

EL ABC DE LA EXTRACCIÓN DE DATOS A PARTIR DE UNA IMAGEN RADIOLÓGICA:

CLAVES PARA ABORDAR UN PROYECTO CIENTÍFICO

María Fresno Vallespir¹, Yensa Rodríguez Álvarez¹, Santiago Medrano Martorell¹, Sofía González Ortiz¹, Camilo Pineda Ibarra¹, Anna Massuet Vilamajor¹, Almudena Pérez Lara², Laura Oleaga Zufiría¹

¹ Hospital Clínic de Barcelona, Barcelona, ² Hospital Regional Universitario Carlos Haya, Málaga

OBJETIVO DOCENTE

- Ser capaz de **identificar oportunidades** para el desarrollo de un proyecto científico a partir de la **extracción de datos de la imagen**, en base a las características de la misma y al campo de estudio.
- Mostrar los **pasos** a seguir para elaborar una base de datos a partir de la segmentación de imágenes.
- **Ilustrar el proceso** con el *ejemplo* de una cohorte de pacientes con cáncer de orofaringe.

CONTENIDOS

1. Introducción
2. Visión general del recorrido
3. Planteamiento del proyecto
4. Preparación de la base de datos
5. Selección de las imágenes
6. Segmentación
7. Extracción de resultados
8. Construcción del modelo de IA
9. Conclusiones
10. Bibliografía

1. INTRODUCCIÓN

- Los avances tecnológicos en radiodiagnóstico, con el advenimiento de la radiómica y la inteligencia artificial (IA), hacen necesario integrar entre los objetivos docentes de los MIR los **flujos de trabajo para el manejo y procesado avanzado de las imágenes**.
- En este trabajo se ilustran los pasos en el proceso de **segmentación** de imágenes, **extracción de datos radiómicos** y **configuración de una base de datos**, a partir de un *ejemplo* con imágenes de TC de una cohorte de pacientes con **cáncer de orofaringe** de nuestro centro.

2. VISIÓN GENERAL DEL RECORRIDO

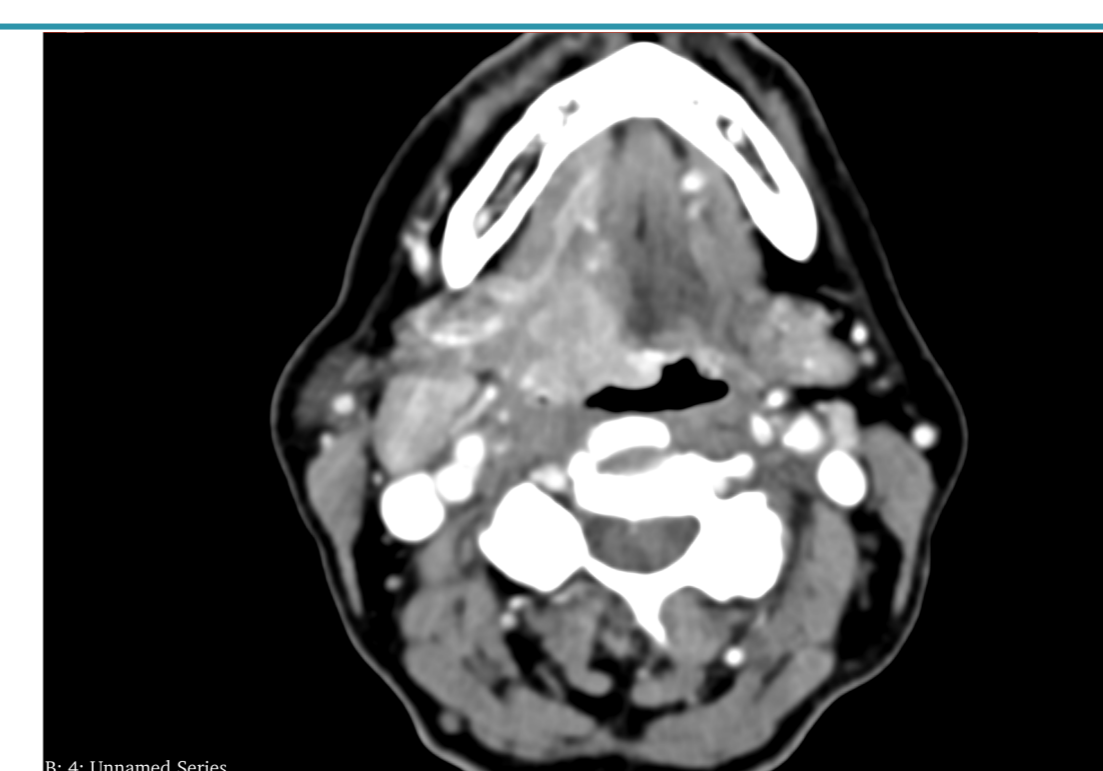
Planteamiento del proyecto



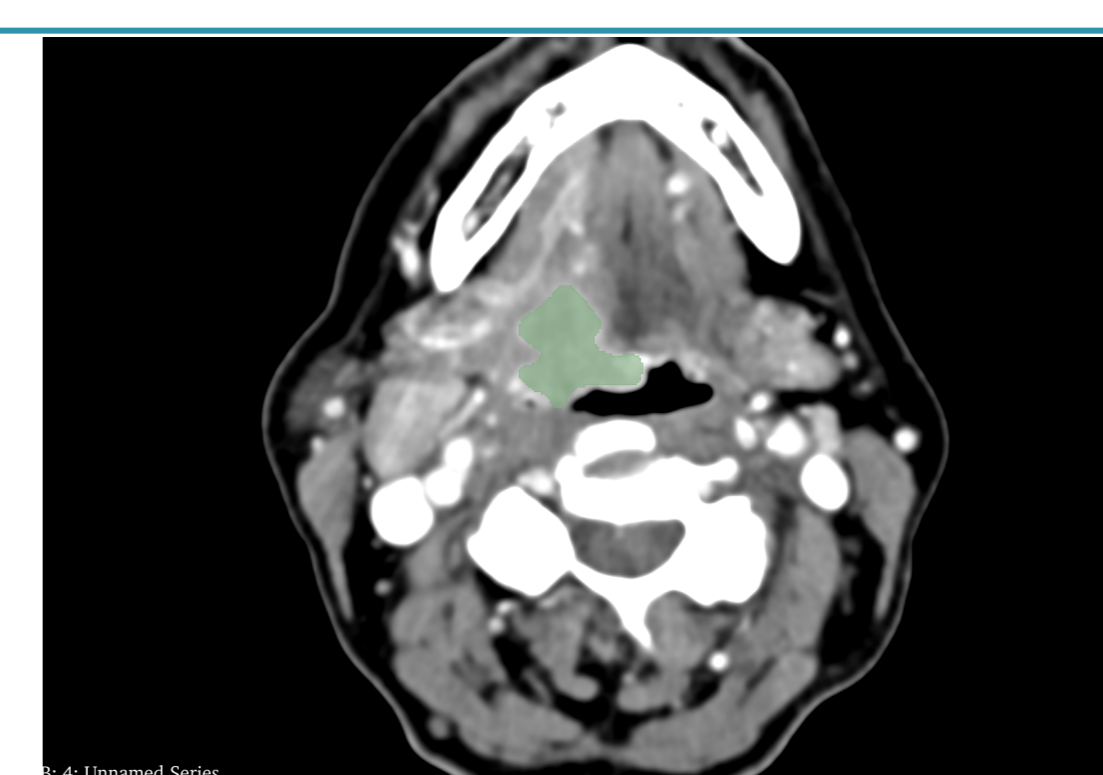
Preparación de la base de datos

Fecha	Valor	Motivo No valorable	Resor de	DEI	Contrast	T4	N2c	Activo	Lateralid	T1	CC	VOL	Patrón de Captaci	UH	Ive	Esféricid	Oni	NY ADP Da	NY ADP
24/10/2022	SI		1,5	SI	Split bolus	T4	N2c	ORT	Izquierda										
12/9/2023	No	No se identifica primario	3,0	No	Única	T1	N1	Cirurgía + CRT	Derecha				necrótico >50%	52,00	Ia			1	1
22/9/2022	SI		2,0	SI	Split bolus	T4b	N1	ORT	Izquierda	31,00	41,00	62,00	23,41 homogéneo	88,00					
22/9/2022	SI		2,0	SI	Split bolus	T4b	N1	ORT	Izquierda	6,00	4,00	12,00	0,15 homogéneo	123,00	III		0,70	3	3
4/8/2022	SI		2,0	SI	Split bolus	T1	N1	ORT	Bilateral	13,00	28,00	19,00	4,20 homogéneo	161,00					
4/8/2022	SI		2,0	SI	Split bolus	T1	N1	ORT	Derecha	13,00	15,00	24,00	3,25 necrótico >50%	99,00	Ia		0,67	2	2
4/8/2022	SI		2,0	SI	Split bolus	T1	N1	ORT	Derecha	9,00	5,00	7,00	0,23 heterogéneo	139,00	Ia		0,74	0	

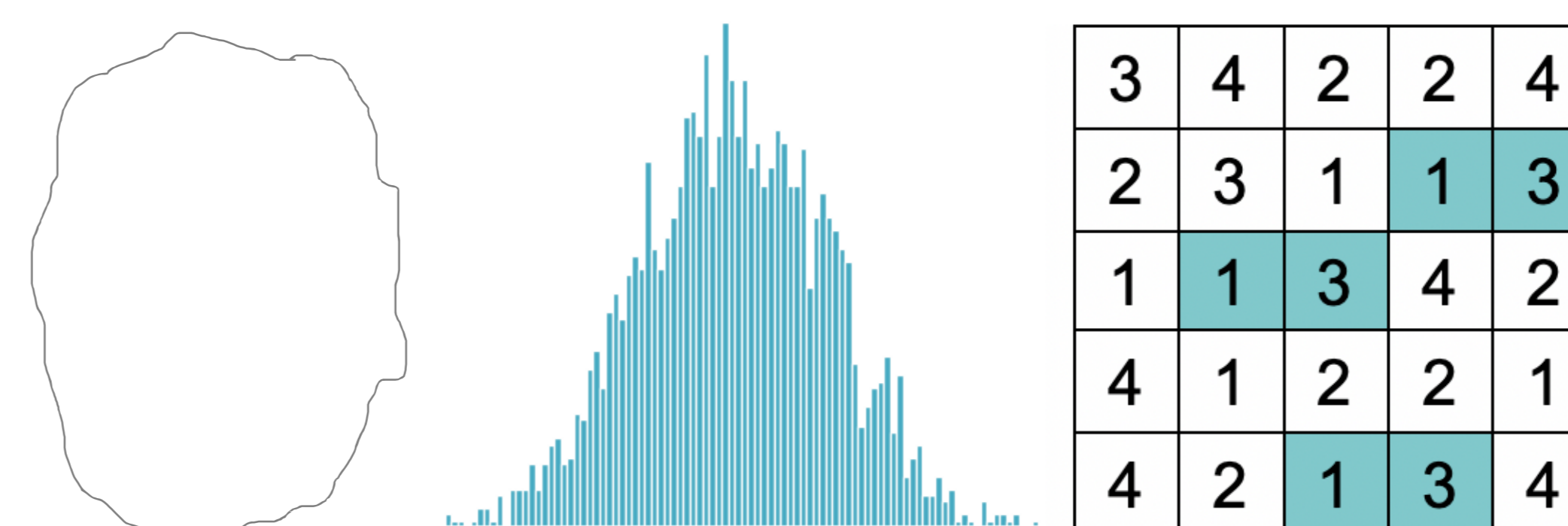
Selección de las imágenes



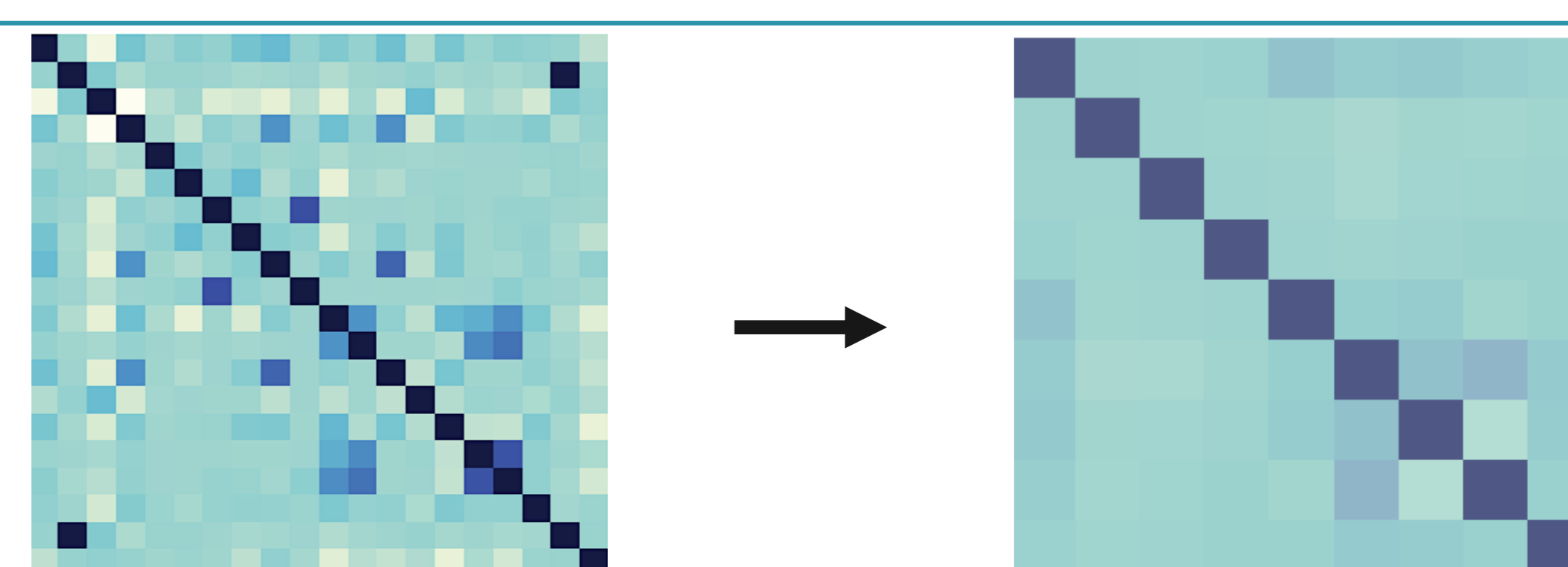
Segmentación



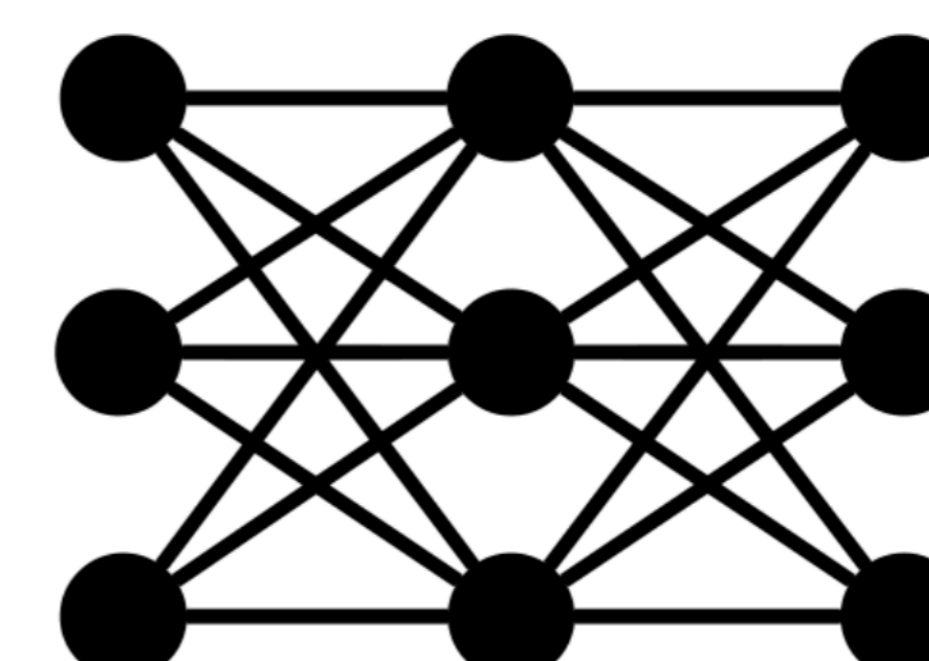
Extracción de resultados



Selección de características radiómicas



Construcción del modelo de IA



3. PLANTEAMIENTO DEL PROYECTO

1. Identificación de la **incógnita/necesidad clínica** y valoración del **potencial** de extracción de datos de la imagen.
 - Validación de la idea a través de **revisión bibliográfica** y puesta en común en un grupo multidisciplinar.
 - Definir tipo de imagen óptima a explotar, tipo de datos que se pueden obtener (datos brutos, histograma, radiómica, volumetría, máscara de segmentación...) y valoración preliminar de su potencial análisis estadístico.
2. Definir la **hipótesis** y el **diseño preliminar** del método de estudio: tipo de estudio, criterios de inclusión y exclusión, características de imagen, análisis estadístico, etc.
3. Si se requiere, validación por el comité ético de la institución.
4. Explorar oportunidades de BECAS / financiación.

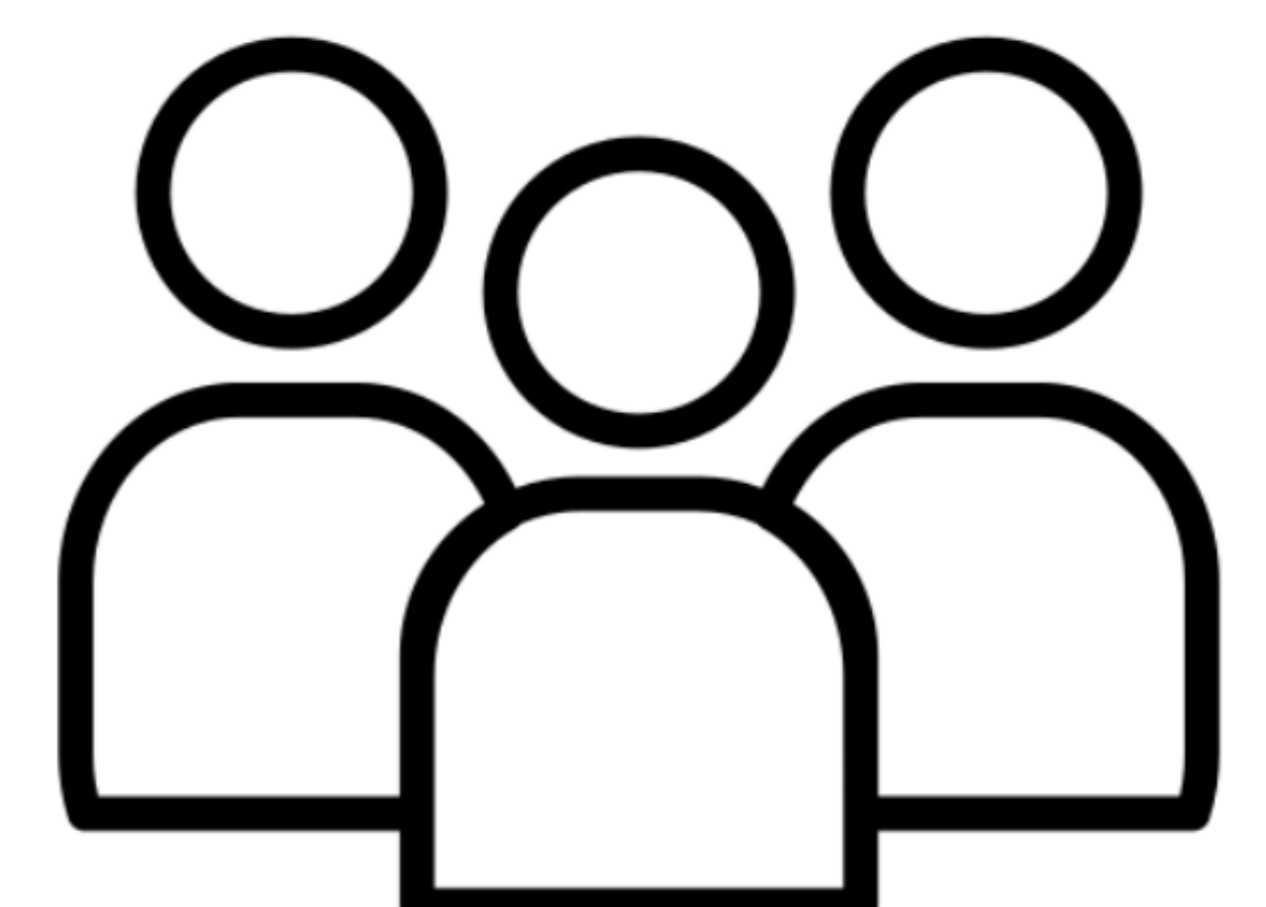


3. PLANTEAMIENTO DEL PROYECTO

La segmentación de imágenes médicas es un proceso complejo y requiere de experiencia en el uso de *software* de imagen médica.

Asimismo, es necesaria una mentorización técnica para el procesamiento de imágenes y el análisis estadístico. Por ello, es importante disponer de un **equipo multidisciplinar**, que debería incluir:

- Investigador principal y colaboradores.
- Mentores: especialistas con experiencia en el campo de estudio.
- Expertos en procesamiento avanzado de la imagen: ingenieros, matemáticos (preferiblemente con experiencia en el marco clínico).
- Expertos en bioestadística.



3. PLANTEAMIENTO DEL PROYECTO

→ *EJEMPLO*: Justificación de un proyecto de segmentación en una cohorte de pacientes con carcinoma de orofaringe.

El objetivo principal es determinar el **estado del virus del papiloma humano** (VPH) a partir de los valores de radiómica extraídos tras la segmentación de tumores de orofaringe en imágenes de TC de cuello.

Otros objetivos son la predicción de la respuesta al tratamiento y la predicción de la supervivencia.

4. PREPARACIÓN DE LA BASE DE DATOS

1. **Obtención** de una base de datos clínica a partir del Sistema de Información en Salud (HIS), grupos de investigación... Revisión y depurado de la **n** (criterios de inclusión y exclusión, selección de imágenes válidas).
2. **Registro** de la información:

DATOS CLINICODEMOGRÁFICOS
Sexo, edad, tipo de tumor,
estadio, tratamiento

DATOS DE LA ADQUISICIÓN
Modalidad, fecha de
adquisición, administración de
contraste, grosor de corte

CARACTERÍSTICAS SEMÁNTICAS
Localización, tamaño, volumen,
patrón de captación, necrosis

CARACTERÍSTICAS RADIÓMICAS
Obtenidas tras la segmentación

3. **Depurado** de la base de datos: preparación de los datos clínicos y de imagen para su análisis y modelado.
 - Normalización de los datos, codificación de variables, gestión de valores perdidos o atípicos.

4. PREPARACIÓN DE LA BASE DE DATOS

→ *EJEMPLO*: Base de datos de tumores de orofaringe. Incluye los siguientes aspectos:

- Datos relacionados con la adquisición de las imágenes
- Datos relacionados con el estadio del tumor y el tratamiento
- Características semánticas del tumor
- Características semánticas de las adenopatías
- Características radiómicas

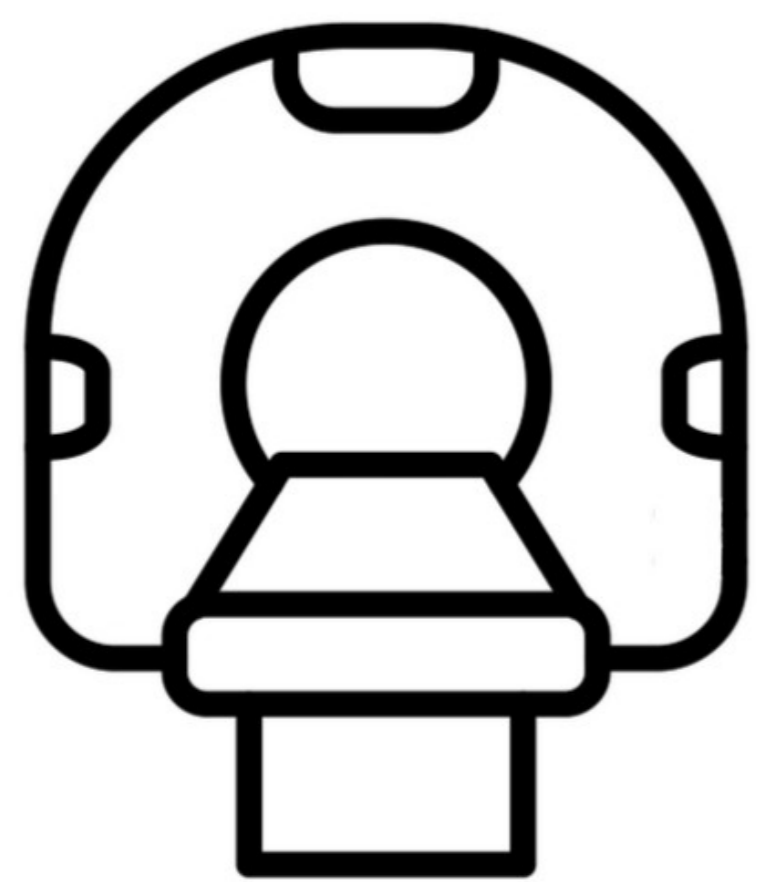
Fecha	Valorab	Motivo No valorable	Grosor de cort	DECT	in contraste	T	N	Actitud	Lateralida
24/10/2022	Sí		1,5	Sí	Split bolus	T4	N2c	QRT	Izquierda
12/9/2023	No	No se identifica primario	3,0	No	Única	T1	N1	Cirugía + QRT	Derecha
22/9/2022	Sí		2,0	Sí	Split bolus	T4b	N1	QRT	Izquierda
22/9/2022	Sí		2,0	Sí	Split bolus	T4b	N1	QRT	Izquierda
4/8/2022	Sí		2,0	Sí	Split bolus	T1	N1	QRT	Bilateral
4/8/2022	Sí		2,0	Sí	Split bolus	T1	N1	QRT	Derecha
4/8/2022	Sí		2,0	Sí	Split bolus	T1	N1	QRT	Derecha

AP	Tr	CC	VOL	Patrón de Captació	UH	Nivel	Esfericida	Total	Nº ADP Derecha	Nº ADP Izquierda	Variables 3D slicer
				necrótico >50%	52,00	Ila		0			
31,00	41,00	62,00	23,41	homogeneo	89,00			1	1		
6,00	4,00	12,00	0,15	homogeneo	123,00	III	0,70	3		3	
13,00	28,00	19,00	4,20	homogeneo	162,00			0			
13,00	15,00	24,00	3,25	necrótico >50%	99,00	Ila	0,67	2	2		
9,00	5,00	7,00	0,23	heterogeneo	139,00	Ila	0,74	0			

Fig. 1. Ejemplo de base de datos de tumores de orofaringe.

5. SELECCIÓN DE LAS IMÁGENES

1. Obtención y selección de imágenes **estandarizadas** y de **alta calidad**. Debe recogerse información sobre el **protocolo de adquisición y de reconstrucción** de las imágenes.



En TC, por ejemplo: voltaje, miliamperaje, grosor de corte, resolución espacial, administración de medio de contraste, matriz de reconstrucción, etc.

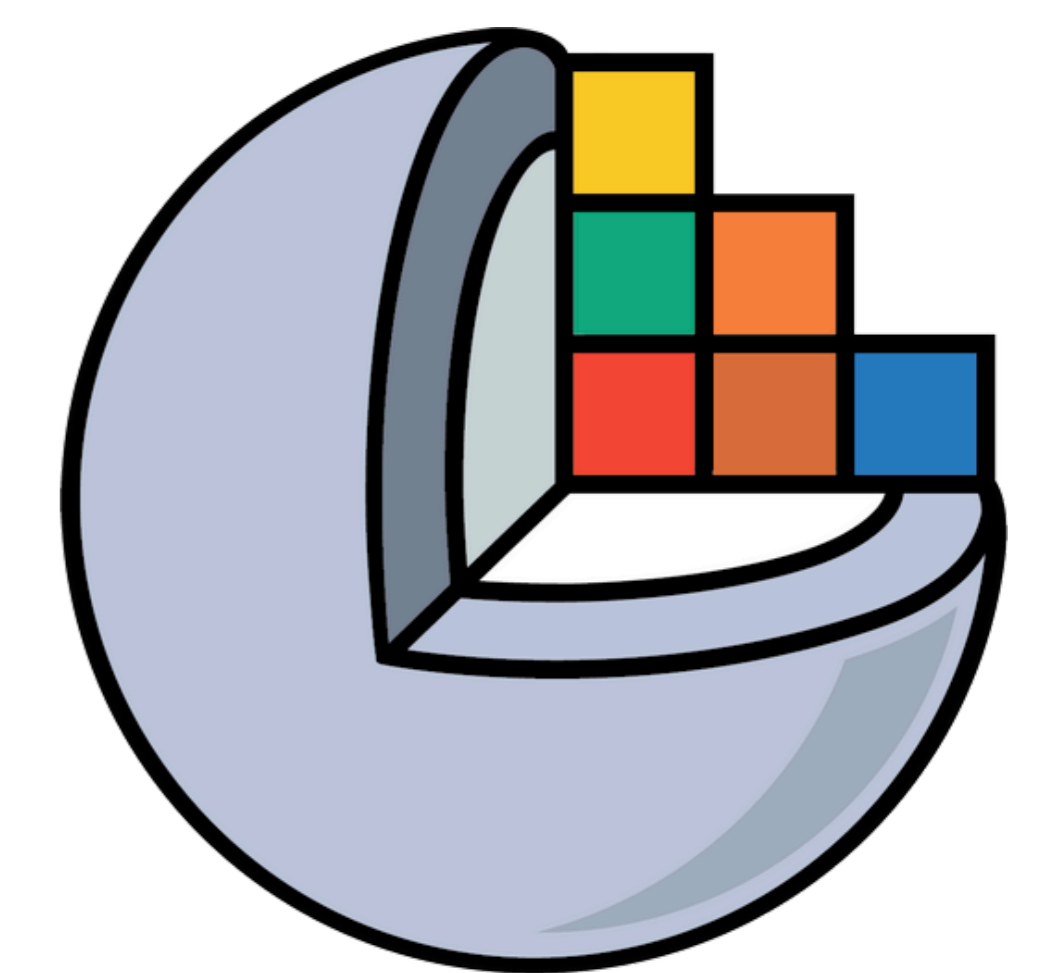
Es recomendable que los parámetros sean **constantes**, pues una gran variabilidad de los mismos puede influir sobre las características radiómicas extraídas y el modelo obtenido a partir de las mismas.

2. **Descarga** de imágenes en disco duro/repositorio seguro.
3. **Anonimización**: eliminar los datos personales identificables de los pacientes en las imágenes (metadatos) y definir un sistema de identificación de las imágenes.

6. SEGMENTACIÓN

Una vez se han seleccionado las imágenes que se van a analizar, el siguiente paso es importarlas al *software* con el que se va a trabajar.

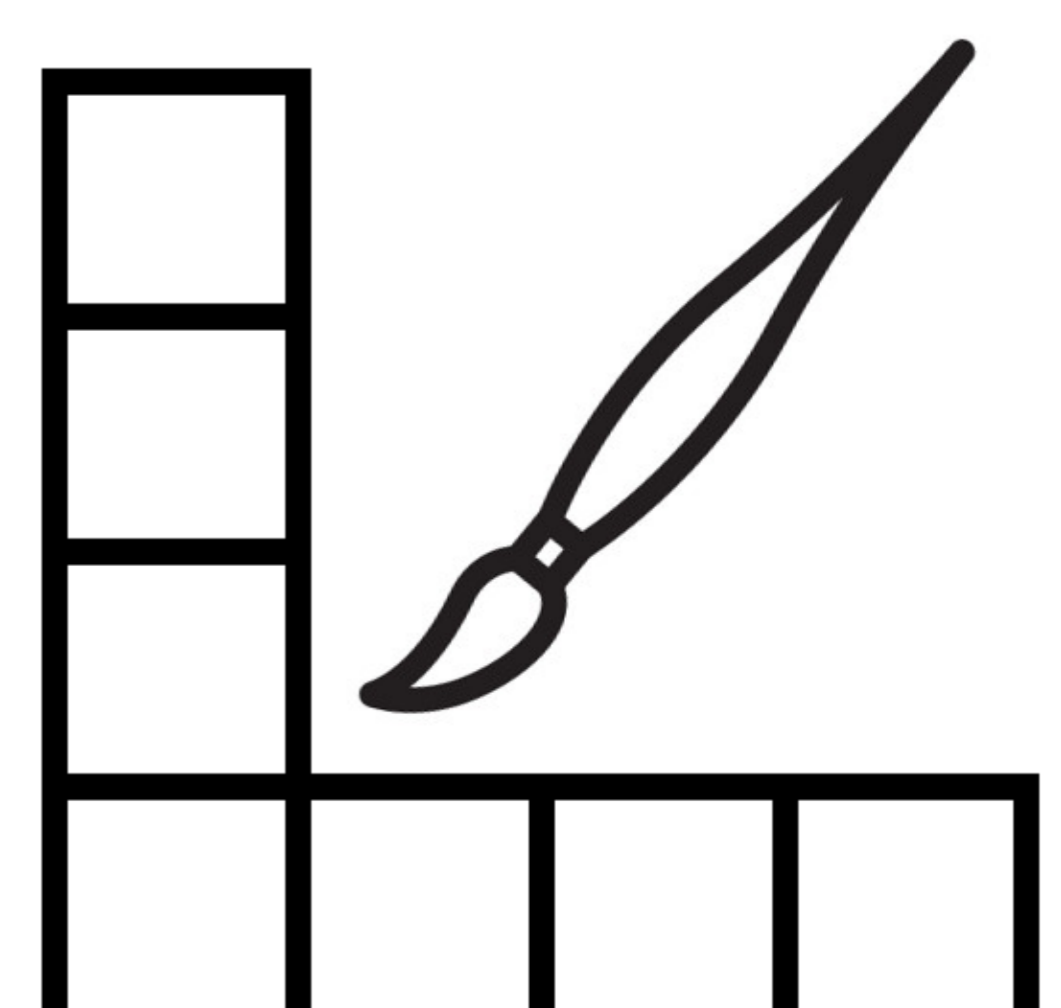
- **3D Slicer**. Paquete de *software* gratuito, de código abierto, disponible para los principales sistemas operativos. Consta de módulos que posibilitan tanto la visualización, el procesamiento y la segmentación de imágenes como el análisis y la extracción de las características radiómicas.



- Otros *softwares* disponibles, ya sea de código abierto o comerciales, son MITK, OsiriX, ITK-SNAP, MeVisLab, Fiji (ImageJ), MIPAV, Freesurfer, LifEx.

6. SEGMENTACIÓN

- Segmentar consiste en identificar y **delinear** el área objetivo en la imagen obteniendo una región de interés (**ROI**) o un volumen de interés (**VOI**), según si el enfoque es bidimensional o tridimensional, respectivamente.
- Los ROI/VOI definen la región en la que se calculan las características radiómicas, por lo que la segmentación debe ser **reproducible y fiable**.
- La técnica de imagen (radiografía, ecografía, TC, TC dual/espectral, RM, PET), así como las características de la estructura o lesión que se desea estudiar (heterogeneidad, definición del contorno) van a influir en la elección del método de segmentación.



6. SEGMENTACIÓN

¿Esta imagen es segmentable?

- Tumor o adenopatía con captación homogénea, de bordes bien definidos, bien contrastados → imagen fácilmente segmentable.

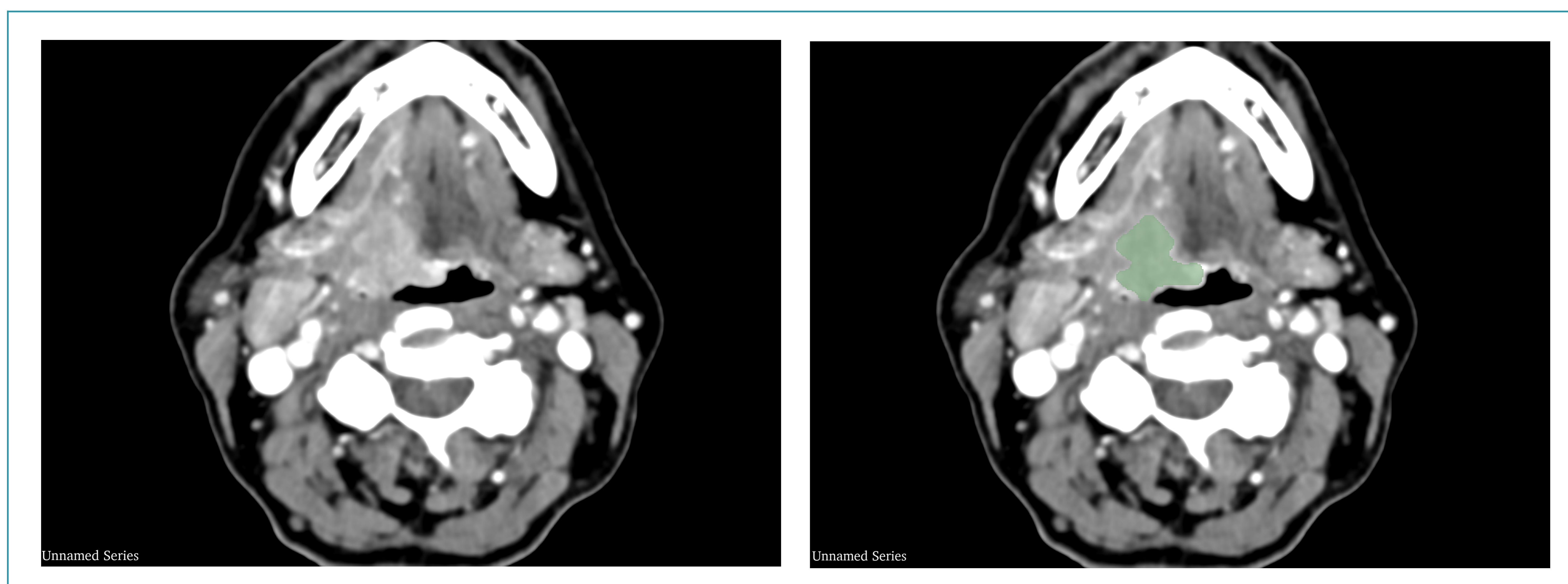


Fig. 2. Ejemplo de carcinoma escamoso de orofaringe e imagen de la segmentación.

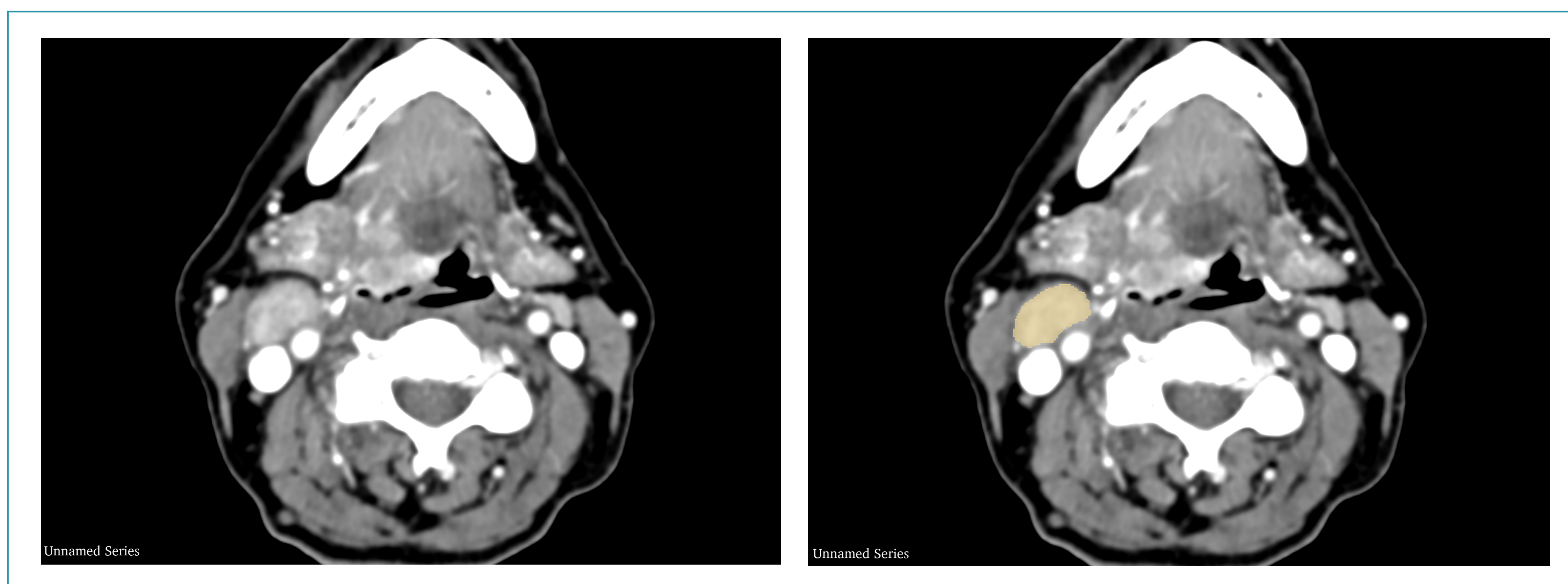


Fig. 3. Ejemplo de adenopatía cervical e imagen de la segmentación.

6. SEGMENTACIÓN

¿Esta imagen es segmentable?

- Tumor de bordes poco definidos, poco contrastado → imagen difícil de segmentar. Precisa ajuste del histograma de intensidades de gris para mejorar el contraste.

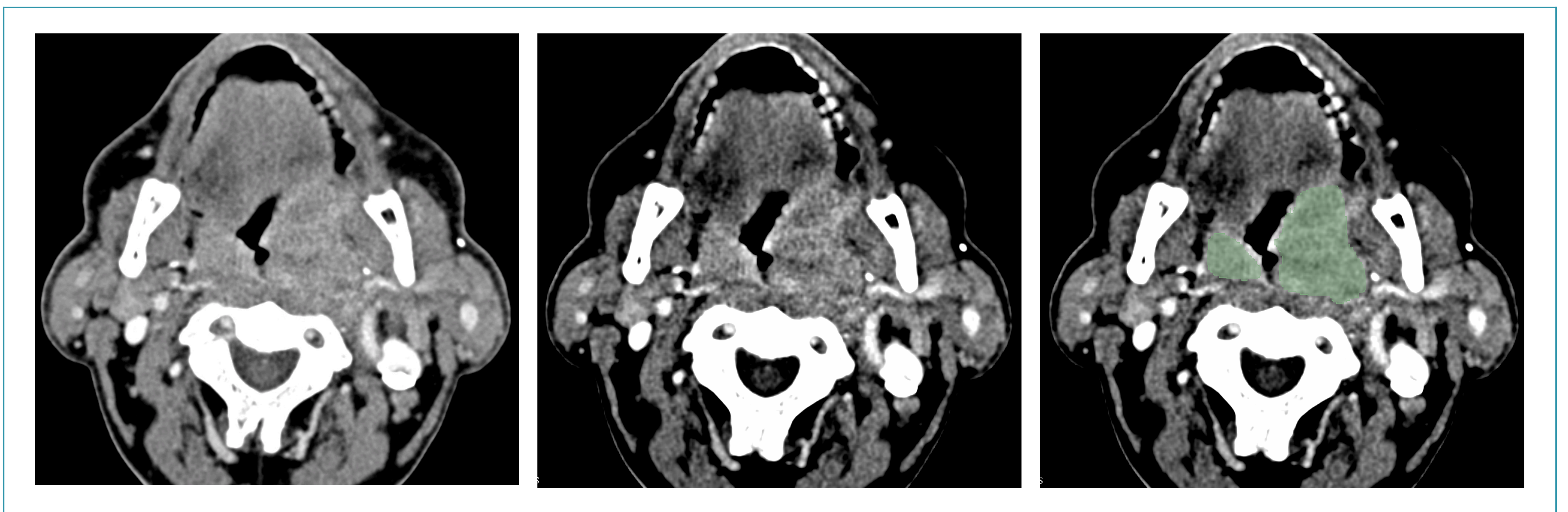


Fig. 4. Ejemplo de carcinoma escamoso de orofaringe. Imagen tras ajuste del histograma e imagen de la segmentación.

6. SEGMENTACIÓN

Existen diferentes métodos de segmentación:

	SEGMENTACIÓN MANUAL	SEGMENTACIÓN SEMIAUTOMÁTICA	SEGMENTACIÓN AUTOMÁTICA
QUIÉN	Observador	Máquina + observador	Máquina
CÓMO	El usuario delimita manualmente el área de interés.	Mediante algoritmos de segmentación (basados en umbrales, en bordes o en regiones), con posterior corrección manual por el usuario.	Mediante algoritmos basados en aprendizaje automático: redes neuronales convolucionales (CNN) y modelos de aprendizaje profundo.
VENTAJAS	<ul style="list-style-type: none"> - Control total por el usuario - Preciso 	<ul style="list-style-type: none"> - Permite correcciones 	<ul style="list-style-type: none"> - Más rápida - Útil en muestras grandes - Resultados más repetibles - Elimina variabilidad inter/intraobservador
INCONVENIENTES	<ul style="list-style-type: none"> - Lenta y laboriosa - Poco adecuada si n elevada - Resultados sujetos a variabilidad inter/intraobservador 	<ul style="list-style-type: none"> - Requiere ciertas características como homogeneidad y bordes bien definidos - Resultados sujetos a variabilidad inter/intraobservador 	<ul style="list-style-type: none"> - Más costosa - Requiere de entrenamiento previo - Falta de generalización de los algoritmos actuales

6. SEGMENTACIÓN

→ EJEMPLO: Segmentación de tumores de orofaringe.

- TC de cuello con contraste endovenoso, en pacientes con carcinoma de orofaringe, previa al tratamiento.
- Segmentación manual del tumor y de todas las adenopatías en cortes contiguos con el *software* 3D Slicer.

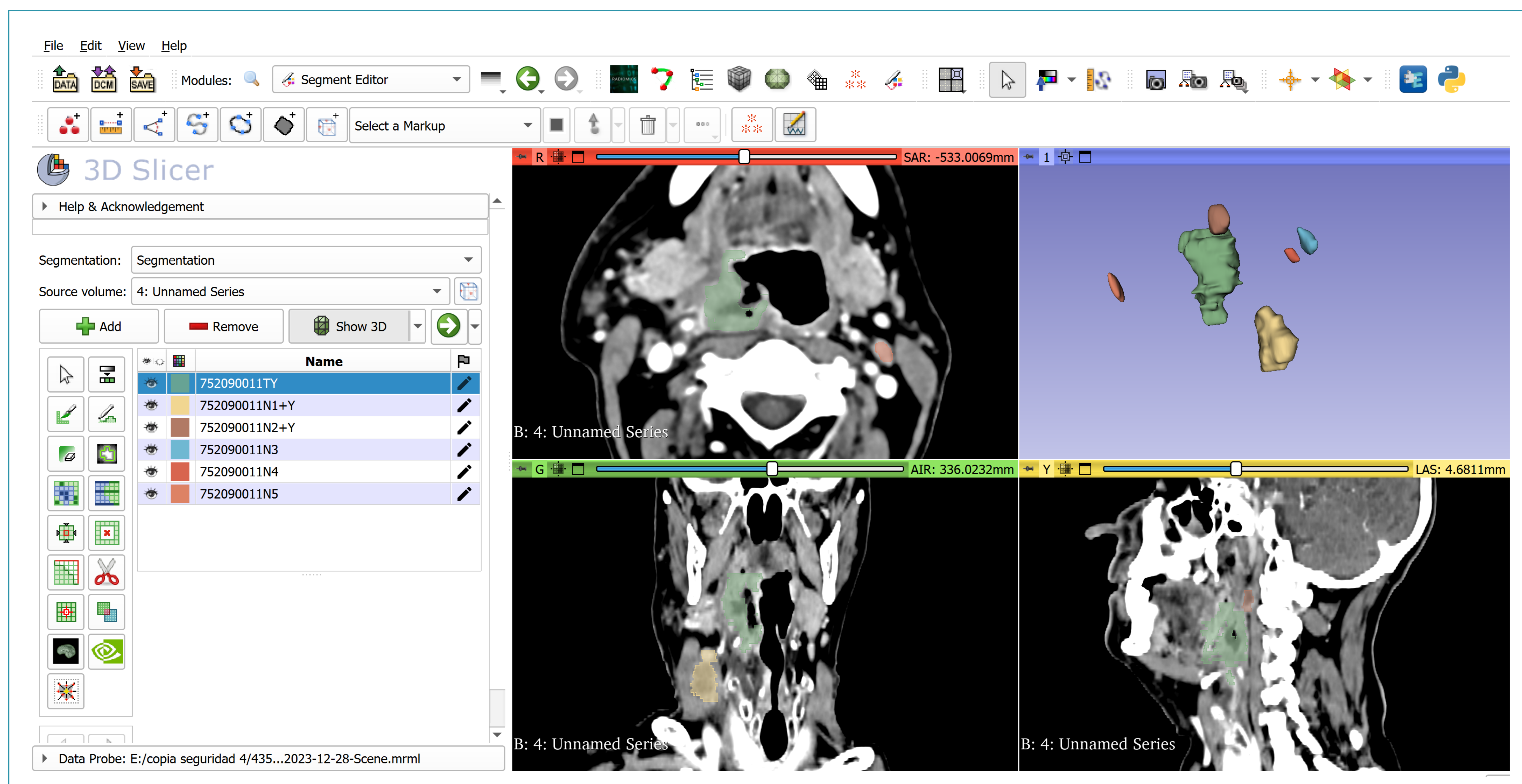


Fig. 5. Vista del proceso de segmentación de un carcinoma de orofaringe y de adenopatías cervicales en el *software* 3D Slicer.

6. SEGMENTACIÓN

En la mayoría de casos es necesario realizar, antes de la segmentación, un **procesamiento de la imagen**, que puede incluir la normalización de la misma.

El objetivo es mejorar la calidad de la imagen y facilitar la detección de las estructuras de interés, ya que todo ello afectará directamente las características extraídas y su reproducibilidad.

Mejora del contraste mediante la modificación de intensidades de gris (histograma).

Eliminación de píxeles/vóxeles del ROI/VOI que quedan fuera de un rango específico de niveles de gris (aire, hueso, vasos).

Eliminación de artefactos o corrección de los mismos mediante reconstrucciones específicas.

Reducción del ruido mediante técnicas de filtrado y suavizado de bordes.

6. SEGMENTACIÓN

→ *EJEMPLO 1*: ante una imagen difícil de segmentar, con un tumor poco contrastado, se puede modificar el histograma de intensidades de gris para mejorar el contraste.

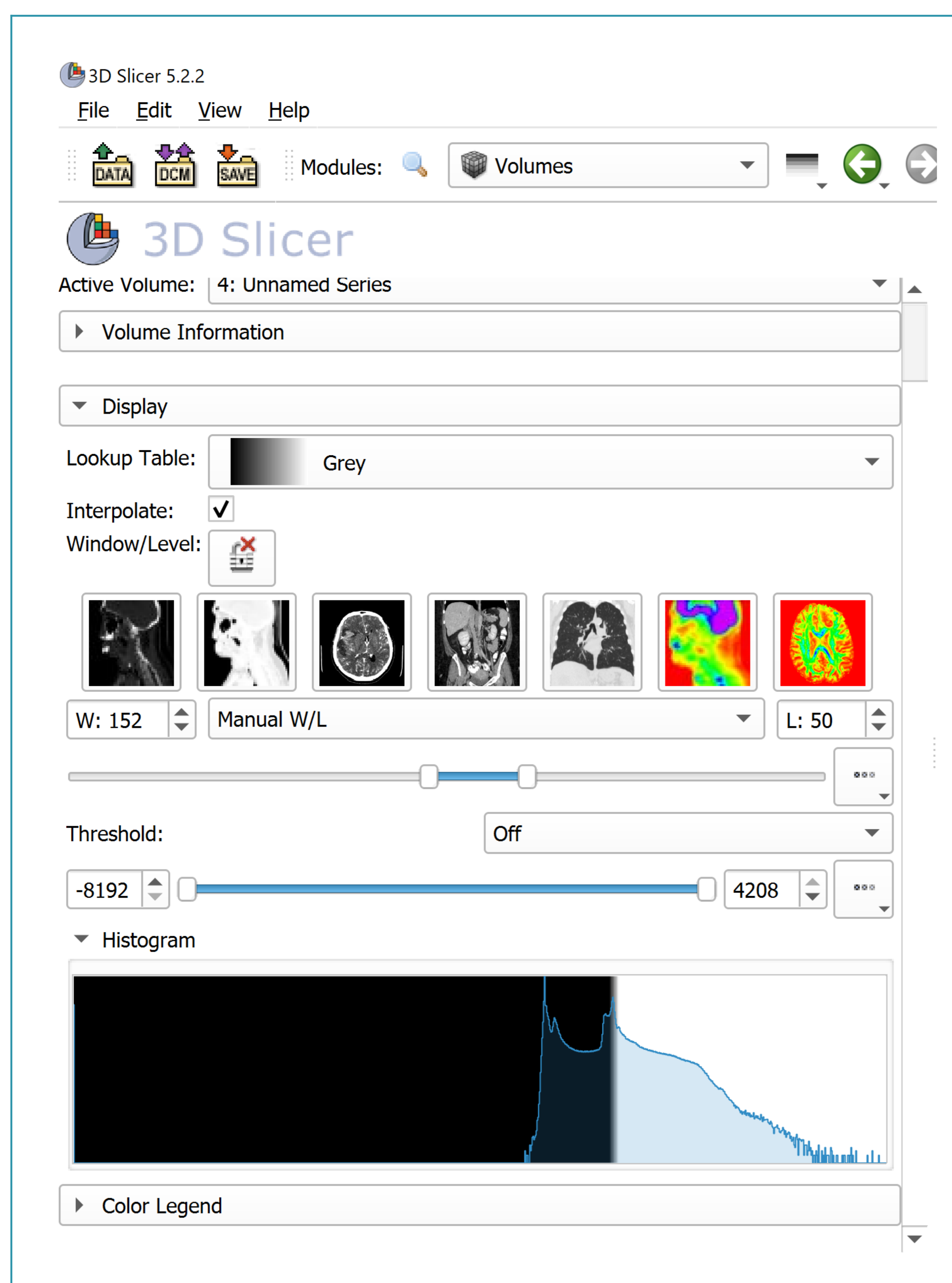


Fig. 6. Vista del ajuste de histograma del *software* 3D Slicer.

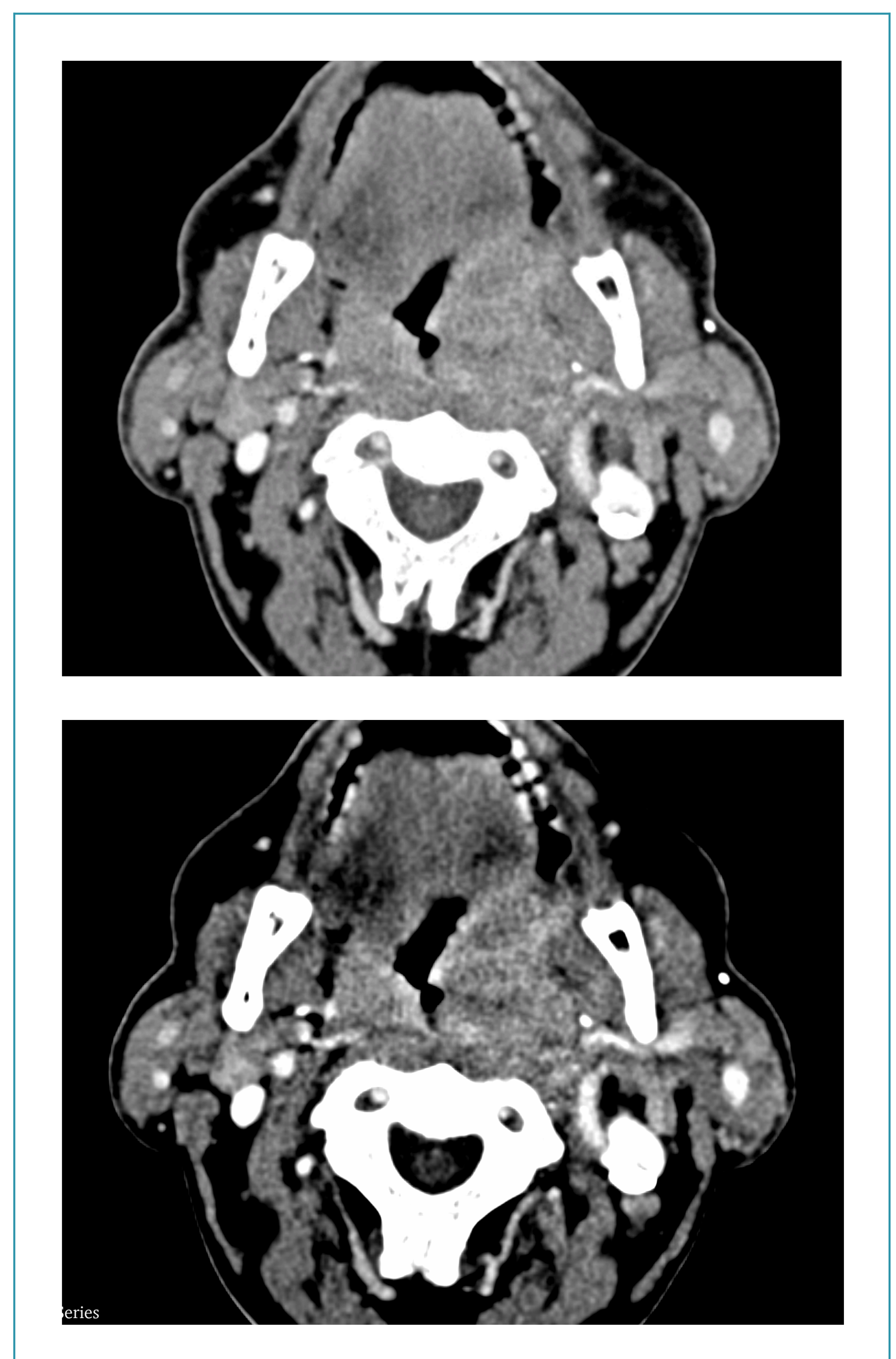


Fig. 7. Ejemplo de carcinoma escamoso de orofaringe e imagen tras el ajuste de histograma.

6. SEGMENTACIÓN

→ *EJEMPLO 2*: los píxeles/vóxeles de aire, hueso y vasos en el interior del área de interés deben ser excluidos de la segmentación, ya que al tener propiedades radiológicas diferentes en comparación con los tejidos blandos, las características radiómicas del tumor pueden verse alteradas.

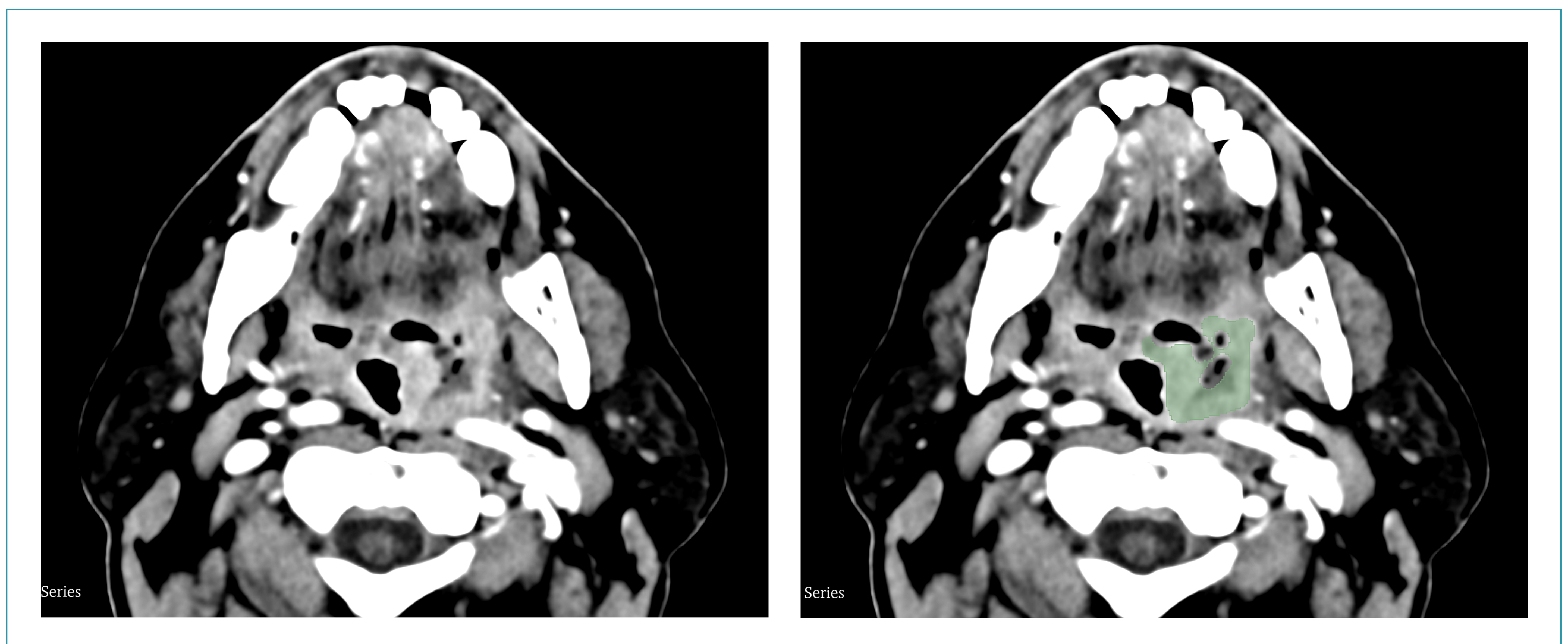


Fig. 8. Ejemplo de carcinoma escamoso de orofaringe con áreas de necrosis e imagen de la segmentación.

6. SEGMENTACIÓN

→ *EJEMPLO 3*: los artefactos (de movimiento por deglución, por material metálico, etc.) dan lugar a una pérdida de contraste dificultando la identificación de las estructuras anatómicas y la delimitación del tumor en algunos cortes. Si no es posible su corrección, las áreas con artefactos deben ser excluidas de la segmentación.

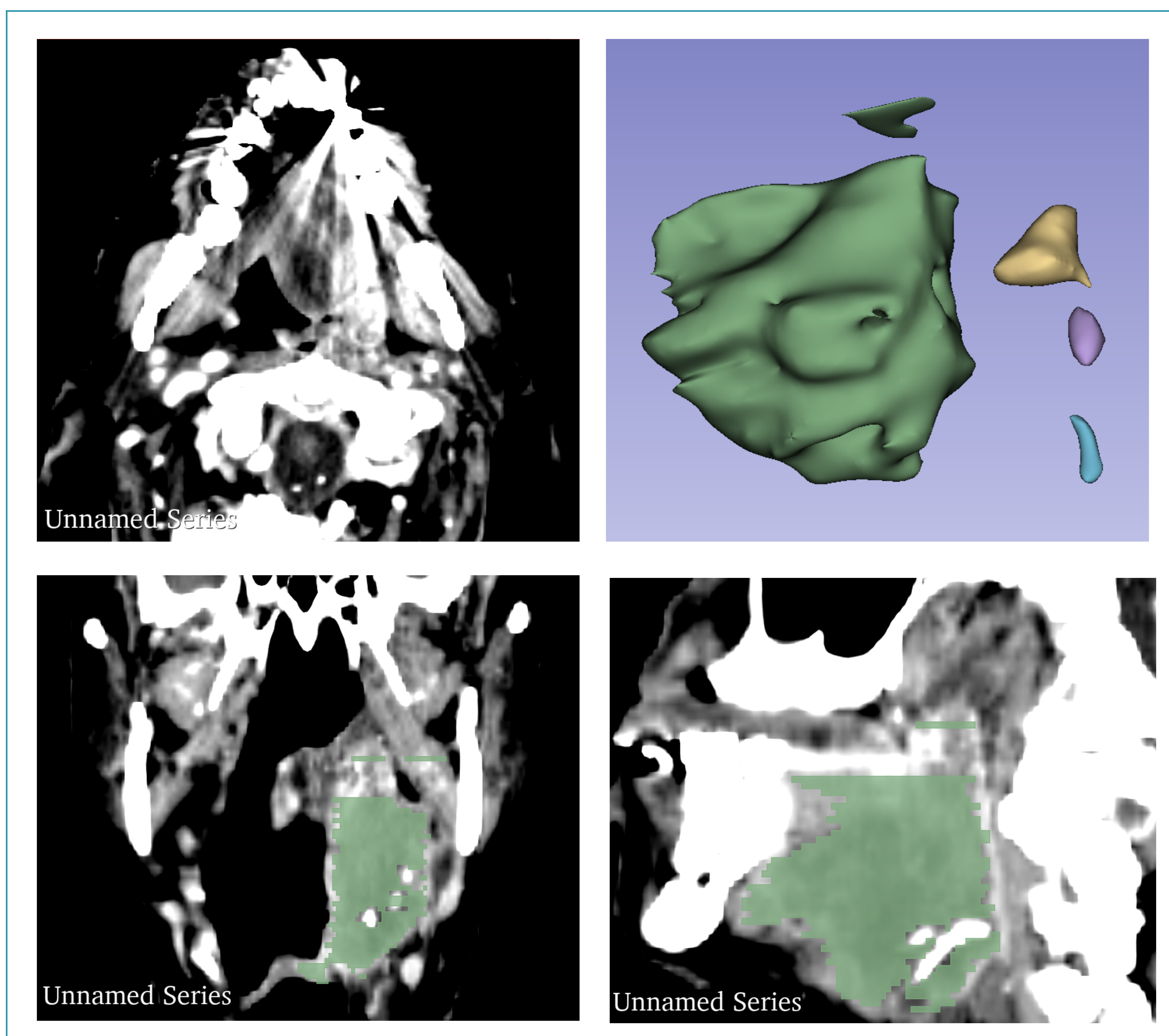


Fig. 9. Ejemplo de carcinoma escamoso de base de lengua, con artefacto en cavidad oral. Imágenes de la segmentación y máscara 3D.

6. SEGMENTACIÓN

- Una vez realizada la segmentación, los ROI/VOI obtenidos pueden utilizarse como **máscaras** para impresión 3D, para planificación y simulación quirúrgica (realidad aumentada) o para el entrenamiento de algoritmos de aprendizaje profundo.
- También pueden extraerse los **datos radiómicos** y utilizarse, de forma aislada o en combinación con otros datos (clínicos, de laboratorio, histológicos, genómicos, proteómicos, etc.), como **apoyo a la toma de decisiones clínicas**. La identificación de patrones/fenotipos, mediante técnicas de aprendizaje automático, podría hacer de ellos potenciales biomarcadores diagnósticos, pronósticos, predictivos, de respuesta al tratamiento y de seguimiento de la enfermedad.

6. SEGMENTACIÓN

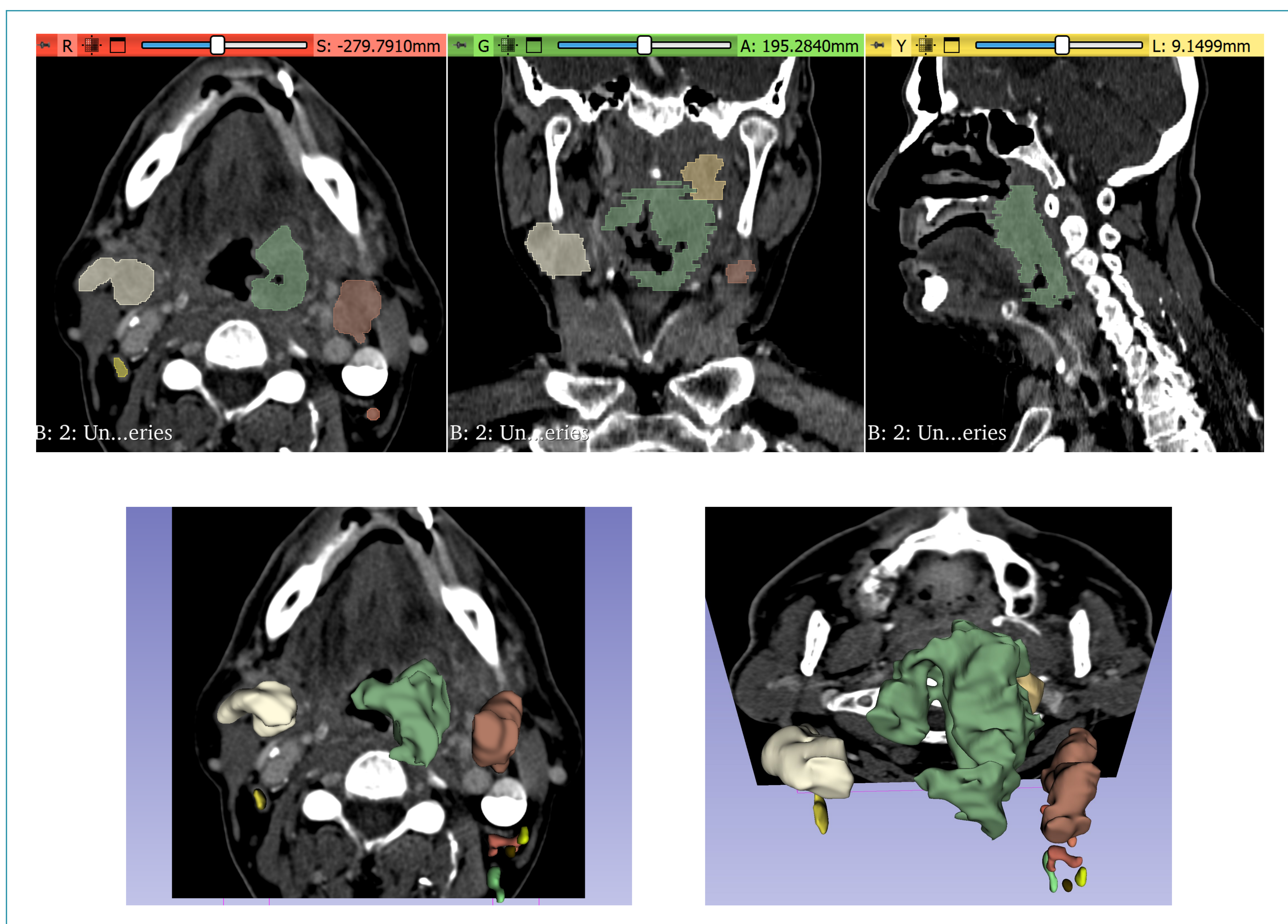


Fig. 10. Imágenes de segmentación de carcinoma de orofaringe y de adenopatías cervicales, con posterior creación de máscara 3D.

7. EXTRACCIÓN DE RESULTADOS

Las características radiómicas extraídas a partir de las imágenes se pueden dividir en cuatro categorías:

SHAPE FEATURES (CARACTERÍSTICAS DE FORMA)

Describen la **forma** de la región de interés y sus propiedades geométricas, tanto bidimensionales como tridimensionales. Incluyen volumen, diámetro máximo, superficie máxima, esfericidad, compacidad, etc.

FIRST-ORDER FEATURES (CARACTERÍSTICAS DE PRIMER ORDEN)

Son propiedades estadísticas que describen la distribución de los valores de **intensidad** de píxeles/vóxeles individuales dentro de la región de interés, sin tener en cuenta las relaciones espaciales. Se basan en histogramas. Incluyen los valores medio, mediano, máximo y mínimo de las intensidades de los píxeles/vóxeles en la imagen, asimetría, curtosis, etc.

SECOND-ORDER FEATURES (CARACTERÍSTICAS DE SEGUNDO ORDEN)

Son propiedades de **textura** que describen la relación de la distribución espacial de los valores de intensidad entre píxeles/vóxeles vecinos. Pueden derivarse de la matriz de coocurrencia de nivel de grises (GLCM), de la matriz de zonas de tamaño de nivel de grises (GLSZM), de la matriz de longitud de recorrido de nivel de grises (GLRLM), entre otras. Aportan información sobre la heterogeneidad de la lesión.

HIGHER-ORDER FEATURES (CARACTERÍSTICAS DE ORDEN SUPERIOR)

Se obtienen después de aplicar filtros o transformaciones matemáticas a las imágenes, con el fin de identificar patrones o eliminar ruido, por ejemplo. Incluyen transformadas de Fourier, transformada wavelet y filtrado Laplaciano del Gaussiano, entre otros.

7. EXTRACCIÓN DE RESULTADOS

A partir de las imágenes se puede obtener un gran número de características radiómicas. La **reducción de la dimensionalidad** consiste en **seleccionar** aquellas más potentes mediante algoritmos de aprendizaje automático y/o métodos estadísticos convencionales.

- Las más **reproducibles** en las segmentaciones manual o semiautomática mediante análisis de reproducibilidad (análisis de correlación intraclase).
- Las más **relevantes** para el propósito deseado.
- Las más **representativas** mediante la creación de grupos de características altamente correlacionadas (*clusters* de correlación) y la eliminación de aquellas que son redundantes.

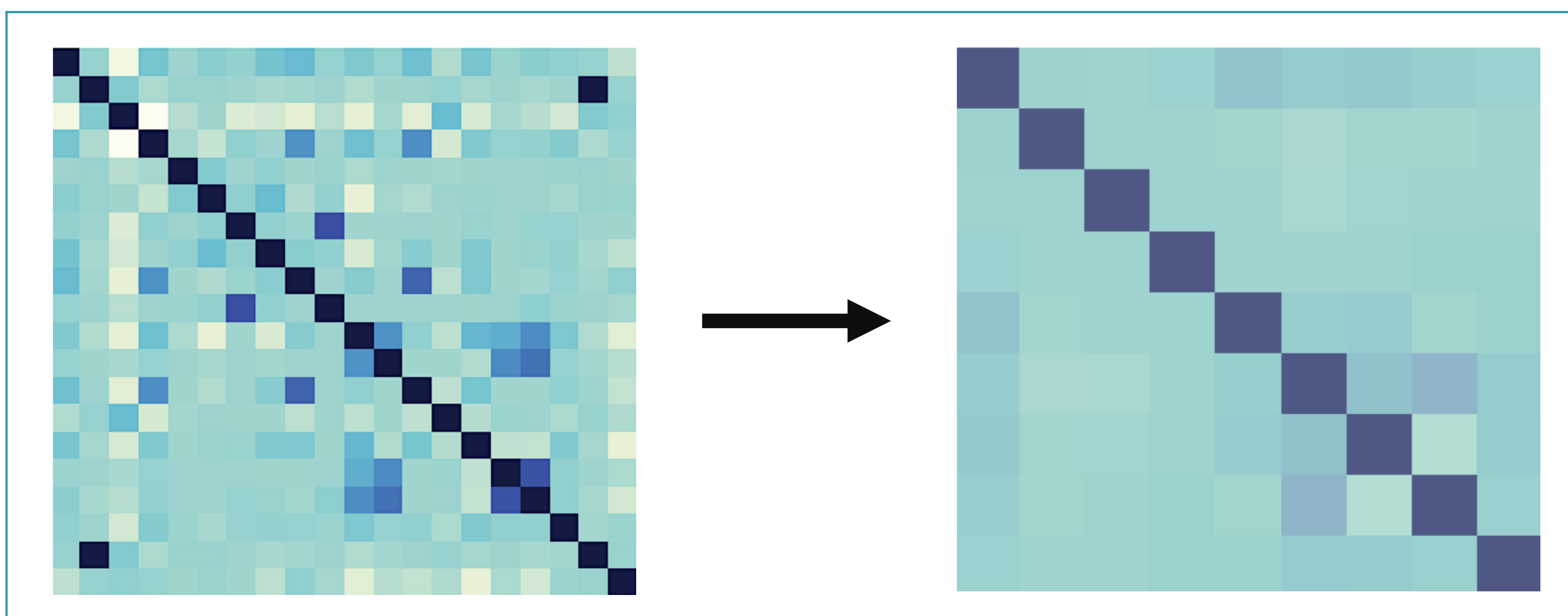


Fig. 11. Representación de la reducción de la dimensionalidad de las características radiómicas.

8. CONSTRUCCIÓN DEL MODELO DE IA

Una vez seleccionadas las características radiómicas más relevantes, se procede a la **construcción del modelo** de aprendizaje automático con el fin de generar **resultados válidos y generalizables**.

1. **Selección del tipo de modelo** según el propósito del estudio. Regresión, clasificación o agrupamiento (*clustering*) son los más comunes.
2. **Selección del algoritmo** de aprendizaje automático para el modelo elegido teniendo en cuenta la cantidad de datos disponibles y sus características.
3. **Entrenamiento**: una parte del conjunto de datos se utiliza para enseñar al modelo los patrones presentes en los datos. Implica ajustar los parámetros del modelo para minimizar los errores entre sus predicciones y los valores reales en los datos de entrenamiento.
4. **Validación**: otra parte del conjunto de datos se utiliza para evaluar el rendimiento del modelo que se ha entrenado y ajustar sus parámetros sin que quede sobreajustado al conjunto de datos de entrenamiento.
5. **Pruebas**: se utilizan conjuntos de datos independientes para evaluar cómo se comporta el modelo ante datos nuevos. Pueden ser precisos ajustes adicionales para mejorarlo, previo a la implementación en un entorno clínico.

9. CONCLUSIONES

- Este trabajo pretende familiarizar a los radiólogos con las tareas de **segmentación** y **extracción de datos** de la imagen para su posterior análisis, así como proporcionar un método de trabajo **multidisciplinar** y **transversal** que estimule el desarrollo de trabajos científicos en la era de la revolución de la IA.

10. BIBLIOGRAFÍA

- Fournier L, Costaridou L, Bidaut L, et al. European Society Of Radiology. Incorporating radiomics into clinical trials: expert consensus endorsed by the European Society of Radiology on considerations for data-driven compared to biologically driven quantitative biomarkers. *Eur Radiol.* 2021;31(8):6001-6012. DOI: 10.1007/s00330-020-07598-8
- Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology.* 2016;278(2):563-77. DOI: 10.1148/radiol.2015151169
- Mayerhoefer EM, Materka A, Langs G, et al. Introduction to Radiomics. *J Nucl Med.* 2020;61(4):488-495. DOI: 10.2967/jnumed.118.222893
- Peng Z, Wang Y, Wang Y, et al. Application of radiomics and machine learning in head and neck cancers. *Int J Biol Sci.* 2021;17(2):475-486. DOI: 10.7150/ijbs.55716
- Rizzo S, Botta F, Raimondi S, et al. Radiomics: the facts and the challenges of image analysis. *Eur Radiol Exp.* 2018;2(1):36. DOI: 10.1186/s41747-018-0068-z
- Van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* 2017;77(21):e104-e7. DOI: 10.1158/0008-5472.CAN-17-0339
- Van Timmeren J, Cester D, Tanadini-Lang S, et al. Radiomics in medical imaging—“how-to” guide and critical reflection. *Insights Imaging.* 2020;11:91. DOI: 10.1186/s13244-020-00887-2
- Wong AJ, Kanwar A, Mohamed AS, et al. Radiomics in head and neck cancer: from exploration to application. *Transl Cancer Res.* 2016;5(4):371-382. DOI: 10.21037/tcr.2016.07.18