



ESCUELA DE INGENIERÍA DE FUENLABRADA

PREDICCIÓN AUTOMÁTICA  
DE LA RESPUESTA A LA QUIMIOTERAPIA  
EN PACIENTES CON OSTEOSARCOMA O SARCOMA DE EWING

TRABAJO FIN DE GRADO

GRADO EN INGENIERÍA BIOMÉDICA

AUTOR: Adrián Manchado Carreira

TUTOR: Ángel Torrado Carvajal

CO-TUTOR: José Acosta Batlle

Curso académico: 2023-2024



# Agradecimientos

Gracias a mi tía, la Doña Modelitos, por aguantarme como buenamente podía en mis momentos más genuinamente inaguantables, manteniéndome lejos del colapso con comidas increíbles; a mi vecino, el Gustaxista, por saturarme un poco más la cabeza con partidas interminables de ajedrez y a mi amigo, Nacho el bichólogo, por recordarme de vez en cuando que hay unas *Droseras rotundifolias* muy cucas, allá por las turberas de Asturias, esperando a ser visitadas una vez restaurada mi vida social. Gracias a mis padres, a mi hermano y a todos mis colegas que, desde la distancia, han sabido desminarme la moral mientras minaba código en la oscuridad de mi habitación.

Gracias a mis tutores, Ángel Torrado Carvajal y José Acosta Batlle, por haberme ofrecido la oportunidad de formar parte de este proyecto y a Borja Rodríguez Vila por acompañarme durante la recta final.





# Resumen

En los últimos años, se ha visto un incremento en la cantidad de imágenes procesadas en los servicios de radiología. El número de radiólogos sigue siendo el mismo, lo que se traduce en un aumento de las tareas, retrasos en el resto de servicios a los que da soporte y deterioro emocional en los profesionales. Surge así la necesidad de crear herramientas que apoyen activamente al equipo clínico, mejorando la toma de decisiones, liberando carga de trabajo y contribuyendo finalmente a la seguridad del paciente. Una posible estrategia es la de crear modelos robustos de inteligencia artificial (IA) utilizando radiómica, entendida como la extracción de características cuantitativas a partir de imágenes médicas.

Concretamente, el objetivo de este proyecto es el de predecir la respuesta al tratamiento quimioterápico neoadyuvante en pacientes con osteosarcomas o sarcomas de Ewing, en colaboración con el servicio de radiología musculoesquelética del Hospital Universitario Ramón y Cajal.

Para ello, se construyó una pequeña base de datos de pacientes con estas patologías, incluyendo imágenes de resonancia magnética y máscaras binarias que delimitan el tumor. Se utilizó para extraer características radiómicas con PyRadiomics, con las que se entrenaron cinco modelos de IA (Regresión Lineal, Random Forest, Perceptrón Multicapa sin normalizar y normalizado y Support Vector Machine) después de haber aplicado tres métodos de selección (Correlación, Información Mutua y Principal Component Analysis).

Los resultados obtenidos destacan las prestaciones del modelo de MLP normalizado. Con una exactitud de 0.84 y un AUC de 0.83, ofrece resultados muy prometedores dado el limitado número de pacientes. Sin embargo, no se logró identificar un biomarcador, por lo que consideramos que los resultados no son lo suficientemente buenos como para trasladar esta herramienta al entorno clínico real. Se asientan así las bases para el desarrollo de herramientas más sofisticadas relacionadas con los tumores óseos que incluyan, entre otras propuestas, el aumento del número de pacientes o la inclusión de características genómicas.



---

# Índice general

<b>Índice de figuras</b>	<b>v</b>
<b>Índice de cuadros</b>	<b>vii</b>
<b>Lista de Acrónimos y Abreviaturas</b>	<b>x</b>
<b>1. Introducción y objetivos</b>	<b>1</b>
1.1. Contexto y motivación . . . . .	1
1.2. Objetivos . . . . .	2
1.3. Metodología . . . . .	3
1.4. Estructura del proyecto . . . . .	4
<b>2. Conocimientos previos</b>	<b>7</b>
2.1. Tumores óseos . . . . .	7
2.1.1. Osteosarcoma . . . . .	7
2.1.2. Sarcoma de Ewing . . . . .	9
2.2. Radiómica . . . . .	10
2.3. Estado del Arte . . . . .	11
<b>3. Materiales</b>	<b>13</b>
3.1. 3D Slicer . . . . .	13
3.2. PyRadiomics . . . . .	14
<b>4. Métodos</b>	<b>15</b>
4.1. Base de Datos . . . . .	15
4.1.1. Búsqueda de Pacientes . . . . .	16
4.1.2. Construcción de la Base de Datos . . . . .	19

4.2.	Extracción de Características Radiómicas . . . . .	20
4.3.	Selección de Características . . . . .	23
4.3.1.	Correlación . . . . .	24
4.3.2.	Información Mutua . . . . .	25
4.3.3.	Principal Component Analysis . . . . .	25
4.4.	Entrenamiento Modelos . . . . .	27
4.4.1.	Cross Validation . . . . .	27
4.4.2.	Regresión Logística . . . . .	28
4.4.3.	Random Forest . . . . .	29
4.4.4.	Multilayer Perceptron . . . . .	30
4.4.5.	Support Vector Machine . . . . .	31
4.4.6.	Búsqueda de hiperparámetros . . . . .	32
4.5.	Explicabilidad . . . . .	33
<b>5.</b>	<b>Experimentación y resultados</b>	<b>37</b>
5.1.	Correlación . . . . .	38
5.1.1.	Características Seleccionadas . . . . .	38
5.1.2.	Resultados . . . . .	39
5.2.	Información Mutua . . . . .	40
5.2.1.	Características Seleccionadas . . . . .	41
5.2.2.	Resultados . . . . .	41
5.3.	PCA . . . . .	43
5.3.1.	Características Seleccionadas . . . . .	43
5.3.2.	Resultados . . . . .	44
5.4.	Explicabilidad . . . . .	46
<b>6.</b>	<b>Discusión</b>	<b>49</b>
6.1.	Figuras de mérito . . . . .	49
6.2.	Limitaciones . . . . .	50



6.3. Dimensionalidad y explicabilidad . . . . .	51
<b>7. Conclusión y Líneas futuras</b>	<b>53</b>
7.1. Conclusión . . . . .	53
7.2. Líneas Futuras . . . . .	54
<b>Bibliografía</b>	<b>57</b>
<b>A. Imágenes RM</b>	<b>67</b>
<b>B. Matrices de Confusión</b>	<b>71</b>



---

# Índice de figuras

1.1. Diagrama de Gantt de desarrollo del proyecto. . . . .	3
2.1. Pasos a seguir en un proyecto radiómico [20]. . . . .	11
3.1. Interfaz gráfica de 3DSlicer. Podemos ver que se está llevando a cabo una segmentación. . . . .	13
3.2. Extracción de características radiómicas con PyRadiomics [31]. . . . .	14
4.1. Diagrama de flujo para la selección de pacientes . . . . .	17
4.2. Formato BIDS para el archivado de imagen médica [34] . . . . .	20
4.3. Secuencias T1 de los 13 fragmentos que forman parte del estudio. . . . .	21
4.4. Secuencia T1, Máscara binaria y ROI para 3 pacientes (sub-009, sub-010 y sub-12). Listado completo en Apéndice A. . . . .	22
4.5. Aumentar el número de características es útil hasta cierto punto. Después, las prestaciones empeoran notablemente [48]. . . . .	24
4.6. Para que el PCA funcione correctamente, los datos deben ser normalizados y centrados, ya que la diferencia entre los rangos dinámicos de las características puede alterar la variabilidad [47]. . . . .	26
4.7. Ejemplo de varianza explicada y varianza acumulada en un problema de 7 dimensiones. Puede verse que casi todos los ejes comparten una variabilidad similar, por lo que son igualmente relevantes. . . . .	27
4.8. División de los conjuntos <i>train</i> (azul) y <i>validation</i> (naranja) en CV de 5-Folds para una repetición [52] . . . . .	28
4.9. Ejemplo de un árbol de decisión . . . . .	29
4.10. Comparación de una neurona real y su modelo artificial . . . . .	30
4.11. Ejemplo de una estructura MLP [58]. . . . .	31
4.12. Ejemplo de fronteras lineales en SVM [61] . . . . .	32
4.13. Ejemplo de frontera no lineal para clasificación binaria [61] . . . . .	32

4.14. Ejemplo de valores SHAP para dos observaciones de un modelo de RF [65] . . .	34
4.15. Ejemplo de valores SHAP para el conjunto de observaciones de un modelo de RF [65] . . . . .	35
5.1. Coeficientes de correlación de las 8 características seleccionadas. . . . .	39
5.2. Curvas ROC y AUC para los modelos entrenados empleando correlación. . . .	40
5.3. Las 35 características con mayor IM. . . . .	41
5.4. Curvas ROC y AUC para los modelos entrenados empleando IM. . . . .	42
5.5. Varianza Explicada y Varianza Acumulada. . . . .	43
5.6. PC más relevantes. . . . .	44
5.7. Curvas ROC y AUC para los modelos entrenados empleando PCA. . . . .	45
5.8. Valores SHAP para los diferentes modelo encontrados utilizando correlación. .	47
5.9. Valores SHAP para los diferentes modelo encontrados utilizando IM. . . . .	48
A.1. Secuencia T1, Máscara binaria y ROI de los pacientes <b>respondedores</b> . . . . .	68
A.2. Secuencia T1, Máscara binaria y ROI de los pacientes <b>no-respondedores</b> . . . .	70
B.1. Matrices de confusión para modelos filtrados con correlación. . . . .	72
B.2. Matrices de confusión para modelos filtrados con IM. . . . .	73
B.3. Matrices de confusión para modelos filtrados con PCA. . . . .	74

---

# Índice de cuadros

4.1.	Lista original de pacientes. El año indica la fecha de creación de la imagen control.	18
4.2.	Secuencias disponibles para cada paciente. T2 SPIR: Secuencia T2 con supresión de grasa, PDW: Densidad Protónica, DWI: Secuencia de difusión, AXIAL: Secuencia de contraste. . . . .	18
4.3.	Listado de pacientes que se utilizarán en el entrenamiento de algoritmos. Solo la secuencia T1 será utilizada. . . . .	19
5.1.	Características seleccionadas mediante correlación. El orden mostrado no indica la importancia. . . . .	38
5.2.	Figuras de mérito de los modelos, utilizando correlación como método de selección de características. . . . .	40
5.3.	Características seleccionadas con IM. El orden mostrado no indica su importancia.	41
5.4.	Figuras de mérito de los modelos utilizando IM como selección de características.	42
5.5.	Figuras de mérito de los modelos utilizando PCA como selección de características. . . . .	44



---

# Lista de Acrónimos y Abreviaturas

**BD** *Base de Datos*

**CV** *Cross Validation*

**GLCM** *Gray-Level Cooccurrence Matrix*

**GLDM** *Neighborhood Gray-Level Dependence Matrix*

**GLRLM** *Gray-Level Run-length Matrix*

**GLSZM** *Gray-Level Size Zone Matrix*

**HCE** *Historia Clínica Electrónica*

**HURC** *Hospital Universitario Ramón y Cajal*

**IA** *Inteligencia Artificial*

**IM** *Información Mutua*

**LoG** *Laplacian of Gaussian*

**MLP-Norm** *Perceptrón Multicapa Normalizado*

**MLP** *Perceptrón Multicapa*

**OS** *Osteosarcoma*

**PCA** *Principal Component Analysis*

**QTN** *Quimioterapia Neoadyuvante*

**RL** *Regresión Logística*

**RM** *Resonancia Magnética*

**ROI** *Región de Interés*

**SE** *Sarcoma de Ewing*

**SHAP** *SHapley Additive exPlanations*

**STD** *Desviación Típica*

**SVD** *Singular Value Decomposition*

**SVM** *Support Vector Machine*

**TC** *Tomografía axial Computarizada*



# Capítulo 1

---

## Introducción y objetivos

En este primer capítulo, se presenta el estado actual del servicio de radiología y las motivaciones detrás de querer predecir la respuesta al tratamiento quimioterápico en pacientes con tumores óseos.

### 1.1 Contexto y motivación

La cantidad de imágenes generada por el servicio de radiología ha aumentado notablemente en las últimas décadas y el número de radiólogos sigue siendo el mismo [1]. Los hospitales cuentan con cada vez más y mejores equipos de imagen, haciendo que las cargas de trabajo sean excesivas al no aumentar el número de radiólogos contratados. Así, los informes se retrasan y con ellos el resto de servicios del hospital. La perspectiva a futuro tampoco parece buena [2], los radiólogos están envejeciendo y no hay personal joven que los sustituya.

Esto se traduce unos índices de *burnout*, definido como el agotamiento físico y mental resultado de un estrés prolongado en el tiempo, sean altos entre los radiólogos [3]. La alta cantidad de tareas y el ambiente de trabajo (sedentario, solitario y oscuro), impacta negativamente en la salud mental de los profesionales, mermando su capacidad de informar casos y comprometiendo finalmente el tratamiento de los pacientes.

Surge así la necesidad de liberar esta gran carga de trabajo. La solución directa es la de contratar a más radiólogos, pero influyen factores socioeconómicos y burocráticos que escapan al alcance de este proyecto. La solución indirecta es la de colaborar activamente con el equipo para crear herramientas de apoyo. El trabajo se vuelve más tolerable y aumenta así la seguridad del paciente.

En este proyecto, trabajamos con el servicio de radiología musculoesquelética para abordar

los **osteosarcomas** (OS) y **sarcomas de Ewing** (SE). Como veremos a continuación, estos tumores son muy agresivos y suelen responder mal al tratamiento quimioterápico. Es motivador pensar que conocer a priori la respuesta que un paciente va a tener a su tratamiento puede tener un impacto muy positivo en su seguimiento, evitando someterle a un proceso dañino que últimamente sería ineficaz.

## 1.2 Objetivos

El objetivo principal de este proyecto es el de predecir la respuesta a la quimioterapia neoadyuvante (QTN) en pacientes con OS o SE. Para alcanzar este objetivo, construiremos una herramienta de apoyo a la decisión clínica basada en inteligencia artificial (IA), en colaboración con el Hospital Universitario Ramón y Cajal (HURC). Podemos descomponer el objetivo principal en los siguientes subobjetivos:

- Buscar e identificar pacientes en la Base de Datos (BD) del HURC que se ajusten al objetivo.
- Crear una BD conformada por la información personal de la historia clínica electrónica (HCE), pruebas imagen de resonancia magnética (RM) y segmentaciones de las lesiones de los pacientes identificados.
- Extraer características radiómicas a partir de las imágenes de RM.
- Seleccionar de manera automática las características más relevantes para la tarea de clasificación.
- Implementar, optimizar y evaluar diferentes modelos de IA basados en las características radiómicas seleccionadas.
- Analizar los resultados, identificando el modelo que mejor se ajuste a nuestro objetivo.

En la Figura 1.1 podemos ver el tiempo invertido en el desarrollo del proyecto.

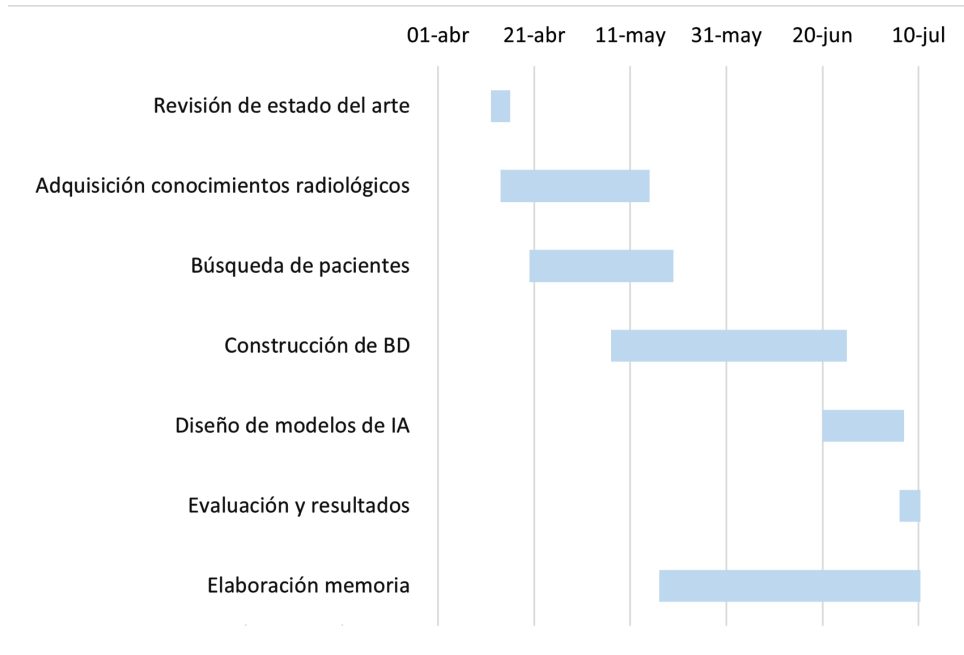


Figura 1.1: Diagrama de Gantt de desarrollo del proyecto.

## 1.3 Metodología

En la elaboración del proyecto se siguió la siguiente metodología:

- Reunirnos con los radiólogos del servicio musculoesquelético del HURC para conocer la realidad y alcance del proyecto.
- Familiarizarnos con el entorno hospitalario, el servicio de radiología, las estaciones de trabajo y los programas informáticos clínicos.
- Buscar proyectos similares que sirvan de inspiración y ampliar el conocimiento sobre los OS y SE.
- Aprender a utilizar las herramientas de software necesarias.
- Aprender a localizar y segmentar las lesiones en las diferentes secuencias de imagen médica.
- Desarrollar criterios de selección de pacientes para el proyecto.

- Construir una BD de pacientes con OS y SE, segmentados y etiquetados según su respuesta a la QTN.
- Extraer las características radiómicas a partir de la imagen de RM.
- Seleccionar las características radiómicas más relevantes para la clasificación.
- Seleccionar diferentes modelos de IA para su posterior evaluación e interpretación de los resultados.
- Elaborar y seguir un diario en el que se recogen las tareas conseguidas, en proceso y por realizar.
- Elaborar una memoria del proyecto.

## 1.4 Estructura del proyecto

El presente documento se encuentra estructurado en 6 capítulos:

- **Capítulo 1. Introducción y objetivos:** Donde hemos explicado el estado actual del servicio de radiología y de sus profesionales, estableciendo una motivación para el desarrollo del proyecto. Presentamos también el objetivo, predecir la respuesta al tratamiento QTN en pacientes con OS y SE, y la metodología a seguir.
- **Capítulo 2. Conocimiento previos:** Donde se ha presentado la información referente a los tumores óseos, OS, SE y al concepto de radiómica. También se presenta el panorama actual de los proyectos radiómicos en el ámbito de la radiología musculoesquelética.
- **Capítulo 3. Materiales:** Donde se explican las herramientas de software utilizadas, 3D slicer y PyRadiomics.
- **Capítulo 4. Métodos:** El capítulo comienza presentando el proceso de construcción de la BD de pacientes con OS y SE seguido de una descripción de las características radiómicas extraídas. Se presentan también los métodos de selección de características utilizadas, los

cinco modelos de IA y cómo serán entrenados y validados. Finalmente, presentamos y definimos el concepto de explicabilidad, utilizando los valores SHAP.

- **Capítulo 5. Experimentación y resultados:** Dónde se muestran las diferentes figuras de mérito obtenidas de cada modelo, para su posterior discusión.
- **Capítulo 6. Discusión:** Intentamos razonar los resultados obtenidos en el capítulo anterior para trasladarlos al entorno clínico. Lo dividimos en tres secciones, según el punto de vista: Figuras de mérito, en el que intentamos buscar el impacto clínico de los resultados obtenidos; Limitaciones, en el que indicamos los que creemos que han sido los mayores inconvenientes y Explicabilidad, en el que estudiamos las características en busca de un biomarcador.
- **Capítulo 7. Conclusión y líneas futuras:** Donde se exponen las conclusiones a las que llegamos una vez completado el proyecto. También, se sugieren investigaciones futuras, como aumentar el número de observaciones en la BD o hacer uso de otras características, como las genómicas.



# Capítulo 2

---

## Conocimientos previos

En este capítulo se presenta el conocimiento previo requerido sobre las patologías y herramientas utilizadas para entender el desarrollo del proyecto.

### 2.1 Tumores óseos

El tumor óseo más frecuente es la metástasis asociada a otros tumores, (como de mama o de pulmón), o el que tiene origen en la médula ósea (leucemia o linfoma) [4]. Únicamente, el 1% de todos los cánceres diagnosticados anualmente tienen origen en el propio hueso [5]. De estos, los más frecuentes son los OS, SE y condrosarcomas y, aunque raros, su alta agresividad hace que los índices de mortalidad y morbilidad sean elevados.

#### 2.1.1 Osteosarcoma

El OS es el tumor óseo más común, afectando anualmente en todo el mundo a 3 personas por cada millón de habitantes. La distribución por edad es bimodal: hay un primer pico de mayor incidencia en niños y adolescentes, siendo en este grupo el tercer cáncer más frecuente, y un segundo pico menor en adultos mayores de 60 años. El tumor se localiza generalmente en las metáfisis de los huesos largos, mostrando predilección por el fémur distal (43%) o la tibia proximal (23%) [6].

Suele manifestarse con dolor intenso en la zona afectada de varios meses de evolución, pudiendo despertar al paciente por las noches. Debido a que la mayoría de pacientes son jóvenes, el dolor suele atribuirse al crecimiento óseo o a actividades deportivas, pudiendo retrasar el diagnóstico. Durante la exploración, se puede percibir un aumento de la temperatura local, piel ligeramente distendida y una masa palpable. En los casos más extremos, este tumor puede

debutar con una fractura patológica. Cuando se sospecha clínicamente un OS, se solicita una prueba de RM, sin y con contraste, para valorar la extensión local [7].

Antes de 1970, la amputación completa del miembro afectado era el único tratamiento disponible para el OS. Tras ella, el 80% de los pacientes fallecían por metástasis, generalmente en el pulmón, demostrando que este tratamiento no era efectivo. La amputación no eliminaba las aparentes micrometástasis, invisibles durante el diagnóstico inicial. Es por ello que se inició la investigación de alternativas hasta llegar al tratamiento utilizado a día de hoy [8].

Actualmente, este tratamiento incluye la QTN (prequirúrgica), la cirugía conservadora del miembro (siempre que sea posible) y la quimioterapia adyuvante (postquirúrgica). Un equipo multidisciplinar de oncólogos, radiólogos, traumatólogos ortopédicos y psicólogos acompañan al paciente desde el momento del diagnóstico [9]. Gracias a este abordaje, la supervivencia es del 65-70% para pacientes con lesión localizada y del 20-30% para aquellos con metástasis [10].

La QTN nos permite tratar de manera temprana esas lesiones micro metastásicas fulminantes y preparar correctamente el miembro para la cirugía conservadora. Esta última, es posible gracias a los avances en prótesis y en los estudios de imagen médica, que identifican los bordes del tumor con mayor precisión [9]. Se administra un cocktail de hasta 4 fármacos que produce efectos secundarios sobre el cuerpo: el daño hepático, el daño renal o la hematotoxicidad son frecuentes, lo que supone un reto añadido a considerar durante el seguimiento [11].

Sin embargo, no todos los pacientes responden al tratamiento, ya que el origen del OS es aún incierto. La alta complejidad del mecanismo genético que causa estos tumores hace que presenten una gran heterogeneidad, tanto a nivel intratumoral como entre los pacientes, así es que la gran mayoría de casos (95%) aparece de manera esporádica. Esto supone un reto a la hora de encontrar nuevos tratamientos o ajustar los ya existentes [12]. Es por ello que es esencial el diseño de herramientas que nos permitan conocer, a priori, la respuesta individualizada de cada paciente a la QTN, permitiendo así la mejor toma de decisiones y evitando los efectos tóxicos que esta produce.



## 2.1.2 Sarcoma de Ewing

El SE es un tumor óseo muy agresivo, miembro de la familia de tumores de Ewing, que afecta principalmente a niños y jóvenes [13]. Aún siendo un tumor raro (menos de un 10% de todas los cánceres), es el segundo más común de estos tumores, con una incidencia anual de hasta 3 casos por cada millón de habitantes [14]. Aparece principalmente en la diáfisis de los huesos largos y ocasionalmente en tejido blando.

La enfermedad suele debutar en fractura patológica, después de un cuadro de dolor intenso que suele atribuirse, como en el OS, al crecimiento de los pacientes o a condiciones asociadas a actividades deportivas. Es por ello que el diagnóstico se ve retrasado hasta 6 meses en el 25 % de los casos [15]. En exploración, se puede percibir una masa dura si la lesión es de un tamaño considerable.

Debido a su comportamiento altamente agresivo, la metástasis es común en estos pacientes, principalmente en el pulmón (50% de los casos). Es el mayor indicador de supervivencia a los 5 años: 75% para pacientes con lesiones localizadas, 50% para lesiones pulmonares localizadas y 30% para metástasis primaria [15].

Cuando aparecen sospechas de SE, el primer paso es solicitar una radiografía convencional, una prueba de resonancia magnética, con y sin contraste, para delimitar la región y una prueba de tomografía axial computarizada (TC) de tórax para estudiar la afectación pulmonar [16].

La particularidad del SE es que es sensible a la radioterapia [17]. En un primer momento, se optó por utilizar la radioterapia como única estrategia ante el SE, pero se descartó al observar que los pacientes fallecían por metástasis. Es por ello que, al igual que en OS, el tratamiento mayormente utilizado es QTN, cirugía conservadora y quimioterapia adyuvante; el mecanismo de acción es similar y conlleva los mismos riesgos y beneficios. La radioterapia se reserva para aquellos casos en los que no es posible realizar cirugía conservadora o esta es insuficiente [18].

## 2.2 Radiómica

En el tratamiento de los tumores óseos, se incluye también una **biopsia** previa a la cirugía conservadora [19]. De esta manera, las características del tumor pueden ser estudiadas en el laboratorio histopatológico, permitiendo al equipo tomar mejores decisiones. Sin embargo, debido a la alta heterogeneidad de los OS y SE, la muestra puede no ser representativa del tumor, provocando un incorrecto diagnóstico y tratamiento. Por otro lado, al ser una técnica invasiva, pueden aparecer efectos no deseados, como la contaminación del canal. Surge así la necesidad de crear una herramienta capaz de mitigar estos efectos inconvenientes.

La radiómica es la extracción de características cuantitativas de las imágenes médicas, pudiendo revelar información importante de patrones invisibles al ojo humano [20]. Su desarrollo comenzó originalmente en oncología, debido a que cada paciente tiene, generalmente, diversas pruebas de imagen médica disponibles durante el continuo de su tratamiento [21]. En conjunto con los datos genómicos de cada paciente, la radiómica puede contribuir notablemente a la medicina personalizada, mediante la búsqueda y validación de biomarcadores [22].

Los proyectos radiómicos siempre se dividen en las mismas etapas (Figura 2.1). Primero se construye una BD de imágenes médicas de los pacientes que se desean estudiar. Después, se delimita y segmenta la región de interés (tumor, ganglio inflamado...). Seguidamente, se extraen las características radiómicas, ya sea manual o automáticamente (mediante herramientas como PyRadiomics). Finalmente, se entrenan y validan los modelos de IA [23]. También, puede incluirse un preprocesado de las imágenes, dependiendo del problema en particular.

Sin embargo, la radiómica no está libre de controversia. La reproducibilidad de las características radiómicas entre los estudios sigue siendo el principal reto, hasta el punto de que ningún estudio ha sido trasladado al ámbito clínico real. Es por ello que la mejora del diseño y validación de los modelos, así como de los estándares, son esenciales para abrir camino a esta disciplina tan prometedora [24].

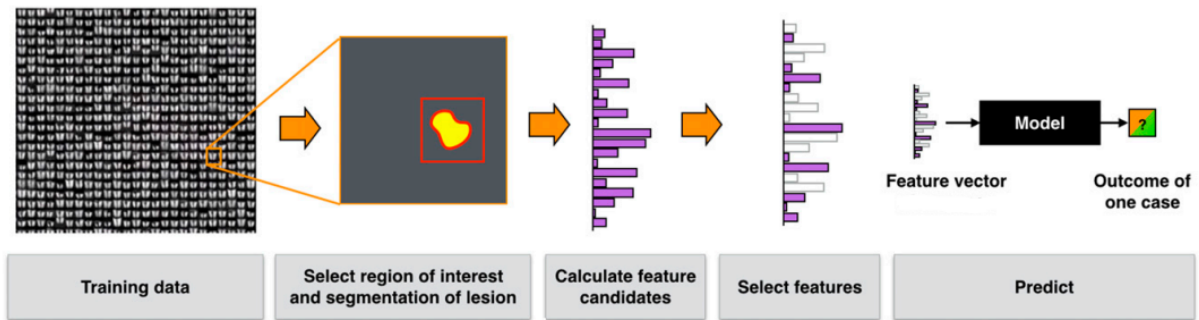


Figura 2.1: Pasos a seguir en un proyecto radiómico [20].

## 2.3 Estado del Arte

En esta sección, presentamos una serie de estudios que resaltan el potencial de la radiómica en el ámbito de la oncología musculoesquelética:

La radiómica puede ser utilizada para conocer el grado de malignidad entre tumores diferentes, como demostraron recientemente Gitto et al. [25] en su trabajo sobre condrosarcomas. El objetivo fue diferenciar los tumores cartilagosos atípicos de los condrosarcomas grado II, tumores que son similares en imagen y que tienen tratamientos muy diferentes (seguimiento y cirugía, respectivamente). Para ello, emplearon un modelo de IA con el que obtuvieron una tasa de acierto del 92 % y concluyeron que su herramienta podría ser trasladada al entorno clínico al compararse con los resultados de un radiólogo experimentado.

También puede ser utilizada para definir el grado de malignidad en un mismo tipo de tumor. El estudio llevado a cabo por Chianca et al. [26] tuvo como objetivo predecir la malignidad de tumores óseos localizados en la columna vertebral, a partir de imágenes de RM. Para ello, emplearon diferentes modelos de IA, obteniendo un 94 % de tasa de acierto en un modelo de clasificación con 2 clases (benigno o maligno), concluyendo que un modelo bien entrenado puede superar los resultados de la biopsia.

Como en este proyecto, la radiómica puede servirnos para predecir la respuesta al tratamiento QTN. Lin et al. [27] buscan en su estudio mejorar la asistencia personalizada mediante

la predicción de la respuesta a la QTN en pacientes con OS. Para ello, construyeron una herramienta de IA, utilizando delta-radiómica con imágenes de TC, que indica la probabilidad de no-respuesta (necrosis inferior al 90 %) para cada paciente. Obtuvieron un AUC del 84 % en el conjunto de validación, asegurando que puede ser una herramienta de gran utilidad para oncólogos y cirujanos.

En definitiva, la radiómica ofrece resultados muy favorables. Los modelos de IA que la utilizan ofrecen soluciones para diversos problemas relacionados con los tumores óseos, que parecen incluso competir con la habilidad de los radiólogos más experimentados, destacando así su valor como herramienta de apoyo a la decisión clínica.

# Capítulo 3

## Materiales

En este breve capítulo, presentamos las dos herramientas de software utilizadas para el desarrollo de este proyecto.

### 3.1 3D Slicer

3DSlicer [28] es una plataforma de código abierto diseñada para dar soporte a tareas relacionadas con imagen médica e investigación biomédica. En continuo desarrollo, es capaz de soportar diversos ficheros DICOM [29] y de ofrecer un gran número de herramientas especializadas gracias a su comunidad, que continúa activamente trabajando en ella cada día, convirtiéndola en la plataforma de referencia para problemas de investigación biomédica.

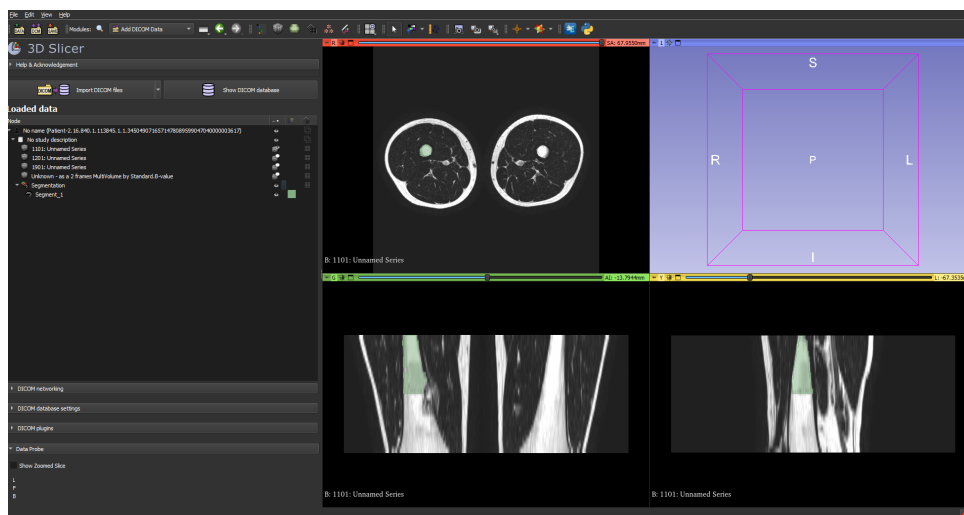


Figura 3.1: Interfaz gráfica de 3DSlicer. Podemos ver que se está llevando a cabo una segmentación.

Gracias a sus múltiples funciones, puede servirnos para realizar segmentaciones (Figura

3.1), manual o automáticamente, de lesiones en las distintas modalidades de imagen médica, para servir de guía en cirugía o para implementar realidad virtual en simulaciones médicas. Además, permite la creación de extensiones personalizadas utilizando Python, cimentando su potencial como herramienta clínica.

## 3.2 PyRadiomics

PyRadiomics [30] es una librería de Python que contiene un gran número de algoritmos diseñados para calcular características radiómicas (Sección 4.2) de imágenes médicas. Hace uso de 3DSlicer para proporcionar una interfaz sencilla. Las características son extraídas automáticamente siguiendo unos pasos básicos [31], cuyos detalles podemos controlar mediante un fichero de parámetros: carga y preprocesado de las imágenes, filtrado, extracción de características (Figura 3.2) y presentación de los resultados.

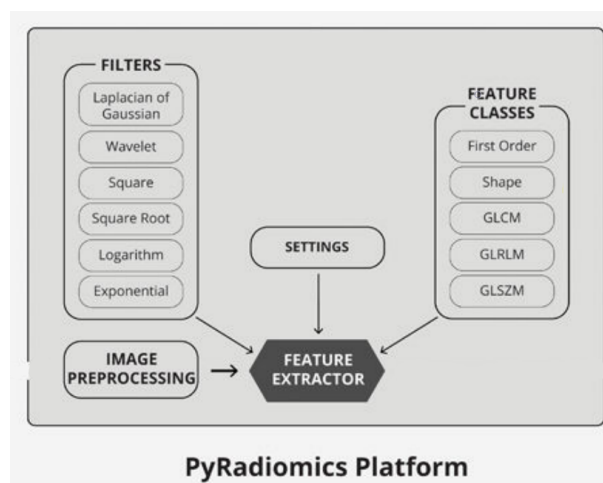


Figura 3.2: Extracción de características radiómicas con PyRadiomics [31].

Múltiples formatos de imagen son aceptados por PyRadiomics, al hacer uso de la librería SimpleITK [32]. También permite hacer un preprocesado básico, como el reescalado de la imagen, que puede ser definido en el fichero de parámetros. También permite seleccionar qué tipo de filtros y características van a ser calculadas.

# Capítulo 4

---

## Métodos

En este capítulo se presentan los métodos empleados para construir la BD, seleccionar características y entrenar los algoritmos de IA.

### 4.1 Base de Datos

En esta sección, presentamos la BD utilizada para el diseño y evaluación de los diferentes modelos de IA. Previamente al desarrollo de este proyecto, no existía una BD de pacientes con OS o SE, por lo que fue necesario crearla.

La respuesta al tratamiento QTN se cuantifica mediante el porcentaje de necrosis o reducción de la masa tumoral en la imagen control, respecto a la imagen inicial. Esta información pudo ser encontrada tanto en la HCE de cada paciente como en el informe radiológico de la imagen control (imagen posterior a la QTN y previa a la cirugía conservadora). Así, dividimos a los pacientes en 3 grupos según su respuesta:

- **No respuesta:** No hay cambios significativos en la imagen de control. La necrosis o reducción es inferior al 35 %.
- **Parcial:** Hay cambios significativos en la imagen de control. La necrosis o reducción es superior al 35 %.
- **Completa:** La necrosis o reducción es superior al 85 % o el tumor ha desaparecido por completo.

Sin embargo, no todos los estudios constan de esta medida, por lo que tuvimos que guiarnos por los comentarios de los radiólogos en la imagen control, como *Hay una disminución de*

*tamaño considerable* (Parcial) o *En conjunto hay persistencia significativa de la masa tumoral* (No respuesta) .

Tras discutirlo con los radiólogos, concluimos que la aproximación que mejor se adaptaba a sus necesidades y a la realidad del proyecto era la de clasificar a los pacientes en dos grupos según esta respuesta al tratamiento QTN. Aquellos que no tienen respuesta forman el grupo de **no-respondedores** y se etiquetan con 0. Aquellos que tengan respuesta Parcial o Completa conformarán un único grupo de **respondedores** y se etiquetaron con 1.

Es importante destacar que la clasificación está basada únicamente en los comentarios de los informes radiológicos y no hacemos uso de los resultados histológicos de la biopsia.

### 4.1.1 Búsqueda de Pacientes

Desde las estaciones de trabajo del HURC, se realizó una búsqueda exhaustiva de pacientes con OS o SE. Se utilizó una herramienta propia al centro que permite filtrar los informes de radiología mediante palabras clave. En nuestro caso, buscamos por *sarcoma*, *Ewing*, *osteosarcoma* o *tumor óseo* para conformar así un listado inicial de **112 pacientes** potencialmente útiles. A continuación, realizamos un estudio individual de cada paciente, explorando su HCE y sus pruebas de imagen de RM, para comprobar si cumplía los criterios de inclusión diseñados para este estudio (Figura 4.1).

Primero, descartamos aquellos pacientes que no presentaban ningún tipo de tumor y que aparecían en nuestra búsqueda debido a sospechas iniciales. Seguidamente, descartamos un alto número de pacientes que no tenían OS o SE. El siguiente filtrado fue eliminar aquellos que o bien fueron intervenidos quirúrgicamente sin recibir QTN o no tenían información respecto a la QTN en su HCE. Después, no tuvimos en cuenta los pacientes que no tenían imagen de RM. De entre los que sí tenían pruebas de RM, eliminamos los que no tenían una imagen de control o información sobre la respuesta a la QTN. Finalmente, no tuvimos en cuenta aquellos pacientes que, aún cumpliendo todos los criterios anteriores, tenían lesiones muy dispersas o de confusa segmentación, para evitar entorpecer el entrenamiento de los modelos. Así, tras realizar este primer filtrado, redujimos el total a **17 pacientes** (Cuadro 4.1).



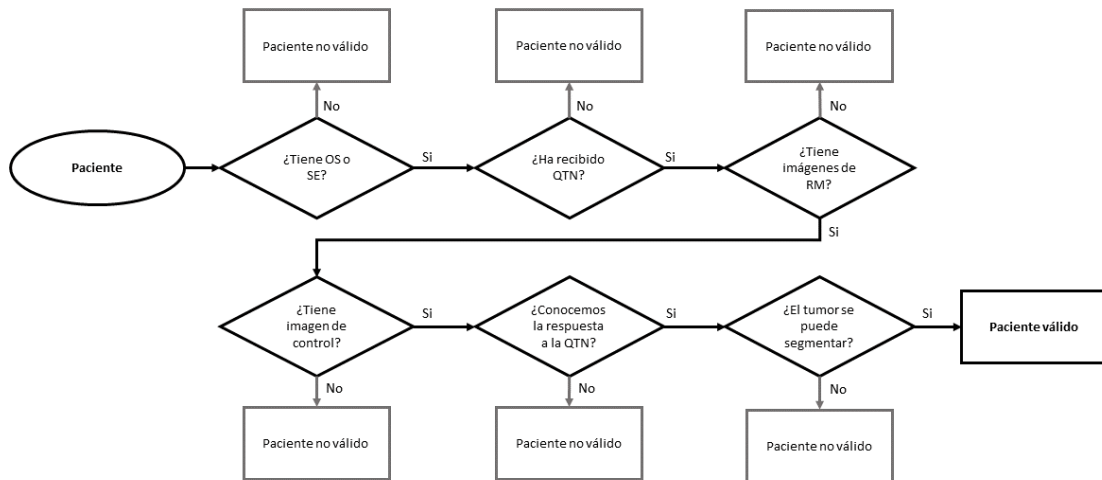


Figura 4.1: Diagrama de flujo para la selección de pacientes

A continuación, homogeneizamos la lista de pacientes. Primero, buscamos que la región afectada fuera lo más similar posible entre los pacientes. Como la mayoría presentaba lesión en el tren inferior, optamos por eliminar a sub-004 y a sub-007, que presentaban la lesión en el tren superior. (Cuadro 4.1) A continuación, buscamos que todos los pacientes tuvieran las mismas secuencias de RM. Observando el Cuadro 4.2, la única secuencia compartida por la gran mayoría de pacientes es T1, por lo que eliminamos los pacientes sub-001 y sub-002.

Tras realizar esta homogeneización, llegamos a un listado de **13 pacientes** (Cuadro 4.3) con el que finalmente podemos construir nuestra BD de imágenes de RM y segmentaciones.

#### 4.1. BASE DE DATOS

Paciente	Año	Edad	Sexo	Lesión	Región	Posición	Respuesta
sub-001	2016	22	F	SE	Peroné	Izq	Parcial
sub-002	2022	23	M	SE	Cadera	Izq	Parcial
sub-003	2022	13	F	SE	Pelvis	Izq	Parcial
sub-004	2015	36	M	SE	Antebrazo	Izq	Completa
sub-005	2013	20	F	SE	Tibia	Der	Parcial
sub-006	2021	17	M	SE	Pelvis	Izq	Parcial
sub-007	2021	9	M	OS	Cervical	Centro	Completa
sub-008	2016	15	M	OS	Rodilla	Der	Parcial
sub-009	2013	65	F	OS	Pelvis	Izq	Parcial
sub-010	2018	10	F	OS	Fémur	Izq	No
sub-011	2017	19	M	OS	Rodilla	Izq	Parcial
sub-012	2020	23	F	OS	Fémur	Der	No
sub-013	2020	41	F	OS	Pelvis	Der	No
sub-014	2017	16	M	OS	Rodilla	Der	Parcial
sub-015	2015	17	M	OS	Fémur	Izq	No
sub-016	2021	35	M	OS	Tibia	Der	No
sub-017	2021	22	F	OS	Pelvis	Der	No

Cuadro 4.1: Lista original de pacientes. El año indica la fecha de creación de la imagen control.

Paciente	T1	T2	T2_SPIR	PDW	DWI	AXIAL
sub-001		✓	✓		✓	
sub-002						
sub-003	✓	✓			✓	
sub-004	✓	✓	✓			✓
sub-005	✓				✓	
sub-006	✓	✓	✓		✓	✓
sub-007	✓	✓				✓
sub-008	✓	✓			✓	✓
sub-009	✓	✓				
sub-010	✓	✓		✓	✓	✓
sub-011	✓			✓		✓
sub-012	✓			✓	✓	✓
sub-013	✓				✓	
sub-014	✓	✓		✓		
sub-015	✓					
sub-016	✓	✓				
sub-017	✓			✓	✓	✓

Cuadro 4.2: Secuencias disponibles para cada paciente. T2 SPIR: Secuencia T2 con supresión de grasa, PDW: Densidad Protónica, DWI: Secuencia de difusión, AXIAL: Secuencia de contraste.

Paciente	Año	Edad	Sexo	Lesión	Región	Posición	Respuesta
sub-003	2022	13	F	SE	Pelvis	Izq	1
sub-005	2013	20	F	SE	Tibia	Der	1
sub-006	2021	17	M	SE	Pelvis	Izq	1
sub-008	2016	15	M	OS	Rodilla	Der	1
sub-009	2013	65	F	OS	Pelvis	Izq	1
sub-010	2018	10	F	OS	Fémur	Izq	0
sub-011	2017	19	M	OS	Rodilla	Izq	1
sub-012	2020	23	F	OS	Fémur	Der	0
sub-013	2020	41	F	OS	Pelvis	Der	0
sub-014	2017	16	M	OS	Rodilla	Der	1
sub-015	2015	17	M	OS	Fémur	Izq	0
sub-016	2021	35	M	OS	Tibia	Der	0
sub-017	2021	22	F	OS	Pelvis	Der	0

Cuadro 4.3: Listado de pacientes que se utilizarán en el entrenamiento de algoritmos. Solo la secuencia T1 será utilizada.

## 4.1.2 Construcción de la Base de Datos

Comenzamos guardando todas las imágenes de cada paciente en formato DICOM [29] utilizando MicroDicom [33], siguiendo la estructura de Brain Imaging Data Structure (BIDS) [34], muy popular para BBDD médicas (Figura 4.2). MicroDicom está ligado al PACS [35] del hospital y nos permite extraer las secuencias directamente. Todas las imágenes fueron anonimizadas previamente a su descarga. Después, fueron transformadas automáticamente a formato NIFTI [36] utilizando Python y su librería *dicom2nifti* [37].

Las imágenes DICOM fueron segmentadas manualmente en 3DSlicer y las segmentaciones fueron guardados en formato STL. Después, se extrajo la máscara binaria en formato NIFTI automáticamente con Python, a partir de las segmentaciones STL de cada paciente y utilizando la secuencia T1 como base (Apéndice A). Los pacientes se guardaron en dos carpetas, según su clase.

Podemos visualizar algunas imágenes en las Figuras 4.3, 4.4. Aunque solo utilicemos la secuencia T1, la BD contiene el resto de secuencias de cada paciente (Cuadro 4.2) para futuros proyectos. Llama la atención que no todos los pacientes tengan la secuencia AXIAL, cuando el

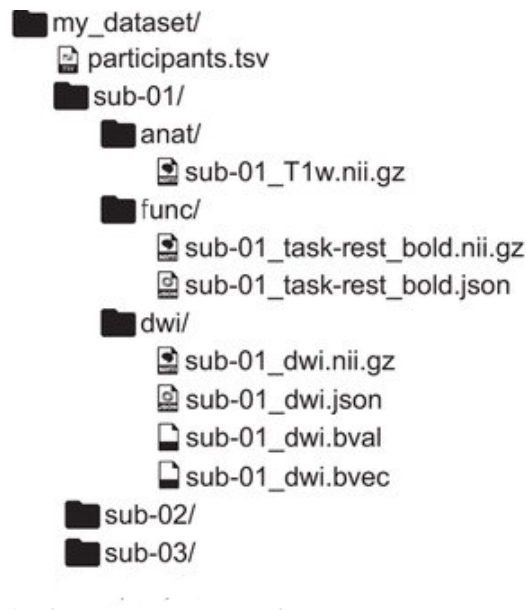


Figura 4.2: Formato BIDS para el archivado de imagen médica [34] .

protocolo de actuación para OS y SE incluye prueba de RM sin y con contraste.

## 4.2 Extracción de Características Radiómicas

A continuación, presentamos los tipos de características radiómicas que extraeremos con Pyradiomics:

- **Características de primer orden:** Son aquellas resultado de analizar el histograma global de la imagen, como el nivel medio de gris o la varianza. Son las más sencillas [20].
- **Gray-Level Cooccurrence Matrix (GLCM):** Analizan la dependencia espacial entre los diferentes niveles de gris en hasta 13 direcciones en imágenes 3D. Entre ellas, se incluye el contraste, una medida de las diferencias entre los niveles de gris entre parejas de píxeles o vóxeles [38].
- **Gray-Level Run-length Matrix (GLRLM):** Se basan en el análisis de conjuntos de

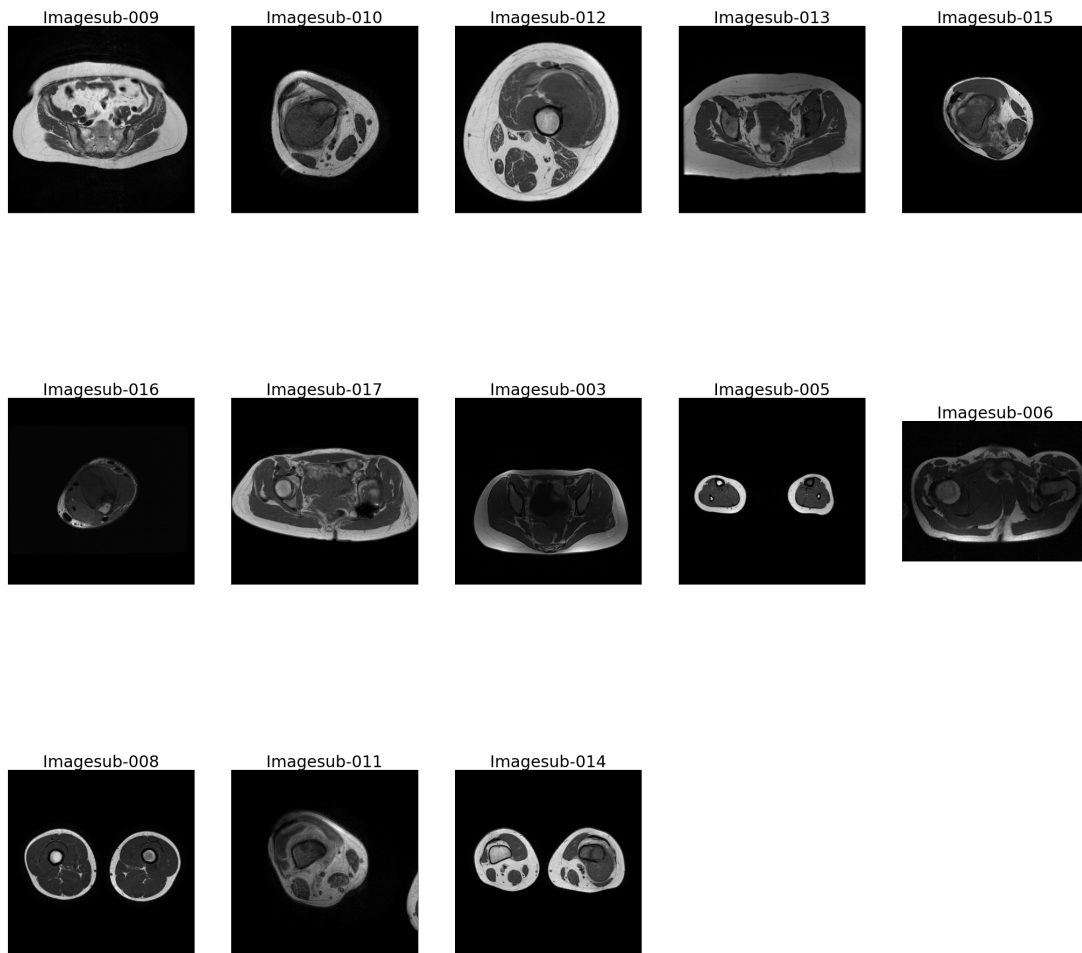


Figura 4.3: Secuencias T1 de los 13 fragmentos que forman parte del estudio.

píxeles o vóxeles adyacentes con un nivel similar de gris, lo que llamamos gray run. Pueden ser calculadas en varias direcciones. Entre ellas, se incluye la uniformidad de los niveles de gris entre los diferentes gray run [39] [40].

- **Gray-Level Size Zone Matrix (GLSZM):** Similares a GLRLM, pero en este caso se tiene en cuenta grupos o zonas de píxeles o vóxeles con el mismo nivel de gris. Se calculan cambiando el tamaño de cada zona. Entre ellas, se encuentra el porcentaje de píxeles o vóxeles que forman parte de una zona [41].
- **Neighborhood Gray-Level Dependence Matrix (GLDM) :** Se basan en las relaciones

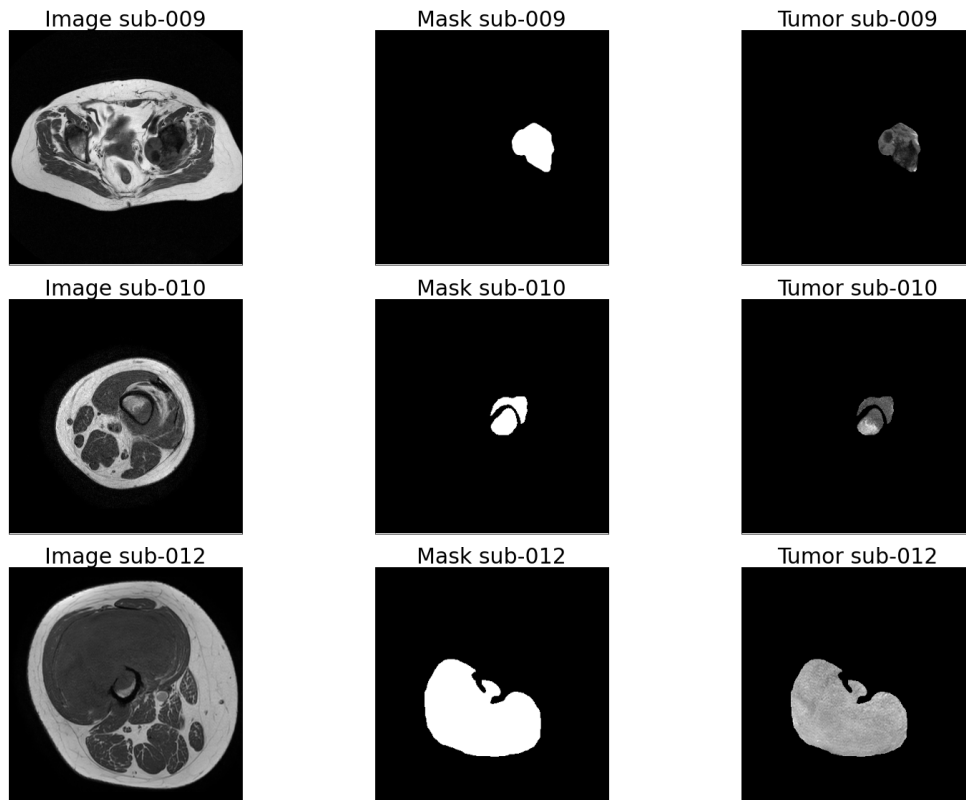


Figura 4.4: Secuencia T1, Máscara binaria y ROI para 3 pacientes (sub-009, sub-010 y sub-12). Listado completo en Apéndice A.

que existen entre un único píxel o vóxel y la región de su alrededor (vecindario). Pueden calcularse características como el énfasis en dependencias altas o la no-uniformidad en los niveles de gris [42].

*Las características GLCM, GLRLM, GLSZM y GLDM aportan información sobre la textura de la imagen, por eso son también conocidas como características de segundo orden.*

- **Shaped-Based Features:** Son aquellas relacionadas con la geometría de la región de interés (ROI) y son conceptualmente muy sencillas. Se incluyen el diámetro, el área o la esfericidad, entre otras [20].
- **Transform-Based Features:** Son aquellas obtenidas filtrando la imagen y llevándolo a un dominio diferente. Se incluyen la transformada de Fourier, de Gabor o Laplacian of Gaussian (LoG), entre otras. Un filtro paso alto permite destacar los bordes y detalles de la imagen y un paso bajo la suaviza [20]. La transformada Wavelet, muy popular en

radiómica por sus buenas prestaciones, realiza un análisis de tiempo y frecuencia que nos permite estudiar todos los patrones de la imagen [43].

En este proyecto, los filtros aplicados a las imágenes para obtener Transform-Based Features son: LoG, Wavelet, Square, SquareRoot, Logarithm y Exponential [44].

## 4.3 Selección de Características

Cuando una BD tiene muchas características, decimos que es de alta dimensionalidad, por ejemplo, un problema con 99 características es de 99 dimensiones. Tiene sentido pensar que a más dimensiones, mejores van a ser las prestaciones de nuestro modelo, ya que tenemos más información disponible sobre nuestro problema. Sin embargo, en la realidad esto no ocurre: Los tiempos de cómputo aumentan y el modelo se sobreentrena (*overfitting*) [45], generando peores resultados al ajustarse demasiado bien a los datos. Es lo que llamamos como **Maldición de la Dimensionalidad** [46].

La demostración matemática detrás de la Maldición no es compleja [47] y se resume en que los datos se vuelven más dispersos a medida que aumenta la dimensionalidad del problema, entorpeciendo la búsqueda de patrones. Por otro lado, una dimensionalidad alta puede incluir características redundantes o irrelevantes.

Para librarnos de esta maldición, podemos aumentar el número de observaciones de nuestra BD para que así los datos estén más agrupados. La regla general es la de tener un mínimo de 10 observaciones por cada característica. Lamentablemente, obtener nuevas observaciones no siempre es posible y los tiempos de cómputo aumentarían.

Encontrar entonces el punto en el que aumentar la dimensionalidad solo condena a nuestros modelos es esencial (Figura 4.5). Es por ello que necesitamos realizar una selección de características, manteniendo aquellas relevantes para el problema. A continuación, presentamos los métodos que hemos utilizado para evitar la Maldición de la Dimensionalidad.

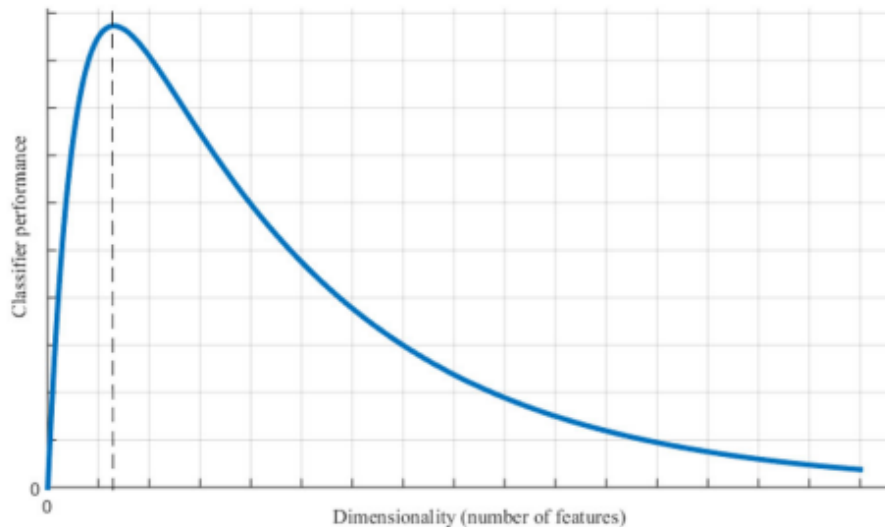


Figura 4.5: Aumentar el número de características es útil hasta cierto punto. Después, las prestaciones empeoran notablemente [48].

### 4.3.1 Correlación

La **correlación** [49] es una medida de la relación lineal que existe entre dos variables. Se interpreta de la siguiente manera:

- $r < 0$  → Correlación negativa: Relación invertida de variables; A valor alto de una variable, a la otra le corresponden bajos, y viceversa. Cuando  $r = -1$ , hablamos de una *correlación negativa perfecta*: relación lineal de pendiente negativa.
- $r > 0$  → Correlación positiva: Relación directa de variables; A valores altos de una variables le corresponden valores altos de la otra, y viceversa. Cuando  $r = 1$ , hablamos de una *correlación positiva perfecta*: relación lineal de pendiente positiva.
- $r = 0$  → Variables incorreladas.

Es decir, a la hora de seleccionar características, nos interesa eliminar una de los componentes de las parejas de características que tengan una correlación cercana a 1, ya que no aportan información diferente. Se presenta en forma de matriz de correlación  $R$ :



$$R = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1n} \\ r_{21} & 1 & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & 1 \end{pmatrix}$$

Donde  $r_{i,j}$  es el coeficiente de correlación de Pearson entre las característica  $i$  y  $j$ .

### 4.3.2 Información Mutua

La **Información Mutua** (IM) [50] es una medida que, a diferencia de la correlación, nos permite conocer si existe cualquier tipo de relación entre las características, ya sea lineal o no-lineal. Para entenderla, primero debemos de conocer qué es la Entropía.

La **Entropía** [50], en teoría de la información, es una medida que nos permite conocer la incertidumbre de un evento. En términos sencillos, equivale a medir la *cantidad de preguntas de si o no que debemos hacer para poder describir un evento en su totalidad*. Es decir, a mayor Entropía, mayor es el número de preguntas que tenemos que hacer y más misterioso es ese evento.

La IM nos dice entonces cuánta incertidumbre reducimos de un evento al conocer detalles sobre otro. Siguiendo la línea de la entropía, la IM puede entenderse mejor como *la cantidad media de preguntas que es capaz de responder un evento sobre ese otro evento misterioso*. El valor mínimo es 0 (no conoce absolutamente nada) y el valor máximo (lo conoce absolutamente todo) depende de la Entropía de las fuentes.

Entonces, aquellas características que tengan mayor IM conocerán más sobre la clase y por ello serán mejores a la hora de entrenar el modelo.

### 4.3.3 Principal Component Analysis

En **Principal Component Analysis** (PCA) [47], los datos se dividen según su variabilidad en lugar de según su proximidad. Para conseguir esto, PCA realiza un Singular Value Decom-

### 4.3. SELECCIÓN DE CARACTERÍSTICAS

position (SVD) [51] que permite identificar los ejes en los que encontramos mayor variabilidad de cada dimensión (Figura 5.5). Posteriormente, los datos se proyectan sobre estos ejes. Tendremos un eje por cada dimensión de los datos originales. Si tenemos más características que observaciones, el número máximo de ejes será igual al número de observaciones.

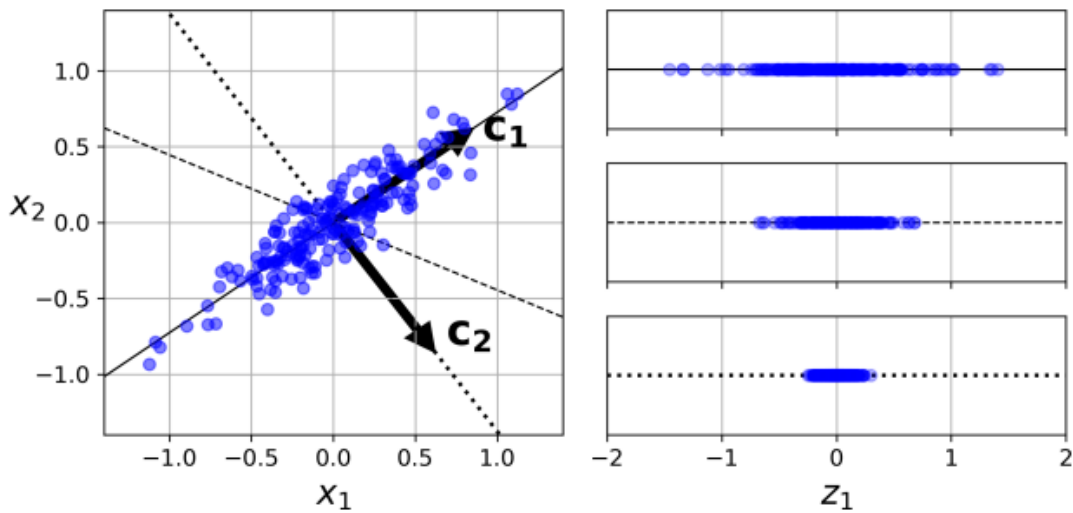


Figura 4.6: Para que el PCA funcione correctamente, los datos deben ser normalizados y centrados, ya que la diferencia entre los rangos dinámicos de las características puede alterar la variabilidad [47].

PCA permite seleccionar características basándose en la idea de que una mayor variabilidad entre los datos entrena mejor los algoritmos, es decir, podemos prescindir de aquellas dimensiones que no aporten variabilidad. Para identificar la proporción de varianza que tiene nuestra BD a lo largo de cada eje, podemos calcular la varianza explicada y la varianza acumulada (Figura 4.7), en la que vemos el porcentaje de 1 Ambas gráficas muestran la misma idea, pero de manera diferente. Generalmente, queremos guardar las dimensiones que representan alrededor del 95 % de la variabilidad.

Sin embargo, debemos de tener cuidado. El PCA únicamente tiene en cuenta la variabilidad de los datos, por lo que pueden existir características con poca variabilidad explicada que sí guarden información relevante para nuestro problema. Además, al transformar los datos, no podemos conocer su naturaleza y perdemos explicabilidad.

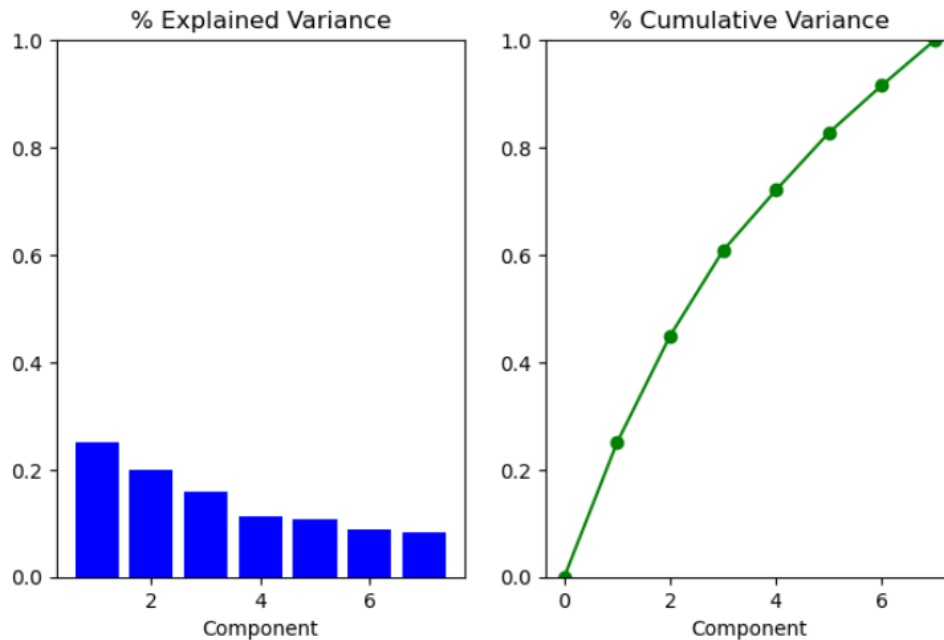


Figura 4.7: Ejemplo de varianza explicada y varianza acumulada en un problema de 7 dimensiones. Puede verse que casi todos los ejes comparten una variabilidad similar, por lo que son igualmente relevantes.

## 4.4 Entrenamiento Modelos

En esta sección, presentamos información sobre los algoritmos de IA que estudiaremos una vez realizada la selección de características, así como de las herramientas relacionadas con su entrenamiento y validación.

### 4.4.1 Cross Validation

**Cross Validation** (CV) [52], o validación cruzada, es un técnica que nos permite mejorar los resultados de nuestros modelos cuando el número de observaciones disponibles en la BD es escaso. El conjunto de datos se divide aleatoriamente en dos: conjunto de entrenamiento (train) y conjunto de validación (validation). El primero entrena el modelo y el segundo sirve para obtener predicciones y figuras de mérito.

De esta manera, no utilizamos todas las observaciones para entrenar el modelo (no tendrí-

mos manera de evaluar el modelo) ni utilizamos observaciones del conjunto de entrenamiento para evaluarlo.

Existen diversas maneras de computar CV. En este proyecto, se utilizará **Repeated Stratified K-Fold**. [53] (Figura 4.8). Los datos se dividen en  $k$  segmentos iguales (folds), utilizando  $k - 1$  para entrenar y el restante para validar y obtener la figura de mérito. El proceso se repite  $k$  veces, alternando los conjuntos de entrenamiento y validación hasta obtener  $k$  figuras de mérito diferentes. Finalmente, se realiza la media de estas figuras de mérito, con la que evaluamos el modelo. Es *Stratified* para asegurarnos de que en cada conjunto las clases estén equilibradas. Es *Repeated* porque el proceso completo se repite las veces que indiquemos, aportando mayor variabilidad.

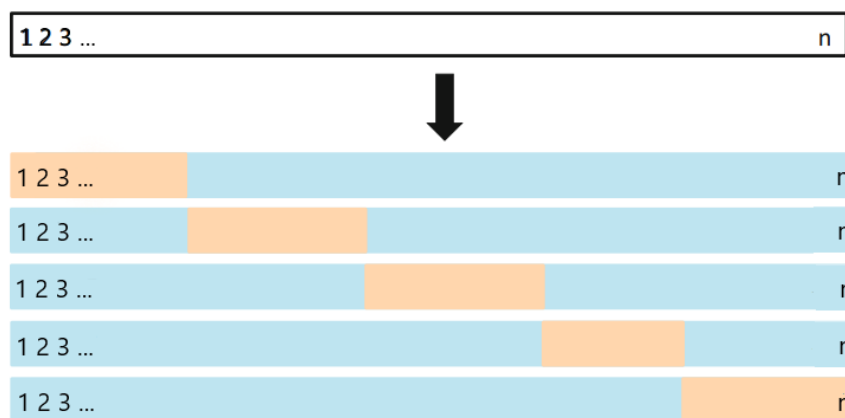


Figura 4.8: División de los conjuntos *train* (azul) y *validation* (naranja) en CV de 5-Folds para una repetición [52]

## 4.4.2 Regresión Logística

El modelo de Regresión Logística (RL) [54] es un algoritmo de IA que nos permite conocer la probabilidad que tiene una observación de pertenecer a una clase. De manera similar a los modelos de regresión lineal [55], RL predice el resultado como la combinación lineal de las características de entrada, pero introduciendo una función logística.

El clasificador se obtiene estableciendo un umbral en la probabilidad obtenida por el RL [55]. Por ejemplo, si la probabilidad de una observación de pertenecer a la clase 1 es superior

al 50 %, se marca 1 y si es inferior, se marca 0.

### 4.4.3 Random Forest

Random Forest (RF)[56] es un modelo de IA utilizado tanto en tareas de clasificación como de regresión. Para entender cómo funciona el bosque debemos entender primero cómo funciona el árbol.

Un árbol de decisión (Decision Tree) [57], como vemos en la Figura 4.9, busca predecir la clase a través de diferentes nodos. A partir del total, preguntamos a cada observación si una de sus características cumplen una condición determinada y las separamos. Así, poco a poco el árbol se construye, hasta que llegamos a un nodo hoja final en el que se identifica la clase.

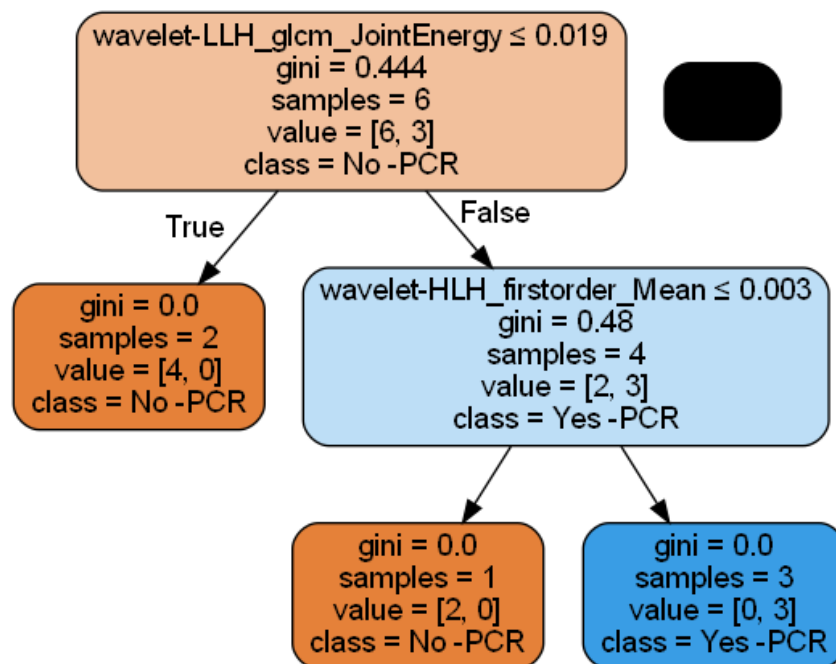


Figura 4.9: Ejemplo de un árbol de decisión

RF surge entonces como la combinación de los resultados de múltiples árboles de decisión, creados de manera aleatoria para explorar todas las combinaciones de preguntas posibles, en busca del mejor resultado. Esto nos permite realizar una selección embebida (*ensemble*) de las características. Es el propio RF el que dictamina intrínsecamente cuáles son las características

más relevantes, convirtiéndolo en uno de los modelos más robustos de IA.

### 4.4.4 Multilayer Perceptron

El Perceptrón Multicapa (MLP) [58] [59] es un modelo de IA que busca imitar el comportamiento de un cerebro vivo mediante la interconexión de cientos de neuronas, tal y como las encontraríamos en la naturaleza. El modelo matemático de neurona artificial más utilizado es el mostrado en la Figura 4.10 ((a)). Tienen 3 partes bien diferenciadas, entrada, cuerpo y salida, que se asemejan a la estructura de una neurona real (Figura 4.10 ((b)) ).

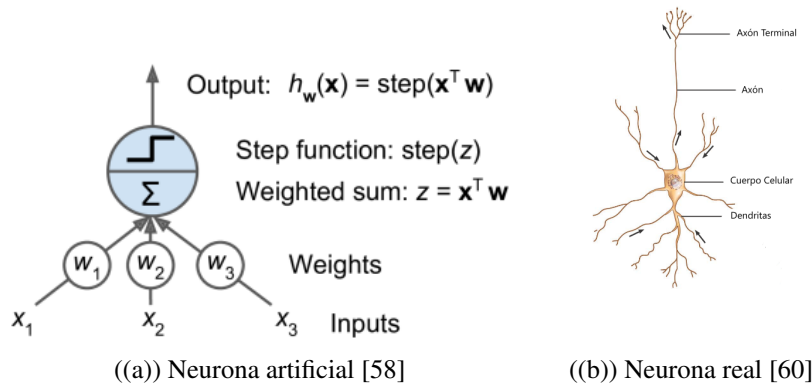


Figura 4.10: Comparación de una neurona real y su modelo artificial

Cada entrada (dendrita) de la neurona artificial es acompañada por un peso que ajusta su valor. El cuerpo (cuerpo celular), computa una suma ponderada que es evaluada por una función escalón (*Step function*). La salida (axón) es el resultado de esta función. Cuando varias neuronas se conectan entre ellas como en la Figura 4.10 ((a)), decimos que forman un único modelo llamado Perceptrón.

Un MPL (Figura 4.11) es entonces un conjunto de varios Perceptrones. Tiene una primera capa de entrada, un número indefinido de capas ocultas y una capa de salida.

MLP aprende mediante el algoritmo iterativo de *backpropagation*. Se tiene en cuenta una pequeña porción de observaciones (*batch*) para cada iteración (*epoch*). En cada *epoch*, el *batch* pasa por todas las capas de neuronas hasta generar una salida (*forward pass*) y se calcula el error de la red. Después, el algoritmo avanza hacia atrás desde la salida, mediante la regla

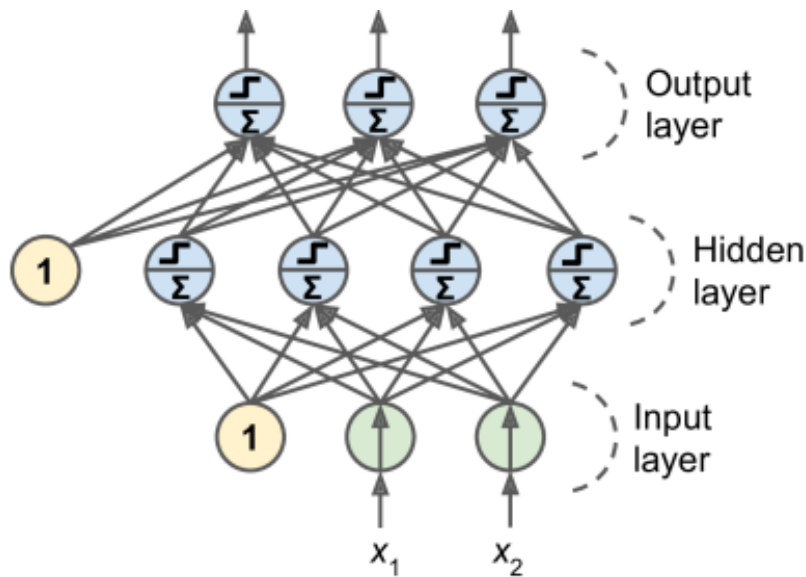


Figura 4.11: Ejemplo de una estructura MLP [58].

de la cadena de cálculo, y es capaz de calcular cuánto ha contribuido cada neurona al error (*backpropagation*). Finalmente, los pesos son ajustados mediante descenso por gradiente para reducir este error y el proceso comienza de nuevo.

#### 4.4.5 Support Vector Machine

**Support Vector Machine** (SVM) [61] [62] es un algoritmo de IA utilizado principalmente en tareas de clasificación. Tal y como vemos en la Figura 4.12 ((a)), SVM funciona creando una frontera entre las observaciones de cada clase, con el objetivo de maximizar el espacio entre ellas. Como los rangos dinámicos de las características influyen en el resultado de SVM, es necesario realizar un normalizado previo.

Idealmente, este espacio lo determinan las observaciones que se encuentran en las fronteras. Sin embargo, en la práctica esto solo lleva a que se produzca un mal ajuste del modelo. Por ello, se introduce el hiperparámetro  $C$ , que regula el compromiso entre el tamaño del espacio y el número de observaciones que quedan dentro de él (Figura 4.12 ((b)) ).

SVM es un modelo muy flexible. Nos permite encontrar los valores atípicos e incluso introducir fronteras no lineales (Figura 4.13) para aquellas BD más complejas.

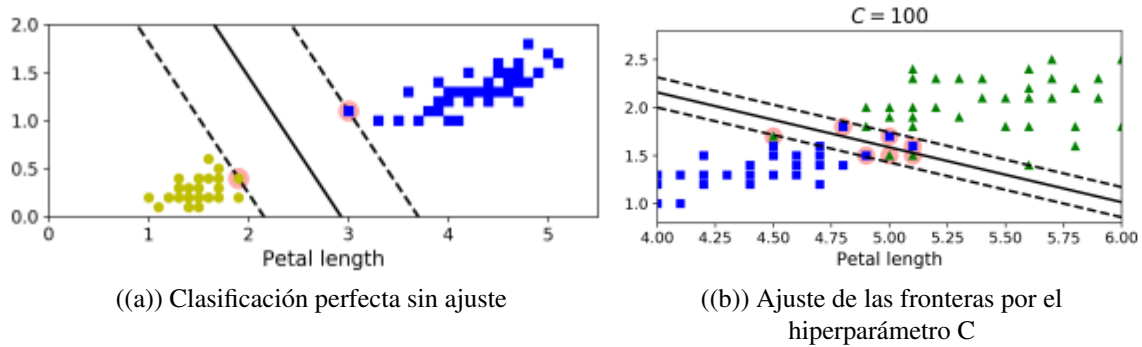


Figura 4.12: Ejemplo de fronteras lineales en SVM [61]

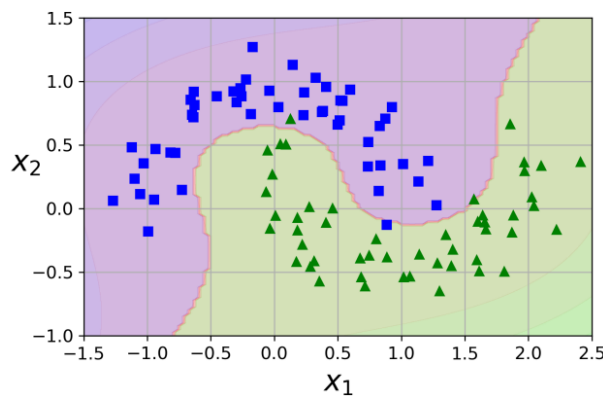


Figura 4.13: Ejemplo de frontera no lineal para clasificación binaria [61]

### 4.4.6 Búsqueda de hiperparámetros

Para afinar las prestaciones de los modelos, es necesario realizar una búsqueda de los hiperparámetros más adecuados para nuestra BD. Para ello, empleamos dos técnicas: *RandomizedSearchCV* y *GridSearchCV*[63]. Para ello, se hace uso de CV (Sección4.4.1).

Primero, *RandomizedSearchCV* realiza una exploración aleatoria de una matriz de hiperparámetros establecida por el usuario, ajusta el modelo y devuelve aquellos hiperparámetros que obtengan la figura de mérito superior. Como no selecciona todos los elementos de la matriz, *RandomizedSearchCV* tiene la función realizar una búsqueda superficial para encontrar los órdenes de magnitud de los hiperparámetros.

A continuación, *GridSearchCV* se encarga de hacer una exploración profunda de lo obtenido



en el paso anterior, ajustando el modelo y calculando las figuras de mérito por cada elemento de la matriz. Los hiperparámetros hallados en esta búsqueda son los utilizados finalmente para ajustar los algoritmos.

## 4.5 Explicabilidad

En ocasiones, entender lo que está ocurriendo internamente dentro de un algoritmo de IA puede ser complicado, haciendo que parezcan cajas negras. La librería SHAP (SHapley Additive exPlanations) [64] de Python es una herramienta que nos permite interpretar las predicciones de cualquier modelo de IA, para hacerlos más transparentes y legibles. El razonamiento matemático detrás de SHAP se escapa al alcance de este proyecto, por lo que solo mencionamos que hace uso de los valores Shapley de Teoría de Juegos cooperativa.[65] [66].

Los valores SHAP permiten cuantificar la aportación de cada característica al resultado, por cada observación introducida. Así, podemos responder a preguntas como *¿Qué característica es la que más influye en la predicción del modelo?* o *¿Qué le ocurre al modelo si cambiamos el valor de una característica?*

Para entender los valores SHAP para todo el conjunto de observaciones, presentamos primero dos ejemplos para una única observación en un problema de predicción de cáncer de cuello de útero (Figura 4.14).

El punto de referencia para los valores SHAP es la media de todas las predicciones. Las características que contribuyen positivamente aumentan la probabilidad de padecer cáncer. Las características que contribuyen negativamente disminuyen la probabilidad de padecer cáncer. Así, vemos que la primera paciente tiene una probabilidad del 0.06 de padecer cáncer de cuello uterino, ya que las características positivas (STDs..number=2.0), que aumentan la probabilidad, se contrarrestan con las negativas (Hormonal.Contraceptives..years=7.0), que la disminuyen. La segunda paciente tiene una probabilidad de 0.71 y vemos que las características que más han intervenido en este resultado, contribuyendo positivamente, son los años como fumadora (Smokes..year=34) y la edad (Age=51).

#### 4.5. EXPLICABILIDAD

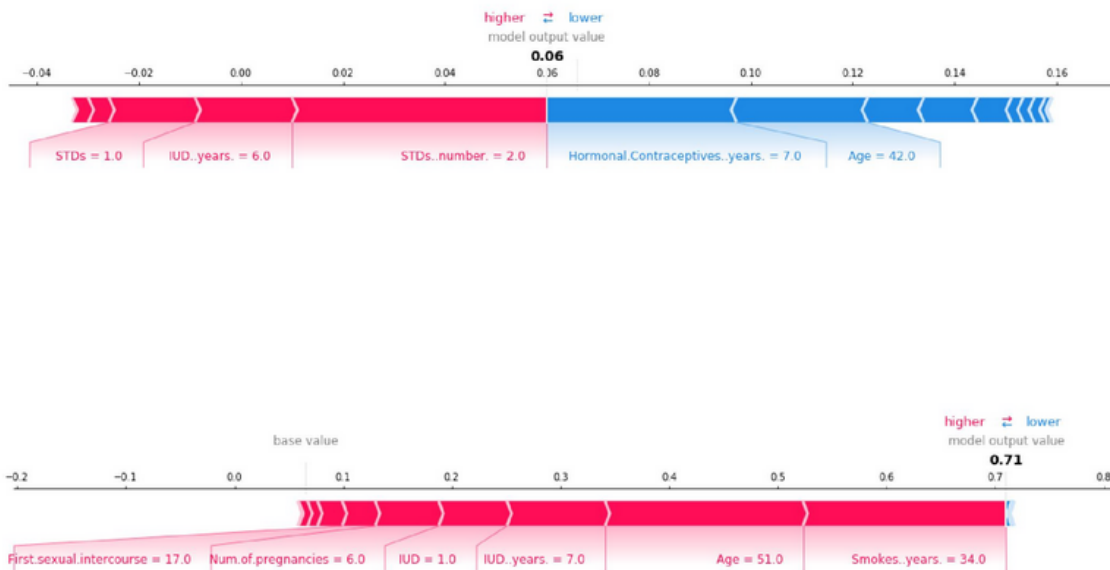


Figura 4.14: Ejemplo de valores SHAP para dos observaciones de un modelo de RF [65]

Para mostrar los resultados del conjunto, utilizamos summary plot (Figura 4.15). El summary plot ordena las características según su importancia y muestra su efecto sobre la predicción del modelo. Cada punto de la gráfica representa un valor SHAP para una característica y una observación. El eje-X muestra los valores SHAP y el eje-Y las características de nuestra BD. Aquellos valores SHAP que se solapan son acumulados a lo largo del eje-Y para mostrar su distribución. El color representa el valor de la característica, rojo para altos y azul para bajos.

En este ejemplo, la característica más importante es Hormonal.Contraceptives..years. Mayoritariamente contribuyen negativamente a padecer cáncer de útero, es decir, disminuye la probabilidad cuando esta característica toma valores bajos. (Cúmulo azul en la gráfica)

Con estos ejemplos, hemos mostrado cómo se pueden explicar los modelos de IA utilizando SHAP. Sin embargo, es importante tener en cuenta que los valores SHAP únicamente aportan información sobre la relevancia de las características, nunca sobre la calidad de los resultados, por lo que necesitamos de otras herramientas para evaluar los modelos al completo.

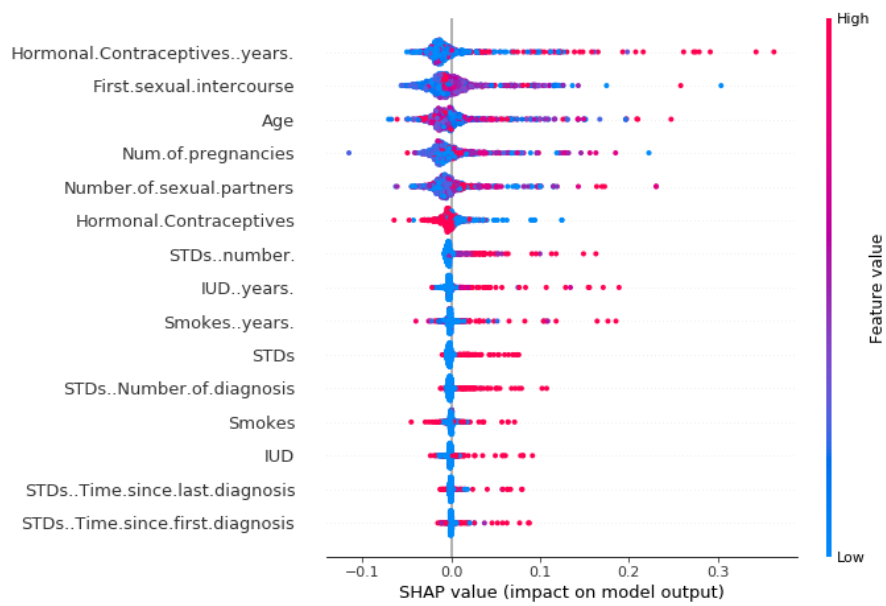


Figura 4.15: Ejemplo de valores SHAP para el conjunto de observaciones de un modelo de RF [65]



# Capítulo 5

---

## Experimentación y resultados

En esta sección, presentamos los resultados obtenidos para cada modelo comentado en la Sección 4.4. A partir de las imágenes de T1 y la máscara correspondiente a cada paciente, extraemos automáticamente, utilizando PyRadiomics, un total de **1391 características** que procedemos a filtrar por los métodos vistos en la Sección 4.3. En el caso de MLP, estudiamos los resultados sin normalizar la entrada (MLP) y normalizando la entrada (MLP-Norm). Las matrices de confusión pueden ser consultadas en el Apéndice B.

Cada algoritmo fue entrenado y validado siguiendo los mismos pasos:

- Dividir el conjunto de datos en características y clase.
- Buscar los mejores hiperparámetros mediante *RandomizedSearchCV* y *GridSearchCV* (Sección 4.4.6).
- Entrenar y validar los algoritmos con el conjunto completo de observaciones, utilizando los hiperparámetros encontrados, mediante *RepeatedStratifiedKFold* (Sección 4.4.1). En nuestro caso, utilizamos *3-folds* y 20 repeticiones.
- Calcular las figuras de mérito por cada *fold*, obteniendo un total de 60.
- Calcular los valores SHAP (Sección 4.5) de cada modelo.
- Calcular la media y la desviación típica (STD) de las figuras de mérito.
- Calcular la matriz de confusión, sumando las matrices de confusión de cada *fold* y dividiendo entre el número de repeticiones.
- Calcular la curva ROC y su área bajo la curva (AUC).

## 5.1 Correlación

Una vez obtenidos los coeficientes de correlación entre cada par de características, establecemos un **umbral de 0.6**, obtenido empíricamente, para seleccionar las características.

### 5.1.1 Características Seleccionadas

Característica	Tipo	Imagen
VoxelVolume	Shape	Original
Sphericity	Shape	Original
10Percentile	Primer Orden	Original
Kurtosis	Primer Orden	LoG (Sigma=0.5)
HLH Mean	Primer Orden	Wavelet
LLH Mean	Primer Orden	Wavelet
LLH Correlation	GLCM	Wavelet
LLH lmc1	GLCM	Wavelet

Cuadro 5.1: Características seleccionadas mediante correlación. El orden mostrado no indica la importancia.

Eliminamos todas aquellas características cuyo coeficiente de correlación absoluto sea superior al umbral, es decir, si por ejemplo *originalshapeVoxelVolume* y *originalshapeMeshVolume* tienen una correlación de 1, guardamos solo una de ellas. Reducimos así el problema a 8 dimensiones. (Cuadro 5.1, Figura 5.1).

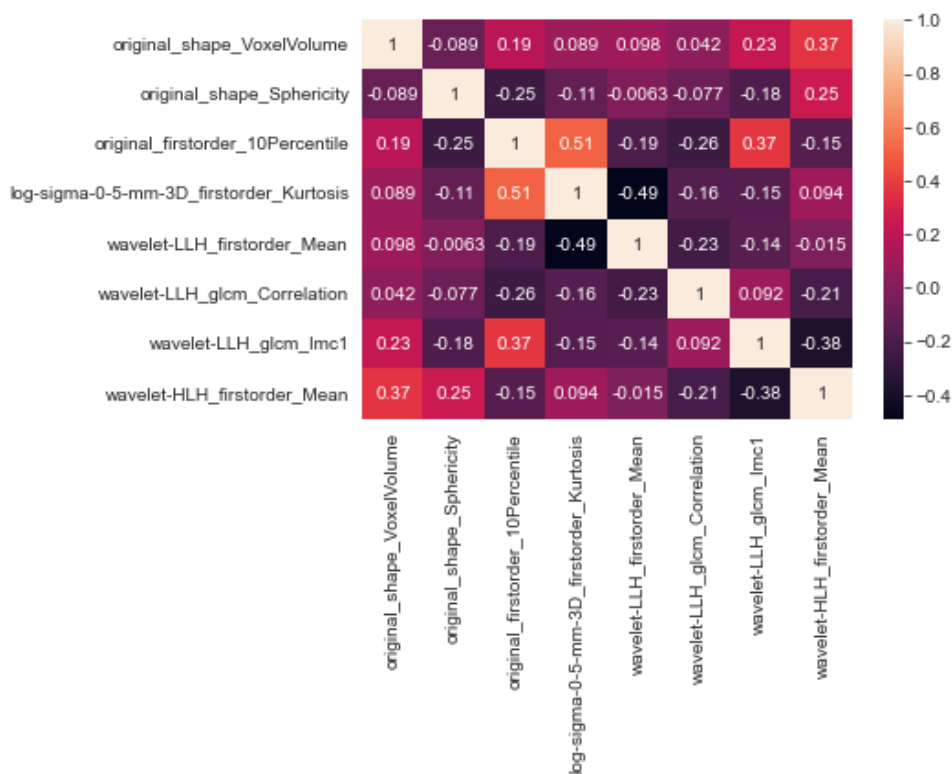


Figura 5.1: Coeficientes de correlación de las 8 características seleccionadas.

## 5.1.2 Resultados

Seguidamente, entrenamos y validamos los algoritmos de IA y obtenemos las figuras de mérito (Cuadro 5.2). En general, los clasificadores tienen una exactitud cercana a 0.5 a la hora de predecir la respuesta al tratamiento QTN. Los modelos muestran menor sensibilidad que especificidad, prediciendo así mejor a los pacientes no-respondedores. Los valores de ambas son bajas, no se llega al 60% de probabilidad de predecir un paciente respondedor como respondedor. La STD es alta para todos los modelos, al tener en cuenta los rangos de magnitud de las figuras de mérito. Los tiempos de cómputo son bajos, no superando el minuto en ninguno de ellos.

El mejor modelo es RF. Presenta una exactitud de 0.63, prediciendo correctamente el 63% de las observaciones. Clasifica mejor aquellos pacientes no-respondedores, al tener una especificidad y valor predictivo negativo (VPN) superiores a la sensibilidad y al valor predictivo

positivo (VPP).

Clasificador	Exactitud		Sensibilidad		Especificidad		Correlación VPP		VPN		rocAUC		Tiempo (s)
	Mean	(STD)	Mean	(STD)	Mean	STD	Mean	STD	Mean	STD	Mean	STD	
RL	<b>0.57</b>	(0.18)	<b>0.53</b>	(0.36)	<b>0.60</b>	(0.30)	<b>0.47</b>	(0.32)	<b>0.60</b>	(0.29)	<b>0.56</b>	(0.18)	<b>0.31</b>
RF	<b>0.63</b>	(0.2)	<b>0.56</b>	(0.35)	<b>0.71</b>	(0.33)	<b>0.59</b>	(0.37)	<b>0.62</b>	(0.37)	<b>0.63</b>	(0.2)	<b>4.86</b>
MLP	<b>0.5</b>	(0.07)	<b>0.39</b>	(0.48)	<b>0.60</b>	(0.48)	<b>0.19</b>	(0.23)	<b>0.33</b>	(0.26)	<b>0.5</b>	(0.03)	<b>11.39</b>
MLP - Norm	<b>0.53</b>	(0.22)	<b>0.53</b>	(0.37)	<b>0.54</b>	(0.33)	<b>0.46</b>	(0.34)	<b>0.54</b>	(0.32)	<b>0.52</b>	(0.22)	<b>39.39</b>
SVM	<b>0.51</b>	(0.18)	<b>0.57</b>	(0.41)	<b>0.47</b>	(0.36)	<b>0.39</b>	(0.29)	<b>0.48</b>	(0.36)	<b>0.52</b>	(0.18)	<b>5.31</b>

Cuadro 5.2: Figuras de mérito de los modelos, utilizando correlación como método de selección de características.

Mostramos las curvas ROC en la Figura 5.7. Los resultados muestran áreas bajo la curva (AUC) cercanas a 0.5, a excepción de RF, indicando que los resultados de la clasificación son resultado del azar y no de modelos robusto.

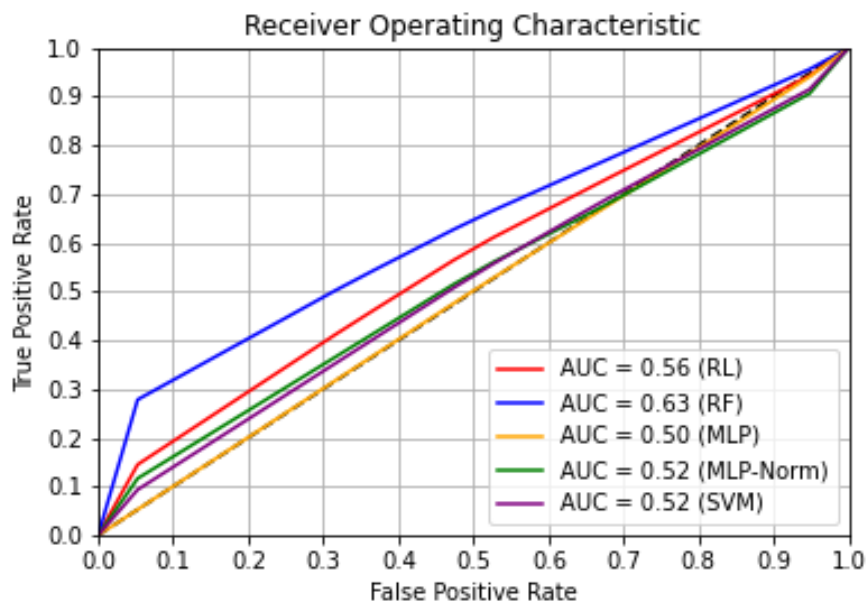


Figura 5.2: Curvas ROC y AUC para los modelos entrenados empleando correlación.

## 5.2 Información Mutua

Calculamos la IM de cada característica en relación a la clase y establecemos un **umbral de 0.266**, obtenido igualmente de manera empírica.



### 5.2.1 Características Seleccionadas

En la Figura 5.3 mostramos las 35 mejores características según IM. Aplicamos el umbral de clasificación y reducimos el problema a 9 dimensiones. (Cuadro 5.3).

Característica	Tipo	Imagen
Skewness	Primer Orden	LoG (Sigma=1)
SmallDependenceLowGrayLevelEmphasis	GLDM	Square
LLH LargeDependenceEmphasis	GLDM	Wavelet
LLH DifferenceVariance	GLCM	Wavelet
LLH DependenceVariance	GLDM	Wavelet
HHL DependenceVariance	GLDM	Wavelet
HHL SumEntropy	GLCM	Wavelet
LHH ClusterShade	GLCM	Wavelet
LLL MaximumProbability	GLCM	Wavelet

Cuadro 5.3: Características seleccionadas con IM. El orden mostrado no indica su importancia.

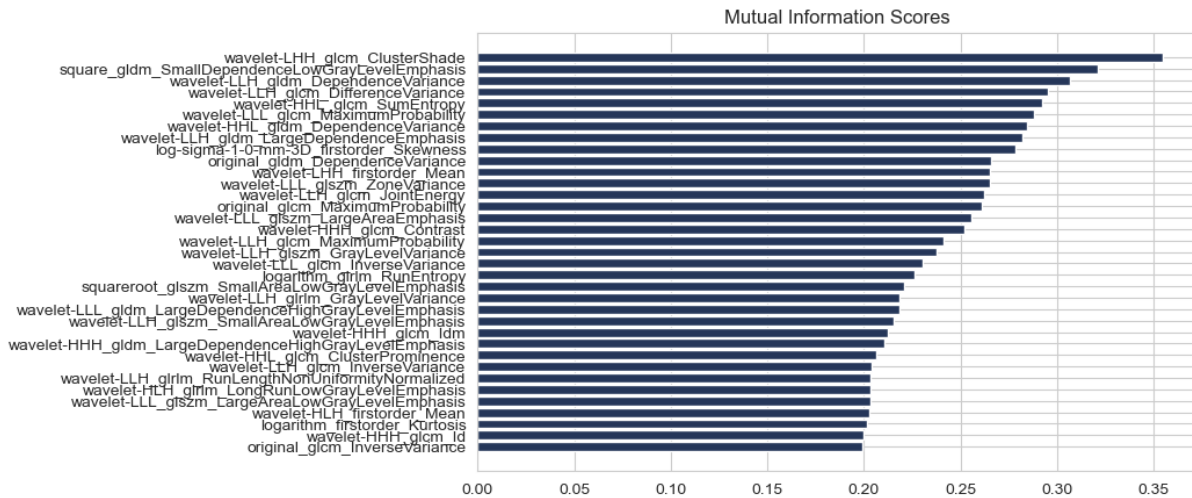


Figura 5.3: Las 35 características con mayor IM.

### 5.2.2 Resultados

Tras entrenar y validar los algoritmos de IA, mostramos los resultados en el Cuadro 5.4. Las figuras de mérito muestran cierta variabilidad entre los modelos. La exactitud presenta valores

## 5.2. INFORMACIÓN MUTUA

cercanos a 0.5, al igual que el método de correlación. RF y SVM tienen una alta sensibilidad, cercana al 80%, prediciendo correctamente a los pacientes respondedores. El valor de la especificidad varía entre los modelos. De nuevo, la STD es alta para estos datos y los tiempos de cómputo son bajos, no superando ninguno de ellos los 30 segundos.

Clasificador	IM										Tiempo (s)		
	Exactitud		Sensibilidad		Especificidad		VPP		VPN(			rocAUC	
	Mean	(STD)	Mean	(STD)	Mean	STD	Mean	STD	Mean	STD		Mean	STD
RL	<b>0.48</b>	(0.22)	<b>0.49</b>	(0.32)	<b>0.46</b>	(0.33)	<b>0.42</b>	(0.29)	<b>47</b>	(0.33)	<b>0.48</b>	(0.22)	<b>0.32</b>
RF	<b>0.71</b>	(0.19)	<b>0.82</b>	(0.27)	<b>0.61</b>	(0.32)	<b>0.67</b>	(0.24)	<b>74</b>	(0.34)	<b>0.71</b>	(0.19)	<b>10.64</b>
MLP	<b>0.56</b>	(0.23)	<b>0.38</b>	(0.40)	<b>0.73</b>	(0.36)	<b>0.38</b>	(0.41)	<b>56</b>	(0.30)	<b>0.56</b>	(0.22)	<b>22.72</b>
MLP - Norm	<b>0.39</b>	(0.19)	<b>0.37</b>	(0.27)	<b>0.42</b>	(0.33)	<b>0.31</b>	(0.25)	<b>37</b>	(0.27)	<b>0.40</b>	(0.19)	<b>18.83</b>
SVM	<b>0.47</b>	(0.16)	<b>0.80</b>	(0.31)	<b>0.19</b>	(0.27)	<b>0.45</b>	(0.17)	<b>28</b>	(0.39)	<b>0.5</b>	(0.16)	<b>8.44</b>

Cuadro 5.4: Figuras de mérito de los modelos utilizando IM como selección de características.

Las curvas ROC son mostradas en la Figura 5.4. Hay una ligera mejora respecto a los resultados obtenidos por correlación. RF es el mejor clasificador, con una AUC de 0.71. MLP y SVM tienen valores cercanos a 0.5, indicando que los resultados obtenidos anteriormente son fruto del azar. RL y MLP-Norm tienen un AUC inferior a 0.5, por lo que clasifican las observaciones al revés.

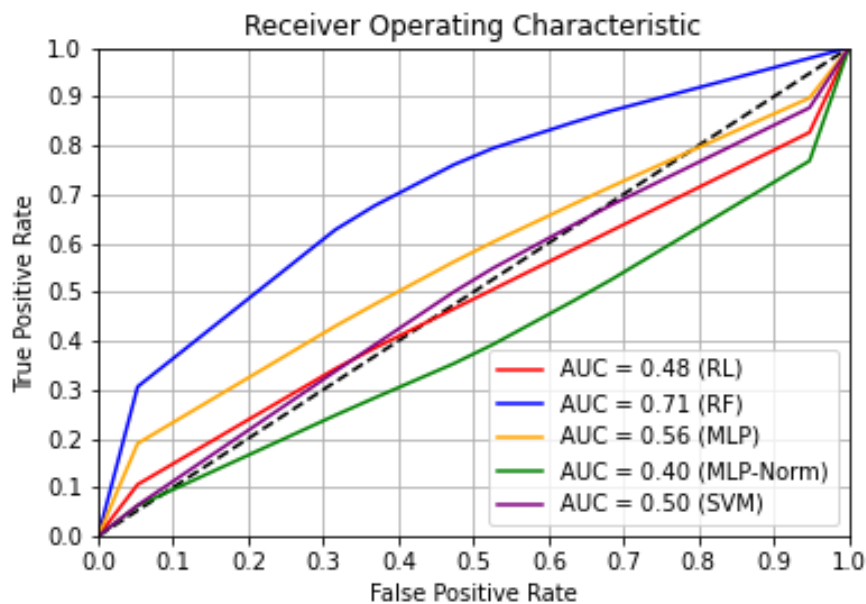


Figura 5.4: Curvas ROC y AUC para los modelos entrenados empleando IM.

## 5.3 PCA

### 5.3.1 Características Seleccionadas

Para PCA, vemos que efectivamente las características no son igualmente relevantes. Tenemos un eje que guarda el 50% de la varianza de los datos, por lo que es muy relevante (Figura 5.5). El resto guarda menos, hasta ver que incluso tenemos 7 características que no aportan casi ninguna varianza a los datos. Si todas las características fueran importantes, presentarían valores similares de varianza explicada, como ejemplificamos en la Figura 4.7. Comprobamos cuáles son los componentes más importantes con IM (Figura 5.6).

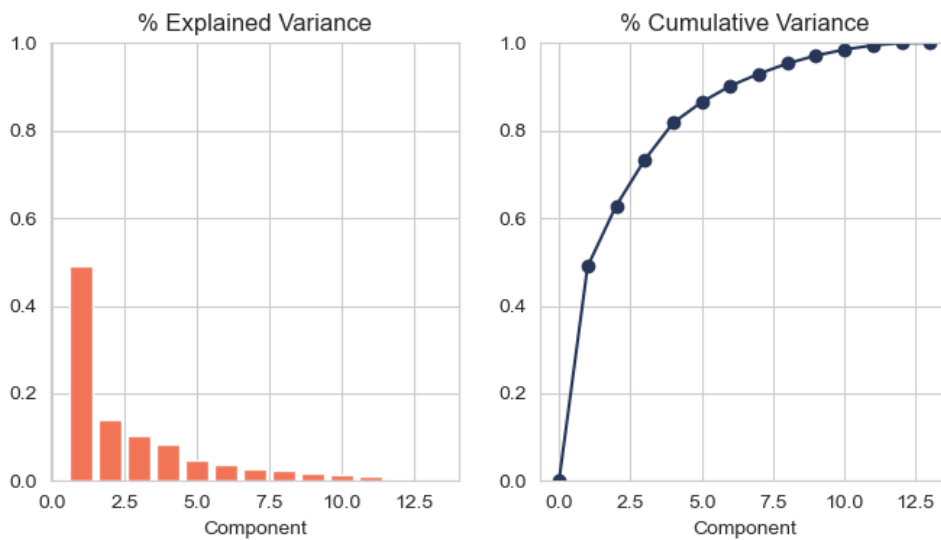


Figura 5.5: Varianza Explicada y Varianza Acumulada.

### 5.3. PCA

Clasificador	PCA											Tiempo (s)	
	Exactitud		Sensibilidad		Especificidad		VPP		VPN		rocAUC		
	Mean	(STD)	Mean	(STD)	Mean	STD	Mean	STD	Mean	STD	Mean		STD
RL	<b>0.74</b>	(0.2)	<b>0.76</b>	(0.30)	<b>0.72</b>	(0.37)	<b>0.76</b>	(0.29)	<b>0.72</b>	(0.34)	<b>0.73</b>	(0.2)	<b>0.3</b>
RF	<b>0.66</b>	(0.22)	<b>0.61</b>	(0.33)	<b>0.70</b>	(0.28)	<b>0.60</b>	(0.33)	<b>0.69</b>	(0.27)	<b>0.65</b>	(0.22)	<b>4.89</b>
MLP	<b>0.77</b>	(0.21)	<b>0.77</b>	(0.25)	<b>0.77</b>	(0.34)	<b>0.83</b>	(0.24)	<b>0.76</b>	(0.31)	<b>0.76</b>	(0.21)	<b>5.66</b>
MLP - Norm	<b>0.84</b>	(0.19)	<b>0.84</b>	(0.25)	<b>0.84</b>	(0.32)	<b>0.87</b>	(0.24)	<b>0.81</b>	(0.20)	<b>0.83</b>	(0.18)	<b>10.98</b>
SVM	<b>0.68</b>	(0.19)	<b>0.72</b>	(0.28)	<b>0.67</b>	(0.38)	<b>0.74</b>	(0.28)	<b>0.65</b>	(0.34)	<b>0.74</b>	(0.18)	<b>1.73</b>

Cuadro 5.5: Figuras de mérito de los modelos utilizando PCA como selección de características.

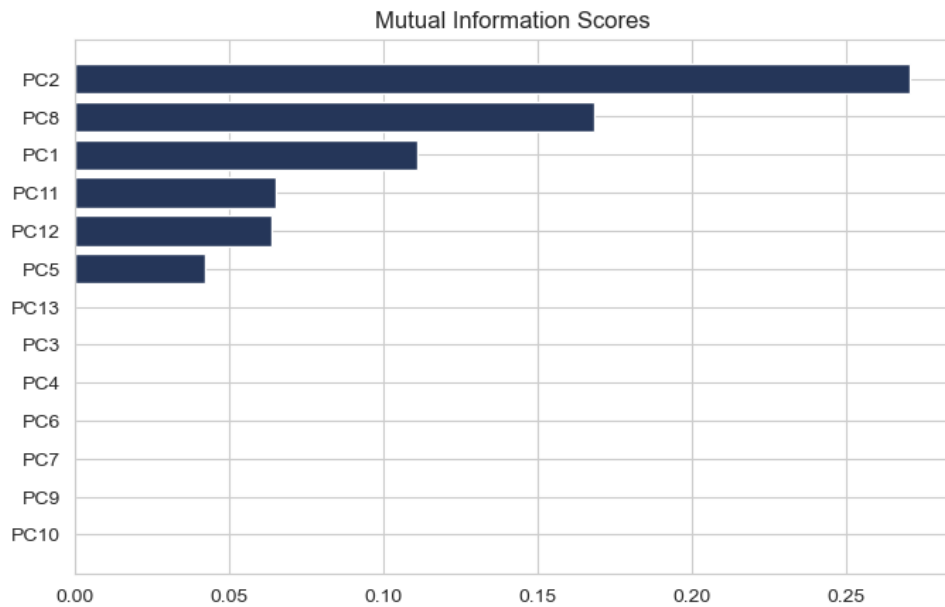


Figura 5.6: PC más relevantes.

### 5.3.2 Resultados

Una vez entrenados y validados los algoritmos de IA, presentamos los resultados obtenidos en el Cuadro 5.5. Obtenemos unos valores de exactitud, sensibilidad y especificidad cercanos a 0.7, muy altos si tenemos en cuenta el número de observaciones con el que se ha trabajado. Al igual que los otros métodos de selección, la STD es elevada. Este método presenta los tiempos de cómputo más corto, al procesarse menos características.

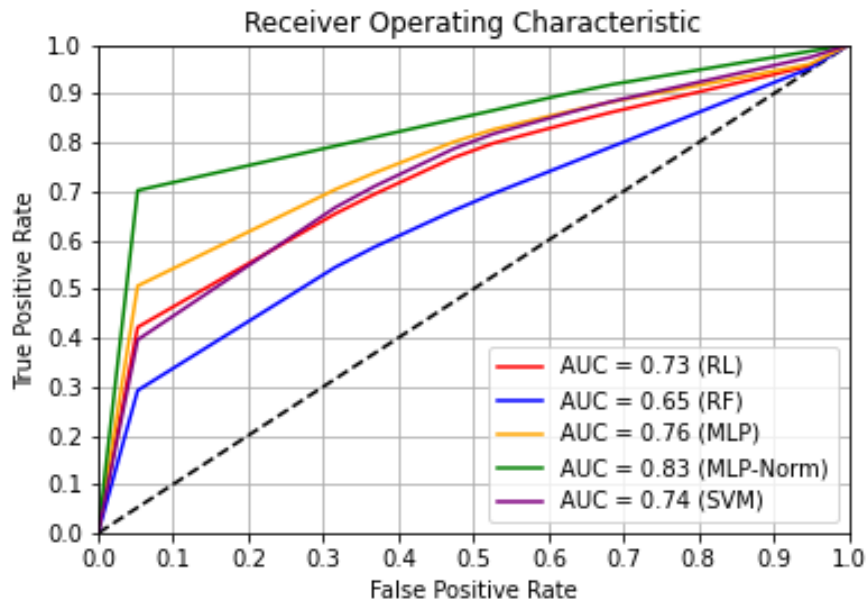


Figura 5.7: Curvas ROC y AUC para los modelos entrenados empleando PCA.

Existe una clara diferencia al utilizar PCA en lugar de correlación o IM. Los valores de AUC son altos, superando 0.7, a excepción de RF. El mejor es **MLP-Norm**, con un AUC de 0.83.

## 5.4 Explicabilidad

En las Figuras 5.8, 5.9, podemos ver los valores SHAP para todos los modelos de correlación e IM. No presentamos los valores SHAP de PCA porque carecen de utilidad explicativa, al haberse hecho una transformación de las características. *Estos valores se han calculado sobre la clase de respondedores, es decir, si un característica contribuye positivamente, el resultado tiende a predecir la observación como respondedora.*

En un vistazo general, parece que todas las características influyen de manera similar a cada uno de sus modelos. Como las característica extraídas son diferentes, no pueden compararse los métodos de selección, pero vemos que la característica más relevante es siempre de tipo Wavelet. La excepción es el modelo SVM de IM, que tiene una característica de tipo Square como más relevante. Llama la atención MLP de correlación, en la que ninguna de las características influye en las decisiones tomadas por el modelo.

En los modelos filtrados por correlación, las características de tipo **Shape** son las que menos influyen al resultado. Es interesante observar que valores altos esfericidad (*originalshapeSphericity*) contribuyen positivamente a los modelos, es decir, las lesiones más redondeadas tienen mejor respuesta a la QTN. Vemos también que si la imagen tiene una cantidad de vóxeles baja (*originalfirstorderVoxelVolume*), el modelo predice mejor a los respondedores.

El inconveniente de las características Wavelet es su interpretabilidad. Extraer conclusiones a partir de estas características es complicado, aún conociendo su razonamiento matemático. Por ejemplo, en los modelos filtrados por IM, *DependenceVariance* es la característica más influyente entre casi todos los modelos. Da información sobre la textura de la imagen al ser de tipo GLDM, concretamente sobre como varía la relación entre los vóxeles y sus vecinos a lo largo de la imagen, pero al haber sido esta característica obtenida a partir de la imagen Wavelet, no podemos establecer a priori una conexión con la imagen original.

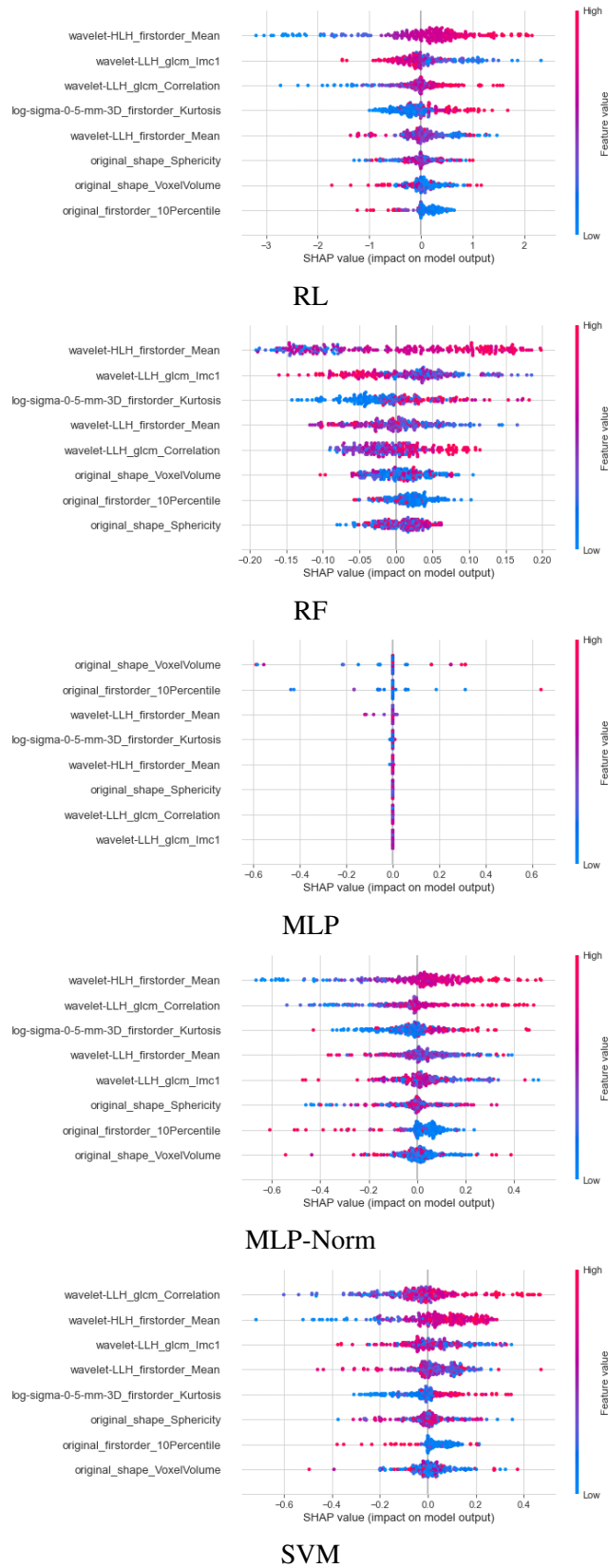
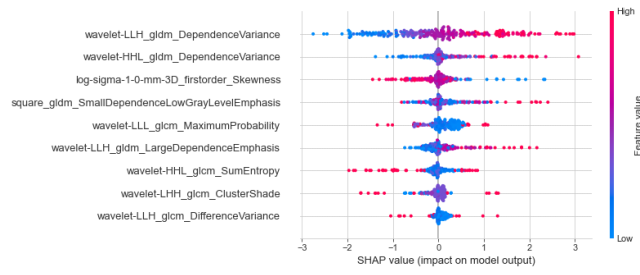
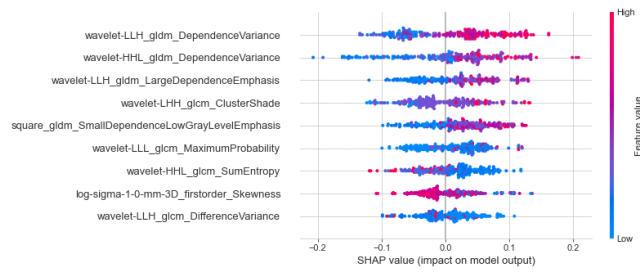


Figura 5.8: Valores SHAP para los diferentes modelo encontrados utilizando correlación.

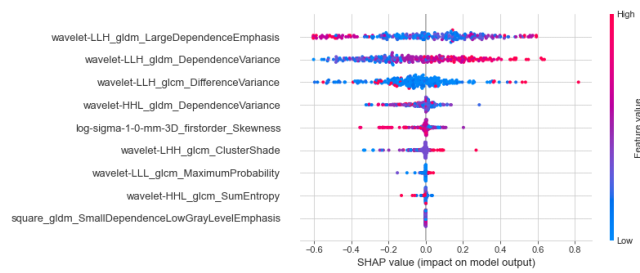
## 5.4. EXPLICABILIDAD



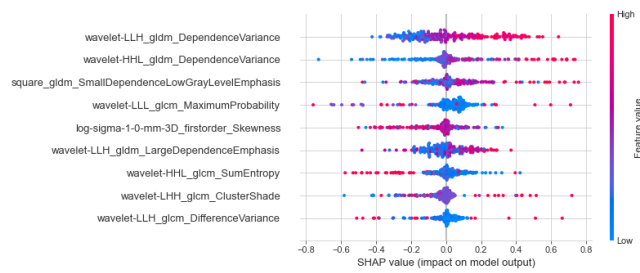
RL



RF



MLP



MLP-Norm



SVM

Figura 5.9: Valores SHAP para los diferentes modelo encontrados utilizando IM.



# Capítulo 6

---

## Discusión

El valor de esta herramienta aumenta si es capaz de predecir mejor aquellos pacientes no-respondedores. Una correcta predicción de los respondedores no aporta información sobre el grado de respuesta, impidiendo la toma de buenas decisiones (una respuesta completa y una respuesta con un 40% de necrosis, por ejemplo, siguen tratamientos muy distintos). Predecir correctamente a los no-respondedores evita administrar un tratamiento QTN ineficaz y permite buscar alternativas desde un primer momento.

### 6.1 Figuras de mérito

Observando los resultados de correlación, podemos ver que la exactitud tiene un valor alrededor de 0.5 para los cinco modelos evaluados, es decir, en la mitad de los casos el resultado no sería el correcto. Vemos que tienen mayor especificidad que sensibilidad, predicen mejor los pacientes no-respondedores que los respondedores. Sin embargo, los valores tampoco son aceptables; rondan el 50% y 60%, respectivamente. El AUC tiene valores cercanos a 0.5, por lo que los resultados anteriores pueden ser un mero fruto del azar. Los tiempos de cómputo son muy buenos para todos los modelos. De entre todos, el mejor es el RF, como es de esperar, al ser un modelo muy robusto. Tiene un AUC de 0.63 y una especificidad del 60%, mejores que el resto de modelos, pero insuficientes para tomar decisiones adecuadas.

En el caso de IM, los resultados son igualmente pobres. Los valores de AUC sugieren el azar en los modelos. RL y MLP-Norm predicen al revés, con AUC de 0.48 y 0.39 respectivamente. Los tiempos son también muy buenos para todos los modelos. Destaca, de nuevo, el modelo RF, con un AUC de 0.72, una sensibilidad del 82% y una especificidad del 61%. Como hemos dicho, es más interesante obtener una mayor especificidad para predecir los no-respondedores,

por lo que este RF tampoco tiene valor clínico.

Seleccionar mediante PCA sí ofrece resultados prometedores. Exceptuando RF, los modelos superan el 0.7 de AUC, insuficiente para la toma de decisiones pero considerablemente alto dado el escaso número de observaciones. Los valores de exactitud también son relativamente altos, cercanos a 0.7, al igual que la sensibilidad y la especificidad. Los tiempos de cómputo son igualmente aceptables. Es interesante ver que, en contra de lo que cabría esperar, RF es el peor modelo.

De entre todos los modelos, el único que parece ofrecer resultados interesantes es el de **MLP-Norm y PCA** como método de selección de características. Con un AUC de 0.84 y sensibilidad y especificidad del 84% para ambos, promete predecir correctamente la respuesta al tratamiento QTN si se entrenase con una cantidad mayor de observaciones. Destaca la diferencia con el modelo MLP, demostrando que los rangos dinámicos de las características son relevantes.

Sin embargo, debemos tener en cuenta la STD de los datos. Viendo los órdenes de magnitud de las figuras de mérito, la STD es muy alta y los resultados de cada *RepeatedStratifiedKFold* se encuentran muy alejados de la media, haciendo que los resultados sean menos sólidos de lo que parecen.

## 6.2 Limitaciones

La principal limitación encontrada en este estudio ha sido la poca cantidad de imágenes disponibles. Que la BD esté conformada por solo 13 pacientes y por una única secuencia de imagen, hace que los resultados obtenidos hayan sido pobres. Realizar *RepeatedStratifiedKFold* no ha sido suficiente para obtener unos resultados que permitan trasladar la herramienta al ámbito clínico real. Este pequeño número de pacientes se debe a la baja incidencia de los OS y SE, como vimos en la Sección 2.1, por lo que la colaboración con otros hospitales es necesaria para aumentar la cantidad de observaciones.

Debemos tener en cuenta también la división de los pacientes, que ha podido ejercer un

efecto negativo sobre los resultados. La BD que hemos utilizado no incluye pacientes con respuesta completa, por lo que puede no existir una diferencia lo suficientemente significativa entre los respondedores y los no-respondedores. Como encontrar pacientes con respuesta completa es complejo debido a la naturaleza agresiva de los tumores, una búsqueda exhaustiva interhospitalaria puede ayudarnos a superar esta limitación.

La secuencia utilizada también ha podido impactar negativamente. Atendiendo a las imágenes de T1 (Figura 4.3), los tumores tienen un nivel de gris similar al resto de estructuras anatómicas de la sección, por lo que las características de textura extraídas pueden no haber sido significativas. Vemos además en las segmentaciones (Apéndice A) que los tumores presentan heterogeneidad en cuanto a forma y tamaño, por lo que las características tipo *shape* pueden ser también irrelevantes.

## 6.3 Dimensionalidad y explicabilidad

Aún habiendo hecho una gran reducción de la dimensionalidad (de 1391 características a 8, 9 y 4 para correlación, IM y PCA, respectivamente), los resultados pueden deberse al número de características del problema.

Si nos fijamos primero en los coeficientes de correlación (Figura 5.1), no todas las parejas de características tienen un valor cercano a 0. Los pobres resultados de estos modelos pueden deberse al umbral establecido. Haber eliminado más características para reducir aún más la dimensionalidad podría haber entrenado mejor los modelos, evitando la Maldición de la Dimensionalidad. (Sección 4.3). En cuanto a la IM, podemos observar en la Figura 5.3 que ocurre un efecto similar. El umbral aplicado guarda 9 características, que pueden empeorar los resultados. Lo anterior puede verse respaldado si nos fijamos en los resultados obtenidos utilizando PCA: con únicamente 4 características, devuelve los mejores resultados.

La IM nos permite conocer más a fondo la relación de las características con la clase. Si bien es cierto que no conocemos la cota superior de la IM para este problema, podemos ver que la mejor solo tiene un valor de 0.35. Referenciando la Sección 4.3.2, esto se traduce en que

solo es capaz de responder una media de 0.35 preguntas de si o no sobre nuestro problema. Es decir, en conjunto con el resto de características seleccionadas, puede no estar conociendo lo suficiente sobre la clase como para crear buenos clasificadores.

Vemos que dependiendo del modelo de selección de características utilizado, obtenemos características diferentes (Cuadros 5.1 y 5.3), por lo que tampoco podemos asegurar que alguna de ellas funcione como biomarcador, es decir, ninguna en particular sirve para la toma de decisiones.

La gran mayoría de características seleccionadas provienen de la imagen Wavelet, demostrando su alta capacidad de extraer características relevantes que no podrían calcularse de otra forma. También vemos que correlación extrae el Volumen de Vóxel (VoxelVolume) y Esfericidad (Sphericity), ambas de tipo shape, características que el radiólogo puede entender y tener en cuenta. Una combinación de estos dos tipos de características puede ser interesante para equilibrar buenos resultados e inteligibilidad.

Refiriéndonos a los valores SHAP (Figura 5.8), vemos que para correlación las características más importantes son siempre las Wavelet, con un aporte similar para todos los modelos. Se demuestra de nuevo la importancia de este tipo de características para los problemas radiómicos. Excepcionalmente, en el modelo MLP, las características no intervienen nada en el resultado. Para IM, ocurre algo similar. Las características más relevantes son aquellas provenientes de Wavelet. Para RL, RF y MLP-Norm, las características parecen realizar una aportación similar al resultado mientras que para MLP, tres de ellas son las más importantes. SVM se diferencia en que su característica principal viene de una imagen Square, en lugar de Wavelet.

Aún ofreciendo los mejores resultados, PCA carece de la explicabilidad que ofrecen los otros dos métodos de selección. Transformar las características hace que perdamos de vista su esencia. Es por ello que mostrar los valores SHAP de PCA no tiene ningún tipo de utilidad explicativa. Es por ello que trasladar un modelo así al ámbito clínico puede ser una tarea complicada, ya que refuerza la fama de *caja negra* que tienen los algoritmos de IA, pudiendo provocar confusión y rechazo entre los profesionales sanitarios.

# Capítulo 7

---

## Conclusión y Líneas futuras

### 7.1 Conclusión

Hemos entrenado y validado 5 algoritmos de IA distintos, utilizando 3 métodos de selección de características (correlación, IM y PCA), con el objetivo de predecir la respuesta al tratamiento QTN en pacientes con OS y SE, a partir de una BD de de secuencias T1 de RM. El escaso número de pacientes y la nula variabilidad entre secuencias han sido los principales obstáculos en el desarrollo de algoritmos beneficiosos.

El método de selección de características utilizado ha sido el factor determinante. Tras una reducción significativa de la dimensionalidad del problema, se ha visto que los resultados han sido muy diferentes según el método utilizado. No se ha encontrado una característica concreta común a todos los métodos, por lo que no se ha podido establecer un biomarcador que permita la toma de decisiones, pero si se ha destacado la importancia de las características Wavelet para la clasificación.

Atendiendo a los figuras de mérito obtenidas (Cuadros 5.2, 5.4 y 5.5), las prestaciones de los modelos proveen, en general, una clasificación muy poco fiable. Si bien es cierto que el modelo MLP-Norm con selección PCA da unos resultados muy prometedores (dado el escaso número de pruebas), no es lo suficientemente bueno como para ser utilizado para tomar decisiones, por lo que concluimos que no puede ser trasladado al entorno clínico real.

## 7.2 Líneas Futuras

Este proyecto sirve entonces para establecer una base para el desarrollo de herramientas de IA para el apoyo a la toma de decisiones clínicas en el sector de la radiología musculoesquelética. La investigación puede ampliarse de varias maneras:

- Aumentar el número de pacientes de la BD, estableciendo una estrategia de colaboración interhospitalaria, para superar la gran limitación que el escaso número de muestras ha supuesto. Encontrar más pacientes que hayan tenido una respuesta completa a la QTN e introducir más secuencias por paciente, como T2 o DWI, para homogeneizar la BD y permitir realizar estudios más extensos.
- Crear una clasificación basada en los resultados histológicos de la biopsia, en lugar de en los comentarios radiológicos.
- Estudiar modelos de generación de imágenes sintéticas a partir de las disponibles en la BD, aportando una mayor variabilidad al escaso número de observaciones.
- Considerar el perfil genético de cada paciente y conformar una BD de características genómicas para entrenar los modelos junto a las radiómicas, para estudiar si los genes juegan un papel importante en el OS y SE. Considerar también otro tipo de características, como edad o antecedentes familiares. Comparar estas características con los modelos entrenados únicamente con datos radiómicos.
- Estudiar otro tipo de modelos de IA basados en aprendizaje profundo, como las redes neuronales convolucionales.
- Establecer un plan en el servicio de radiología para estandarizar los informes de las imágenes control. Eliminar la subjetividad de los comentarios del radiólogo en cuanto a la respuesta al tratamiento QTN, que limitan el desarrollo de proyectos radiómicos y generan confusión. Buscar alternativas al porcentaje de necrosis o reducción del tumor respecto a la imagen original, estableciendo un biomarcador más representativo de las patologías.

- Entrenar modelos de clasificación empleando 3 clases en lugar de 2: no-respondedores, respondedores-parciales y respondedores-completos. Considerar modelos de regresión que devuelvan la probabilidad de responder al tratamiento QTN, para ofrecer una herramienta más intuitiva, capaz de guiar mejor el tratamiento de los pacientes.
- Estudiar otros métodos de selección de características, ya que hemos demostrado que es uno de los factores más influyentes para los resultados.





---

# Bibliografía

- [1] A. Rimmer, “Radiologist shortage leaves patient care at risk, warns royal college,” *BMJ*, vol. 359, 2017. [Online]. Available: <https://www.bmj.com/content/359/bmj.j4683>
- [2] V. Markotić, T. Pojužina, D. Radančević, M. Miljko, and V. Pokrajcic, “The radiologist workload increase; where is the limit?: Mini review and case study,” *Psychiatria Danubina*, vol. 33, p. 217, 06 2021.
- [3] B. Spieler and N. Baum, “Burnout: A mindful framework for the radiologist,” *Current Problems in Diagnostic Radiology*, vol. 51, no. 2, pp. 155–161, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0363018821001687>
- [4] G. D. Roodman, “Mechanisms of bone metastasis,” *New England Journal of Medicine*, vol. 350, no. 16, pp. 1655–1664, 2004, pMID: 15084698. [Online]. Available: <https://doi.org/10.1056/NEJMra030831>
- [5] J. Ferguson and S. Turner, “Bone cancer: Diagnosis and treatment principles,” *Am Fam Physician*, vol. 98, no. 4, pp. 205–213, Aug 2018, pMID: 15084698. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/30215968/>
- [6] M. S. Isakoff, S. S. Bielack, P. Meltzer, and R. Gorlick, “Osteosarcoma: Current treatment and a collaborative pathway to success,” *Journal of Clinical Oncology*, vol. 33, no. 27, pp. 3029–3035, Sep 2015. [Online]. Available: <https://doi.org/10.1200/JCO.2014.59.4895>
- [7] E. Simpson and H. L. Brown, “Understanding osteosarcomas,” *JAAPA*, vol. 31, pp. 15–19, Aug 2018. [Online]. Available: <https://doi.org/10.1097/01.jaa.0000541477.24116.8d>
- [8] G. Ottaviani and N. Jaffe, *The Epidemiology of Osteosarcoma*. Boston, MA: Springer US, 2010, pp. 3–13. [Online]. Available: [https://doi.org/10.1007/978-1-4419-0284-9\\_1](https://doi.org/10.1007/978-1-4419-0284-9_1)

- [9] G. Bacci and S. Lari, “Current treatment of high grade osteosarcoma of the extremity: Review,” *Journal of Chemotherapy*, vol. 13, no. 3, pp. 235–243, 2001, pMID: 11450880. [Online]. Available: <https://doi.org/10.1179/joc.2001.13.3.235>
- [10] C. Meazza and P. Scanagatta, “Metastatic osteosarcoma: a challenging multidisciplinary treatment,” *Expert Review of Anticancer Therapy*, vol. 16, no. 5, pp. 543–556, 2016. [Online]. Available: <https://doi.org/10.1586/14737140.2016.1168697>
- [11] H. Khalid, D. Ibrahim, B. Abdulrahman, and et al., “Overview of methotrexate toxicity: A comprehensive literature review,” *Cureus*, vol. 14, Sep 2022. [Online]. Available: <https://doi.org/10.7759%2Fcureus.29518>
- [12] I. Corre, F. Verrecchia, V. Crenn, F. Redini, and V. Trichet, “The osteosarcoma microenvironment: A complex but targetable ecosystem,” *Cells*, vol. 9, no. 4, p. 976, Apr 2020. [Online]. Available: <http://dx.doi.org/10.3390/cells9040976>
- [13] B. Esen, G. M *et al.*, “Ewing sarcoma in an infant and review of the literature,” *The Turkish Journal of Pediatrics*, vol. 61, no. 5, pp. 760–764, 2019. [Online]. Available: <https://doi.org/10.24953/turkjped.2019.05.016>
- [14] K. A. Ross, N. A. Smyth, C. D. Murawski, and J. G. Kennedy, “The biology of ewing sarcoma,” *The Turkish Journal of Pediatrics*, vol. 2013, p. 7, 2013. [Online]. Available: <http://dx.doi.org/10.1155/2013/759725>
- [15] L. Haveman *et al.*, “High-dose chemotherapy followed by autologous haematopoietic cell transplantation for children, adolescents, and young adults with primary metastatic ewing sarcoma,” *Cochrane Database of Systematic Reviews*, no. 9, 2021. [Online]. Available: <https://doi.org//10.1002/14651858.CD011405.pub2>
- [16] G. Bacci, A. Ballardelli, C. Forni, A. Longhi, M. Serra, N. Fabbri, M. Alberghini, S. Ferrari, M. S. Benassi, and P. Picci, “Ewing’s sarcoma family tumours,” *The Journal of Bone Joint Surgery British Volume*, vol. 89-B, no. 9, pp. 1229–1233, 2007. [Online]. Available: <https://doi.org/10.1302/0301-620X.89B9.19422>

- [17] S. S. Donaldson, “Ewing sarcoma: Radiation dose and target volume,” *Pediatric Blood & Cancer*, vol. 42, no. 5, pp. 471–476, 2004. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/pbc.10472>
- [18] D. D. Moore and R. C. Haydon, *Ewing’s Sarcoma of Bone*. Cham: Springer International Publishing, 2014, pp. 93–115. [Online]. Available: [https://doi.org/10.1007/978-3-319-07323-1\\_5](https://doi.org/10.1007/978-3-319-07323-1_5)
- [19] P. G. Casali *et al.*, “Bone sarcomas: Clinical practice guidelines for diagnosis, treatment and follow-up,” *Annals of Oncology*, vol. 29, no. 4, pp. iv79–iv95, 2018. [Online]. Available: <http://doi.org/10.1093/annonc/mdy310>
- [20] M. E. Mayerhoefer, A. Materka, G. Langs, I. Häggström, P. Szczypiński, P. Gibbs, and G. Cook, “Introduction to radiomics,” *Journal of Nuclear Medicine*, vol. 61, no. 4, pp. 488–495, 2020. [Online]. Available: <https://jnm.snmjournals.org/content/61/4/488>
- [21] R. J. Gillies, P. E. Kinahan, and H. Hricak, “Radiomics: Images are more than pictures, they are data,” *Radiology*, vol. 278, no. 2, pp. 563–577, 2016, pMID: 26579733. [Online]. Available: <https://doi.org/10.1148/radiol.2015151169>
- [22] A. Zwanenburg, M. Vallières, M. A. Abdalah *et al.*, “The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping,” *Radiology*, vol. 295, no. 2, pp. 328–338, 2020, pMID: 32154773. [Online]. Available: <https://doi.org/10.1148/radiol.2020191145>
- [23] Y. Mintz and R. Brodie, “Introduction to artificial intelligence in medicine,” *Minimally Invasive Therapy & Allied Technologies*, vol. 28, no. 2, pp. 73–81, 2019, pMID: 30810430. [Online]. Available: <https://doi.org/10.1080/13645706.2019.1575882>
- [24] C. Fanciullo, S. Gitto, E. Carlicchi, D. Albano, C. Messina, and L. M. Sconfienza, “Radiomics of musculoskeletal sarcomas: A narrative review,” *Journal of Imaging*, vol. 8, no. 2, 2022. [Online]. Available: <https://www.mdpi.com/2313-433X/8/2/45>

- [25] M. Imbriaco, A. Luzzati *et al.*, “Mri radiomics-based machine learning classification of atypical cartilaginous tumour and grade ii chondrosarcoma of long bones,” *eBioMedicine*, vol. 75, p. 103757, 2022. [Online]. Available: <http://doi.org/10.1016/j.ebiom.2021.103757>
- [26] V. Chianca, R. Cuocolo, S. Gitto, D. Albano, I. Merli, J. Badalyan, M. C. Cortese, C. Messina, A. Luzzati, A. Parafioriti, F. Galbusera, A. Brunetti, and L. M. Sconfienza, “Radiomic machine learning classifiers in spine bone tumors: A multi-software, multi-scanner study,” *European Journal of Radiology*, vol. 137, p. 109586, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0720048X21000668>
- [27] P. Lin, P.-F. Yang, S. Chen, Y.-Y. Shao, L. Xu, Y. Wu, W. Teng, X.-Z. Zhou, B.-H. Li, C. Luo, L.-M. Xu, M. Huang, T.-Y. Niu, and Z.-M. Ye, “A delta-radiomics model for preoperative evaluation of neoadjuvant chemotherapy response in high-grade osteosarcoma,” *Cancer Imaging*, vol. 20, no. 1, p. 7, January 2020. [Online]. Available: <https://doi.org/10.1186/s40644-019-0283-8>
- [28] S. Community. (2023) 3d slicer documentation. [Online]. Available: <https://slicer.readthedocs.io/en/latest/>
- [29] M. Mantri, S. Taran, and G. Sunder, “Dicom integration libraries for medical image interoperability: A technical review,” *IEEE Reviews in Biomedical Engineering*, vol. 15, pp. 247–259, 2022. [Online]. Available: <https://doi.org/10.1109/RBME.2020.3042642>
- [30] Pyradiomics-Community. (2016) Pyradiomics documentation. [Online]. Available: <https://pyradiomics.readthedocs.io/en/latest/>
- [31] J. J. van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R. G. Beets-Tan, J.-C. Fillion-Robin, S. Pieper, and H. J. Aerts, “Computational Radiomics System to Decode the Radiographic Phenotype,” *Cancer Research*, vol. 77, no. 21, pp. e104–e107, Oct 2017. [Online]. Available: <https://doi.org/10.1158/0008-5472.CAN-17-0339>
- [32] NumFOCUS. (2020) Simpleitk documentation. [Online]. Available: <https://simpleitk.readthedocs.io/en/master/gettingStarted.html#python-binary-files>

- [33] MicroDicom. (2023) Microdicom dicom viewer. [Online]. Available: <https://www.microdicom.com/>
- [34] K. Gorgolewski, T. Auer *et al.*, “The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments,” *Scientific data*, vol. 3, p. 160044, 6 2016. [Online]. Available: <https://doi.org/10.1038/sdata.2016.44>
- [35] I. Zhovannik, S. Pai, T. A. da Silva Santos, L. L. van Driel, A. Dekker, R. Fijten, A. Traverso, J. Bussink, and R. Monshouwer, “Radiomics integration into a picture archiving and communication system,” *Physics and Imaging in Radiation Oncology*, vol. 20, pp. 30–33, Oct 2021. [Online]. Available: <https://doi.org/10.1016/j.phro.2021.09.007>
- [36] W. T. Clarke, T. K. Bell, U. E. Emir, M. Mikkelsen, G. Oeltzschner, A. Shamaei, B. J. Soher, and M. Wilson, “Nifti-mrs: A standard data format for magnetic resonance spectroscopy,” *Magnetic Resonance in Medicine*, vol. 88, no. 6, pp. 2358–2370, 2022. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.29418>
- [37] icometrix. (2016) dicom2nifti. [Online]. Available: <https://dicom2nifti.readthedocs.io/en/latest/index.html>
- [38] R. M. Haralick, K. Shanmugam, and I. Dinstein, “Textural features for image classification,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, 1973. [Online]. Available: <https://doi.org/10.1109/TSMC.1973.4309314>
- [39] S. Dash and M. R. Senapati, “Gray level run length matrix based on various illumination normalization techniques for texture classification,” *Evolutionary Intelligence*, vol. 14, no. 2, pp. 217–226, Jun 2021. [Online]. Available: <https://doi.org/10.1007/s12065-018-0164-2>
- [40] M. M. Galloway, “Texture analysis using gray level run lengths,” *Computer Graphics and Image Processing*, vol. 4, no. 2, pp. 172–179, 1975. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0146664X75800086>

- [41] G. Thibault, J. Angulo, and F. Meyer, “Advanced statistical matrices for texture characterization: Application to cell classification,” *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 3, pp. 630–637, 2014. [Online]. Available: <https://doi.org/10.1109/TBME.2013.2284600>
- [42] C. Sun and W. G. Wee, “Neighboring gray level dependence matrix for texture classification,” *Computer Vision, Graphics, and Image Processing*, vol. 23, no. 3, pp. 341–352, 1983. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0734189X83900324>
- [43] A. Abbasian Ardakani, N. J. Bureau, E. J. Ciaccio, and U. R. Acharya, “Interpretation of radiomics features—a pictorial review,” *Computer Methods and Programs in Biomedicine*, vol. 215, p. 106609, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0169260721006830>
- [44] PyRadiomicsCommunity. (2016) Pipeline modules. [Online]. Available: <https://pyradiomics.readthedocs.io/en/latest/radiomics.html#radiomics-imageoperations-label>
- [45] D. Chicco, “Ten quick tips for machine learning in computational biology,” *BioData Min*, vol. 10, p. 35, Dec 2017. [Online]. Available: <https://doi.org/10.1186/2Fs13040-017-0155-3>
- [46] L. Chen, *Curse of Dimensionality*. Boston, MA: Springer US, 2009, pp. 545–546. [Online]. Available: [https://doi.org/10.1007/978-0-387-39940-9\\_133](https://doi.org/10.1007/978-0-387-39940-9_133)
- [47] A. Géron, *Hands-on Machine Learning with Scikit-Learn, Keras TensorFlow*, 2nd ed. O’Reilly Media, Inc., Sep 2019, ch. 8, pp. 215–225.
- [48] W. Jia, M. Sun, J. Lian, and S. Hou, “Feature dimensionality reduction: a review,” *Complex & Intelligent Systems*, vol. 8, no. 3, pp. 2663–2693, Jun 2022. [Online]. Available: <https://doi.org/10.1007/s40747-021-00637-x>
- [49] A. Leon-Garcia, *Probability, Statistics, and Random Processes for Electrical Engineering*. Pearson Education, Inc., 2008, ch. 5, pp. 258–261.

- 
- [50] R. M. Gray, *Entropy and Information Theory*, 2nd ed. Springer New York, NY, Feb 2011, ch. 3, pp. 82–90. [Online]. Available: <https://link.springer.com/book/10.1007/978-1-4419-7970-4>
- [51] M. Akay, *Detection and Estimation Methods for Biomedical Signals*. ACADEMIC PRESS, INC, 1996, ch. 5, pp. 73–77.
- [52] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, 1st ed. Springer New York, NY, Jun 2013, ch. 5, pp. 175–186. [Online]. Available: <https://doi.org/10.1007/978-1-4614-7138-7>
- [53] ScikitLearnDevelopers. (2023) Repeatedstratifiedkfold. [Online]. Available: [https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.RepeatedStratifiedKFold.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.RepeatedStratifiedKFold.html)
- [54] ScikitLearnDeveloper. (2023) 1.1. linear models. [Online]. Available: [https://scikit-learn.org/stable/modules/linear\\_model.html](https://scikit-learn.org/stable/modules/linear_model.html)
- [55] A. Géron, *Hands-on Machine Learning with Scikit-Learn, Keras TensorFlow*, 2nd ed. O’Reilly Media, Inc., Sep 2019, ch. 4, pp. 144–146.
- [56] T. K. Ho, “Random decision forests,” *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, vol. 14-16, p. 278–282, Aug 1995. [Online]. Available: <https://web.archive.org/web/20160417030218/http://ect.bell-labs.com/who/tkh/publications/papers/odt.pdf>
- [57] A. Géron, *Hands-on Machine Learning with Scikit-Learn, Keras TensorFlow*, 2nd ed. O’Reilly Media, Inc., Sep 2019, ch. 6, pp. 179–183.
- [58] A.Géron, *Hands-on Machine Learning with Scikit-Learn, Keras TensorFlow*, 2nd ed. O’Reilly Media Inc., Sep 2019, ch. 10, pp. 277–288.
- [59] S. learn Developers. (2023) 1.17. neural network models (supervised). [Online]. Available: [https://scikit-learn.org/stable/modules/neural\\_networks\\_supervised.html](https://scikit-learn.org/stable/modules/neural_networks_supervised.html)

- [60] G. J. Tortora and B. Derrickson, *Principios de Anatomía y Fisiología*, 13th ed. Editorial Médica Panamericana S.A., May 2013, ch. 12, pp. 447–453.
- [61] A. Géron, *Hands-on Machine Learning with Scikit-Learn, Keras TensorFlow*, 2nd ed. O'Reilly Media, Inc., Sep 2019, ch. 5, pp. 155–163.
- [62] S. learn Developers. (2023) 1.4. support vector machines. [Online]. Available: <https://scikit-learn.org/stable/modules/svm.html>
- [63] ScikitLearnDevelopers. (2023) Randomizedsearchcv. [Online]. Available: [https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.RandomizedSearchCV.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.RandomizedSearchCV.html)
- [64] S. Lundberg. (2018) Shap documentation. [Online]. Available: <https://shap.readthedocs.io/en/latest/index.html>
- [65] C. Molnar, *Interpretable Machine Learning: A Guide For Making Black Box Models Explainable*, Feb 2022, ch. 9. [Online]. Available: <https://christophm.github.io/interpretable-ml-book/shap.html>
- [66] A. Shrikumar, P. Greenside, and A. Kundaje, “Learning important features through propagating activation differences,” in *Proceedings of the 34th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, D. Precup and Y. W. Teh, Eds., vol. 70. PMLR, 06–11 Aug 2017, pp. 3145–3153. [Online]. Available: <https://proceedings.mlr.press/v70/shrikumar17a.html>
- [67] J. Guiot, A. Vaidyanathan *et al.*, “A review in radiomics: Making personalized medicine a reality via routine imaging,” *Medicinal Research Reviews*, vol. 42, no. 1, pp. 426–440, 2022. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/med.21846>
- [68] A. Crombé, D. Fadli, A. Italiano, X. Saut, Olivierand Buy, and M. Kind, “Systematic review of sarcomas radiomics studies: Bridging the gap between concepts and clinical applications?” *European Journal of Radiology*, vol. 132, p. 109283, 2020. [Online]. Available: <http://doi.org/10.1016/j.ejrad.2020.109283>



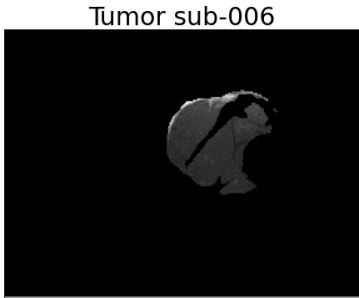
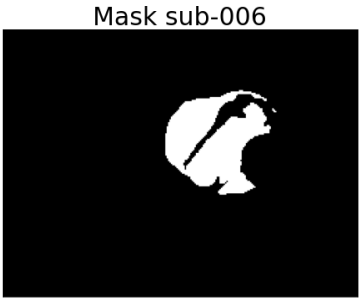
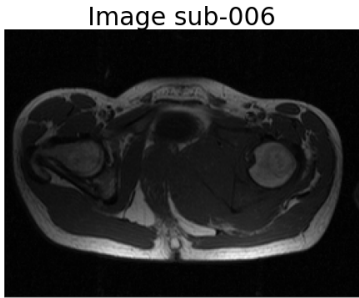
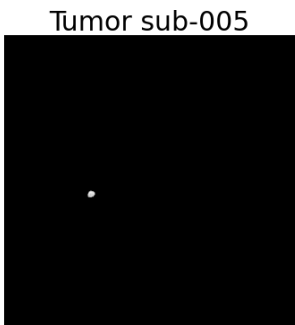
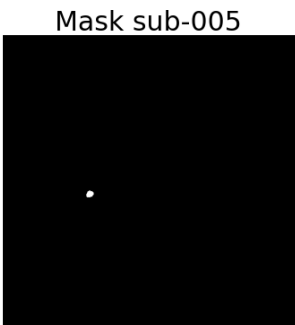
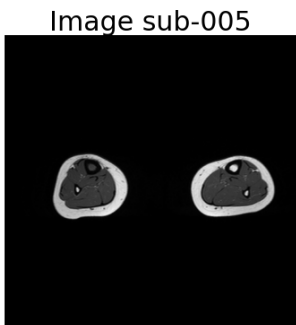
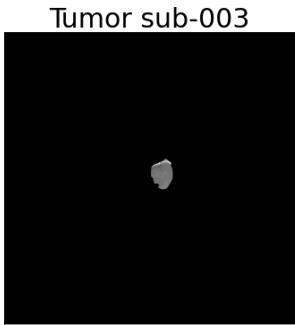
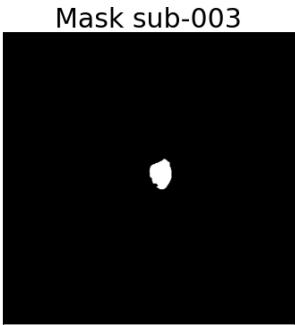
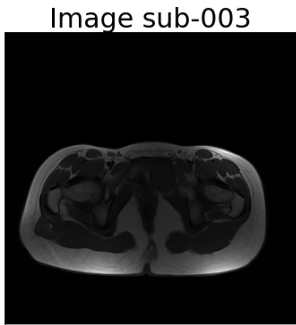
- [69] ScikitLearnDeveloper. (2023) Gridsearchcv. [Online]. Available: [https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.GridSearchCV.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html)



# Apéndice A

---

## Imágenes RM



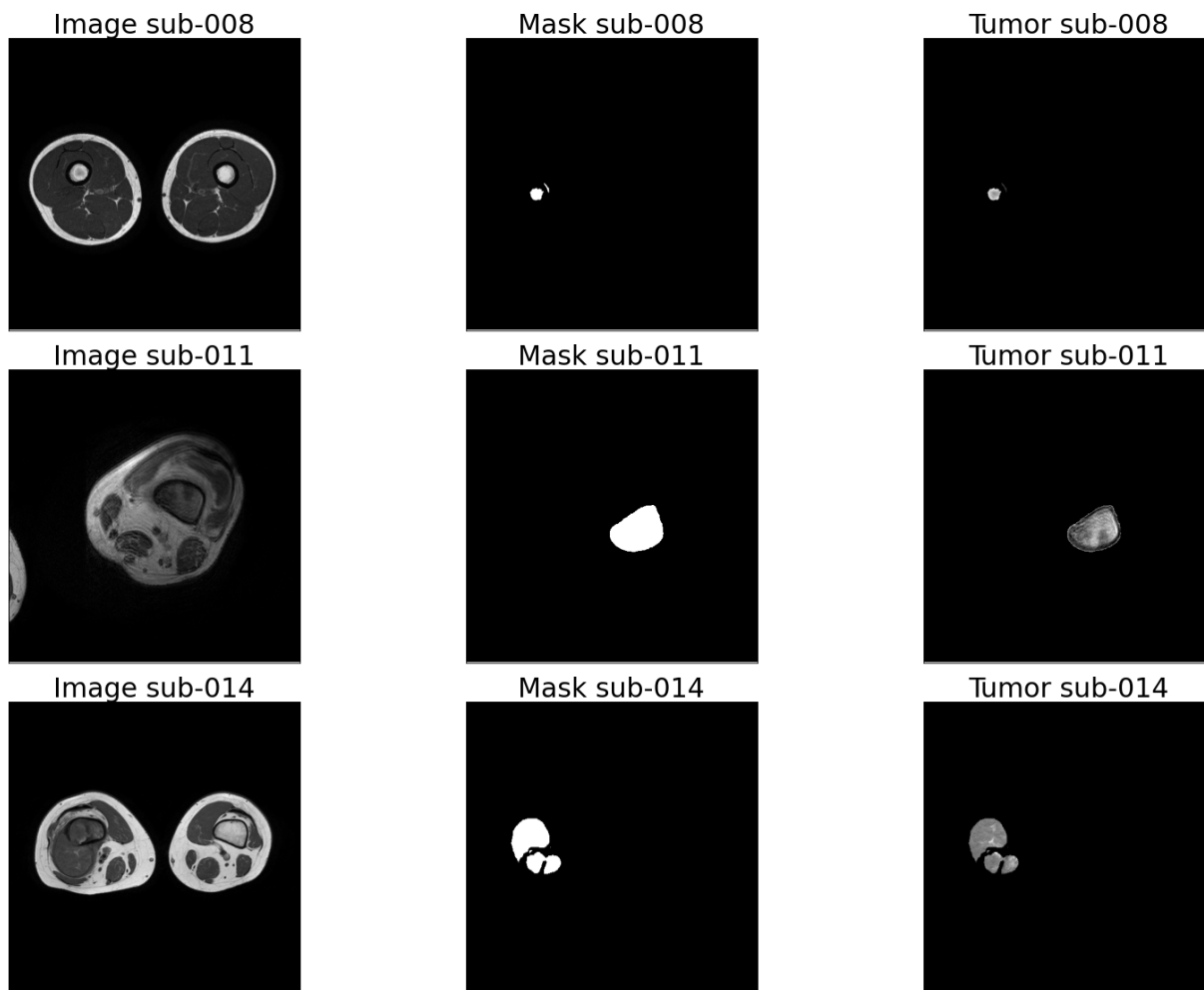
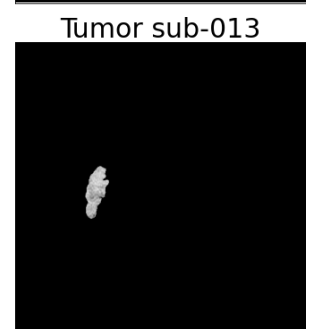
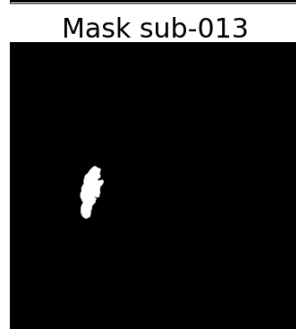
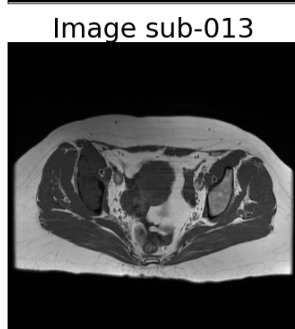
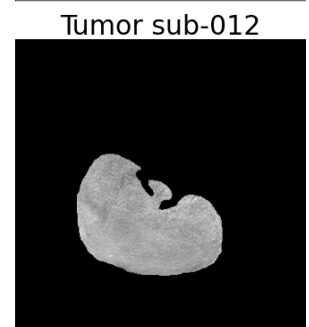
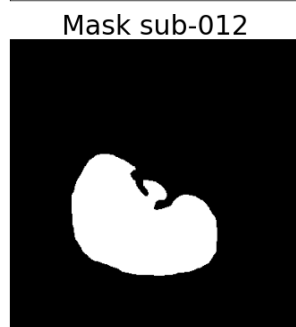
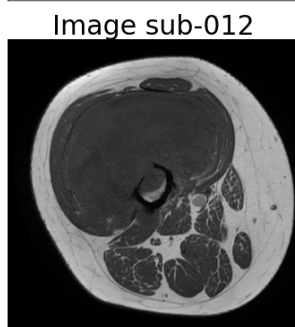
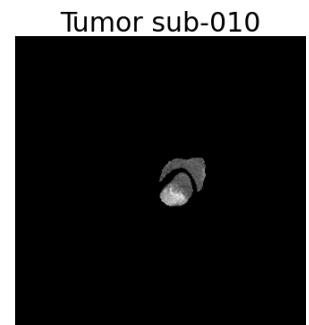
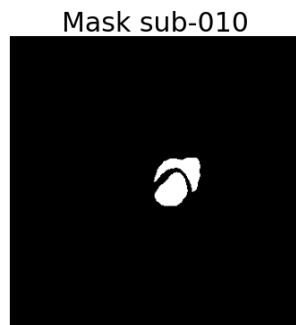
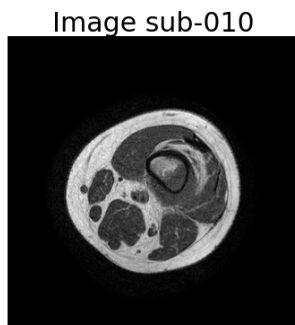
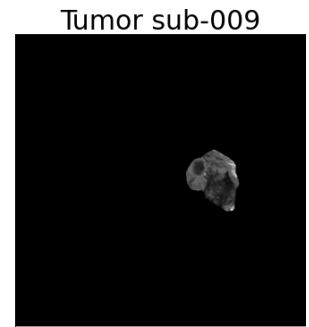
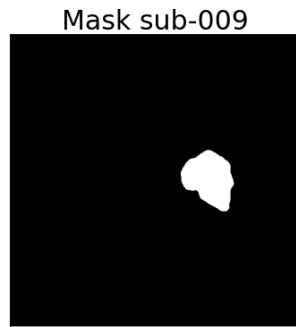
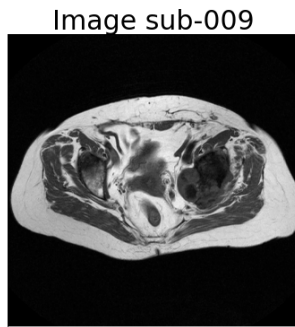


Figura A.1: Secuencia T1, Máscara binaria y ROI de los pacientes **respondedores**.



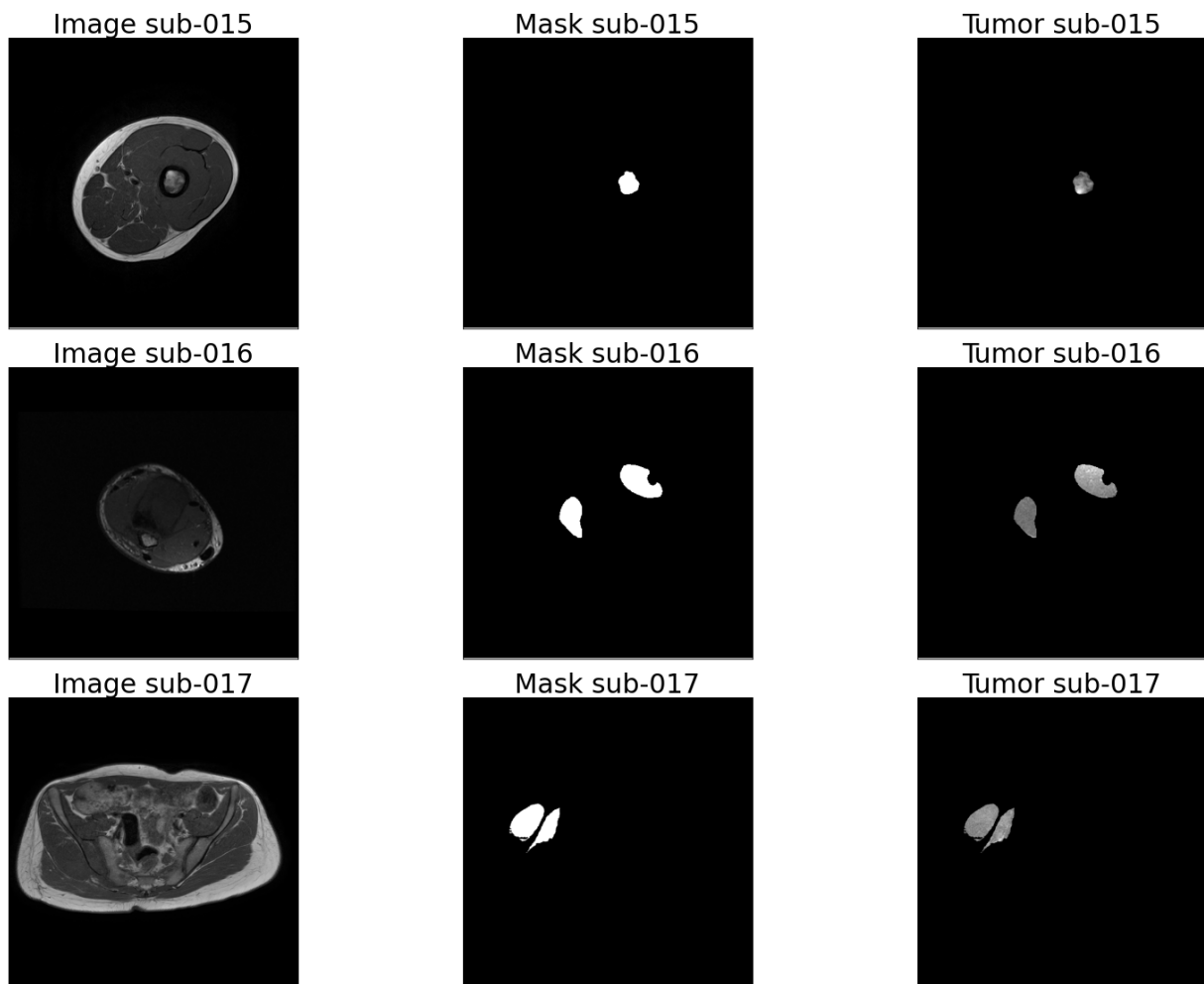


Figura A.2: Secuencia T1, Máscara binaria y ROI de los pacientes **no-respondedores**.

# Apéndice B

---

## Matrices de Confusión

En este Apéndice se muestran la matrices de confusión de los modelos, que por cuestiones de extensión decidimos no introducir en la Sección 5.



Figura B.1: Matrices de confusión para modelos filtrados con correlación.





Figura B.2: Matrices de confusión para modelos filtrados con IM.



Figura B.3: Matrices de confusión para modelos filtrados con PCA.