspondiente a la implementación de los filtros utilizando *ITK*, que se han descrito en la sección anterior, junto con una colección de archivos necesarios para compilar el código y la carpeta *Segmentation-application* donde aparece el código en *Python*.

¹⁵Concretamente, Ubuntu 16.04

A continuación, se presentan las instrucciones para proceder con la compilación y la instalación. Se dan también *scripts* de ejemplo que dan lugar a segmentaciones de pulmón para diferentes modelos animales:

1. Instalación de herramientas auxiliares: Es necesario contar con una serie de paquetes pre-instalados para poder realizar el proceso con éxito. Todos ellos se pueden obtener de forma simple desde los repositorios de *Ubuntu* con la siguiente línea, que además se puede encontrar en el script *auxiliares.sh*:

Listing 1: bash version

```

1  #!/bin/bash
2  sudo apt-get install g++ gcc git-core git-svn
3  libfontconfig-dev cmake-qt-gui libgl1-mesa-dev
4  libglu1-mesa-dev libncurses5-dev libosmesa6-dev
5  libx11-dev libxrender-dev libxt-dev make python-dev
6  python-numpy subversion libbz2-dev

```

2. Compilación de los filtros desarrollados con *C++* e *ITK*: Los filtros implementados en *C++* dependen de varias librerías que no se encuentran incluidas en la distribución estándar. Evidentemente, *ITK* es una de ellas, pero también *Boost* o *OpenCV*.

Por esta razón, se facilita un archivo para la compilación con *CMake*, *CMakeLists.txt*. Dicho archivo permite realizar en un primer paso, la compilación de todas las librerías que son dependencias y en un segundo paso, la compilación de los filtros implementados (ver figuras 19 y 22). Se recomienda el uso de la herramienta *cmake-gui* como se muestra en la figura 19.

Para realizar el proceso de compilación de forma directa, se introduce como fuente de código (*where is the source code*) la ruta a la carpeta *TLS*, mientras que los binarios (*Where to build the binaries*) se situarán en el mismo nivel que esta última, dentro de una carpeta que se debe llamar *TLS_BUILD* (fig.19)¹⁶. Cuando se ha realizado dicha configuración, se pulsa en *Configure*, se acepta la creación de la nueva carpeta *TLS_BUILD* (pulsar "yes") y se especifica el uso de los *Unix Makefiles* y *Default native Compilers* (ver figura 20). Entonces, se pulsa *Finish*. Cuando la configuración aparece como *Done*, se expande la variable *USE* y se seleccionan todas las librerías (ver figura 19), dejando sin seleccionar la variable *TLS_SECOND_STEP*.

Tras esta simple configuración, se vuelve a pulsar en *Configure* y a continuación, en *Generate*. Esto último genera el archivo *Makefile*, necesario para construir los binarios de las librerías, que además contiene la información para descargar las fuentes directamente desde sus repositorios y configurar de la forma adecuada las dependencias entre estas. Este proceso es facilitado utilizando el software proporcionado de forma libre por Ali Uneri en su repositorio de *GitHub*.

Con el archivo de entrada de *CMake* generado, basta con llamar al compilador desde dentro de la carpeta *TLS_BUILD* (fig.21) con el comando *make -jN*, donde *N* es el parámetro que indica el número de procesadores a utilizar durante la compilación.

Se trata de librerías "pesadas" por lo que el proceso de descarga y compilación no es inmediato y depende de la máquina utilizada.

Cuando la compilación de todas las librerías termina, se realiza el segundo paso. Para esto, se vuelve a generar el archivo *Makefile* pero en esta ocasión con la variable

¹⁶Es crítico configurar bien los nombres para correr la aplicación de manera correcta

TLS_SECOND_STEP seleccionada (ver figura 22). Es decir, se selecciona la variable y se pulsa consecutivamente *Configure* y *Generate*. Seguidamente, se vuelve a compilar con *make*, en este caso, únicamente se compilarán los filtros implementados por lo que el proceso es mucho más rápido.

3. Pre-requisitos para *Python*: Como se comentó anteriormente, para la implementación se ha utilizado *python 2.7*. Además se utilizan los paquetes *Numpy* y *SimpleITK* que no se encuentran integrados en la versión estándar de *Python* pero que se pueden instalar fácilmente utilizando, por ejemplo, *pip* (*pip install simpleitk* y *pip install numpy*).

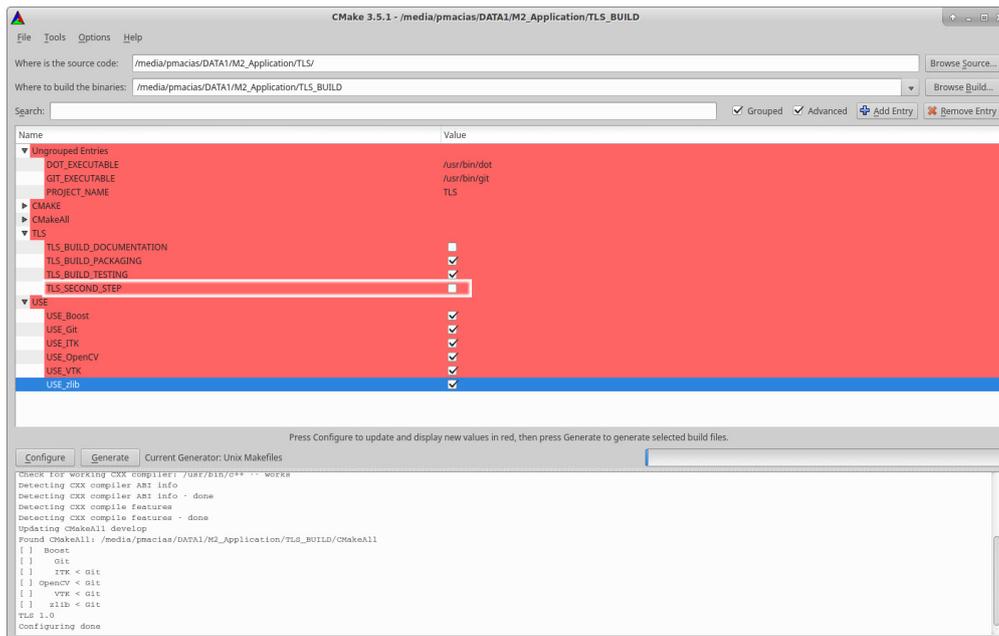


Figura 19: Configuración de *CMake* para la compilación del conjunto de librerías utilizadas en la construcción de los filtros. Apreciar que todas las librerías bajo la variable *USE* están seleccionadas, mientras que la variable *TLS_SECOND_STEP* no lo está.

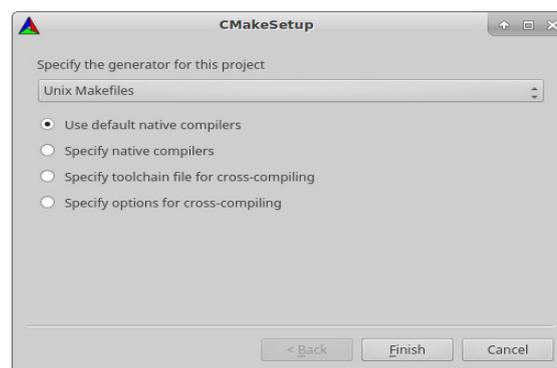
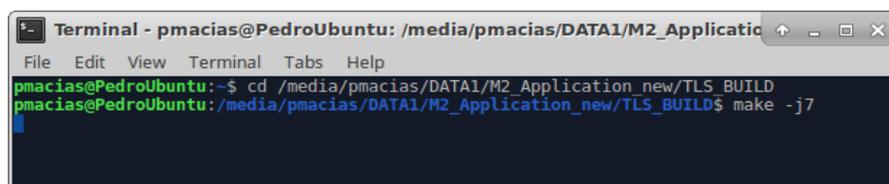


Figura 20: Selección de los *Unix Makefiles* para compilar.



```
Terminal - pmacias@PedroUbuntu: /media/pmacias/DATA1/M2_Application_new/TLS_BUILD
File Edit View Terminal Tabs Help
pmacias@PedroUbuntu:~$ cd /media/pmacias/DATA1/M2_Application_new/TLS_BUILD
pmacias@PedroUbuntu: /media/pmacias/DATA1/M2_Application_new/TLS_BUILD$ make -j7
```

Figura 21: Comando para la generación de los archivos binarios.

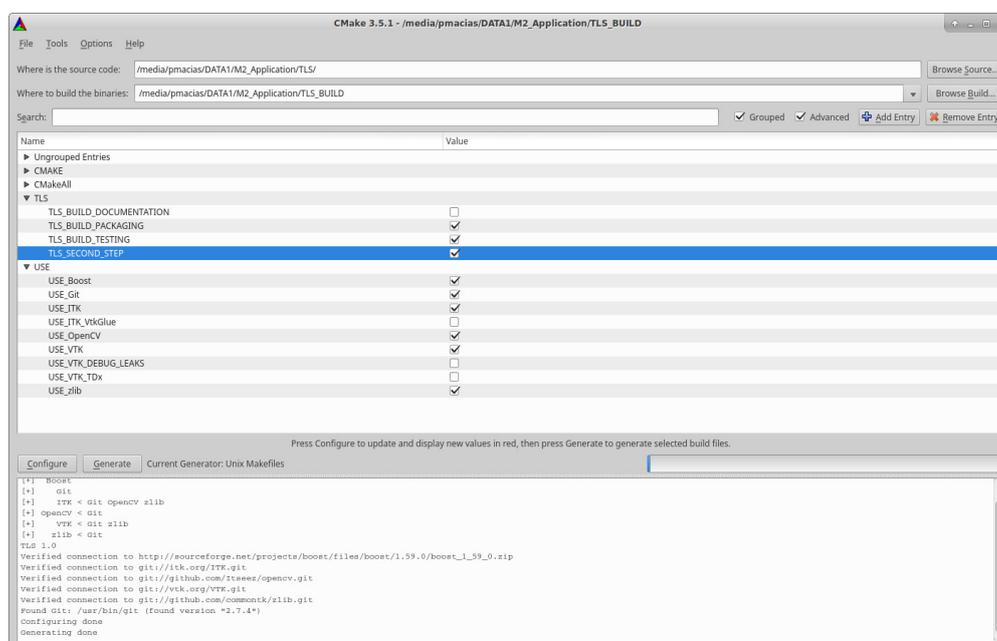


Figura 22: Compilación de los filtros implementados. Dese cuenta de que en esta ocasión *TLS_SECOND_STEP* está seleccionada.

6.2 Software para la extracción de características de textura

La cuantificación correcta de la textura que identifica cada una de las regiones de los pulmones segmentados, como ya se indicó en la sección de metodología (ver sec.5.2.1), resulta fundamental para el propósito de la aplicación propuesta en este trabajo.

Existen funcionalidades incluidas dentro de los paquetes de procesamiento de imagen más populares, que permiten el cálculo de características de textura derivadas de las GLCMs. Sin embargo, los más robustos y testados presentan limitaciones. Por ejemplo, la versión para la extracción de características incorporada en *scikit-image* [128], sólo permite el cálculo de textura en imágenes de dos dimensiones. Mientras que la versión para ITK [126] únicamente realiza el cálculo de ocho indicadores y además, realiza simplificaciones en los cálculos que implican que el rango de valores obtenidos en algunas características sea demasiado pequeño para poder expresar diferencias entre las mismas.

Por esta razón, se ha implementado una nueva aplicación que permite el cálculo de las matrices GLCM y sus características asociadas de forma robusta y computacionalmente eficiente.

La aplicación se ha desarrollado enteramente en *Python* y se puede encontrar en el repositorio de código en el interior de la carpeta de *Texture*.

La herramienta se estructura sobre dos clases fundamentales, la clase *Image_To_GLCM*,

recogida en *GLCM_Filter_Histogram_based.py*, y la clase *Texture_Features* que se puede encontrar en *GLCM_Features_Calc.py*. A continuación, se resumen los detalles de las mismas:

- *Image_To_GLCM*: Esta clase realiza el cálculo de la GLCM de una imagen, o una región de esta cuando se proporciona la máscara de la misma, para un desplazamiento determinado, mediante la llamada a su método *glcm()*.
Al contrario que las implementaciones habituales, que recorren todos los vóxeles de una imagen y todos los vecinos de este para un desplazamiento especificado, esta clase adapta la estructura de la imagen para hacer uso de la implementación, ya optimizada, para el cálculo de histogramas entre vectores proporcionado por la librería de *Python*, *numpy*.
- *Texture_Features*: La implementación de esta clase permite extraer las características de textura definidas en el Apéndice B, para un conjunto de regiones de una imagen y cada uno de los desplazamientos especificados, incluyendo el valor medio de las características en cada uno de estos.
Dado que dependiendo de la aplicación, el número de regiones y desplazamientos a evaluar puede ser elevado, se ha incluido la posibilidad de paralelizar el cálculo de forma que la computación resulte más eficiente.

7 Resultados Experimentales

De nuevo, los resultados se dividen en bloques para ilustrar con mayor claridad las ventajas, desventajas y eficacia de las propuestas implementadas para la extracción del parénquima pulmonar (ver Resultados de la segmentación) y la clasificación de los diferentes tejidos que se pueden encontrar en este.

7.1 Resultados para la segmentación del pulmón

Se utilizó la herramienta implementada (desarrollo de software) para segmentar automáticamente las imágenes descritas en la sección de Materiales. Los parámetros de las diferentes técnicas se fijaron de manera heurística, guiándose por la literatura existente.

Los parámetros utilizados en ambos modelos animales resultan similares, salvo excepciones. Por ejemplo, los referentes a la configuración del algoritmo de propagación implementado para la extracción de vía aérea, ya que debido a que las imágenes del modelo-B son más difusas y con menos diferencias en las unidades *Hounsfield* que representan los diferentes tejidos (bajo contraste), algunos parámetros toman valores más conservadores.

Los parámetros utilizados fueron:

- Umbralización: Se utilizó el método de *Otsu* para el modelo A. Las imágenes del modelo B resultaron ser bastante difusas, con menores diferencias entre zonas "aireadas" y tejido, con unidades *Hounsfield* inesperadas. Por ello, fue necesario utilizar un filtro de *Hu* modificado para poder desplazar el umbral obtenido automáticamente por 275 HU.
- Perímetro esperado de la traquea: Entre 11.5 y 33.5 *mm* para el modelo A y entre 1.2 y 4 *mm* para el modelo animal B.
- *Time Step*: 0.8 para las imágenes del modelo animal A y 0.5 para el modelo animal B.
- Factor de propagación α (ec. 8): 1.4 para el modelo animal A y 1 para el modelo B.
- Valor inicial para el umbral variable de intensidad, T_i (ecuación 8): -625 HU
- Umbral para el operador de *Sobel*: 2.5
- Factor β (ecuación 9): 2
- Umbral para la tasa de crecimiento (Ecuación 10): 1
- Umbral para la compactibilidad (Ecuación 11): 0.72

Los parámetros de control para el refinamiento final de la segmentación (sec. 5.1.7) sólo son válidos para el modelo animal A, dado que la naturaleza especialmente difusa de las imágenes del modelo animal B no es adecuada para el uso de esta metodología.

- $\alpha = 1.5$
- $\beta = 0.25$
- $\gamma = 1.0$

Los resultados obtenidos se han dividido en dos apartados. Una primera parte de resultados cuantitativos (sec.7.1.1) donde se realiza la comparación entre las máscaras obtenidas automáticamente y las máscaras segmentadas manualmente para el modelo A (ver sección de Materiales). Dado que no se cuenta con segmentaciones manuales para el modelo B, se incluye una segunda parte, (resultados cualitativos), donde se muestran los resultados de manera visual para ambos modelos.

7.1.1 Resultados cuantitativos

En la tabla 3 se presentan la media y la desviación típica de los coeficientes de similitud entre las segmentaciones automáticas y las manuales correspondientes al modelo animal A. Mientras que los diagramas de cajas de las figuras 23 y 24, muestran la variación de cada uno de los coeficientes en función de cada sujeto y en función de cada semana, respectivamente.

Analizando las medidas correspondientes al solapamiento entre volúmenes, es decir el DSC, FNE y FPE, se puede observar un parecido alto entre ambas segmentaciones, con un DSC medio del $93.4\% \pm 1$ distribuido de manera similar entre sujetos y diferentes semanas como se puede observar en los diagramas de cajas. Es destacable que la fuente de error debida al FNE, que presenta una tasa media del 13.4%, es mucho mayor que la debida a los errores por falsos positivos situada en apenas un 2%, lo que denota una serie de características de la aplicación que resultaran ventajosas o perjudiciales en función de la aplicación concreta.

Examinado las distancias de *Hausdorff*, que proporcionan información sobre la calidad de la segmentación de las superficies, mediante la distancia existente entre estas, se obtiene una HD media cercana a los 4mm, mientras que la HDA se sitúa en valores bajos localizados alrededor de 0.038mm.

Aunque la distribución, para ambas medidas es también bastante constante entre los sujetos y las semanas. Se puede apreciar la existencia de algunos valores que se alejan bastante de la tendencia, debido a que la medida es muy sensible a cambios bruscos en la forma de las superficies, lo que supone que en términos generales la HD presente un coeficiente variación alto.

De forma general, las medidas son independientes tanto del sujeto de estudio como del momento temporal al que corresponden. No aparecen tendencias que evidencien una mejor o peor segmentación cuando un individuo se encuentra más o menos afectado por la patología.

Sujeto	Semana	DSC	FNE	FPE	HD	HDA
Subj-1	0	0.947	0.126	0.002	3.001	0.034
Subj-1	16	0.939	0.113	0.000	8.574	0.029
Subj-1	28	0.935	0.144	0.000	3.543	0.042
Subj-2	0	0.921	0.176	0.001	4.889	0.049
Subj-2	16	0.941	0.129	0.001	2.001	0.041
Subj-2	28	0.932	0.138	0.008	4.003	0.048
Subj-3	0	0.941	0.121	0.002	1.991	0.030
Subj-3	16	0.934	0.131	0.002	5.156	0.037
Subj-3	28	0.917	0.121	0.003	3.845	0.036
Media	-	0.934	0.134	0.002	4.111	0.038
Std.	-	0.009	0.018	0.002	2.007	0.007

Tabla 3: Coeficientes de similitud entre las máscaras segmentadas utilizando la herramienta implementada y las obtenidas de forma manual. Las columnas muestran el sujeto, la semana después de la infección a la que se realiza la adquisición de la imagen, DSC (*Dice Similarity Coefficient*), FNE (*False Negative Error*), FPE (*False Positive Error*), HD (*Hausdorff Distance*) y HDA (*Hausdorff Distance Average*).

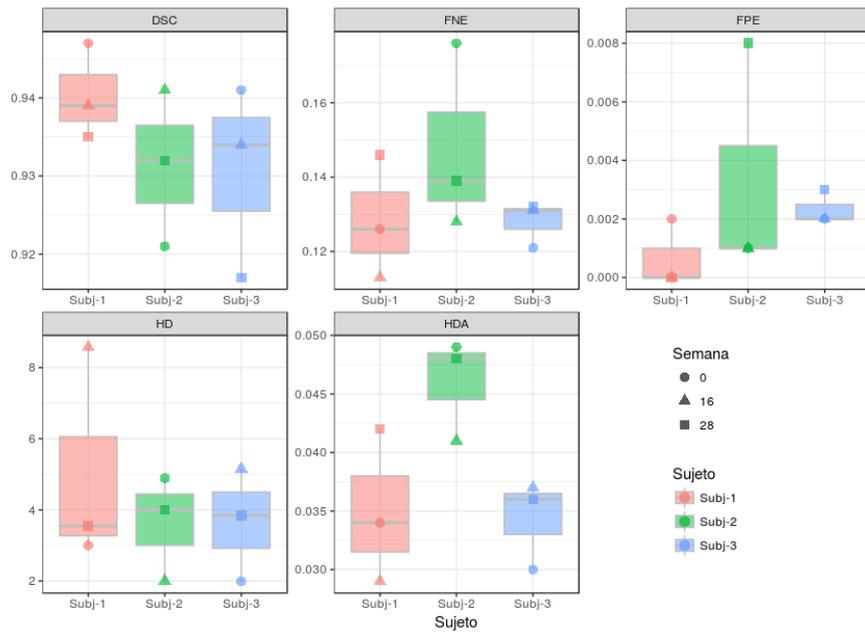


Figura 23: Diagrama de cajas para cada sujeto en cada una de las tres semanas del estudio para los cinco coeficientes de similitud utilizados. Sobre las cajas se representan los puntos correspondientes a las tres semanas del estudio.

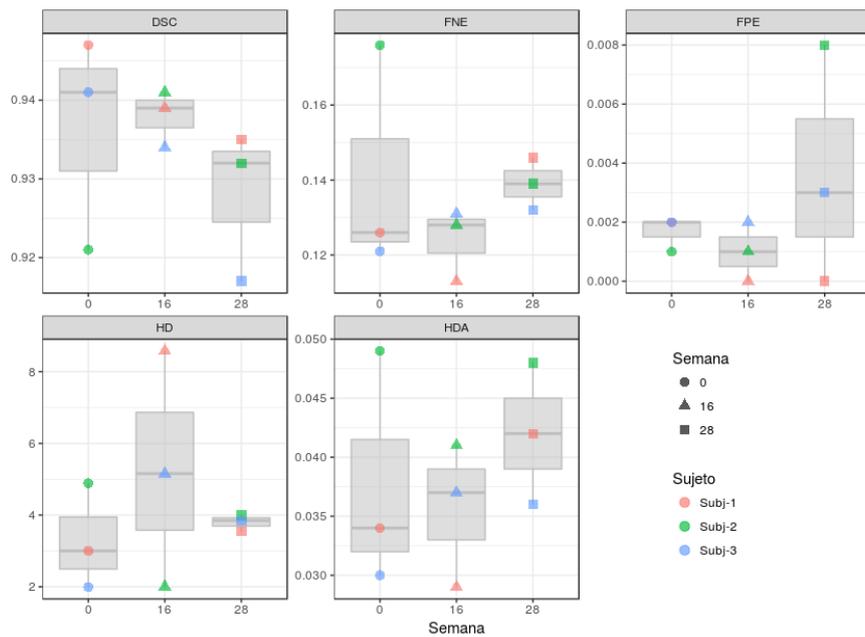


Figura 24: Diagrama de cajas para cada semana para los tres sujetos del estudio correspondiente al modelo animal A, representados con diferentes formas, en los cinco coeficientes de similitud utilizados.

7.1.2 Resultados cualitativos

Las figuras 25 y 26 muestran una serie de imágenes resultantes del proceso de segmentación, comparadas con las imágenes reales y además, en el caso del modelo A, con las segmentaciones manuales evaluadas en la sección anterior.

Estas imágenes exponen resultados representativos del conjunto total de las segmentaciones obtenidas. Para ello, a continuación se han incluido ilustraciones de los fenómenos más repetidos en las segmentaciones de los dos conjuntos de datos.

En la figura 25, se puede observar una tendencia entre las segmentaciones manuales, a incluir bordes que se extienden hasta zonas difusas de la imagen, generalmente debidas a el artefacto de movimiento introducido por la respiración y que claramente no forman parte del parénquima pulmonar.

Este efecto se puede observar para la totalidad del tejido pulmonar, por lo que sus consecuencias sobre las medidas de calidad de la segmentación mostradas anteriormente deben ser tenidas en cuenta en la discusión de los resultados.

Por otra parte, la segmentación automática es poco agresiva cerca de los bordes del parénquima pulmonar. Esto provoca que no se incluyan bordes difusos pero implica que algunas lesiones situadas en la zona límite no sean incluidas, como se puede apreciar en la tercera imagen de la última fila, incluso cuando se utiliza el proceso de refinado expuesto en la sección 5.1.7. Además, la política menos expansiva de la segmentación automática no evita que en ocasiones se incluyan pequeños pedazos de hueso pertenecientes a las costillas que forman la caja torácica y protegen los pulmones (primera imagen de la última columna). Sin embargo, a pesar de estas limitaciones, la segmentación automática incluye la mayor parte de las lesiones creadas por el *Mycobacterium tuberculosis*, como las indicadas en el interior de los círculos azules de la imagen.

Las vías aéreas son retiradas con éxito como se puede apreciar en la imagen de la segunda columna, donde en la imagen original se aprecia la bifurcación de la traquea en los bronquios, algo que no es recogido por la segmentación.

En cuanto a las imágenes del modelo B, mostradas en la Figura 26, se observa un comportamiento similar al descrito anteriormente. Aunque como se puede observar, la calidad de estas imágenes es menor, y la definición entre tipos de tejido resulta mas incierta. Por ejemplo, en algunas segmentaciones se producen efectos como el que se aprecia en la segmentación de la última fila, donde zonas difusas causan una bajada importante de la calidad de la segmentación. Aunque este efecto es evidente, la cantidad de rodajas donde se produce es pequeña en comparación con el volumen total.

A pesar de esto, las lesiones características de la tuberculosis se recogen de manera exitosa en gran parte de las segmentaciones. Caso aparte son algunas regiones no incluidas, indicadas mediante círculos amarillos, donde resulta complicado diferenciar sin conocimientos específicos anatómicos sobre el animal, si se trata de vasos, lesiones o lesiones unidas a los vasos.

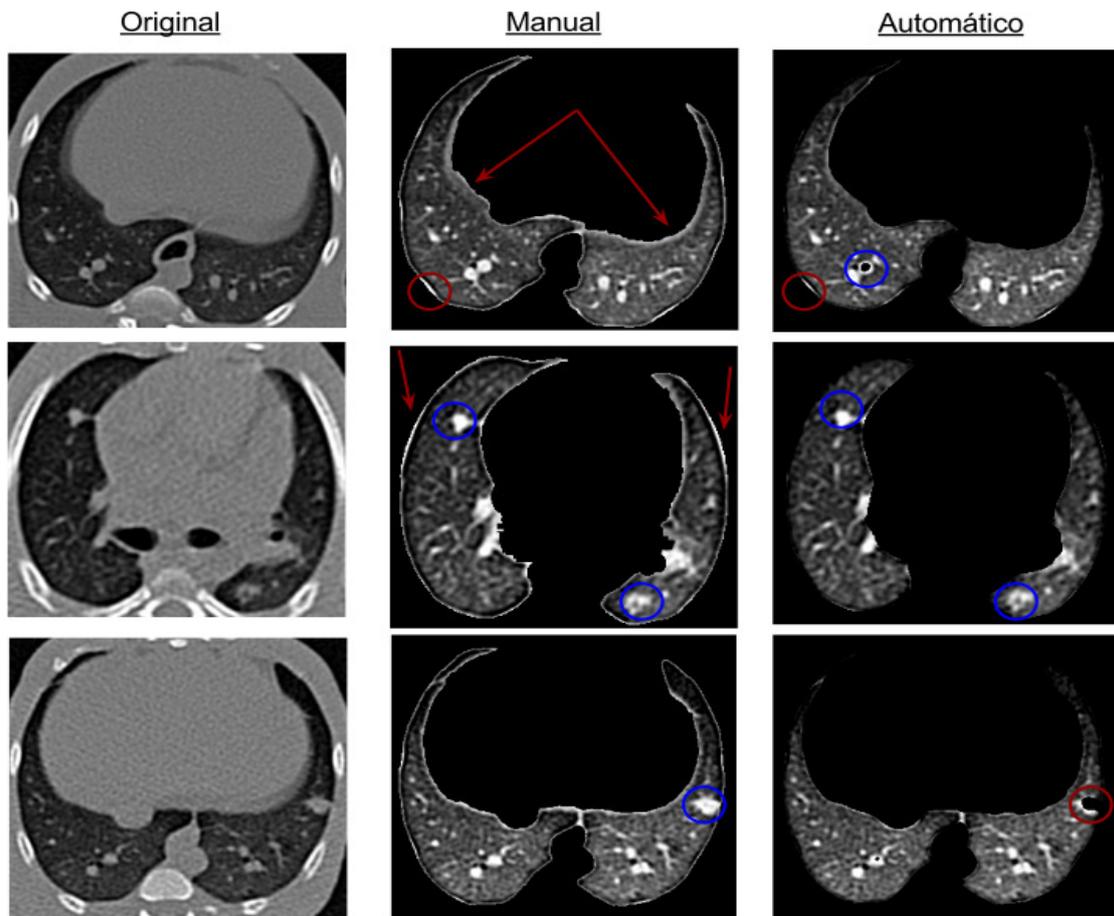


Figura 25: Imágenes representativas de la segmentación manual y automática del pulmón en el modelo A de TB. En la columna de la izquierda se muestran las imágenes originales, en la central las segmentaciones manuales, mientras que en la derecha las segmentaciones obtenidas automáticamente. Los círculos azules denotan lesiones causadas por la enfermedad bien segmentadas, las flechas rojas hacen referencia a zonas difusas mal segmentadas, nódulos no incluidos en la segmentación o huesos huesos erróneamente incluidos como parte del parénquima pulmonar.

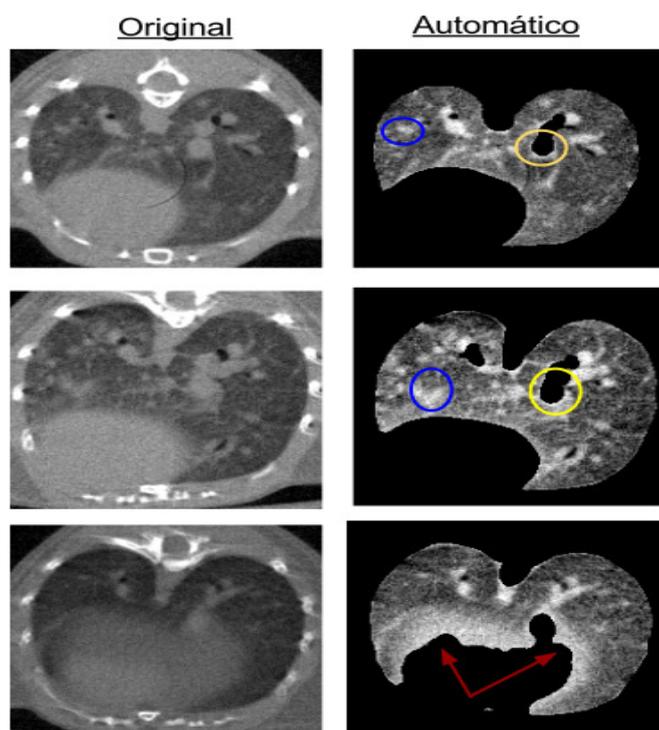


Figura 26: Imágenes representativas para la segmentación automática del pulmón en el modelo B de TB. En la columna de la izquierda se muestran las imágenes originales y en la de la derecha las segmentaciones obtenidas. Los círculos azules denotan lesiones causadas por la enfermedad bien segmentadas, las flechas rojas hacen referencia a zonas difusas mal segmentadas. Finalmente, los círculos amarillos, denotan zonas ambiguas, en estas es complicado establecer si se trata de vasos grandes, lesiones o ambas sin la ayuda de un experto. Es importante apreciar la diferencia en el contraste entre las adquisiciones del modelo A y las imágenes del modelo B, que tiene una apariencia mucho más difusa.

7.2 Resultados de la clasificación de lesiones

Los pasos descritos en la metodología (ver sec.5.2), se llevaron a cabo utilizando diferentes implementaciones. La extracción de características de textura de las regiones se realizó utilizando el software implementado desde cero para este trabajo, descrito en la sección 6.2. Las técnicas de muestreo se realizaron a partir de pequeñas modificaciones de la librería scikit-learn-contrib, así como la implementación del proyecto principal de scikit para el resto de funcionalidades.

7.2.1 Selección de características

Como se especificó en el bloque de Métodos, el proceso de selección de características se realiza combinando el criterio MI-mRMR de filtrado, que establece un *ranking* para formalizar subconjuntos de los indicadores a introducir en un clasificador LDA donde se balancean los datos de entrenamiento.

Por motivos de claridad se continua con este esquema, presentando en primer lugar los resultados del ranking.

Los resultados completos de la clasificación jerarquizada, se muestran el mapa de calor de la figura 27, donde se pueden observar las posiciones que ocupan las características

ordenadas por el criterio de MI-mRMR para cada uno de los niveles de cuantización. En esta matriz, los colores más oscuros se utilizan para las posiciones más bajas en el ranking (características menos relevantes).

El mapa presenta una distribución a nivel local bastante homogénea, lo que indica que las mismas características ocupan posiciones de relevancia semejantes. De la misma forma se pueden observar columnas enteras con colores muy similares o iguales lo que indica que hay características más invariantes al nivel de cuantización como puede ser el caso de la variable $F26$ (suma de cuadrados) que siempre ocupa la última posición.

A partir de la clasificación establecida para cada nivel de gris, se realizó una validación cruzada de 30 particiones para cada subconjunto de características. Los resultados obtenidos para este proceso se resumen en la figura 28 y aparecen pormenorizados en el apéndice A.

La figura 28 muestra el valor medio de la medida F1 (ver sec.5.1.8) obtenido por un clasificador LDA en el proceso de validación cruzada. En este gráfico, se representan los seis niveles de cuantización a diferentes tonalidades.

En esta gráfica se puede observar, valoraciones aparte sobre la calidad de la clasificación en sí misma, como de manera general, existe una marcada tendencia al aumento de la precisión a medida que se aumenta el número de características empleadas.

Como se puede apreciar, cuando se utiliza únicamente un indicador, los resultados son muy pobres. En el mejor de los casos (empleando 256 niveles de gris), apenas se alcanza un 35%, valor que se reduce hasta un 0.05 y 0.17 para 32 y 8 niveles de gris. Sin embargo, al emplear las diez características de textura más informativas se alcanzan valores por encima del 70% cuando se utilizan 256 niveles de gris, aún cuando se usa un clasificador no optimizado.

Este comportamiento es claramente visible, para algunas curvas donde se detiene de forma significativa la precisión obtenida cuando se utilizan las cuantizaciones que aportan mayor nitidez a las imágenes de CT (64, 128, 256). Este fenómeno es especialmente notable en el caso de la cuantización a 256 niveles de gris, donde el proceso de aprendizaje parece converger; convergencia o disminución drástica del aumento de la precisión que también se observa cuando se utilizan subconjuntos con más características en los casos con 128, 62, 32 e incluso 16 niveles de gris.

Resulta evidente, que las características de textura calculadas empleando 256 niveles de gris, aportan mayor información y hacen que las medidas de calidad obtenidas en este caso sean siempre mayores, sea cual sea el número de características usadas. Esto produce una diferencia casi constante entre la curva de 256 niveles de gris y la curva de 128.

Estas diferencias entre curvas, proporcionan una idea inmediata de que niveles de cuantización proporcionan mayor información y en qué magnitud. Así, por ejemplo, se aprecia una diferencia entre la curva de 128 y 64, considerable aunque mucho menor que la descrita anteriormente. Sin embargo, estas diferencias constantes desaparecen cuando se utilizan niveles de cuantización menores.

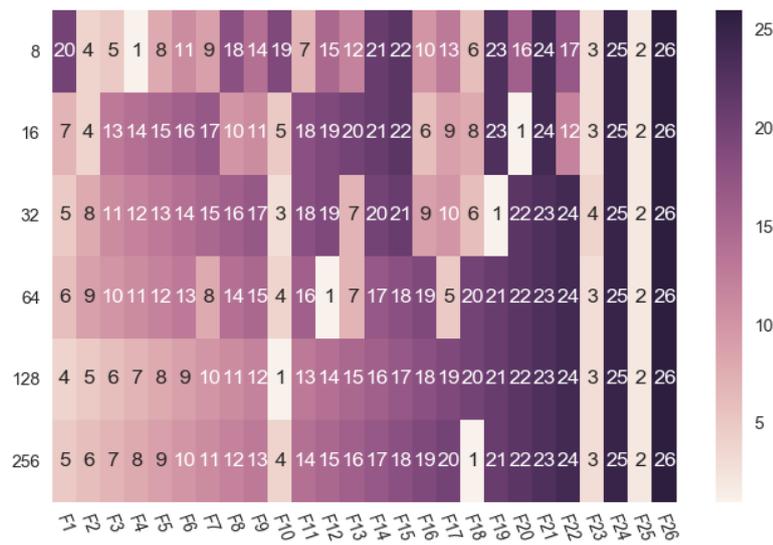


Figura 27: La matriz muestra la posición que obtuvo en el ranking por el criterio mRMR, cada una de las características de textura (9), en cada uno de los niveles de cuantización.

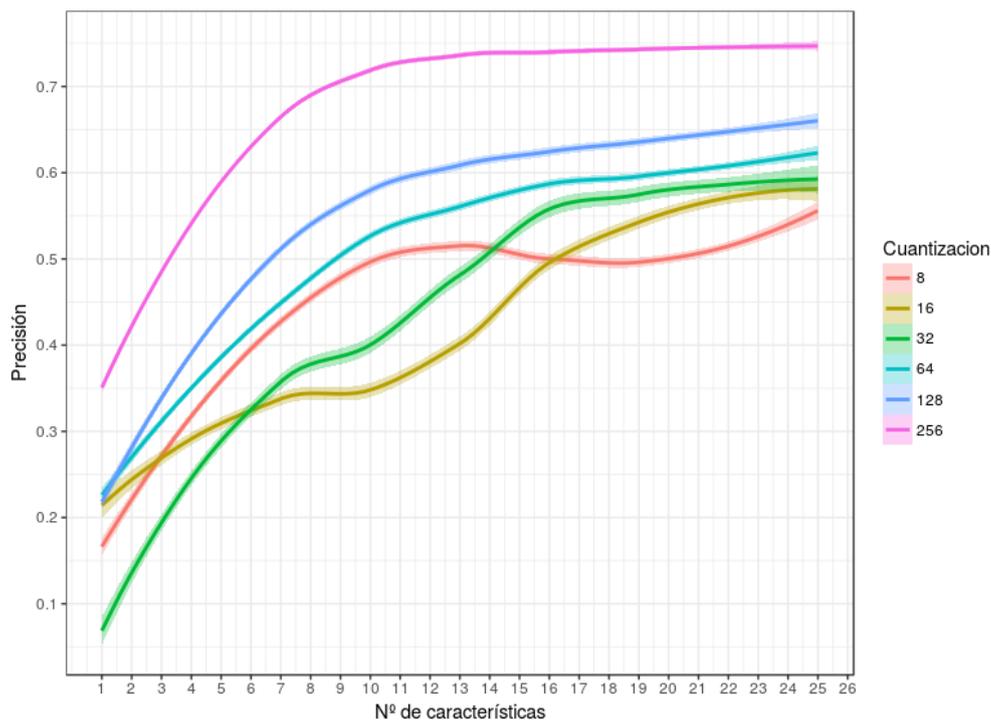


Figura 28: Precisión obtenida con un clasificador simple (*Lineal Discriminant Analysis*, LDA) [120]) para los seis niveles de cuantización definidos utilizando las características ordenadas mediante mRMR. La parte sombreada representa el intervalo de confianza (95%) de la medida sobre 30 particiones.

La tabla 4 presenta cuantitativamente las diferencias significativas entre los resultados de clasificación cuando se emplean cuantizaciones y subconjuntos de características

diferentes, previamente identificadas en un análisis visual.

Para esta evaluación se ha utilizado el test de análisis de varianza *ANalysis Of VAriance* para tres factores: el nivel de cuantización, el número de características y el conjunto de datos utilizados en la validación, (ANOVA). El resultado del test muestra que efectivamente existen diferencias significativas ($p < 0.05$) entre las precisiones logradas para los niveles de cuantización y las diferentes características usadas. Además, se puede decir que no hay evidencia de las particiones utilizadas en la validación introduzcan diferencias en el valor F1 obtenido.

Para examinar que niveles de cuantización introducen mayores o menores diferencias, el ANOVA se acompaña de las comparaciones entre los niveles de cuantización por pares teniendo en cuenta la precisión para cada subconjunto de características. Estos resultados (tabla 5) se obtienen aplicando el test de diferencias significativas de *Tukey*, donde se puede observar que efectivamente existe diferencia en los resultados obtenidos entre pares para todos los casos excepto para 16 y 32 niveles de cuantización.

Dependencia	F	Pr(>F)
Nivel Cuant.	939.53	$< 2 \times 10^{-16}$
Num. Caract.	6876.40	$< 2 \times 10^{-16}$
Partición	0.10	0.75

Tabla 4: Resultado de un ANOVA (*ANalysis Of VAriance*) de tres factores: el nivel de cuantificación utilizado, el número de características utilizadas y la partición de datos. Se muestra el valor para el estadístico *F*.

Comparación	Diferencia medias	95% Intervalo confianza	p-valor
8/16	-0.02	[-0.04,-0.01]	$< 2 \times 10^{-7}$
8/32	-0.01	[-0.02,0.00]	0.43
8/64	0.06	[0.05,0.07]	$< 2 \times 10^{-8}$
8/128	0.10	[0.09,0.11]	$< 2 \times 10^{-8}$
8/256	0.23	[0.22,0.24]	$< 2 \times 10^{-8}$
16/32	0.02	[0.00,0.03]	0.004
16/64	0.08	[0.07,0.10]	$< 2 \times 10^{-8}$
16/128	0.13	[0.11,0.14]	$< 2 \times 10^{-8}$
16/256	0.25	[0.24,0.26]	$< 2 \times 10^{-8}$
32/64	0.07	[0.05,0.08]	$< 2 \times 10^{-8}$
32/128	0.11	[0.10,0.12]	$< 2 \times 10^{-8}$
32/256	0.24	[0.22,0.25]	$< 2 \times 10^{-8}$
64/128	0.04	[0.03,0.06]	$< 2 \times 10^{-8}$
64/256	0.17	[0.16,0.18]	$< 2 \times 10^{-8}$
128/256	0.13	[0.11,0.14]	$< 2 \times 10^{-8}$

Tabla 5: Resultados del test de *Tukey* para evaluar la diferencia entre entre niveles de cuantización por pares.

A partir de los resultados anteriores, se establece que el subconjunto de características a emplear a efectos comparativos en la optimización, sea el formado por **10** indicadores de textura para todos los niveles de cuantificación.

Los resultados para este subconjunto se presentan en la tabla 6, mientras que en la figura

29 aparecen los diagramas de cajas que muestran la escasa dispersión presentada por la medida en este caso.

Cuant.	8	16	32	64	128	256
	0.51±0.01	0.36±0.03	0.38±0.02	0.51±0.01	0.59±0.01	0.72±0.01

Tabla 6: Resultados de un clasificador LDA cuando se utilizan 10 características ordenadas mediante el criterio de MI-mRMR (*Mutual Information-minimal Redundancy Maximal Relevance*)5.2.2

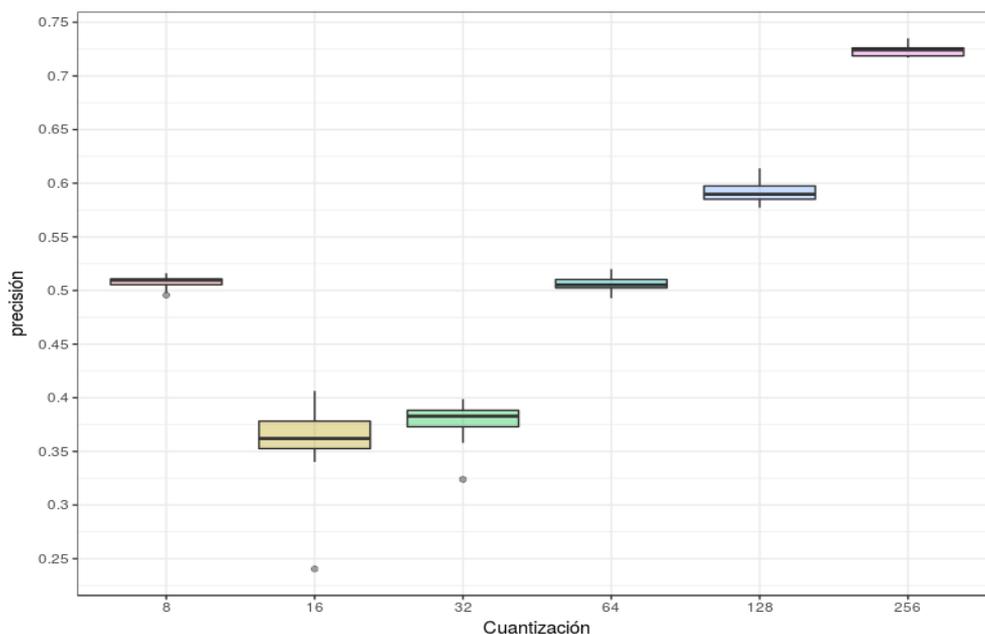


Figura 29: Diagrama de cajas de los resultados de clasificación utilizando 10 características mRMR (inicio de la convergencia para el clasificador) ordenadas mediante el criterio mRMR.

7.2.2 Resultados del proceso de optimización

En las tablas 7 y 8 se exponen los resultados obtenidos para el valor F1 de la SVM optimizada en cada uno de los casos, para una validación cruzada con 30 particiones. De forma correspondiente el diagrama de cajas de la figura 30 presenta estos resultados y su dispersión gráficamente. Además, en la tabla 9, se muestra el resultado del análisis de varianza del valor de F1 obtenido para los factores presentes en el experimento, es decir, el nivel de cuantización, el número de características utilizadas (10 o 26) y el balanceado de los datos.

Estos resultados muestran una clara superioridad de los aprendizajes realizados utilizando datos balanceados, que en media aumentan la calidad de la clasificación un 4%; a la par que, la mejora en el rendimiento cuando se emplean las cuantizaciones de mayor magnitud, como ya sucediera con el clasificador simple en la etapa de selección de características. No en vano para ambos factores, el nivel de significación estadística se encuentra muy por debajo del clásico 0.05 comúnmente empleado (9.43×10^{-13} y 2×10^{-16}).

Por otro lado, el número de características empleado no genera grandes diferencias en las clasificaciones ($p = 0.852$).

La calidad de la clasificación obtenida, llega a alcanzar valores del 88% cuando se utilizan 256 niveles de gris, utilizándose tan solo las 10 características más informativas para el modelo que aprende de los datos balanceados. Incluso cuando se utilizan 64 niveles de cuantización, lo cual facilita el cálculo computacional de las múltiples GLCMs, se obtiene un valor del 86% que podría resultar suficiente para una primera aproximación a los modelos de clasificación.

	N.Cuant Carac.	8		16		32	
		10	26	10	26	10	26
Balanceado	No	0.80 ± 0.02	0.80 ± 0.01	0.81 ± 0.01	0.80 ± 0.01	0.82 ± 0.01	0.81 ± 0.01
	Si	0.84 ± 0.02	0.85 ± 0.02	0.85 ± 0.02	0.85 ± 0.02	0.86 ± 0.01	0.86 ± 0.01

Tabla 7: Resultados de la validación cruzada del proceso de optimización de SVMs de kernel gaussiano cuando se utilizan 8,16 y 32 niveles de gris, el conjunto de las 10 características más informativas y el conjunto total, y cuando se realiza el balanceo de los datos de entrenamiento y cuando no.

	N.Cuant Carac.	64		128		256	
		10	26	10	26	10	26
Balanceado	No	0.82 ± 0.02	0.83 ± 0.01	0.84 ± 0.02	0.84 ± 0.01	0.84 ± 0.02	0.85 ± 0.02
	Si	0.86 ± 0.02	0.86 ± 0.03	0.87 ± 0.02	0.87 ± 0.02	0.88 ± 0.02	0.88 ± 0.02

Tabla 8: Resultados de la validación cruzada del proceso de optimización de SVMs de kernel gaussiano cuando se utilizan 64,128 y 256 niveles de gris, el conjunto de las 10 características más informativas y el conjunto total, y cuando se realiza el balanceo de los datos de entrenamiento y cuando no.

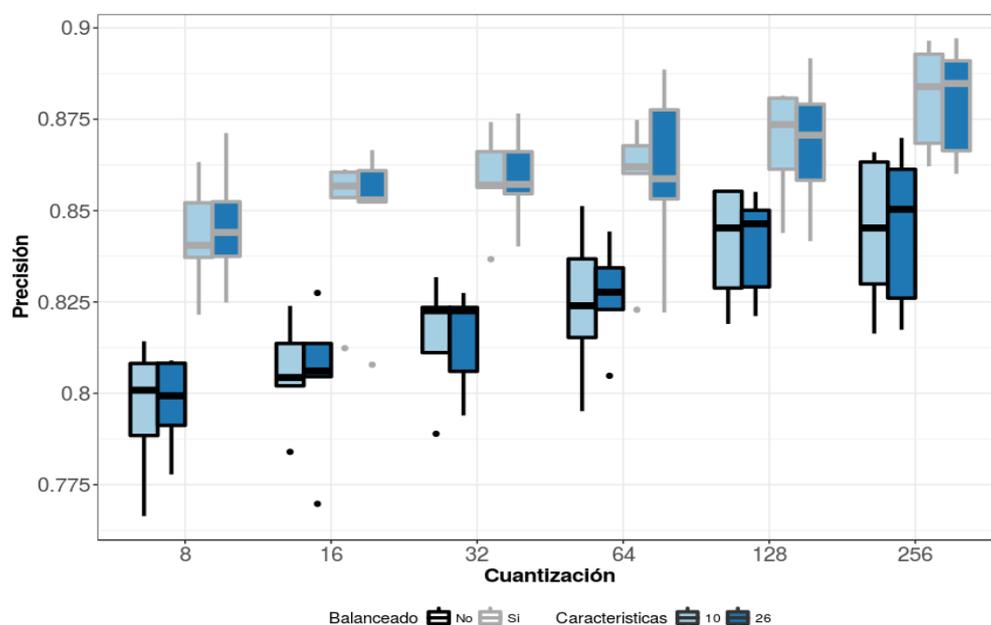


Figura 30: Resultados para las 30 particiones de la SVM óptima para cada caso contemplado en el experimento, esto es, 6 niveles de cuantización (mostrados en el eje horizontal), 10 y el total de las 26 características de textura (representados por el color de la caja) y el balanceo o no de los datos de entrenamiento (representados por el color del borde de la caja).

Rel	F	Pr(>F)
Nivel Cuantización	17.458	$< 2e - 16$
Num. Caract.	0.035	0.852
Balanceado	151.869	$9.43e - 13$

Tabla 9: Resultado de un ANOVA (*ANalysis Of VAriance*) de tres factores: el nivel de cuantificación utilizado, el número de características utilizadas y el balanceo. Se muestra el valor para el estadístico F .

8 Discusión

8.1 Discusión de los resultados de segmentación

Los resultados de la segmentación muestran una enorme discrepancia entre las fuentes de error, con tasas de falsos negativos que rondan el 13% y con falsos positivos de apenas un 2%. Estos datos permiten explicar el origen de la disimilitud entre los volúmenes de pulmón segmentados (7% en media) reflejada en el DSC. Debiéndose ésta fundamentalmente a zonas no incluidas como parte del parénquima pulmonar cuando se realiza la segmentación automática. Este hecho concuerda con la visualización de los resultados cualitativos, donde se evidencia que la configuración de los parámetros resulta en una segmentación del pulmón conservadora.

Hay que tener en cuenta, que la configuración se ha elegido para ser cuidadoso en el proceso de umbralización inicial. Esto es debido a que el análisis de la inclusión de zonas difusas en imágenes fuertemente afectadas por los artefactos de movimiento, resulta complicado cuando se utilizan algoritmos básicos de umbralización que normalmente carecen de mecanismos para conseguir una segmentación precisa. Hay que tener en cuenta que las técnicas de umbralización persiguen una aproximación básica y computacionalmente eficiente por lo que generalmente no emplean demasiada información, como por ejemplo la distribución espacial de los diferentes niveles de gris en una imagen. Por esta razón, es preferible realizar una umbralización conservadora y seguir una estrategia de inclusión de tejidos empleando filtros más sofisticados *a posteriori*, aunque este procedimiento no llegue a incluir algunas regiones menores, evidenciadas en los datos presentados por la tasa de falsos negativos.

Aunque idealmente se pretende extraer todo el parénquima pulmonar (sano, anormal o lesiones), la configuración presentada en este trabajo proporciona, a la vista de los resultados, un compromiso aceptable entre el tejido extraído y la limitación del efecto de sobre-segmentación que supondría incluir ruido en análisis posteriores de cuantificación. Además, como se ha comentado en la sección de resultados cualitativos, hay que tener en cuenta que las segmentaciones realizadas manualmente incluyen cierta variabilidad en la segmentación de regiones similares, debida a la subjetividad del experto, la calidad de la imagen, la resolución del monitor o la herramienta empleado para la segmentación manual de los volúmenes.

El hecho comentado anteriormente justifica razonablemente los resultados obtenidos para las medidas de las distancias entre las superficies segmentadas, tanto para la distancia máxima existente entre dos puntos de ambas, la HD, como para la distancia media entre las mismas, la HDA.

Pueden existir variaciones en la segmentación que no necesariamente implican grandes cambios en la inclusión de lesiones. Por ejemplo, la inclusión de una zona difusa completa

como las mostrada en la figura 26, implican distancias de *Hausdorff* de magnitud relativamente grande. Sin embargo, dado que la variación se da a nivel muy local, la distancia media máxima entre la superficies, la HDA, resulta significativamente menor, lo que implica que la segmentación si estaría extrayendo la mayor parte del parénquima y el error introducido por la región en un proceso de cuantificación podría ser compensado.

Además, para intentar minimizar los casos en los que las diferencias se deben a la falta de sensibilidad para incluir lesiones, se introduce el método de refinado basado en el modelo morfológico esperado.

Es importante apreciar, que este refinamiento se realiza al final del proceso y de nuevo con parámetros conservadores. Esto implica que en un caso extremo de pérdida de sensibilidad en este paso final de inclusión de lesiones, se podría optar por una ejecución semi-automática de la herramienta. En este modo de ejecución, un usuario experto en identificación de lesiones, podría volver a procesar las imágenes segmentadas solo para una versión más agresiva de este último proceso.

Por último, es importante señalar que los resultados de segmentación obtenidos muestran independencia del sujeto y la semana para la cual se realiza la evaluación. Esto sugiere que la herramienta funciona de forma robusta y con independencia del nivel de infección y por tanto, nivel de deterioro pulmonar de un sujeto. Aunque evidentemente se necesitaría de un mayor número de muestras para poder establecer conclusiones estadísticamente significativas.

8.2 **Discusión de los resultados de clasificación**

De forma global, los resultados obtenidos sobre la caracterización de regiones anómalas en el parénquima pulmonar utilizando imágenes de CT se puede analizar desde diferentes perspectivas.

Centrándose únicamente en el *ranking* de características (con el criterio MI-mRMR) y aún sin utilizar los resultados de la clasificación de subconjuntos, se puede comprobar que la afirmación aparecida en la literatura sobre el uso de únicamente indicadores de textura de primer orden (medía, mínimo, desviación típica, etc.), que enuncia la importancia de estas pero su insuficiencia, es cierta. Las características F_1 , F_2 , F_3 y F_4 , que representan el máximo, la media, el mínimo y la desviación típica respectivamente, generalmente ocupan posiciones medias entre los 10 primeros puestos del ranking para todos los niveles de cuantización (ver figura 27), aunque no siempre aparecen las cuatro en estas posiciones. Hay que tener en cuenta que estas características se encuentran correladas, especialmente, en los casos de las cuantizaciones de menor magnitud, donde todos los vóxeles presentan unidades más similares, lo que implica que suelen aparecer el mínimo y la media pero no el máximo y la desviación típica (el mínimo y la media son analizadas antes en el algoritmo MI-mRMR por mero orden de llegada), dado la relación entre estos indicadores.

Las características de primer orden son superadas en el ranking por las características que contienen información espacial y resultan menos redundantes entre las de segundo orden, lo que concuerda con lo esperado teóricamente. Concretamente, las características F_{25} y F_{23} que ponderan, bien en media o bien en varianza, la probabilidad conjunta de co-ocurrencia, aparecen siempre entre las primeras posiciones del *ranking*. Esto remarca que las lesiones están caracterizadas por patrones específicos, distinguibles entre las diferentes clases y no solo por la distribución más o menos ordenada del brillo en la región de interés, algo que viene a indicar la característica F_{10} (correlación a nivel espacial).

Además, como se comentó brevemente en la presentación de los resultados, el mapa de la figura 27 muestra como las características entre niveles de cuantización adyacentes se sitúan en valores similares, intercambiando de forma general una o dos posiciones a lo sumo. Este

hecho, indica que tanto las características, como la técnica de filtrado, realizan los cálculos de forma robusta. Dado que, los indicadores que son informativos para, por ejemplo 128 niveles de gris, de manera general también deberían de serlo para 256 o en caso de no serlo se sustituyen por características fuertemente relacionadas entre si.

Cuando los conjuntos de características ordenadas proporcionados por el filtro, alimentan el clasificador simple con el fin de encontrar que subgrupo de indicadores proporciona información suficiente, se observa la ya comentada superioridad del uso de 256 niveles. Esto indica que los patrones existentes en la imagen se suavizan (cuantizaciones de menor magnitud) degradando demasiado las texturas, lo que hace imposible distinguir entre, por ejemplo, consolidaciones y granulomas. Es decir, se pierde información válida para la diferenciación entre clases.

Esta degradación a niveles de cuantización bajos, explica la forma irregular de las curvas correspondientes a 8,16 y, en menor medida, 32 niveles. En estos casos, la degradación produce una redundancia mucho mayor entre las características, por lo que el método de filtrado sitúa estas en puesto bajos del ranking, como ocurre por ejemplo con la energía o la entropía que a una resolución baja aportan información similar a la suma promedio a la diferencia de entropías. Sin embargo, cuando se incluyen en la clasificación, no solo no mejoran el rendimiento si no que en algunos casos lo empeoran (ver figura 28 para 8 niveles de gris). La introducción de estas características produce que las funciones de distribución que caracterizan a cada una de las clases presentes, se solapan entre si, lo que supone que el clasificador LDA establezca fronteras más difusas e incorrectas entre las clases.

Este efecto es una de las desventajas que presentan los clasificadores no paramétricos como el LDA, aunque en este caso resulta beneficiosa dado que nos permite obtener información valiosa sobre los indicadores. Sin embargo, es razonable realizar experimentos de clasificación cuando se utiliza un modelo optimizado, por esta razón se presentaron los resultados para la clasificación ofrecida por SVMs de *kernel* radial en cada uno de los casos de clasificación. Los resultados al optimizar, no muestran efectos de degradación cuando se utilizan todas las características, presumiblemente por la capacidad de las SVMs para penalizar ciertas variables, sin embargo, queda patente que no aportan nueva información y que se pueden realizar análisis más sencillos con el conjunto de 10 características. Además, la optimización aclara que el uso de datos balanceados en la fase de entrenamiento es altamente recomendable para este problema sea cual sea el nivel de cuantización utilizado.

En cuanto a este último punto, el nivel de cuantificación a utilizar para obtener una calidad aceptable, todo indica que emplear 256 niveles, es lo acertado. Sin embargo, el cálculo de las GLCMs crece de forma exponencial con el nivel de cuantización, de manera que en una aplicación real, donde se necesiten extraer parámetros de miles de regiones esto puede resultar demasiado costoso en términos de tiempo de computación. Por lo que sería recomendable realizar estas estimaciones a la hora de ocuparse de un problema de mayor dimensión que el tratado en este trabajo.

9 Conclusiones

A continuación, se manifiestan las principales aportaciones realizadas con la consecución de este trabajo:

- Metodología para la segmentación completamente automática, eficaz y robusta de los pulmones en diferentes modelos animales.
- Distribución de la herramienta de *software* implementada para la segmentación de pulmones.
- Metodología para la clasificación de regiones anómalas en el parénquima pulmonar, con capacidad para identificar indicadores de textura informativos y clasificar utilizando conjuntos de datos de naturaleza desbalanceada.
- Distribución de la herramienta para el computo de las características de textura derivadas de la matriz GLCM a partir de imágenes volumétricas.

Es importante apreciar, que las aportaciones resumidas en los puntos anteriores, tienen su nicho de aplicación en un entorno muy interdisciplinar, dentro del cual, diferentes técnicas del campo de la inteligencia artificial proporcionan herramientas al investigador biomédico. La aplicación de las mismas permitirá una mejor interpretación de la Tuberculosis y con ello, facilitará la aparición de nuevos tratamientos y protocolos que posibiliten la futura erradicación de la patología.

Apéndice A: Tablas de resultados para la selección de características

Cuant.	1	2	3	4	5
8	0.16±0.04	0.19±0.04	0.35±0.01	0.34±0.01	0.33±0.01
16	0.18±0.04	0.19±0.03	0.35±0.01	0.33±0.01	0.34±0.02
32	0.04±0.00	0.18±0.04	0.23±0.08	0.34±0.01	0.33±0.01
64	0.24±0.00	0.27±0.01	0.34±0.01	0.32±0.01	0.36±0.01
128	0.22±0.08	0.24±0.03	0.38±0.01	0.44±0.01	0.43±0.01
256	0.31±0.01	0.41±0.01	0.55±0.01	0.55±0.01	0.62±0.01

Tabla 10: Precisión obtenida por un clasificador LDA, durante el proceso de selección de características, sobre 30 particiones de datos, cuando se utilizan entre 1 y 5 características ordenadas mediante mRMR

Cuant.	6	7	8	9	10
8	0.34±0.01	0.35±0.02	0.52±0.01	0.51±0.01	0.51±0.01
16	0.34±0.03	0.34±0.03	0.36±0.03	0.36±0.03	0.36±0.03
32	0.37±0.02	0.39±0.01	0.38±0.01	0.38±0.01	0.38±0.02
64	0.36±0.02	0.51±0.01	0.52±0.01	0.51±0.01	0.51±0.01
128	0.43±0.01	0.48±0.00	0.59±0.01	0.59±0.01	0.59±0.01
256	0.64±0.01	0.64±0.00	0.69±0.01	0.72±0.01	0.72±0.01

Tabla 11: Precisión obtenida por un clasificador LDA, durante el proceso de selección de características, sobre 30 particiones de datos, cuando se utilizan entre 6 y 10 características ordenadas mediante mRMR

Cuant.	11	12	13	14	15
8	0.51±0.01	0.51±0.01	0.51±0.01	0.50±0.01	0.50±0.01
16	0.36±0.04	0.35±0.04	0.35±0.04	0.35±0.04	0.54±0.01
32	0.37±0.02	0.36±0.02	0.56±0.01	0.56±0.01	0.56±0.01
64	0.50±0.01	0.57±0.01	0.58±0.01	0.57±0.01	0.57±0.01
128	0.59±0.01	0.59±0.01	0.60±0.01	0.60±0.01	0.61±0.01
256	0.73±0.01	0.73±0.00	0.73±0.00	0.73±0.01	0.73±0.01

Tabla 12: Precisión obtenida por un clasificador LDA, durante el proceso de selección de características, sobre 30 particiones de datos, cuando se utilizan entre 11 y 15 características ordenadas mediante mRMR

Cuant.	16	17	18	19	20
8	0.50±0.01	0.50±0.01	0.50±0.01	0.50±0.01	0.50±0.01
16	0.54±0.01	0.54±0.01	0.54±0.01	0.54±0.01	0.54±0.01
32	0.56±0.01	0.56±0.01	0.56±0.01	0.56±0.01	0.56±0.00
64	0.57±0.01	0.60±0.01	0.60±0.01	0.60±0.01	0.60±0.01
128	0.64±0.01	0.64±0.01	0.64±0.01	0.64±0.01	0.64±0.01
256	0.74±0.01	0.75±0.01	0.75±0.01	0.74±0.01	0.74±0.01

Tabla 13: Precisión obtenida por un clasificador LDA, para la selección de características, sobre 30 particiones, cuando se utilizan entre 16 y 20 características ordenadas por mRMR

Cuant.	21	22	23	24	25
8	0.50±0.01	0.50±0.01	0.50±0.01	0.50±0.01	0.61±0.01
16	0.54±0.01	0.55±0.01	0.54±0.01	0.54±0.01	0.65±0.00
32	0.57±0.01	0.57±0.01	0.56±0.01	0.56±0.01	0.65±0.01
64	0.60±0.01	0.59±0.01	0.59±0.01	0.59±0.01	0.67±0.01
128	0.64±0.01	0.64±0.01	0.63±0.01	0.63±0.01	0.70±0.01
256	0.74±0.01	0.74±0.01	0.74±0.01	0.74±0.01	0.76±0.00

Tabla 14: Precisión obtenida por un clasificador LDA, durante el proceso de selección de características, sobre 30 particiones de datos, cuando se utilizan entre 20 y 25 características ordenadas mediante mRMR

Ranking de características

Rank.	8	16	32	64	128	256
1	f4	f20	f19	f12	f10	f18
2	f25	f25	f25	f25	f25	f25
3	f23	f23	f10	f23	f23	f23
4	f2	f2	f23	f10	f1	f10
5	f3	f10	f1	f17	f2	f1
6	f18	f16	f18	f1	f3	f2
7	f11	f1	f13	f13	f4	f3
8	f5	f18	f2	f7	f5	f4
9	f7	f17	f16	f2	f6	f5
10	f16	f8	f17	f3	f7	f6
11	f6	f9	f3	f4	f8	f7
12	f13	f22	f4	f5	f9	f8
13	f17	f3	f5	f6	f11	f9
14	f9	f4	f6	f8	f12	f11
15	f12	f5	f7	f9	f13	f12
16	f20	f6	f8	f11	f14	f13
17	f22	f7	f9	f14	f15	f14
18	f8	f11	f11	f15	f16	f15
19	f10	f12	f12	f16	f17	f16
20	f1	f13	f14	f18	f18	f17
21	f14	f14	f15	f19	f19	f19
22	f15	f15	f20	f20	f20	f20
23	f19	f19	f21	f21	f21	f21
24	f21	f21	f22	f22	f22	f22
25	f24	f24	f24	f24	f24	f24
26	f26	f26	f26	f26	f26	f26

Tabla 15: Ranking de selección de características para cada nivel de cuantización

Apéndice B: Definición de características de textura y expresiones utilizadas para el cálculo

1. Máximo:

$$f_1 = \max I(i, j)$$

2. Media:

$$f_2 = \frac{1}{N+M} \sum_i^N \sum_j^M I(i, j)$$

3. Mínimo:

$$f_3 = \min I(i, j)$$

4. Desviación típica:

$$f_4 = \frac{1}{N+M} (I(i, j) - f_2)^{\frac{1}{2}}$$

5. Autocorrelación:

$$f_5 = \sum_i \sum_j (ij) p(i, j)^2$$

6. *Cluster Prominance*:

$$f_6 = \sum_i \sum_j (i + j - \mu_x - \mu_y)^4 p(i, j)$$

7. *Cluster Shade*:

$$f_7 = \sum_i \sum_j (i + j - \mu_x - \mu_y)^3 p(i, j)$$

8. Contraste:

$$f_8 = \left| \sum_i \sum_j \right|^2 p(i, j)$$

9. Correlación 1:

$$f_9 = \sum_i \sum_j \frac{(i - \mu_x)(j - \mu_y) p(i, j)}{\sigma_x \sigma_y}$$

10. Correlación 2:

$$f_{10} = \frac{\sum_{i=1} \sum_{j=1} (ij) p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y}$$

11. Diferencia de entropía:

$$f_{11} = - \sum_{i=0}^{L-1} p_{x-y}(i) \log(p_{x-y}(i))$$

12. Diferencia de varianza:

$$f_{12} = \sum_{i=0}^{L-1} i^2 p_{x-y}(i)$$

13. Disimilitud:

$$f_{13} = \left| \sum_i \sum_j p(i, j) \right|$$

14. Energy:

$$f_{14} = \sum_i \sum_j p(i, j)^2$$

15. Entropía:

$$f_{15} = - \sum_{i=1}^L \sum_{j=1}^L p(i, j) \log(p(i, j))$$

16. Homogeneidad 1:

$$f_{16} = \sum_i \sum_j \frac{p(i, j)}{1 + |i + j|}$$

17. Homogeneidad 2:

$$f_{17} = \sum_i \sum_j \frac{p(i, j)}{1 + |i + j|^2}$$

18. Medida de Información de la Correlación 1:

$$f_{18} = \frac{f_9 - HXY1}{\max(HX, HY)}$$

19. Medida de Información de la Correlación 2:

$$f_{19} = [1 - \exp(-2(HXY2 - f_9))]^{1/2}$$

20. Diferencia inversa normalizada:

$$f_{20} = \sum_{i=1}^L \sum_{j=1}^L \frac{1}{1 + |i - j|^2 / L} p(i, j)$$

21. Momento de diferencia inversa normalizado:

$$f_{21} = \sum_{i=1}^L \sum_{j=1}^L \frac{1}{1 + (i - j)^2 / L} p(i, j)$$

22. Probabilidad Máxima:

$$f_{22} = \max_{i, j} p(i, j)$$

23. Suma Promedio:

$$f_{23} = \sum_{i=2}^{2L} i p_{x+y}(i)$$

24. Suma de Entropía:

$$f_{24} = - \sum_{i=2}^{2L} p_{x+y}(i) \log(p_{x+y}(i))$$

25. Suma de Varianzas:

$$f_{25} = \sum_{i=2}^{2L} (i - f_8)^2 p_{x+y}(i)$$

26. Suma de Cuadrados:

$$f_{26} = \sum_i \sum_j (i - v)^2 p(i, j)$$

Expresiones utilizadas en el cálculo

- L : Nivel de cuantización
- $p(i, j)$: entrada (i, j) de la matriz de Co-ocurrencia
- $v = \frac{1}{L} \sum_i \sum_j p(i, j)$
- $p_x(i) = \sum_{j=1}^L p(i, j)$
- $p_y(j) = \sum_{i=1}^L p(i, j)$
- $p_{x+y}(k) = \sum_{i=1, i+j=k}^L \sum_{j=1}^L p(i, j), \quad k = 2, 3, \dots, 2L$
- $p_{x-y}(k) = \sum_{i=1, |i-j|=k}^L \sum_{j=1}^L p(i, j), \quad k = 0, 1, \dots, L-1$
- $HX = - \sum_i p_x(i) \log(p_x(i))$
- $HY = - \sum_j p_y(j) \log(p_y(j))$
- $HXY = - \sum_i \sum_j p(i, j) \log(p(i, j))$
- $HXY1 = - \sum_i \sum_j p(i, j) \log(p_x(i) p_y(j))$
- $HXY2 = - \sum_i \sum_j p_x(i) p_y(j) \log(p_x(i) p_y(j))$

Bibliografía

- [1] Burrill, J. *et al.* Tuberculosis: A Radiologic Review. *Radiographics* **27**, 1255–1273 (2007). DOI 10.1148/rg.275065176.
- [2] Coleman, M. T. *et al.* PET/CT imaging reveals a therapeutic response to oxazolidinones in macaques and humans with tuberculosis. *Science Translational Medicine* **265**, 265ra167–265ra167 (2014). DOI 10.1126/scitranslmed.3009500.
- [3] Trouiller, P. *et al.* Drug development for neglected diseases: a deficient market and a public-health policy failure. *The Lancet* **359**, 2188–2194 (2002). URL <http://www.sciencedirect.com/science/article/pii/S0140673602090967>. DOI [http://dx.doi.org/10.1016/S0140-6736\(02\)09096-7](http://dx.doi.org/10.1016/S0140-6736(02)09096-7).
- [4] World Health Organization and others. Global tuberculosis report 2016. Tech. Rep. (2016).
- [5] Barry 3rd, C. E. *et al.* The spectrum of latent tuberculosis: rethinking the goals of prophylaxis. *Nature reviews. Microbiology* **7**, 845–855 (2009). DOI 10.1038/nrmicro2236.
- [6] Dye, C. & Williams, B. G. Eliminating human tuberculosis in the twenty-first century. *Journal of the Royal Society Interface* **5**, 653–662 (2008). DOI 10.1098/rsif.2007.1138.
- [7] World Health Organization. Global tuberculosis report 2015. Tech. Rep. (2015).
- [8] Pai, M. *et al.* Tuberculosis. *Nature Reviews Disease Primers* **2** (2016). DOI dx.doi.org/10.1038/nrdp.2016.76.
- [9] Nachiappan, A. C. *et al.* Pulmonary Tuberculosis: Role of Radiology in Diagnosis and Management. *RadioGraphics* **37**, 52–72 (2017). DOI 10.1148/rg.2017160032.
- [10] Galbán, C. J. *et al.* Computed tomography-based biomarker provides unique signature for diagnosis of COPD phenotypes and disease progression. *Nature medicine* **18**, 1711–1715 (2012). DOI 10.1038/nm.2971.
- [11] Chen, R. Y. *et al.* PET/CT imaging correlates with treatment outcome in patients with multidrug-resistant tuberculosis. *Science translational medicine* **6**, 166–265 (2014). DOI 10.1126/scitranslmed.3009501.
- [12] Sharpe, S. A. *et al.* Determination of lesion volume by MRI and stereology in a macaque model of tuberculosis. *Tuberculosis* **89**, 405–416 (2009). DOI 10.1016/j.tube.2009.09.002.
- [13] Udupa, J. K. *et al.* A framework for evaluating image segmentation algorithms. *Computerized Medical Imaging and Graphics* **30**, 75–87 (2006). DOI 10.1016/j.compmedimag.2005.12.001.
- [14] Rubin, G. D. Data explosion: the challenge of multidetector-row CT. *European journal of radiology* **36**, 74–80 (2000). DOI 10.1016/S0720-048X(00)00270-9.
- [15] Andreu, N., Elkington, P. T. & Wiles, S. Molecular imaging in TB: from the bench to the clinic. In *Understanding Tuberculosis-Global Experiences and Innovative Approaches to the Diagnosis* (InTech, 2012). URL <http://dx.doi.org/10.5772/59479%0A>.

- [16] Panicker, R. O. o., Soman, B., Saini, G. & Rajan, J. A Review of Automatic Methods Based on Image Processing Techniques for Tuberculosis Detection from Microscopic Sputum Smear Images. *Journal of medical systems* **40**, 1–13 (2016). DOI 10.1007/s10916-015-0388-y.
- [17] Kak, A. C. & Slaney, M. *Principles of computerized tomographic imaging* (SIAM, 2001).
- [18] Schneider, U., Pedroni, E. & Lomax, A. The calibration of CT Hounsfield units for radiotherapy treatment planning. *Physics in Medicine & Biology* **41**, 111 (1996). URL <http://stacks.iop.org/0031-9155/41/i=1/a=009>.
- [19] Knechel, N. A. Tuberculosis: Pathophysiology, Clinical Features, and Diagnosis. *Critical Care Nurse* **29**, 34–43 (2009). URL <http://ccn.aacnjournals.org/content/29/2/34.short>. DOI 10.4037/ccn2009968.
- [20] Pal, N. R. & Pal, S. K. A review on image segmentation techniques. *Pattern Recognition* **26**, 1277–1294 (1993). URL <http://www.sciencedirect.com/science/article/pii/003132039390135J>. DOI [http://dx.doi.org/10.1016/0031-3203\(93\)90135-J](http://dx.doi.org/10.1016/0031-3203(93)90135-J).
- [21] Pham, D. L., Xu, C. & Prince, J. L. Current Methods in Medical Image Segmentation. *Annual Review of Biomedical Engineering* **2**, 315–337 (2000). URL <https://doi.org/10.1146/annurev.bioeng.2.1.315>. DOI 10.1146/annurev.bioeng.2.1.315.
- [22] McInerney, T. & Terzopoulos, D. Deformable models in medical image analysis: a survey. *Medical Image Analysis* **1**, 91–108 (1996). URL <http://www.sciencedirect.com/science/article/pii/S1361841596800077>. DOI [http://dx.doi.org/10.1016/S1361-8415\(96\)80007-7](http://dx.doi.org/10.1016/S1361-8415(96)80007-7).
- [23] Cremers, D., Rousson, M. & Deriche, R. A Review of Statistical Approaches to Level Set Segmentation: Integrating Color, Texture, Motion and Shape. *International Journal of Computer Vision* **72**, 195–215 (2007). URL <https://doi.org/10.1007/s11263-006-8711-1>. DOI 10.1007/s11263-006-8711-1.
- [24] Bezdek, J. C., Hall, L. O. & Clarke, L. P. Review of MR image segmentation techniques using pattern recognition. *Medical Physics* **20**, 1033–1048 (1993). URL <http://dx.doi.org/10.1118/1.597000>. DOI 10.1118/1.597000.
- [25] Noble, J. A. & Boukerroui, D. Ultrasound image segmentation: a survey. *IEEE Transactions on Medical Imaging* **25**, 987–1010 (2006). DOI 10.1109/TMI.2006.877092.
- [26] Zhou, J. *et al.* Automated compromised right lung segmentation method using a robust atlas-based active volume model with sparse shape composition prior in CT. *Computerized Medical Imaging and Graphics* (2015). DOI 10.1016/j.compmedimag.2015.07.003.
- [27] Birkbeck, N. *et al.* Lung Segmentation from CT with Severe Pathologies Using Anatomical Constraints. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 804–811 (2014).
- [28] El-Baz, A. *et al.* Computer-aided diagnosis systems for lung cancer: Challenges and methodologies. *International Journal of Biomedical Imaging* **46** (2013). DOI 10.1155/2013/942353.

- [29] Van Ginneken, B., Katsuragawa, S., Ter Haar Romeny, B. M., Doi, K. & Viergever, M. A. Automatic detection of abnormalities in chest radiographs using local texture analysis. *IEEE Transactions on Medical Imaging* **21**, 139–149 (2002). DOI 10.1109/42.993132.
- [30] Lo, P. *et al.* Extraction of Airways From CT (EXACT'09). *IEEE Transactions on Medical Imaging* **31**, 2093–2107 (2012). DOI 10.1109/TMI.2012.2209674.
- [31] Mansoor, A. *et al.* A Generic Approach to Pathological Lung Segmentation. *IEEE Transactions On Medical Imaging* **33**, 2293–2310 (2014). DOI 10.1109/TMI.2014.2337057.
- [32] Mansoor, A. *et al.* Segmentation and Image Analysis of Abnormal Lungs at CT: Current Approaches, Challenges, and Future Trends 1. *RadioGraphics* **35**, 1056–1076 (2015). DOI 10.1148/rg.2015140232.
- [33] Hu, S., Hoffman, E. A. & Reinhardt, J. M. Automatic Lung Segmentation for Accurate Quantitation of Volumetric X-Ray CT Images. *IEEE Transactions on Medical Imaging* **20**, 490–498 (2001). DOI 10.1109/42.929615.
- [34] Roy, A. S., Armato III, S. G., Wilson, A. & Drukker, K. Automated detection of lung nodules in CT scans: false-positive reduction with the radial-gradient index. *Medical physics* **33**, 1133–1140 (2006). DOI 10.1118/1.2178450.
- [35] Hojjatoleslami, S. A. & Kittler, J. Region growing: A new approach. *IEEE Transactions on Image Processing* (1998). DOI 10.1109/83.701170.
- [36] Grady, L. Random Walks for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**, 1768–1783 (2006). DOI 10.1109/TPAMI.2006.233.
- [37] Li, B., Christensen, G. E., Hoffman, E. A., McLennan, G. & Reinhardt, J. M. Establishing a normative atlas of the human lung: Intersubject warping and registration of volumetric CT images. *Academic Radiology* **10**, 255–265 (2003). DOI 10.1016/S1076-6332(03)80099-5.
- [38] Artaechevarria, X. *et al.* Longitudinal study of a mouse model of chronic pulmonary inflammation using breath hold gated micro-CT. *Eur Radiol* (2010). DOI 10.1007/s00330-010-1853-0.
- [39] LeCun, Y., Bengio, Y. & Hinton, G. Deep Learning. *Nature* **521**, 436–444 (2015). DOI doi:10.1038/nature14539.
- [40] Yao, J., Dwyer, A., Summers, R. M. & Mollura, D. J. Computer-aided Diagnosis of Pulmonary Infections Using Texture Analysis and Support Vector Machine Classification. *Academic Radiology* (2011). DOI 10.1016/j.acra.2010.11.013.
- [41] Song, Y., Cai, W., Zhou, Y. & Feng, D. D. Feature-based image patch approximation for lung tissue classification. *IEEE Transactions on Medical Imaging* **32**, 797–808 (2013). DOI 10.1109/TMI.2013.2241448.
- [42] Soliman, A. *et al.* Accurate Lungs Segmentation on CT Chest Images by Adaptive Appearance-Guided Shape Modeling. *IEEE Transactions on Medical Imaging* **PP**, 263–276 (2016). DOI 10.1109/TMI.2016.2606370.

- [43] Greenspan, H., Ginneken, B. v. & Summers, R. M. Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique. *IEEE Transactions on Medical Imaging* **35**, 1153–1159 (2016). DOI 10.1109/TMI.2016.2553401.
- [44] Messay, T., Hardie, R. C. & Tuinstra, T. R. Segmentation of pulmonary nodules in computed tomography using a regression neural network approach and its application to the Lung Image Database Consortium and Image Database Resource Initiative dataset. *Medical Image Analysis* **22**, 48–62 (2015). DOI 10.1016/j.media.2015.02.002.
- [45] Shen, S., Bui, A. A. T., Cong, J. & Hsu, W. An automated lung segmentation approach using bidirectional chain codes to improve nodule detection accuracy. *Computers in Biology and Medicine* **57**, 139–149 (2015). DOI 10.1016/j.compbiomed.2014.12.008.
- [46] Tao, Y. *et al.* Multi-level ground glass nodule detection and segmentation in CT lung images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention. MICCAI*, 715–723 (2009).
- [47] Farag, A. A., Abd, H. E., Munim, E., Graham, J. H. & Farag, A. A. A Novel Approach for Lung Nodules Segmentation in Chest CT Using Level Sets. *IEEE Transactions on Image Processing* **22**, 5202–5213 (2013). DOI 10.1109/TIP.2013.2282899.
- [48] Sluimer, I., Schilham, A., Prokop, M. & Van Ginneken, B. Computer analysis of computed tomography scans of the lung: A survey. *IEEE Transactions on Medical Imaging* **25**, 385–405 (2006). DOI 10.1109/TMI.2005.862753.
- [49] Vos, T. *et al.* Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet* **386**, 743 (2015).
- [50] Bourke, S. J. Interstitial lung disease: progress and problems. *Postgraduate Medical Journal* **82**, 494–499 (2006). URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2585700/>. DOI 10.1136/pgmj.2006.046417.
- [51] Uppaluri, R., Mitsa, T., Sonka, M., Hoffman, E. A. & McLennan, G. Quantification of pulmonary emphysema from lung computed tomography images. *American journal of respiratory and critical care medicine* **156**, 248–254 (1997). DOI 10.1164/ajrccm.156.1.9606093.
- [52] Prasad, M. N. & Sowmya, A. Multilevel emphysema diagnosis of HRCT lung images in an incremental framework. In *Medical Imaging 2004*, 42–50 (International Society for Optics and Photonics, 2004).
- [53] Kauczor, H.-U. *et al.* Automatic detection and quantification of ground-glass opacities on high-resolution CT using multiple neural networks: comparison with a density mask. *American Journal of Roentgenology* **175**, 1329–1334 (2000). DOI 10.2214/ajr.175.5.1751329.
- [54] Uchiyama, Y. *et al.* Quantitative computerized analysis of diffuse lung disease in high-resolution computed tomography. *Medical Physics* **30**, 2440–2454 (2003). DOI 10.1118/1.1597431.

- [55] Delorme, S., Keller-Reichenbecher, M.-A., Zuna, I., Schlegel, W. & Van Kaick, G. Usual interstitial pneumonia: quantitative assessment of high-resolution computed tomography findings by computer-assisted texture-based image analysis. *Investigative radiology* **32**, 566–574 (1997).
- [56] Shamsheeva, A. & Sowmya, A. Tuning kernel function parameters of support vector machines for segmentation of lung disease patterns in high-resolution computed tomography images. In *Proc. of SPIE Vol.*, vol. 5370, 1549 (2004).
- [57] Zheng, B., Leader, J. K., Fuhrman, C. R., Scirba, F. C. & Gur, D. Automated detection and classification of interstitial lung diseases from low-dose CT images. In *Proceedings of SPIE*, vol. 5370, 849–856 (2004).
- [58] Xu, Z. *et al.* Computer-aided detection and quantification of cavitary tuberculosis from CT scans. *Medical Physics* **48**, 113701 (2013). DOI 10.1118/1.4824979.
- [59] Yan, D., Wang, P., Knudsen, B. S., Linden, M. & Randolph, T. W. Statistical Methods for Tissue Array Images – Algorithmic Scoring and Co-training. *The annals of applied statistics* **6**, 1280–1305 (2012). URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3441061/>.
- [60] Pang, H. *et al.* Quantitative evaluation methods of skin condition based on texture feature parameters. *Saudi Journal of Biological Sciences* **24**, 514–518 (2017). URL <http://dx.doi.org/10.1016/j.sjbs.2017.01.021>. DOI 10.1016/j.sjbs.2017.01.021.
- [61] Lambin, P. *et al.* HHS Public Access. *Proceedings of SPIE—the International Society for Optical Engineering* **73**, 389–400 (2015). URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4749027&tool=pmcentrez&rendertype=abstract>. DOI 10.1016/j.ejca.2011.11.036.Radiomics.
- [62] Yip, S. S. F. & Aerts, H. J. W. L. Applications and limitations of radiomics. *Physics in medicine and biology* **61**, 150–66 (2016). URL <http://stacks.iop.org/0031-9155/61/i=13/a=R150?key=crossref.134478778713970aff90f16abe110608%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/27269645%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4927328>. DOI 10.1088/0031-9155/61/13/R150.
- [63] Ginneken, B. V., Romeny, B. M. T. H. & Viergever, M. A. Computer-aided diagnosis in chest radiography: a survey. *IEEE Transactions on Medical Imaging* **20**, 1228–1241 (2001). DOI 10.1109/42.974918.
- [64] Kassner, A. & Thornhill, R. E. Texture Analysis: A Review of Neurologic MR Imaging Applications. *American Journal of Neuroradiology* **31**, 809–816 (2010). URL <http://www.ajnr.org/content/31/5/809>. DOI 10.3174/ajnr.A2061.
- [65] Prati, R. C., Batista, G., Monard, M. C. & others. Class imbalances versus class overlapping: an analysis of a learning system behavior. In *MICAI*, vol. 4, 312–321 (Springer, 2004).
- [66] Ganesan, K. *et al.* Computer-Aided Breast Cancer Detection Using Mammograms: A Review. *IEEE Reviews in Biomedical Engineering* **6**, 77–98 (2013). DOI 10.1109/RBME.2012.2232289.

- [67] Ramesh, A. N., Kambhampati, C., Monson, J. R. T. & Drew, P. J. Artificial intelligence in medicine. *Annals of the Royal College of Surgeons of England* **86**, 334–338 (2004). URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1964229/>. DOI 10.1308/147870804290.
- [68] He, H. & Garcia, E. A. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering* (2009). DOI 10.1109/TKDE.2008.239.
- [69] Tomek, I. An experiment with the edited nearest-neighbor rule. *IEEE Transactions on systems, Man, and Cybernetics* **SMC-6**, 448–452 (1976). DOI 931915.
- [70] Lloyd, S. P. Least Squares Quantization in PCM. *IEEE Transactions on Information Theory* **28**, 129–137 (1982). DOI 10.1109/TIT.1982.1056489.
- [71] Laurikkala, J. Improving identification of difficult small classes by balancing class distribution. In *Artificial Intelligence in Medicine*, 63–66 (Springer, 2001).
- [72] Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research* **16**, 321–357 (2002). DOI 10.1613/jair.953.
- [73] Keogh, E. & Mueen, A. Curse of Dimensionality. In Sammut, C. & Webb, G. I. (eds.) *Encyclopedia of Machine Learning*, 257–258 (Springer US, Boston, MA, 2010). URL http://dx.doi.org/10.1007/978-0-387-30164-8_192.
- [74] Lu, Y., Cohen, I., Zhou, X. S. & Tian, Q. Feature Selection Using Principal Feature Analysis. In *Proceedings of the 15th ACM International Conference on Multimedia*, 301–304 (2007).
- [75] Guyon, I., Elisseeff, A. & De, A. M. An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research* **3**, 1157–1182 (2003). DOI 10.1016/j.aca.2011.07.027.
- [76] Tourassi, G. D., Frederick, E. D., Markey, M. K. & Floyd, C. E. Application of the mutual information criterion for feature selection in computer-aided diagnosis. *Medical physics* **28**, 2394–402 (2001). URL <http://www.ncbi.nlm.nih.gov/pubmed/11797941>. DOI 10.1118/1.1418724.
- [77] Kohavi, R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In *IJCAI*, March 2001, 1137–1145 (1995).
- [78] Kohavi, R. & John, G. H. Wrappers for feature subset selection. *Artificial Intelligence* **97**, 273–324 (1997). URL <http://linkinghub.elsevier.com/retrieve/pii/S000437029700043X>. DOI 10.1016/S0004-3702(97)00043-X.
- [79] Vergara, J. R. & Estévez, P. A. A review of feature selection methods based on mutual information. *Neural Computing and Applications* 175–186 (2014). DOI 10.1007/s00521-013-1368-0.
- [80] Pudil, P., Novovičová, J. & Kittler, J. Floating search methods in feature selection. *Pattern recognition letters* **15**, 1119–1125 (1994). URL www.sciencedirect.com/science/article/pii/0167865594901279. DOI 10.1016/0167-8655(94)90127-9.

- [81] Sun, Y., Babbs, C. F. & Delp, E. J. A comparison of feature selection methods for the detection of breast cancers in mammograms: adaptive sequential floating search vs. genetic algorithm. In *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, 6532–6535 (IEEE, 2006).
- [82] Wu, W., Zhou, Z., Wu, S. & Zhang, Y. Automatic Liver Segmentation on Volumetric CT Images Using Supervoxel-Based Graph Cuts DOI 10.1155/2016/9093721.
- [83] Barrett, J. F. & Keat, N. Artifacts in CT: Recognition and Avoidance. *RadioGraphics* **24**, 1679–1691 (2004). URL <http://pubs.rsna.org/doi/10.1148/rg.246045065>. DOI 10.1148/rg.246045065.
- [84] Ritchie, C. J. *et al.* Predictive respiratory gating: a new method to reduce motion artifacts on CT scans. *Radiology* **190**, 847–852 (1994). URL <http://dx.doi.org/10.1148/radiology.190.3.8115638>. DOI 10.1148/radiology.190.3.8115638.
- [85] Sawant, A., Zeman, H., Muratore, D., Samant, S. & Dibianca, F. An adaptive median filter algorithm to remove impulse noise in x-ray and CT images and speckle in ultrasound images (1999). URL <http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=980864>.
- [86] Coggins, J. M. & Jain, A. K. A spatial filtering approach to texture analysis. *Pattern Recognition Letters* **3**, 195–203 (1985). URL <http://www.sciencedirect.com/science/article/pii/0167865585900534>. DOI [http://dx.doi.org/10.1016/0167-8655\(85\)90053-4](http://dx.doi.org/10.1016/0167-8655(85)90053-4).
- [87] Yu, Y. & Acton, S. T. Speckle reducing anisotropic diffusion. *IEEE Transactions on Image Processing* **11**, 1260–1270 (2002). DOI 10.1109/TIP.2002.804276.
- [88] Feeman, T. G. *The mathematics of medical imaging* (Springer, 2010).
- [89] Ricón, M. & Bachiller, M. *Visión Artificial* (2009).
- [90] Otsu, N. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics* **9**, 62–66 (1979). DOI 10.1109/TSMC.1979.4310076.
- [91] Lehmann, G. Label object representation and manipulation with ITK. *The Insight Journal* **8** (2007). URL <http://hdl.handle.net/1926/584>.
- [92] Irving, B., Taylor, P. & Todd-Pokropek, A. 3D segmentation of the airway tree using a morphology based method. In *Proc. of Second International Workshop on Pulmonary Image Analysis*, 297–307 (2009).
- [93] Lo, P., Sporning, J. & de Bruijne, M. Multiscale vessel-guided airway tree segmentation. In *Proc. of Second International Workshop on Pulmonary Image Analysis*, 323–332 (2009).
- [94] Lee, J. & Reeves, A. P. Segmentation of the airway tree from chest CT using local volume of interest. In *Proc. of Second International Workshop on Pulmonary Image Analysis*, 273–284 (2009).
- [95] Schlathoelter, T., Lorenz, C., Carlsen, I. C., Renisch, S. & Deschamps, T. Simultaneous segmentation and tree reconstruction of the airways for virtual bronchoscopy. In *Proc. SPIE*, 103–113 (2002).

- [96] Bülow, T., Lorenz, C. & Renisch, S. A General Framework for Tree Segmentation and Reconstruction from Medical Volume Data. In Barillot, C., Haynor, D. R. & Hellier, P. (eds.) *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 533–540 (Springer Berlin Heidelberg, Saint-Malo, 2004).
- [97] Caselles, V., Kimmel, R. & Sapiro, G. Geodesic Active Contours. *International Journal of Computer Vision* **22**, 61–79 (1997). DOI 10.1023/A:1007979827043.
- [98] Deschamps, T. & Cohen, L. D. Fast extraction of minimal paths in 3D images and applications to virtual endoscopy. *Medical Image Analysis* **5**, 281–299 (2001). DOI 10.1016/S1361-8415(01)00046-9.
- [99] Sobel, I. History and definition of the sobel operator. *Retrieved from the World Wide Web* (2014).
- [100] Santago, P. & Gage, H. D. Statistical models of partial volume effect. *IEEE Transactions on Image Processing* **4**, 1531–1540 (1995). DOI 10.1109/83.469934.
- [101] Ceresa, M., Artaechevarria, X., Munoz-Barrutia, A. & Ortiz-De-Solorzano, C. Automatic Leakage Detection and Recovery For Airway Tree Extraction in Chest CT Images. In *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 568–571 (IEEE, 2010).
- [102] Bribiesca, E. An easy measure of compactness for 2D and 3D shapes. *Pattern Recognition* **41**, 543–554 (2008). DOI 10.1016/j.patcog.2007.06.029.
- [103] Liepa, P. Filling Holes in Meshes. In *Proceedings of the 2003 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing*, SGP '03, 200–205 (Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 2003). URL <http://dl.acm.org/citation.cfm?id=882370.882397>.
- [104] Janaszewski, M., Couprie, M. & About, L. Hole filling in 3D volumetric objects. *Pattern Recognition* **43**, 3548–3559 (2010). DOI 10.1016/j.patcog.2010.04.015.
- [105] Suzuki, K. *et al.* Computer-aided measurement of liver volumes in CT by means of geodesic active contour segmentation coupled with level-set algorithms. *Medical physics* **37**, 2159–2166 (2010). DOI 10.1118/1.3395579.
- [106] Sethian, J. A. & others. Level set methods and fast marching methods. *Journal of Computing and Information Technology* **11**, 1–2 (2003). DOI 0-471-80599-8.
- [107] Noor, N. M. *et al.* Performance Evaluation of Lung Segmentation. In *Medical Imaging Technology: Reviews and Computational Applications*, chap. 5, 111–127 (Springer, Singapore, 2015).
- [108] Mansoor, A. *et al.* CIDI-lung-seg: A single-click annotation tool for automatic delineation of lungs from CT scans. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 1087–1090 (2014).
- [109] Sørensen, T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Biol. Skr.* **5**, 1–34 (1948).
- [110] Dice, L. R. Measures of the Amount of Ecologic Association Between Species. *Ecology* **26**, 297–302 (1945). URL <http://dx.doi.org/10.2307/1932409>. DOI 10.2307/1932409.

- [111] Peck, R. & Devore, J. L. *Statistics: The exploration and analysis of data* (Cengage Learning, 2011).
- [112] Huttenlocher, D. P., Klanderman, G. A. & Rucklidge, W. J. Comparing Images Using the Hausdorff Distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **15**, 850–863 (1993). DOI 10.1109/34.232073.
- [113] Haralick, R. M., Shanmugan, K. & Dinstein, I. H. Textural features for image classification. *Systems, Man and Cybernetics* **6**, 610–621 (1973). DOI 10.1109/TSMC.1973.4309314.
- [114] Haralick, R. M. Statistical and structural approaches to texture. *Proceedings of the IEEE* **67**, 786–804 (1979). DOI 10.1109/PROC.1979.11328.
- [115] Depeursinge, A., Foncubierta-Rodriguez, A., Van De Ville, D. & Müller, H. Three-dimensional solid texture analysis in biomedical imaging: Review and opportunities. *Medical Image Analysis* **18**, 176–196 (2014). URL <http://dx.doi.org/10.1016/j.media.2013.10.005>. DOI 10.1016/j.media.2013.10.005.
- [116] Soh, L. K. & Tsatsoulis, C. Texture analysis of sar sea ice imagery using gray level co-occurrence matrices. *IEEE Transactions on Geoscience and Remote Sensing* **37**, 780–795 (1999). DOI 10.1109/36.752194.
- [117] Peng, H., Long, F. & Ding, C. Feature selection based on mutual information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2005). DOI 10.1109/TPAMI.2005.159.
- [118] Howarth, P. & Rüger, S. M. Evaluation of texture features for content-based image retrieval. In *International Conference on Image and Video Retrieval*, vol. 3115, 326–334 (Springer, 2004).
- [119] Bannasar, M., Hicks, Y. & Setchi, R. Feature selection using Joint Mutual Information Maximisation. *Expert Systems with Applications* **42**, 8520–8532 (2015). URL <http://dx.doi.org/10.1016/j.eswa.2015.07.007>. DOI 10.1016/j.eswa.2015.07.007.
- [120] FISHER, R. A. The use of multiple measurements in taxonomic problems. *Annals of Eugenics* **7**, 179–188 (1936). URL <http://doi.wiley.com/10.1111/j.1469-1809.1936.tb02137.x>. DOI 10.1111/j.1469-1809.1936.tb02137.x.
- [121] Lefebvre, F., Meunier, M., Thibault, F., Laugier, P. & Berger, G. Computerized ultrasound B-scan characterization of breast nodules. *Ultrasound in medicine & biology* **26**, 1421–1428 (2000). DOI 10.1016/S0301-5629(00)00302-1.
- [122] Drukker, K., Pesce, L. & Giger, M. Repeatability in computer-aided diagnosis: Application to breast cancer diagnosis on sonography. *Medical physics* **37**, 2659–2669 (2010). DOI 10.1118/1.3427409.
- [123] Moutoussis, M., Fearon, P., El-Deredy, W., Dolan, R. J. & Friston, K. J. Bayesian inferences about the self (and others): A review. *Consciousness and Cognition* **25**, 67–76 (2014). URL <http://dx.doi.org/10.1016/j.concog.2014.01.009>. DOI 10.1016/j.concog.2014.01.009.
- [124] Gal, Y. & Ghahramani, Z. Dropout as a Bayesian Approximation : Representing Model Uncertainty in Deep Learning. In *International Conference on Machine Learning*, 1050–1059 (2016).

-
- [125] Vapnik, V. *The nature of statistical learning theory* (Springer science & business media, 2013).
- [126] Johnson, H. J., McCormick, M. M. & Ibañez, L. The ITK Software Guide. Tech. Rep. (2017). URL <https://itk.org/ItkSoftwareGuide.pdf>.
- [127] Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat Meth* **9**, 671–675 (2012). URL <http://dx.doi.org/10.1038/nmeth.2089>.
- [128] van der Walt, S. *et al.* scikit-image: image processing in Python. *PeerJ* **2**, e453 (2014). URL <http://dx.doi.org/10.7717/peerj.453>. DOI 10.7717/peerj.453.