



VNIVERSITAT
DE VALÈNCIA

Facultat de Medicina i Odontologia

Departamento de Medicina

Programa de Doctorado de Medicina 3139 (R. D. 99/2011)

Estudio de la agresividad biológica de tumores de estirpe neuroblástica en edad pediátrica mediante el análisis de las imágenes de Resonancia Magnética

TESIS DOCTORAL

Doctoranda:

Diana Veiga Canuto

Dirigida por:

Dr. Luis Martí Bonmatí

Leonor Cerdá Alberich

Arantxa Blasco Serra

Tutor:

Juan Manuel Sanchis García

Valencia, Abril de 2024

Don Luis Martí Bonmatí, Doctor en Medicina y Cirugía por la Universitat de València.

CERTIFICA:

Que la presente tesis doctoral titulada [Estudio de la agresividad biológica de tumores de estirpe neuroblástica en edad pediátrica mediante el análisis de las imágenes de Resonancia Magnética](#) presentada por Doña Diana Veiga Canuto, ha sido realizada bajo mi dirección y cumple los requisitos para que su autora pueda obtener con ella el título de Doctora por la Universitat de València.

Para que conste, firmo el presente certificado, en Valencia, a

Fdo. D. Luis Martí Bonmatí

Doña Leonor Cerdá Alberich, Doctora en Física por la Universitat de València y Profesora Asociada del Departamento de Estadística e Investigación Operativa Aplicadas y Calidad de la Universidad Politécnica de Valencia.

CERTIFICA:

Que la presente tesis doctoral titulada [Estudio de la agresividad biológica de tumores de estirpe neuroblástica en edad pediátrica mediante el análisis de las imágenes de Resonancia Magnética](#) presentada por Doña Diana Veiga Canuto, ha sido realizada bajo mi dirección y cumple los requisitos para que su autora pueda obtener con ella el título de Doctora por la Universitat de València.

Para que conste, firmo el presente certificado, en Valencia, a

Fdo. D. Leonor Cerdá Alberich

Doña Arantxa Blasco Serra, Doctora en Neurociencias por la Universitat de València y Profesora Ayudante Doctor del Departamento de Anatomía y Embriología de la Universitat de València

CERTIFICA:

Que la presente tesis doctoral titulada [Estudio de la agresividad biológica de tumores de estirpe neuroblástica en edad pediátrica mediante el análisis de las imágenes de Resonancia Magnética](#) presentada por Doña Diana Veiga Canuto, ha sido realizada bajo mi dirección y cumple los requisitos para que su autora pueda obtener con ella el título de Doctora por la Universitat de València.

Para que conste, firmo el presente certificado, en Valencia, a

Fdo. D. Arantxa Blasco Serra

Don Juan Manuel Sanchis García, Doctor en Medicina por la Universitat de València y Profesor Asociado del Departamento de Medicina de la Universitat de València

CERTIFICA:

Que la presente tesis doctoral titulada [Estudio de la agresividad biológica de tumores de estirpe neuroblástica en edad pediátrica mediante el análisis de las imágenes de Resonancia Magnética](#) presentada por Doña Diana Veiga Canuto, ha sido realizada bajo mi dirección y cumple los requisitos para que su autora pueda obtener con ella el título de Doctora por la Universitat de València.

Para que conste, firmo el presente certificado, en Valencia, a

Fdo. D. Juan Manuel Sanchis García

Una flor d'avui es marceix just a l'endemà
Cal que neixin flors a cada instant

A mis padres Rosa y José, y a mi hermana Nadia, por enseñarme el valor del esfuerzo y la constancia y por el apoyo en el camino de esta Tesis Doctoral.

AGRADECIMIENTOS

A Luís Martí Bonmatí, por su excepcional habilidad para transformar motivación en sabiduría.

A Leonor Cerdá Alberich, por fomentar el progreso científico.

A Arantxa Blasco Serra, por animarme y ayudarme a avanzar.

A Juan Manuel Sanchis García, por su buena disposición.

Al Instituto de Investigación Sanitaria La Fe por permitirme llevar a cabo este proyecto de Tesis Doctoral como personal investigador.

A mis compañeros del Grupo de Investigación Biomédica en Imagen del Instituto de Investigación Sanitaria La Fe por su constante apoyo y amistad. A Matías Fernández y a María Beser por su sincera ayuda.

A los colaboradores del proyecto PRIMAGE por su inestimable aportación. Al servicio de Oncología Pediátrica del Hospital Universitari i Politècnic La Fe, al St. Anna Children's Hospital and Children's y al Cancer Research Institute (CCRI) (Viena, Austria), a los Departamentos de Oncología Pediátrica de los siguientes centros sanitarios: Hospital Universitario de Niños de Colonia, Hospital Universitario de Pisa, Instituto Gaslini (Pisa, Italia), Hospital La Paz y Hospital Niño Jesús (Madrid, España), Hospital Son Espases (Palma de Mallorca, España), y Hospital Reina Sofía (Córdoba, España).

A Ana Jiménez por estar siempre dispuesta a colaborar.

A mis compañeros de la sección de Radiología Pediátrica del Área de Imagen Médica del Hospital Universitari i Politècnic La Fe, por acompañarme en este proceso. A Cinta Sangüesa por su ayuda con las segmentaciones de tumores neuroblásticos.

A Joan Carreres, per encoratjar-me sempre.

ÍNDICE GENERAL

INTRODUCCIÓN.....	33
1.1. El problema clínico que queremos abordar: heterogeneidad de los tumores neuroblásticos	33
1.2. Patrones de referencia para la agresividad tumoral: estratificación del riesgo	37
1.2.1. INSS.....	38
1.2.2. INRGSS	39
1.2.3. INRG.....	41
1.3. Minimizando limitaciones: biomarcadores de imagen.....	43
1.4. Análisis avanzado de imagen: preparación de la imagen, segmentación, radiómica y radiómica profunda	45
1.5. Finalidad del estudio: estimación de la agresividad	49
HIPÓTESIS.....	52
OBJETIVOS	53
METODOLOGÍA.....	55
3.1. Diseño del proyecto de investigación	55
3.2. Población de pacientes.....	57
3.2.1. Estimación del tamaño muestral.....	57
3.2.2. Reclutamiento: proceso de selección y criterios de elegibilidad.....	60
3.2.3. Distribución de pacientes.....	61
3.3. Base de datos	62
3.3.1. Privacidad de datos: seudonimización	62
3.3.2. Seguridad de los datos: ubicaciones de bases de datos	63
3.4. Plataforma.....	63
3.5. Evaluación clínica y biomarcadores.....	67
3.5.1. Variables clínicas	68
3.5.2. Variables analíticas.....	68
3.5.3. Variables de imagen.....	69
3.5.4. Variables de estadificación.....	70
3.5.5. Variables clínicas establecidas que definen la agresividad.....	70
3.6. Imagen de RM.....	71
3.6.1. Protocolos de imagen de RM	71

3.6.2.	Selección de secuencias	72
3.7.	Preparación de la imagen RM	72
3.7.1.	Filtros de ruido (<i>denoising</i>)	73
3.7.2.	Eliminación de inhomogeneidades.....	75
3.7.3.	Normalización de intensidad	76
3.7.4.	Cambio de tamaño y registro (remuestreo espacial)	77
3.8.	Segmentación manual en RM.....	77
3.8.1.	Metodología de segmentación.....	77
3.8.2.	Criterios restrictivos de segmentación	80
3.9.	Metodología estadística	81
3.9.1.	Segmentación.....	81
3.9.2.	Reproducibilidad de la radiómica.....	86
3.9.3.	Relación de la radiómica y la radiómica profunda con variables analíticas y de estadificación	87
3.9.4.	Modelos predictivos de supervivencia.....	90
RESULTADOS.....	94
4.1.	Características de la población.....	94
4.2.	Segmentación manual: análisis de la variabilidad interobservador	96
4.3.	Segmentación automática en RM.....	101
4.3.1.	Herramienta para la segmentación: nnU-Net.....	102
4.3.2.	Desarrollo de una herramienta de detección y segmentación automática.....	105
4.3.3.	Validación de la herramienta de detección y segmentación automática en RM...	118
4.4.	Estudio de radiómica	131
4.4.1.	Extracción de parámetros de radiómica.....	131
4.4.2.	Estabilidad de la radiómica.....	136
4.4.3.	Relación de la radiómica con variables analíticas y de estadificación	157
4.5.	Estudio de radiómica profunda.....	165
4.5.1.	Extracción de radiómica profunda	165
4.5.2.	Reducción de variables de radiómica profunda.....	168
4.5.3.	Relación de la radiómica profunda con variables analíticas y de estadificación...	169
4.6.	Modelo predictivo de supervivencia global.	174
4.6.1.	Entrenamiento de un modelo predictivo de supervivencia global.....	176
4.6.2.	Validación de un modelo de supervivencia global.....	178
4.6.3.	Entrenamiento de otros modelos de supervivencia global.....	181
4.6.4.	Validación de otros modelos de supervivencia global	182

4.7. Explicabilidad de los modelos de supervivencia global desarrollados	183
4.7.1. Posibles biomarcadores de imagen para predicción de supervivencia global en tumores neuroblásticos.....	183
4.7.2. Variables más predictivas del modelo de supervivencia	186
DISCUSIÓN	191
5.1. Resultados más importantes y comparación con otros trabajos	192
5.1.1. Segmentación automática	192
5.1.2. Reproducibilidad de la radiómica.....	197
5.1.3. Relación de la radiómica y radiómica profunda con variables analíticas y de estadificación.....	199
5.1.4. Modelos de predicción de supervivencia global.....	201
5.2. Limitaciones y líneas de futuro	204
CONCLUSIONES	207
BIBLIOGRAFÍA	210
ANEXOS	226
ANEXO 1	228
ANEXO 2	238
ANEXO 3	242
ANEXO 4	249

LISTA DE SÍMBOLOS, ABREVIATURAS Y SIGLAS

ADF: Filtro de Difusión Anisotrópica
ANOVA: Análisis de la Varianza
AUC: Área debajo de la Curva
CCC: Coeficiente de Correlación de Concordancia
CFF: Filtro de Flujo de Curvatura
CNN: Red Neuronal Convolutiva
CoV: Coeficiente de Variación
CSD: Coeficiente de Similitud Dice
DCE: Realce Dinámico de Contraste
DICOM: Imagen Digital y Comunicación en Medicina
DW: Imagen potenciada en Difusión
FN: Falsos Negativos
FP: Falsos Positivos
FS: Supresión Grasa
GF: Filtro Gaussiano de eliminación de ruido
GLCM: Matriz de Co-ocurrencia de Niveles de Gris
GLDM: Matriz de Dependencia de Niveles de Gris
GLRLM: Matriz de Longitud de Segmentos Consecutivos de Píxeles con el mismo Nivel de Gris
GLSZM: Matriz de Tamaño de Zona de Niveles de Gris
HD: Distancia de Hausdorff
IA: Inteligencia Artificial
IDRF: Factores de Riesgo Definidos por Imagen
INRG: Grupo Internacional de Riesgo del Neuroblastoma
INRGSS: Sistema de Estadificación del Grupo Internacional de Riesgo del Neuroblastoma
IPW: *Inverse Probability Weighting* (Ponderación de probabilidad inversa)
IQR: Rango Intercuartílico
IR: Recuperación de la Inversión
JAC: índice de Jaccard
NGTDM: Matriz de Diferencia de Tono de Gris Vecino
NIFTI: Iniciativa de Tecnología Informática de Neuroimagen

NLMF: Filtro de Medias No Locales

PACS: Sistema de Archivo y Comunicación de Imágenes

PRIMAGE: PRedictive In-silico Multiscale Analytics to support cancer personalized diaGnosis and prognosis, Empowered by imaging biomarkers

RM: Resonancia Magnética

ROC: Característica Operativa del Receptor

ROI: Región de Interés

DE: Desviación Estándar

SE: *Spin Echo*

STIR: *Short inversion Time Inversion Recovery*

TC: Tomografía Computarizada

TFN: Tasa de Falsos Negativos

TFPm: Tasa de Falsos Positivos modificada

T: Tesla

T2wSE: T2 spin echo

T2*wGE: T2* eco de gradiente

VN: Verdaderos Negativos

VP: Verdaderos Positivos

ÍNDICE DE TABLAS Y FIGURAS

Tablas

Tabla 1. Sistema de Estadificación Internacional del Neuroblastoma (INSS).....	38
Tabla 2. Factores de Riesgo Definidos por Imagen (IDRF)	39
Tabla 3. Sistema de estadificación del Grupo Internacional de Riesgo del Neuroblastoma (INRGSS)	41
Tabla 4. Sistema de clasificación de riesgo del Grupo Internacional de Riesgo del Neuroblastoma (INRG).....	42
Tabla 5. Estimación del tamaño muestral requerido para realizar un modelo de supervivencia para neuroblastoma	58
Tabla 6. Tamaño muestral en las diferentes variables analizadas para el análisis de radiómica y radiómica profunda.	87
Tabla 7. Distribución de pacientes en función de las diferentes variables analíticas y de estadificación analizadas para el análisis de radiómica y radiómica profunda y de la edad	88
Tabla 8. Características basales de la población estudiada	95
Tabla 9. Características basales de la muestra empleada para la comparación de segmentaciones manuales entre dos observadoras	99
Tabla 10. Variabilidad interobservador para la segmentación manual de tumores neuroblásticos considerando diferentes métricas de rendimiento.....	100
Tabla 11. Características basales de la muestra empleada para el desarrollo de la herramienta de detección y segmentación automática.....	106
Tabla 12. División equilibrada y estratificada de los casos para cada uno de los subconjuntos para el entrenamiento y validación interna de la herramienta de segmentación automática	107

Tabla 13. Métricas de rendimiento para los 5 subconjuntos de entrenamiento mediante validación cruzada	110
Tabla 14. Solución de conjunto del modelo de segmentación aplicada a todos los casos de entrenamiento-ajuste.....	113
Tabla 15. Modelo lineal generalizado sobre la variable respuesta Coeficiente de Similitud Dice considerando como factores categóricos las variables localización y campo magnético y como covariables la edad y volumen tumoral.....	114
Tabla 16. Métricas de rendimiento para el conjunto de validación interna del modelo de segmentación automática	117
Tabla 17. Comparación de las características de la muestra en el conjunto de datos de entrenamiento original y el de validación.	122
Tabla 18. Métricas de rendimiento para el conjunto de validación independiente, comparando las máscaras obtenidas automáticamente con las corregidas manualmente	124
Tabla 19. Métricas de rendimiento para el conjunto de validación independiente del modelo de segmentación automática estratificadas en función del momento de realización de la RM, la ubicación, el campo magnético y la secuencia segmentada ...	127
Tabla 20. Análisis de la varianza para el conjunto de validación independiente de la herramienta de segmentación automática.	128
Tabla 21. Características basales de la muestra empleada para para la extracción y posterior análisis de radiómica	133
Tabla 22. Lista de las 107 características radiómicas de forma, de primer orden y de segundo orden extraídas de tumores neuroblásticos en imágenes de resonancia magnética potenciada en T2/T2*	135
Tabla 23. Esquema de las modificaciones de imagen aplicadas para el análisis de la estabilidad de la radiómica.....	139

Tabla 24. Número de características de radiómica y porcentaje de aquellas que presentaron un nivel de acuerdo excelente, bueno, moderado o pobre en cuanto al CCC para las 8 series con modificaciones en comparación con la verdad fundamental	142
Tabla 25. Número y porcentaje de características de radiómica de primer y segundo orden que presentaron un excelente nivel de acuerdo para las 6 series con modificaciones en el procesamiento de imágenes (series 2 a 7) en comparación con la serie de referencia	144
Tabla 26. Nombres de las características de radiómica que mostraron un $CCC > 0.90$ después de la modificación de filtros (series 2 a 5).....	145
Tabla 27. Características radiómicas que mostraron un $CCC > 0.90$ después de la aplicación de remuestreo y normalización combinadas (series 6 y 7).....	147
Tabla 28. Número y porcentaje de características de radiómica de primer y segundo orden que presentaron un excelente nivel de acuerdo para las 2 series con modificaciones en la forma de la máscara (series 8 y 9) en comparación con la serie de referencia	149
Tabla 29. Características radiómicas que mostraron un $CCC > 0.90$ después de la aplicación de dilataciones y erosiones en las máscaras (series 8 y 9).	150
Tabla 30. Características radiómicas que mostraron un nivel pobre de acuerdo en todas las series en comparación con la verdad fundamental.	152
Tabla 31. Análisis de la varianza para las series 2 a 7	153
Tabla 32. Comparaciones post hoc de las series 2 a 7	154
Tabla 33. Análisis de la varianza para las series 8 y 9.	156
Tabla 34. Características radiómicas que mostraron un Coeficiente de Variación < 0.10	156
Tabla 35. Características basales de la población estudiada en los análisis finales de esta Tesis Doctoral	159
Tabla 36. Distribución de la muestra en función del MYCN	160

Tabla 37. Variables de radiómica con diferencias significativas en la amplificación de MYCN en <18 meses.	160
Tabla 38. Distribución de la muestra en función del INRGSS.	161
Tabla 39. Variables de radiómica con diferencias significativas en función de INRGSS en pacientes <18 meses.....	161
Tabla 40. Variables de radiómica profunda con diferencias significativas en función de INRGSS en pacientes ≥ 18 meses.	162
Tabla 41. Distribución de la muestra en función del INRGSS.	163
Tabla 42. Variables de radiómica con diferencias significativas en función de INRG en pacientes ≥ 18 meses.	163
Tabla 43. Resumen de los resultados de las relaciones estadísticamente significativas en cada subgrupo según la edad y las variables analíticas y de estadificación con las variables de radiómica.	164
Tabla 44. Variables de radiómica profunda con diferencias significativas en la amplificación de MYCN en <18 meses.....	170
Tabla 45. Variables de radiómica profunda con diferencias significativas en la amplificación de MYCN en ≥ 18 meses.	170
Tabla 46. Variables de radiómica profunda con diferencias significativas en función de INRGSS en pacientes <18 meses.....	171
Tabla 47. Variables de radiómica profunda con diferencias significativas en función de INRGSS en pacientes ≥ 18 meses.	171
Tabla 48. Variables de radiómica profunda con diferencias significativas en función de INRG en pacientes <18 meses.	172
Tabla 49. Variables de radiómica profunda con diferencias significativas en función de INRG en pacientes ≥ 18 meses.	172
Tabla 50. Resumen de los resultados de las relaciones estadísticamente significativas en cada subgrupo según la edad y las variables analíticas y de estadificación con la radiómica profunda.....	173

Tabla 51. Resultados del modelo de supervivencia Cox para el conjunto de datos de entrenamiento	177
Tabla 52. Resultados de las métricas empleadas para evaluar el modelo de supervivencia de Cox en el conjunto de datos de validación.....	179
Tabla 53. Resultados de los modelos de supervivencia Gradient Boosting, Support Vector Machines for Survival Analysis y Neural Networks for Survival Analysis para el conjunto de datos de entrenamiento y comparación con el modelo Cox.	182
Tabla 54. Resultados de los modelos de supervivencia Gradient Boosting, Support Vector Machines for Survival Analysis y Neural Networks for Survival Analysis para el conjunto de datos de validación y comparación con el modelo Cox.....	183

Figuras

Figura 1. Variabilidad en la ubicación de los tumores neuroblásticos en imágenes de Resonancia Magnética.	34
Figura 2. Pruebas de imagen para la evaluación de la extensión local y a distancia....	36
Figura 3. Esquema del sistema de estadificación actual y clasificación en grupos de riesgo del Grupo Internacional de Riesgo del Neuroblastoma.	43
Figura 4. Descripción del flujo de trabajo para la elaboración de modelos a partir de imágenes fuente.....	46
Figura 5. Gráfico lineal para la estimación del tamaño muestral requerido para realizar un modelo de supervivencia para neuroblastoma.....	59
Figura 6. Evolución cronológica del tamaño muestral y la distribución de pacientes para cada subanálisis.	62
Figura 7. Esquema del flujo de datos en la plataforma.	65
Figura 8. Ejemplo de la vista de la plataforma PRIMAGE.....	66
Figura 9. Ejemplo de la vista de la plataforma PRIMAGE.....	67
Figura 10. Flujo de preparación de la imagen.	73
Figura 11. Imagen original y tras la aplicación de diversos filtros de denoising	75
Figura 12. Efecto de corrección del sesgo por campo magnético	76
Figura 13. Distribución de los valores de las intensidades de cada vóxel en las imágenes originales y tras la aplicación de Z-score.	77
Figura 14. Ejemplo de segmentación manual de un neuroblastoma pélvico.	78
Figura 15. Entorno de segmentación ITK-SNAP®.	79
Figura 16. Criterios restrictivos de segmentación manual.....	81
Figura 17. Representación visual del Coeficiente de Similitud de Dice.	82
Figura 18. Representación visual del índice de Jaccard	83
Figura 19. Representación visual de la distancia de Hausdorff.....	83

Figura 20. Representación visual de falsos positivos, verdaderos positivos y falsos negativos	85
Figura 21. Pruebas de igualdad de medias entre dos grupos.....	89
Figura 22. Pruebas de igualdad de medias entre más de dos grupos.	90
Figura 23. Flujo de trabajo: análisis de la variabilidad interobservador.....	97
Figura 24. Flujo de comparación de variabilidad interobservador en la segmentación manual de tumores neuroblásticos.....	98
Figura 25. Comparación de dos casos segmentados por Radióloga 1 y Radióloga 2 y superposición y comparación de máscaras	100
Figura 26. Flujo de trabajo: entrenamiento y validación interna de una herramienta de segmentación automática.	101
Figura 27. Configuración del método automatizado nnU-Net para la segmentación de imágenes biomédicas basada en aprendizaje profundo	102
Figura 28. Entrenamiento de la red neuronal convolucional en 3 de los subconjuntos empleados para el entrenamiento.....	104
Figura 29. Diagrama del entrenamiento con validación cruzada y ajuste del modelo de segmentación basado en nnU-Net	109
Figura 30. Ejemplos de segmentaciones realizadas automáticamente por nnU-Net y comparación con segmentaciones realizadas manualmente por la Radióloga 1	111
Figura 31. Distribución de los 106 pacientes del grupo de entrenamiento en función del valor de Coeficiente de Similitud de Dice.....	112
Figura 32. Ejemplo de segmentación con Coeficiente de Similitud de Dice < 0.01	113
Figura 33. Diagramas de cajas y bigotes representando el Coeficiente de Similitud de Dice de cada uno de los subconjuntos de entrenamiento y del conjunto de validación.	116
Figura 34. Flujo de trabajo: validación externa de la herramienta de detección y segmentación automática.	119

Figura 35. Diseño del estudio para la validación independiente de la herramienta de detección y segmentación automática de tumores neuroblásticos.....	121
Figura 36. Esquema del proceso llevado a cabo para la validación independiente de la herramienta de segmentación.....	124
Figura 37. Ejemplos de máscaras de segmentación automática antes y después de la edición manual en cuatro casos con variabilidad en la ubicación y parámetros de adquisición.....	126
Figura 38. Ejemplos del rendimiento de la herramienta de segmentación automática y la edición manual en dos casos con variabilidad en el momento de adquisición, la ubicación y parámetros de adquisición.	130
Figura 39. Flujo de trabajo: extracción de radiómica y análisis de su estabilidad y reproducibilidad.	132
Figura 40. Flujo de pacientes seleccionados para la extracción de parámetros de radiómica.	132
Figura 41: Esquema del flujo para evaluar la estabilidad y reproducibilidad de la radiómica. Modificación de diferentes parámetros de preparación de la imagen y edición de las máscaras y comparación con la serie original.	137
Figura 42. Representación gráfica del impacto de los pasos de procesamiento	140
Figura 43. Gráficos de cajas y bigotes y <i>raincloud</i> para el Coeficiente de Correlación de Concordancia en las series 2 a 7.....	153
Figura 44. Medias marginales con intervalos de confianza para los valores de Coeficiente de Correlación de Concordancia obtenidos en las series 2 a 7	154
Figura 45. Gráficos de cajas y bigotes y <i>raincloud</i> para el Coeficiente de Correlación de Concordancia de las series 8 y 9.....	155
Figura 46. Flujo de trabajo: relación de radiómica con variables analíticas y de estadificación.....	158
Figura 47. Flujo de trabajo: extracción de radiómica profunda y relación con variables analíticas y de estadificación.	166

Figura 48. Esquema de funcionamiento de una red neuronal convolucional.....	168
Figura 49. Flujo de trabajo: construcción de un modelo de supervivencia global.....	174
Figura 50. Área bajo la curva en función del tiempo de los grupos de entrenamiento de validación cruzada para el modelo de supervivencia global.....	177
Figura 51. Curvas Kaplan-Meier del modelo en función del tiempo para los 392 pacientes del grupo de entrenamiento, divididos en 5 subconjuntos de datos.....	178
Figura 52. Curva Kaplan-Meier del modelo en función del tiempo para los 116 pacientes del grupo de validación	180
Figura 53. Diagrama de barras representando la distribución de la variable tiempo de supervivencia de los pacientes desde el momento de diagnóstico.....	180
Figura 54. Representación mediante <i>SHAP values</i> del impacto de cada una de las variables más representativas del modelo Cox basado en radiómica y radiómica profunda para el conjunto de datos de validación.	185

1. INTRODUCCIÓN

INTRODUCCIÓN

1.1. El problema clínico que queremos abordar: heterogeneidad de los tumores neuroblásticos

Los tumores neuroblásticos constituyen las neoplasias sólidas extracraneales más frecuentes en edad pediátrica [1]. Su prevalencia es de aproximadamente 1 caso por cada 7.000 nacidos vivos [2].

Los tumores neuroblásticos son neoplasias derivadas de células de la cresta neural, y comprenden el ganglioneuroma, el ganglioneuroblastoma y el neuroblastoma. El ganglioneuroma está compuesto por gangliocitos y estroma maduro, siendo el más benigno de los tres tipos de tumor. El ganglioneuroblastoma está formado por gangliocitos maduros y neuroblastos inmaduros y tiene un potencial maligno intermedio. Por su parte el neuroblastoma, el tipo más frecuente de los tumores neuroblásticos, es más inmaduro e indiferenciado, y constituye aproximadamente el 7% de todos los tumores pediátricos [3].

Son tumores que afectan principalmente a niños pequeños. La mediana de edad en el momento del diagnóstico de los tumores neuroblásticos es de aproximadamente 16 meses, con una incidencia de 35 por millón en niños menores de 1 año, constituyendo el tipo histológico de cáncer pediátrico más frecuente durante el primer año de vida. La incidencia disminuye con la edad a alrededor de 1 por millón entre los 10 y 14 años. El 95% de los casos se diagnostican antes de los 7 años de edad, y únicamente un 5% de pacientes son diagnosticados durante la adolescencia o edad adulta [4, 5].

Estos tumores se presentan en una amplia variedad de ubicaciones debido a su origen a partir de las células del sistema nervioso simpático suprarrenal, paravertebral y paraaórtico (Figura 1). Las localizaciones más comunes son la región suprarrenal (48%), el retroperitoneo extraadrenal (25%) y el tórax (16%), seguidos por el cuello (3%) y la pelvis (3%) [6]. Además, muestran una alta variabilidad en su tamaño, forma y límites, lo que resulta en una tarea desafiante para diferenciarlos de las estructuras vecinas.



Figura 1. Variabilidad en la ubicación de los tumores neuroblásticos en imágenes de Resonancia Magnética. A. Tumor suprarrenal derecho. B. Tumor paravertebral torácico izquierdo. C. Tumor paravertebral derecho con extensión intracanal. D. Tumor torácico izquierdo con extensión foraminal dorsal.

El neuroblastoma puede presentar una clínica muy diversa, siendo bien asintomático o causando una amplia variedad de síntomas, como fiebre, dolor, irritabilidad, pérdida de peso, hipertensión, estreñimiento, o, en caso de presentar una localización torácica, síndrome de Horner y síntomas respiratorios [4]. La extensión epidural o intradural ocurre en el 5% al 15% de los pacientes y puede resultar en compresión de la médula espinal [7].

El neuroblastoma presenta un pronóstico heterogéneo, con un espectro que varía desde la regresión espontánea en algunos casos hasta un curso clínico agresivo, fatal a pesar de un tratamiento multimodal, constituyendo el 10% de la mortalidad oncológica infantil [2]. El 35% de pacientes presentan diseminación ganglionar regional en el momento del diagnóstico y el 50% de pacientes debutan con enfermedad metastásica, siendo las localizaciones más comunes de metástasis el hueso, la médula ósea y el hígado [8].

El diagnóstico de la enfermedad se realiza en base a la clínica, exploración física y exploraciones complementarias como las pruebas de imagen, análisis de laboratorio, evaluación de la médula ósea y la biopsia del tumor. La imagen es fundamental para el diagnóstico. En la actualidad, se emplea el Sistema de Estadificación del Grupo Internacional de Riesgo de Neuroblastoma (INRGSS, que se expondrá en el siguiente apartado 1.2.), que permite realizar una estadificación preterapéutica en diferentes grupos mediante las imágenes de Resonancia Magnética (RM) o Tomografía Computarizada (TC) [8]. Esta estadificación local permite valorar la resecabilidad tumoral y decidir el tipo de procedimiento a realizar (biopsia o resección completa) en función de la evaluación de diferentes Factores de Riesgo Definidos por Imagen (IDRF) [6, 9], lo cual presenta impacto en la supervivencia de estos pacientes [10, 11]. Por tanto, la TC o RM son fundamentales para la estadificación tumoral local. La RM es la técnica de elección, ya que no irradia, presenta mayor resolución entre tejidos, permite valorar la invasión medular y la afectación de la pared costal con mayor contraste tisular que la TC, además de permitir obtener información sobre la celularidad con la secuencia de difusión (DW) o sobre la captación dinámica de contraste con la secuencia de perfusión (DCE). Los estudios sobre el papel de la RM de cuerpo entero en el neuroblastoma en el momento del depósito de esta Tesis Doctoral son escasos, basados en un número limitado de pacientes y, en algunos casos, con resultados contradictorios [12].

La extensión local y la presencia de metástasis se evalúan mediante aspirados y biopsias de médula ósea y estudios con metaiodobenzilguanidina [¹²³I]MIBG [13], un análogo de la norepinefrina que se concentra selectivamente en el tejido nervioso simpático y que no presenta captación fisiológica en el hueso o en la

médula ósea. Este hecho permite detectar tumores neuroblásticos primarios y metástasis con una sensibilidad y especificidad del 90% y del 100%, respectivamente [14, 15]. En la actualidad se realiza la gammagrafía complementada con un estudio 3D de Tomografía por Emisión de Fotón Simple (SPECT) tras la administración de [^{123}I]MIBG, que permite combinar la imagen funcional y morfológica [16] (Figura 2). El Grupo Europeo de Neuroblastoma de la Sociedad Internacional de Oncología Pediátrica (SIOPEN) ha elaborado un sistema de puntuación esquelética semi-cuantitativa de [^{123}I]MIBG que constituye un indicador pronóstico fiable en el estadiaje del neuroblastoma [17, 18]. La Tomografía por Emisión de Positrones con fluorodesoxiglucosa [^{18}F]FDG está indicada en los casos [^{123}I]MIBG negativos [12, 19].

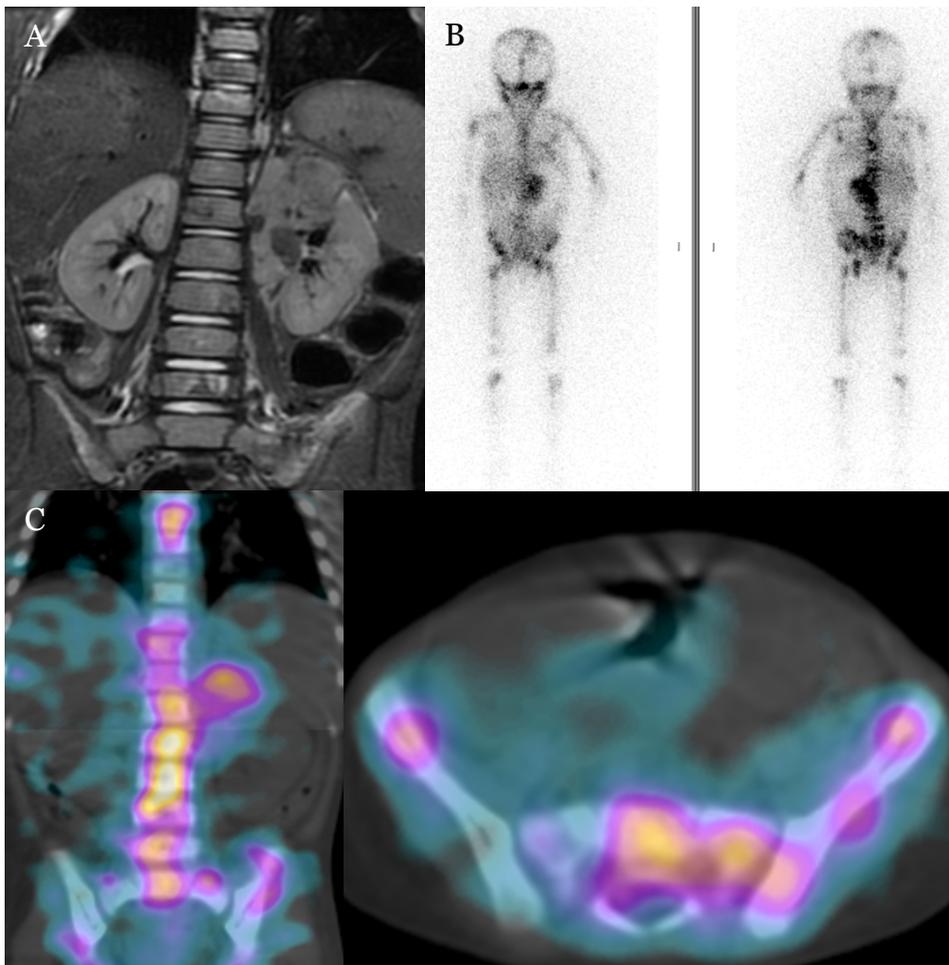


Figura 2. Pruebas de imagen para la evaluación de la extensión local y a distancia. (A) Resonancia Magnética en plano coronal que permite la evaluación de la extensión local de un tumor suprarrenal izquierdo. (B y C) Evaluación de la extensión a distancia

mediante gammagrafía planar de cuerpo completo con [¹²³I]MIBG y SPECT en plano coronal y transversal.

La biopsia del tumor permite llevar a cabo estudios biológicos y genéticos. El patrón genómico del neuroblastoma es complejo, estando fuera del marco teórico de esta Tesis Doctoral. Sin embargo, cabe destacar algunas alteraciones genéticas, como el gen MYCN, un oncogén (localizado en el cromosoma 2p) cuya amplificación es un factor de riesgo en el neuroblastoma. Se encuentra en alrededor del 20% de los casos de neuroblastoma y frecuentemente está asociado con histología desfavorable, y otras alteraciones genéticas (como diploidía, delección de 1p y ganancia de 17q). Se ha relacionado con un estadio de enfermedad más avanzado al diagnóstico y una progresión acelerada de la enfermedad [20, 21]. Por otro lado, la pérdida de 11q se considera la aberración más significativa en los tumores con MYCN no amplificado [22].

Está muy reconocido que los tumores neuroblásticos presentan una gran heterogeneidad en términos de edad, ubicación, características biológicas, genéticas, clínicas y pronósticas, lo cual dificulta la identificación de patrones comunes de agresividad tumoral, surgiendo así la necesidad de establecer patrones de referencia y sistemas de estandarización para aproximar el riesgo de los tumores y permitir la toma de decisiones terapéuticas.

1.2. Patrones de referencia para la agresividad tumoral: estratificación del riesgo

La estratificación del neuroblastoma según factores de riesgo es fundamental y se ha llevado a cabo durante décadas con la finalidad de establecer esquemas de referencia que permitan estandarizar las decisiones terapéuticas y el manejo de estos pacientes [20, 23]. En los últimos años se han propuesto diferentes sistemas de estadificación para tal fin. El Comité del Sistema de Estadificación Internacional del Neuroblastoma (INSS) desarrolló un método de clasificación quirúrgico/patológico en 1988, que fue el primer sistema de estadificación internacionalmente aceptado para el neuroblastoma [24]. En 2009, el Grupo de

Trabajo del Grupo Internacional de Riesgo del Neuroblastoma (INRG) propuso un nuevo sistema de estadificación (INRGSS) basado en imágenes [8]. A continuación, se exponen las características fundamentales de estos sistemas de estadificación.

1.2.1. INSS

El Sistema de Estadificación Internacional del Neuroblastoma se basa en la evaluación clínica, radiológica y postquirúrgica de niños con neuroblastoma, estableciendo estadios 1, 2A, 2B, 3, 4 y 4S en función de la extensión anatómica del tumor y su diseminación en el momento del diagnóstico [24], como se detalla en la Tabla 1.

Tabla 1. Sistema de Estadificación Internacional del Neuroblastoma (INSS). Adaptado de Brodeur et al.[24]

Estadio 1	Tumor localizado con escisión macroscópica completa; ganglios linfáticos ipsilaterales microscópicamente negativos
Estadio 2A	Tumor con escisión macroscópica incompleta; ganglios linfáticos ipsilaterales negativos
Estadio 2B	Tumor localizado con escisión macroscópica completa o incompleta; ganglios linfáticos ipsilaterales positivos. Ganglios linfáticos contralaterales agrandados microscópicamente negativos
Estadio 3	Tumor infiltrante más allá de la línea media con afectación de ganglios linfáticos regionales o sin ésta; o tumor unilateral localizado con afectación de ganglios linfáticos regionales contralaterales; o tumor en la línea media con afectación bilateral de ganglios linfáticos
Estadio 4	Tumor primario con diseminación a ganglios linfáticos distantes, hueso, médula ósea, hígado, piel u otros órganos con excepción de lo definido para el estadio 4S
Estadio 4S	Tumor primario localizado con diseminación a la piel, el hígado o la médula ósea en niños menores de 1 año

1.2.2. INRGSS

En 2009 se elaboró un nuevo sistema de estadificación del Grupo Internacional de Riesgo del Neuroblastoma (INRGSS) para estratificar a los pacientes en el momento del diagnóstico, previamente a cualquier tratamiento.

A diferencia de la clasificación INSS, no se incluyeron los resultados quirúrgicos ni la extensión a ganglios linfáticos [8, 25], sino que el INRGSS emplea únicamente los resultados de las pruebas de imagen realizadas previamente a la cirugía. La extensión de la enfermedad se valora mediante los factores de riesgo definidos por imágenes (IDRF), que se obtienen a partir de imágenes de TC o RM. Estos parámetros permiten evaluar la extensión del tumor a diferentes regiones anatómicas y su relación con diferentes estructuras, como los vasos, algunas raíces nerviosas o la tráquea [6] (Tabla 2).

Tabla 2. Factores de Riesgo Definidos por Imagen (IDRF). Adaptado de Brisse et al. [6]

Región anatómica	Descripción
Múltiples compartimentos corporales	Extensión tumoral ipsilateral dentro de dos compartimentos del cuerpo (cuello-tórax, tórax-abdomen, abdomen-pelvis)
Cuello	Tumor que engloba la arteria carótida, arteria vertebral y/o la vena yugular interna Tumor que se extiende hasta la base del cráneo Tumor que comprime la tráquea
Unión cervicotorácica	Tumor que engloba las raíces del plexo braquial Tumor que engloba los vasos subclavios y/o la arteria vertebral y/o carótida Tumor que comprime la tráquea
Tórax	Tumor que engloba la aorta y/o sus ramas principales Tumor que comprime la tráquea y/o los bronquios principales Tumor mediastínico inferior que infiltra la unión costovertebral entre los niveles vertebrales T9 y T12

Unión toracoabdominal	Tumor que engloba la aorta y/o la vena cava
Abdomen/pelvis	<p>Tumor que infiltra la porta hepática y/o el ligamento hepatoduodenal</p> <p>Tumor que engloba ramas de la arteria mesentérica superior en la raíz mesentérica</p> <p>Tumor que engloba el origen del eje celíaco y/o de la arteria mesentérica superior</p> <p>Tumor que invade uno o ambos pedículos renales</p> <p>Tumor que engloba la aorta y/o la vena cava</p> <p>Tumor que engloba los vasos ilíacos</p> <p>Tumor pélvico que cruza la escotadura ciática</p>
Extensión del tumor intraespinal	En cualquier ubicación, siempre que más de un tercio del canal espinal en el plano axial esté invadido, los espacios leptomeníngeos perimedulares no sean visibles y/o la intensidad de señal de la médula espinal sea anormal
Infiltración de órganos/estructuras adyacentes	Pericardio, diafragma, riñón, hígado, bloque duodenopancreático y mesenterio
Las condiciones que deben registrarse pero que no se consideran IDRF son los tumores primarios multifocales, el derrame pleural con o sin células malignas y la ascitis con o sin células malignas	

La enfermedad locorregional se determina por la ausencia o presencia de IDRF, siendo estos estadios L1 y L2, respectivamente. Se emplea el estadio M para la enfermedad ampliamente diseminada, y el MS describe el neuroblastoma metastásico limitado a piel, hígado y médula ósea sin compromiso óseo cortical en niños de 0 a 18 meses con tumores primarios L1 o L2 (Tabla 3).

Tabla 3. Sistema de estadificación del Grupo Internacional de Riesgo del Neuroblastoma (INRGSS). Adaptado de Cohn et al.[8]

Estadio	Descripción
L1	Tumor localizado sin afectación de estructuras vitales definidas en la lista IDRF y confinado a un compartimento corporal
L2	Tumor locorregional con presencia de 1 o más IDRF
M	Enfermedad metastásica a distancia (excepto estadio MS)
MS	Enfermedad metastásica en niños menores de 18 meses, con metástasis en la piel, hígado y/o médula ósea

El Sistema de Estadificación del INRG ejemplifica una estrategia de estratificación que ignora el tamaño en el momento del diagnóstico. En lugar de basarse en el tamaño absoluto y la extensión, como otros sistemas de estadificación ampliamente empleados en imágenes radiológicas [26], se basa en relaciones anatómicas que principalmente definen la probabilidad (en el momento del diagnóstico inicial) de que el tumor primario pueda ser completamente extirpado. Por otro lado, no se ha definido un tamaño acordado para evaluar los ganglios linfáticos metastásicos según el grupo de edad y la región anatómica [27].

1.2.3. INRG

El neuroblastoma puede responder de manera diferente al tratamiento en distintos pacientes, lo que se conoce como el comportamiento clínico de la enfermedad. Algunos niños son curados solo con cirugía o cirugía con quimioterapia, mientras que otros presentan una enfermedad muy agresiva que es resistente al tratamiento. Para poder adaptar el tratamiento, se emplea un sistema de clasificación del INRG que combina diferentes factores clínicos, patológicos y genéticos que permiten predecir la agresividad del tumor, su comportamiento clínico y cómo responderá al tratamiento, siendo estos factores [8]:

- Estadio de la enfermedad según el INRGSS

- Edad en el momento del diagnóstico
- Categoría histológica
- Grado o diferenciación de las células del tumor
- Estado del gen MYCN (amplificación o no amplificación)
- Estado del cromosoma 11q (aberración o no aberración)
- Ploidía de las células tumorales (contenido de ADN de las células tumorales, diploide o hiperdiploide)

Tras la evaluación de estos factores, el neuroblastoma se clasifica en una de cuatro categorías por grupo de riesgo: muy bajo riesgo, bajo riesgo, riesgo intermedio o alto riesgo (Tabla 4).

Tabla 4. Sistema de clasificación de riesgo del Grupo Internacional de Riesgo del Neuroblastoma (INRG). Extraída de Cohn et al. [8].

INRG Stage	Age (months)	Histologic Category	Grade of Tumor Differentiation	MYCN	11q Aberration	Ploidy	Pretreatment Risk Group
L1/L2		GN maturing; GNB intermixed					A Very low
L1		Any, except GN maturing or GNB intermixed		NA			B Very low
				Amp			K High
L2	< 18	Any, except GN maturing or GNB intermixed		NA	No		D Low
					Yes		G Intermediate
	≥ 18	GNB nodular; neuroblastoma	Differentiating	NA	No		E Low
			Poorly differentiated or undifferentiated		Yes		H Intermediate
				Amp		N High	
M	< 18			NA		Hyperdiploid	F Low
	< 12			NA		Diploid	I Intermediate
	12 to < 18			NA		Diploid	J Intermediate
	< 18			Amp			O High
	≥ 18						P High
MS				NA	No		C Very low
					Yes		Q High
	< 18			Amp			R High

Por todo ello, en la actualidad se lleva a cabo la estadificación por imagen INRGSS (que tiene en cuenta diferentes IDRF) y esta, junto a diferentes marcadores pronósticos clínicos, patológicos y genéticos, permite la estratificación en grupos de riesgo INRG, como se resume en la Figura 3:

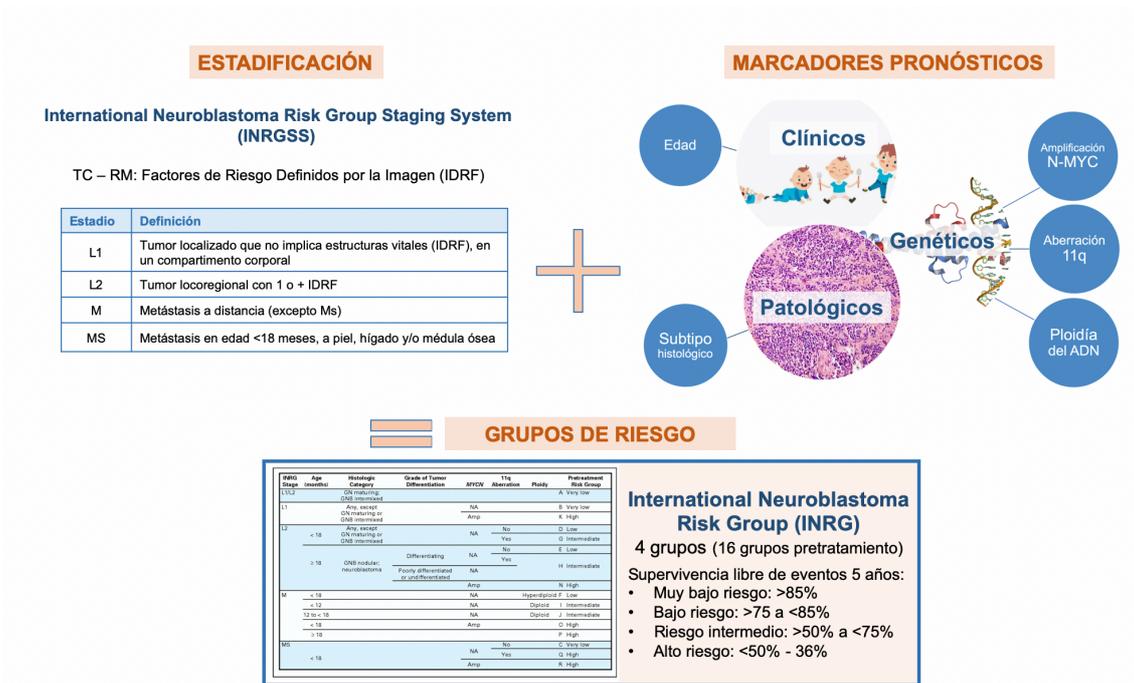


Figura 3. Esquema del sistema de estadificación actual y clasificación en grupos de riesgo del Grupo Internacional de Riesgo del Neuroblastoma.

Los sistemas actuales de estadificación presentan limitaciones debido a la variabilidad en el comportamiento de los tumores, que en ocasiones resulta impredecible. No existe un consenso internacional sobre las variables pronósticas a evaluar para asignar grupos de riesgo y estratificar el tratamiento, de modo que los factores seleccionados por varios grupos cooperativos expertos en neuroblastoma alrededor del mundo para definir el riesgo no son uniformes [8, 28]. Por ello, surge la necesidad de establecer nuevos biomarcadores adicionales para mejorar la precisión en la estratificación del riesgo tumoral.

1.3. Minimizando limitaciones: biomarcadores de imagen

Una forma de minimizar las limitaciones de los patrones de referencia actuales es el uso de biomarcadores de imagen. Cada vez más, la radiología se basa en datos objetivos y cuantificables extraídos de biomarcadores de imagen cuantitativos, indicadores generados a partir de características estructurales, funcionales, fisiológicas o biológicas de lesiones patológicas que proporcionan información adicional sobre la anatomía, la fisiología y la morfología del tumor. Con estos parámetros se pretende mejorar la precisión en la identificación fenotípica de los patrones de agresividad tumoral *in vivo* [29].

En el desarrollo del flujo de trabajo para la obtención de biomarcadores de imagen, diferentes funciones y modelos computacionales complejos extraen atributos cuantitativos de las imágenes, como características radiómicas, para correlacionarlos con el fenotipo o las firmas genéticas de las lesiones [30]. Estos análisis permiten detectar alteraciones para predecir el pronóstico o evaluar de forma no invasiva la respuesta al tratamiento [31 – 33].

La *Quantitative Imaging Biomarker Alliance* (QIBA) de la Sociedad de Radiología de Norteamérica (RSNA) y la *European Imaging Biomarker Alliance* (EIBALL) de la Sociedad Europea de Radiología (ESR) apoyan firmemente la adopción de este nuevo paradigma basado en biomarcadores de imagen, habiendo elaborado diversas directrices para desarrollar y validar los biomarcadores [34].

La ESR reconoce diferentes biomarcadores de imagen para diversas patologías. En el campo de la oncología pediátrica existen diversos marcadores de imagen reconocidos como biomarcadores potenciales. En el caso del neuroblastoma, el estudio de las secuencias de DW ha permitido establecer el Coeficiente de Difusión Aparente (ADC) como biomarcador potencial, ya que se ha descrito correlación entre la restricción en secuencias de DW y diferentes subtipos histológicos [35 – 38]

Asimismo, considera el Score SIOPEN para la evaluación de la captación de [¹²³I]MIBG como biomarcador de imagen con valor pronóstico del neuroblastoma [17, 20]. Finalmente, incluye la captación de [¹⁸F]-FDG en PET-TC (medida mediante su *Standardized Uptake Value*, SUV) como biomarcador correlacionado con la estadificación tumoral [19, 39].

Sin embargo, la aplicación de técnicas novedosas para la extracción de características cuantitativas a partir de las imágenes de RM (imágenes que se obtienen de manera rutinaria para el diagnóstico y estadificación de pacientes con tumores neuroblásticos), como características de radiómica, no ha sido explorada para la evaluación de los tumores neuroblásticos. La extracción de parámetros cuantitativos de las imágenes de RM podría constituir un biomarcador de imagen potencial, que permitiría la mejora en la capacidad de estratificación del riesgo y predicción de agresividad en pacientes con tumores

neuroblásticos, permitiendo una mejor tipificación de la enfermedad de cada paciente para realizar un manejo individualizado y preciso.

1.4. Análisis avanzado de imagen: preparación de la imagen, segmentación, radiómica y radiómica profunda

El procesamiento de imágenes desempeña un papel fundamental en diversas aplicaciones, ya que permite extraer características clave y realizar tareas importantes de clasificación, detección o segmentación. En este contexto, diferentes herramientas de inteligencia artificial, como el *machine learning* (aprendizaje automático) y el *deep learning* (aprendizaje profundo) permiten analizar grandes conjuntos de datos y son capaces de producir modelos con un rendimiento excepcional [40]. Las técnicas de aprendizaje automático permiten el desarrollo de algoritmos que son entrenados para desarrollar tareas mediante el aprendizaje de patrones a partir de datos, sin requerir programación explícita, y se han convertido en un instrumento fundamental de la inteligencia artificial. Por otro lado, el aprendizaje profundo, una clase de algoritmos de aprendizaje automático que emplea redes neuronales con múltiples capas, destaca por su capacidad para aprender representaciones complejas y abstracciones profundas de los datos. En esta Tesis Doctoral se emplean algoritmos de aprendizaje automático y aprendizaje profundo.

Para extraer biomarcadores de imagen de manera precisa y eficiente y garantizar la mejor usabilidad de las imágenes, es esencial desarrollar un proceso de procesamiento de imágenes robusto y reproducible. Este proceso incluye la preparación de la imagen, aplicando diferentes etapas que permitan homogeneizar las imágenes fuente. Posteriormente, se requiere llevar a cabo la segmentación de imágenes, y finalmente se realiza la extracción de radiómica y radiómica profunda, que será integrada con variables clínicas para poder extraer predicciones sobre la agresividad tumoral (Figura 4).

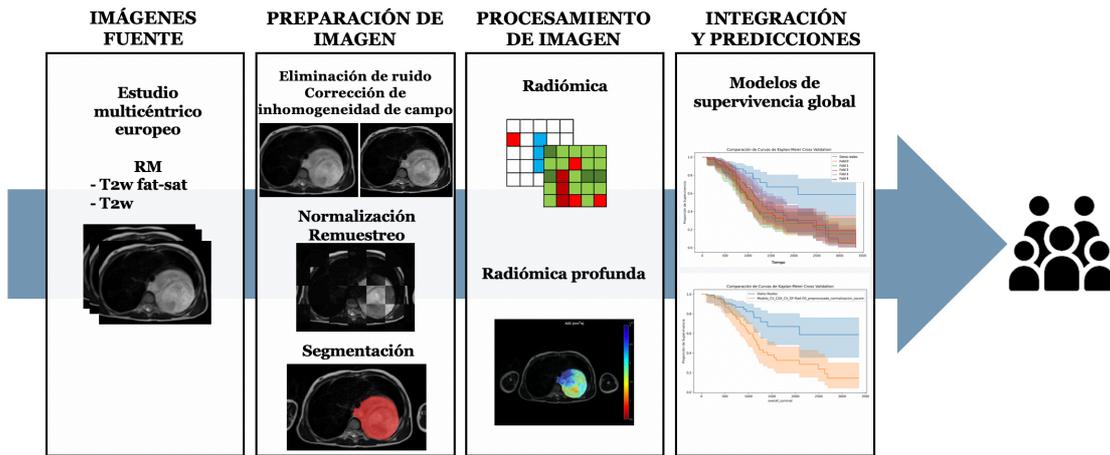


Figura 4. Descripción del flujo de trabajo para la elaboración de modelos a partir de imágenes fuente.

Para la extracción de modelos robustos y fiables, se requieren varios pasos preliminares de preparación de imágenes, como eliminación de ruido, normalización de señal y remuestreo del tamaño de los vóxeles, que garanticen la homogenización de las imágenes para que sean comparables [33, 41, 42].

Uno de los pasos fundamentales en el flujo de trabajo previo a la extracción de modelos es la segmentación, que consiste en colocar una Región de Interés (ROI) en un área específica (como un tumor), delimitando y etiquetando aquéllos vóxeles en la imagen que correspondan a la ROI. La segmentación del tumor se puede realizar de tres maneras: manual, semiautomática y automática. La segmentación manual generalmente la realiza un radiólogo experimentado, delimitando manualmente el tumor en cada una de las imágenes del estudio. Se trata de un método fiable pero que consume tiempo y dificulta el flujo de trabajo, especialmente en casos de procesamiento masivo de datos. Sin embargo, la segmentación manual depende del observador que la realiza, y puede mostrar una amplia variabilidad inter e intraobservador [43, 44]. Esta variabilidad está influenciada por algunos factores objetivos, como las características del órgano/tumor o su contorno, y por algunos factores subjetivos relacionados con el observador, como su experiencia o coordinación [43].

La segmentación semiautomática intenta resolver algunos de los problemas relacionados con la segmentación manual [45], facilitando la segmentación con algoritmos, por ejemplo, ampliando la segmentación sobre una región para disminuir la necesidad de una segmentación corte a corte. Permite reducir el esfuerzo y el tiempo requeridos por el usuario. Sin embargo, la variabilidad entre observadores sigue presente, ya que la parte manual de la segmentación y la configuración del algoritmo influyen en el resultado. En el caso de los tumores neuroblásticos, varios estudios han explorado el desarrollo de algoritmos de segmentación semiautomática. Se han realizado en imágenes de TC o RM, utilizando morfologías matemáticas, conectividad difusa y otras herramientas de procesamiento de imágenes [46 – 48]. Estas herramientas se desarrollaron incluyendo un número muy bajo de casos y los hallazgos mostraron poca mejoría con respecto a la segmentación manual.

Hoy en día, la mayoría de las herramientas avanzadas están diseñadas para utilizarse como métodos de segmentación automática que, por definición, no dependen de la interacción del usuario. Estas soluciones están construidas con algoritmos de segmentación de aprendizaje profundo [49], por lo general, basadas en redes neuronales convolucionales (CNN). Las CNN utilizan varias operaciones de convolución y agrupación secuenciales en las que las imágenes son procesadas para extraer características y reconocer patrones, utilizando las propias imágenes para ser entrenadas durante el proceso de aprendizaje [50]. Una de las arquitecturas más comúnmente utilizadas es la red U-Net, que consiste en un algoritmo con una ruta de contracción y una ruta de expansión simétrica, lo que logra un rendimiento muy bueno en la segmentación de diferentes tipos de cáncer [51, 52]. No obstante, se ha observado que su aplicabilidad al análisis de imágenes específicas y su reproducibilidad en diferentes estructuras o lesiones es limitada [53]. Recientemente se ha propuesto una nueva solución basada en algoritmos CNN llamada nnU-Net. Consiste en un marco de segmentación basado en aprendizaje profundo que se configura automáticamente, incluyendo preprocesamiento, arquitectura de red, entrenamiento y procesamiento, y se adapta a cualquier conjunto de datos nuevo, superando la mayoría de los enfoques existentes [53, 54]. En nuestro conocimiento, no se ha desarrollado hasta la fecha una solución sólida y

generalizable para la segmentación de neuroblastoma, por lo que en esta Tesis Doctoral se pretende elaborar una herramienta de segmentación automática con validación manual que permita facilitar el flujo de segmentación.

Por su parte, la validación externa es un paso fundamental antes de que cualquier solución de inteligencia artificial (IA) sea aplicable en la práctica clínica, ya que verifica la reproducibilidad y generalización de las herramientas desarrolladas, aplicándolas a diferentes pacientes, instituciones y escáneres. Consiste en aplicar el modelo desarrollado en un conjunto de pacientes nuevos y diferentes a los empleados para el entrenamiento para determinar si la herramienta funciona en un grado aceptable, ya que la solución podría haber sido sobreajustada y haber aprendido los detalles específicos de los datos de entrenamiento sin ser capaz de generalizar bien a nuevos datos [55]. Por lo general, el desempeño en la validación externa es peor que en el proceso de entrenamiento y validación interna [56]. A pesar del desarrollo de varias herramientas de segmentación en los últimos años, solo unas pocas han sido validadas externamente [57 – 59], mientras que la mayoría están diseñadas como estudios de viabilidad metodológica para demostrar un concepto [54]. Por ello, la validación externa de la herramienta de segmentación automática desarrollada en esta Tesis Doctoral resulta fundamental para garantizar su generalización y reproducibilidad.

La radiómica implica la extracción matemática de características cuantitativas de regiones de interés en imágenes, proporcionando información sobre relaciones de forma, textura y vóxel [41, 60]. Varios estudios han demostrado que la radiómica puede ayudar en la caracterización de tumores [61, 62]. Sin embargo, la estabilidad y reproducibilidad de las características son limitaciones clave en la generalización de este tipo de estudios. La estabilidad del análisis radiómico es un desafío importante debido a las variaciones inherentes en los parámetros de adquisición y reconstrucción de imágenes, así como a las modificaciones en el flujo de trabajo de radiómica.

Se han realizado varios intentos para estandarizar y homogeneizar el proceso de extracción radiómica, y mejorar así la calidad del procesamiento y la presentación de radiómica, pero aún no se ha alcanzado un compromiso en el proceso de

estandarización del análisis radiómico [63, 64]. Se han llevado a cabo numerosos estudios para evaluar la repetibilidad (es decir, el enfoque de escaneo y repetición) [65 – 70], y los cambios inducidos por las diferentes máscaras de segmentación realizadas por diferentes lectores [71 – 73]). La robustez de la extracción de las características radiómicas es crucial para su exitosa traducción al entorno clínico.

Las características de radiómica profunda son descriptores numéricos que se obtienen de una CNN y que se utilizan ampliamente para clasificación y reconocimiento, y contienen información abstracta y compleja sobre la textura y forma de las imágenes [74]. No existe un método de denominación específico para estas características, aunque se indican mediante el número de columna de características (posición de una neurona en una capa oculta de una CNN) [75].

1.5. Finalidad del estudio: estimación de la agresividad

La diversidad de las características cuantitativas extraídas de las imágenes médicas permite una comprensión más profunda de la heterogeneidad tumoral, proporcionando información detallada sobre la arquitectura, textura y dinámica del tumor. Estos datos se pueden emplear para desarrollar modelos de supervivencia mediante diversas técnicas estadísticas y de aprendizaje automático, que integran variables clínicas y radiómicas para predecir la probabilidad de supervivencia de los pacientes, y permiten mejorar la estratificación del riesgo y el manejo de los pacientes [76].

En este estudio se aplicarán diversas técnicas avanzadas de procesamiento de imagen sobre imágenes de RM, técnica fundamental no invasiva realizada en la práctica clínica habitual y que constituye una herramienta para la tipificación tumoral y la estadificación pronóstica de los tumores neuroblásticos.

Se pretende establecer un flujo robusto y reproducible de homogeneización de las imágenes, y la obtención de una herramienta de segmentación automática de tumores neuroblásticos que permita mejorar el flujo actual de detección y delimitación tumoral, para extraer factores radiómicos y de radiómica profunda

de los tumores neuroblásticos. Finalmente, se llevará a cabo la estimación de la agresividad tumoral mediante la elaboración de modelos predictivos de la supervivencia global, a partir de la extracción de biomarcadores de imagen precisos y eficientes que representen los diversos procesos patológicos que acontecen en el tejido de los tumores neuroblásticos y que permitan mejorar la capacidad de diagnóstico y estratificación del riesgo en pacientes con neuroblastoma.

2. HIPÓTESIS Y OBJETIVOS

HIPÓTESIS

El análisis de las imágenes de RM en tumores neuroblásticos permitirá determinar la agresividad tumoral y las características de la imagen que mayor impacto tengan sobre esta. La agresividad tumoral se considera como la capacidad del tumor de extenderse en vecindad y a distancia, y de acortar la expectativa de vida, considerando la supervivencia global del paciente como un buen marcador de la agresividad del tumor. El análisis de las imágenes de RM se obtiene mediante distintas técnicas avanzadas de preparación y procesamiento de imagen (radiómica, radiómica profunda) y aporta información sobre las características del tumor, permitiendo extraer correlaciones con parámetros clínicos y desarrollar modelos que permitan predecir la agresividad tumoral.

Esta Tesis Doctoral pretende establecer una metodología novedosa para la detección y segmentación automática, la extracción de características radiómicas y la predicción de la agresividad de los tumores neuroblásticos a partir del análisis de imágenes de RM en pacientes en edad pediátrica.

OBJETIVOS

El objetivo principal de esta Tesis Doctoral es extraer las características y descriptores de imagen cuantitativos mediante algoritmos avanzados de procesamiento de las imágenes de RM, que permitan evaluar y predecir la agresividad biológica y la heterogeneidad de los tumores de estirpe neuroblástica en pacientes de edad pediátrica.

Los objetivos secundarios son:

- Evaluar los métodos de preparación y procesamiento de las imágenes de RM para aumentar la reproducibilidad y robustez de su análisis posterior.
- Determinar una metodología para la segmentación manual de los tumores neuroblásticos mediante RM y elaborar una herramienta de segmentación automática basada en redes neuronales convolucionales. Validar de forma independiente esta herramienta de segmentación.
- Extraer parámetros cuantitativos de radiómica y evaluar su estabilidad y reproducibilidad.
- Extraer parámetros cuantitativos de radiómica profunda y realizar una reducción de las variables obtenidas.
- Correlacionar las variables de radiómica y radiómica profunda con diversas variables analíticas y de estadificación.
- Relacionar los parámetros cuantitativos extraídos de la imagen (radiómica y radiómica profunda) con la agresividad tumoral, definida por la supervivencia global como evento clínico final, y extraer modelos predictores de la supervivencia global. Realizar una validación externa independiente de las estimaciones de la agresividad tumoral.
- Definir el impacto de los parámetros cuantitativos de imagen para la predicción de agresividad tumoral.

3. METODOLOGÍA

METODOLOGÍA

3.1. Diseño del proyecto de investigación

Para desarrollar los objetivos de esta Tesis Doctoral se ha diseñado un estudio analítico observacional de cohorte retrospectivo multicéntrico, con la finalidad de crear modelos predictivos. El estudio se realizó entre mayo de 2019 y junio de 2023. La población diana de esta Tesis Doctoral fueron pacientes pediátricos con diagnóstico final de tumor neuroblástico. Los pacientes incluidos en el estudio se recopilaron a partir de la base de datos del proyecto europeo PRIMAGE (PRedictive In-silico Multiscale Analytics to support cancer personalized diaGnosis and prognosis, Empowered by imaging biomarkers), comprendiendo pacientes incluidos en dos ensayos clínicos europeos, llevados a cabo por la Society of Paediatric Oncology European Neuroblastoma Network (SIOOPEN). Estos ensayos fueron el Protocolo Europeo de Neuroblastoma de Riesgo Bajo e Intermedio de SIOOPEN (LINES/SIOOPEN) (NCT01728155), liderado por el Hospital Universitari i Politènic La Fe (Valencia, España), y el Estudio de Neuroblastoma de Alto Riesgo de la Sociedad Europea de Oncología Pediátrica (HR-NBL1/SIOOPEN) (NCT00030719), liderado por St. Anna Children's Hospital and Children's Cancer Research Institute (CCRI) (Viena, Austria) (ambos ensayos serán considerados en esta Tesis Doctoral como centros de procedencia únicos dada la heterogeneidad de la procedencia de los casos en cada uno de los ensayos). Adicionalmente, se incluyeron pacientes que no formaban parte de los ensayos y estudios clínicos mencionados, reclutados en diferentes instituciones europeas colaboradoras con el proyecto PRIMAGE: pacientes del Registro Alemán de Neuroblastoma NB2016 de la Sociedad de Oncología y Hematología Pediátrica, dirigido por el Departamento de Oncología Pediátrica del Hospital Universitario de Niños de Colonia, pacientes del Hospital Universitario de Pisa, Instituto Gaslini (Pisa, Italia), así como pacientes del Hospital La Paz y Hospital Niño Jesús (Madrid, España), Hospital Son Espases (Palma de Mallorca, España), y Hospital Reina Sofía (Córdoba, España).

Las pruebas de imagen (RM y TC) se realizaron en la institución de procedencia de cada paciente. Las imágenes en el formato estándar DICOM (Imagen Digital y Comunicación en Medicina) se almacenaron en la Plataforma del proyecto

PRIMAGE, una adaptación específica de la plataforma *QUIBIM Precision®* (*SME Instrument Phase 2, project number 778064*) [77].

Para la recogida de datos se empleó el programa informático Microsoft Excel versión 16.16.27. El análisis estadístico se realizó con el programa IBM SPSS Statistics 26.0. Los cálculos del tamaño muestral para el modelado del tiempo hasta el evento se implementaron en R utilizando el paquete *pmsampsize*.

El estudio se adhirió a los principios éticos de la Declaración de Helsinki de 1975, como se recoge en la aprobación del Comité Ético de Investigación Clínica (CEIC) de cada institución participante. En el caso del Hospital Universitario y Politécnico La Fe el número de registro del Dictamen del Comité de Ética de la Investigación con Medicamentos del Hospital Universitario y Politécnico La Fe para el Proyecto de Tesis Doctoral “Estudio de la agresividad biológica de tumores de estirpe neuroblástica en edad pediátrica mediante el análisis de las imágenes de Resonancia Magnética” fue 2021-666-1, y el número de registro del Dictamen del Comité de Ética de la Investigación con Medicamentos del Hospital Universitario y Politécnico La Fe para el Proyecto de Investigación “PRIMAGE”: 2018/0228. Los dictámenes del resto de centros se encuentran en la memoria descriptiva del proyecto PRIMAGE (Deliverable 11.2 – H - Requirement No. 2) (Anexo 1).

En el momento de adquisición de las imágenes, todos los pacientes o tutores legales (en caso de minoría de edad) firmaron el documento de consentimiento informado para la realización de la prueba de imagen. En la resolución del CEIC del Hospital La Fe se obtuvo la exención de la obligación de obtener consentimiento informado para la participación en Proyecto de Tesis Doctoral al tratarse de un estudio retrospectivo con datos anonimizados previamente almacenados en registros hospitalarios, en base a las Pautas éticas internacionales para la investigación relacionada con la salud con seres humanos, elaboradas por el Consejo de Organizaciones Internacionales de Ciencias Médicas (CIOMS) en colaboración con el Organización Mundial de la Salud (OMS) en el que se afirma que: “Un comité de ética de la investigación podrá aprobar una modificación o renuncia al consentimiento informado para la investigación: si la investigación no sería factible o practicable de llevarse a cabo sin la renuncia o

modificación; la investigación tiene un valor social importante; la investigación no plantea más que riesgos mínimos para los participantes” [78].

La presentación de los datos de investigación se ha realizado siguiendo las directrices de la lista de verificación para inteligencia artificial en imágenes médicas (CLAIM) [79] (Anexo 2). Asimismo, se implementó la lista de verificación MAIC-10 (Must Artificial Intelligence Criteria-10) con 10 ítems como guía para diseñar estudios y evaluar publicaciones relacionadas con la IA en el campo de las imágenes médicas [80] (Anexo 2).

3.2. Población de pacientes

3.2.1. Estimación del tamaño muestral

La estimación del tamaño muestral es fundamental para realizar y validar con veracidad los modelos de predicción clínica de una forma robusta. En un estudio realizado por el Grupo de investigación Biomédica en Imagen (GIBI2³⁰) en el marco del proyecto PRIMAGE, se desarrolló una solución práctica que proporciona un conjunto de algoritmos para calcular el tamaño muestral necesario para construir y validar modelos predictivos paramétricos [81]. Esta solución permite calcular el tamaño de muestra necesario para desarrollar un modelo predictivo computacional *in silico* y fue empleada para el cálculo del tamaño muestral necesario para los modelos predictivos desarrollados en esta Tesis Doctoral.

La metodología se fundamenta en datos epidemiológicos conocidos, ajustando el tamaño de la muestra obtenida a la problemática específica de [82]. Los cálculos del tamaño muestral para el modelo de tiempo hasta el evento se implementaron en R utilizando el paquete *pmsampsize*. Los datos de los eventos clínicos necesarios para la realización de los cálculos se obtuvieron a partir de la literatura consultada: se empleó una tasa de supervivencia a 5 años del 30.7%, un tiempo mediano hasta evento de 24,2 meses, un tiempo de seguimiento de 60 meses, y una prevalencia de recaídas/progresión de 25,8% [83, 84].

Se calculó el tamaño muestral requerido para desarrollar modelos confiables para neuroblastoma, obteniendo la variabilidad en el tamaño de la muestra en relación con el número de variables o parámetros predictores incluidos en el desarrollo del modelo. Para un modelo con 5 variables predictoras el tamaño muestral requerido sería de 180 pacientes, mientras que para un modelo de 30 parámetros predictores se requerirían 1077 pacientes (Tabla 5) (Figura 5).

Tabla 5. Estimación del tamaño muestral requerido para realizar un modelo de supervivencia para neuroblastoma, realizada a partir de la solución práctica desarrollada en el Grupo de Investigación Biomédica en Imagen [81].

Parámetros del modelo	Tamaño muestral
5	180
10	359
15	539
20	718
25	898
30	1077

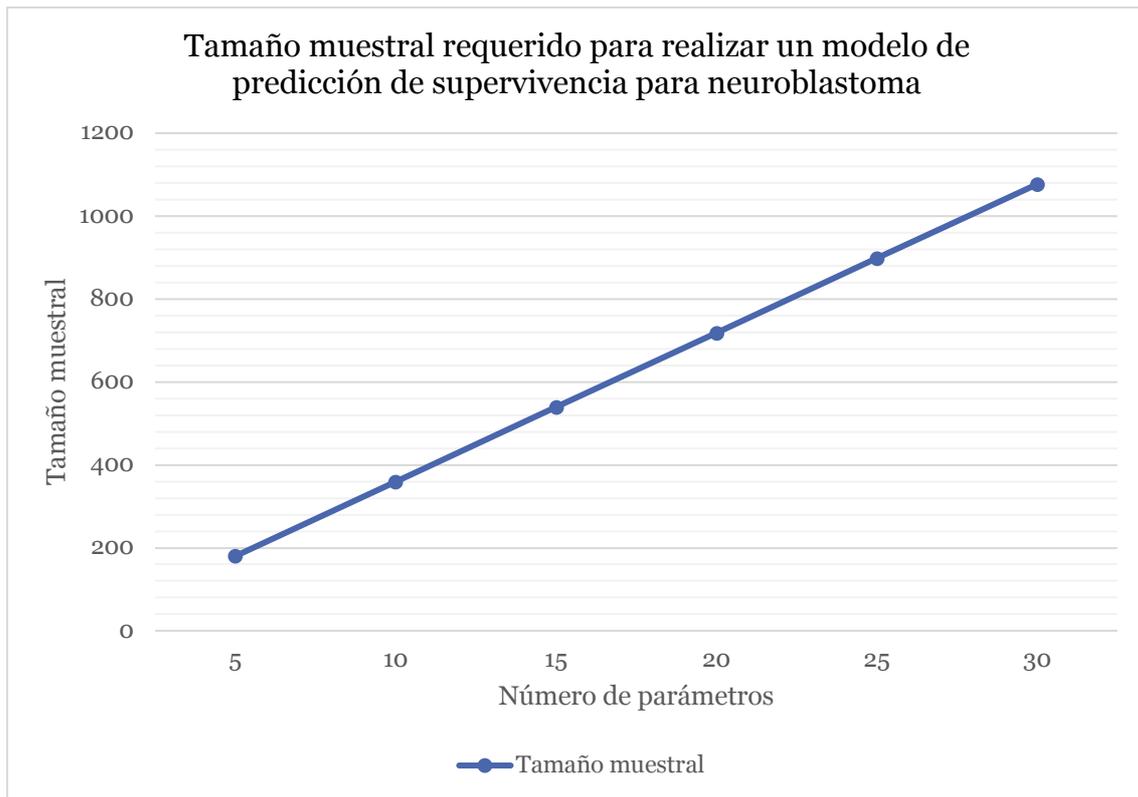


Figura 5. Gráfico lineal para la estimación del tamaño muestral requerido para realizar un modelo de supervivencia para neuroblastoma.

En el caso de esta Tesis Doctoral, como se verá en los posteriores capítulos, el tamaño muestral final fue de 508 pacientes (392 pacientes para entrenamiento de modelos de supervivencia, 116 para validación), y los modelos finales se elaboraron para la predicción de la supervivencia global, seleccionando 10 variables predictoras como representativas.

En base a la estimación del tamaño muestral realizada, para la elaboración de modelos predictivos con 10 parámetros se requerirían 359 pacientes. Por tanto, el tamaño muestral empleado en esta Tesis Doctoral para el entrenamiento de un modelo predictivo para neuroblastoma (392 pacientes), es superior al tamaño muestral necesario para establecer un modelo confiable (359 pacientes).

Por otro lado, la validación del modelo tiene un papel esencial para demostrar que un algoritmo de estimación es reproducible y puede aplicarse de manera consistente en la práctica clínica. Por consenso, se considera que el tamaño muestral mínimo para una validación robusta debería tener al menos 100 eventos

[85]. En este estudio se emplearon 116 pacientes para la validación del modelo de supervivencia.

3.2.2. Reclutamiento: proceso de selección y criterios de elegibilidad

Se revisó de manera retrospectiva una serie de pacientes con diagnóstico confirmado de tumor neuroblástico (neuroblastoma, ganglioneuroma, ganglioneuroblastoma), durante un periodo de 18 años, desde 2005 hasta 2023, que hubiesen sido incorporados en la plataforma del proyecto Horizon 2020 PRIMAGE.

Se seleccionaron aquellos pacientes con una RM que incluyese el tumor al diagnóstico, y/o tras la primera aplicación de quimioterapia neoadyuvante (durante el primer ciclo de tratamiento, anterior a la posible cirugía).

El motivo para la realización de la exploración de RM al diagnóstico fue la sospecha de tumor neuroblástico tras haber identificado una lesión sospechosa mediante ecografía, radiografía de tórax o TC, con la finalidad de tipificar y estudiar la lesión previamente a cualquier decisión terapéutica. En los casos en los que se realizó una RM tras el tratamiento inicial con quimioterapia, el motivo del estudio fue la obtención de un control de imagen del tumor.

Pese a que los pacientes se reclutaron de diferentes ensayos clínicos, estudios y práctica clínica habitual, en esta Tesis Doctoral se han seguido unos criterios de inclusión y exclusión para construir la cohorte de estudio.

- Criterios de inclusión:
 - pacientes menores de 18 años,
 - diagnosticados de neuroblastoma por anatomía patológica o biopsia líquida (en el caso de neonatos con masas suprarrenales, se requirió biopsia líquida positiva para neuroblastoma o estudio [¹²³I]MIBG positivo en los primeros 3 meses de edad),
 - con realización de RM al diagnóstico y/o tras primera aplicación de quimioterapia,

- cuyas imágenes estuviesen disponibles en el sistema PACS (Sistema de Archivo y Comunicación de Imágenes),
- con estadificación completa según INRGSS,
- con estudios biológicos y genéticos completos según SIOPEN.
- Criterios de exclusión:
 - imágenes de calidad inadecuada para el diagnóstico tras la evaluación visual de las imágenes de RM llevada a cabo por personal facultativo especialista en radiodiagnóstico.

3.2.3. Distribución de pacientes

El estudio constó de diferentes fases realizadas cronológicamente, y cada fase tuvo un número de casos diferentes debido a la sucesión de su realización.

En un primer estudio se incluyeron 46 pacientes con una secuencia de RM potenciada en T2w al diagnóstico para el cálculo de la variabilidad interobservador de la segmentación manual. Posteriormente, se amplió el tamaño muestral a 132 pacientes con RM al diagnóstico con secuencias potenciadas en T2w con la finalidad de entrenar y validar internamente la herramienta de segmentación automática de tumores neuroblásticos. A continuación, se evaluaron 300 pacientes con 535 secuencias potenciadas en T2w en las que se llevó a cabo la validación independiente de la herramienta de segmentación automática. Posteriormente, la muestra ascendió a 419 pacientes con un total de 746 secuencias T2w para la extracción de radiómica y el cálculo de estabilidad y reproducibilidad de los parámetros radiómicos extraídos. Finalmente, la base de datos aumentó a 508 pacientes con RM al diagnóstico, de los que se obtuvieron parámetros de radiómica, radiómica profunda y datos clínicos y biológicos para obtener modelos de supervivencia global. La Figura 6 describe la evolución cronológica por años de la distribución de pacientes para cada subanálisis, especificando el número de pacientes incluidos, el momento de realización de la RM, la finalidad de cada subanálisis y la procedencia de los casos para cada uno de los estudios.

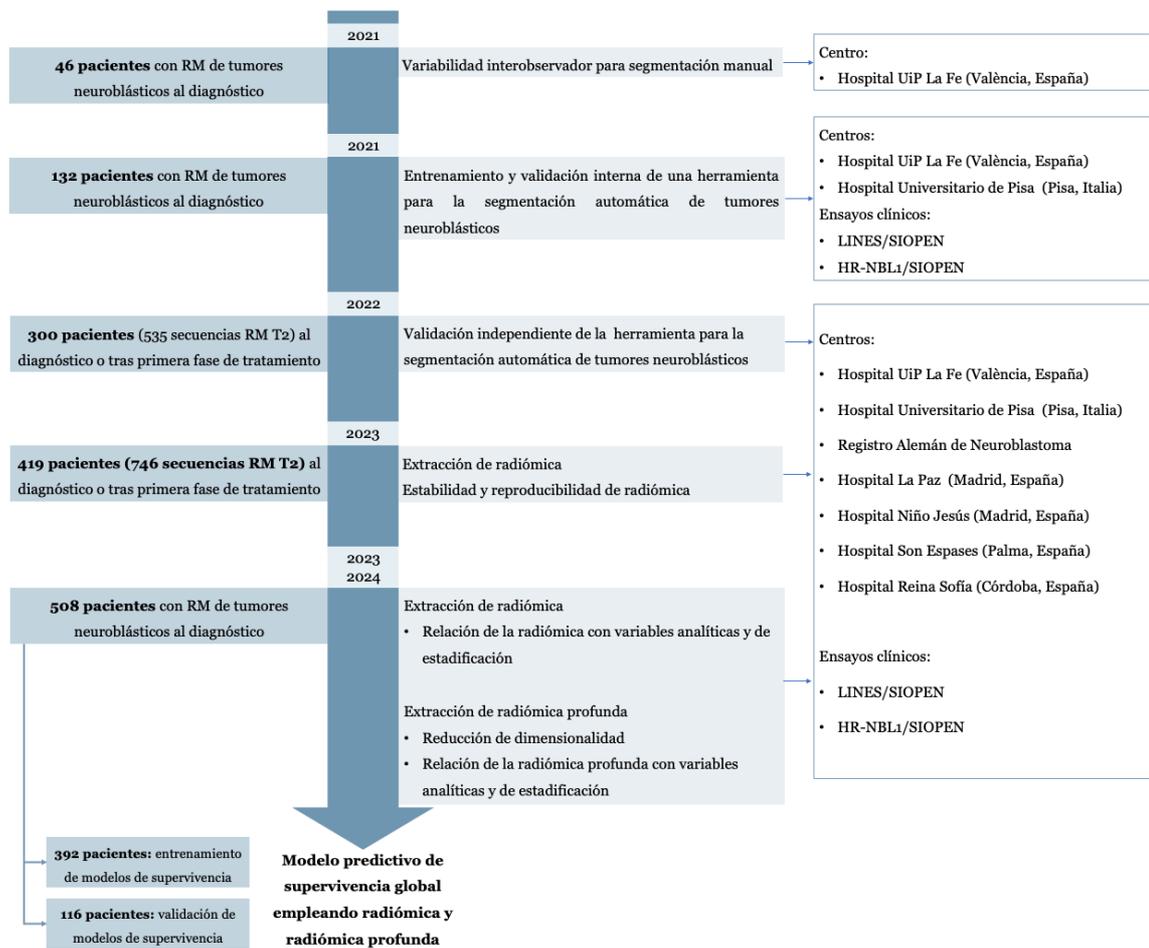


Figura 6. Evolución cronológica del tamaño muestral y la distribución de pacientes para cada subanálisis.

En el apartado de Resultados se detallarán las características de la muestra en cada uno de los subanálisis (apartados 4.1., 4.2., 4.3.2, 4.3.3., 4.4.1., 4.4.3.).

3.3. Base de datos

3.3.1. Privacidad de datos: seudonimización

La base de datos se elaboró siguiendo un modelo de estructura estandarizada (*Common Data Elements*) con la definición de cada variable del estudio. Se utilizó una plantilla de Excel de referencia elaborada por el GIBI²³⁰ para el manejo estructurado de los datos. La adquisición de éstos fue llevada a cabo

revisando los datos disponibles en la plataforma PRIMAGE, así como la información obtenida dentro del protocolo de estudio de RM.

Para garantizar el manejo de datos personales de acuerdo con las reglas de privacidad y protección de datos, todas las imágenes y los datos clínicos fueron seudonimizados. La seudonimización de los datos consiste en el “tratamiento de datos personales de manera tal que ya no puedan atribuirse a un interesado sin utilizar información adicional, siempre que dicha información adicional figure por separado y esté sujeta a medidas técnicas y organizativas destinadas a garantizar que los datos personales no se atribuyan a una persona física identificada o identificable”, como se recoge en el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos) [86]. Para llevar a cabo este proceso, se empleó el *European Patient Identity Management* (EUPID), una herramienta paneuropea de seudonimización desarrollada por el *Austrian Institute of Technology* [87].

3.3.2. Seguridad de los datos: ubicaciones de bases de datos

Los datos seudonimizados se almacenaron en la plataforma del proyecto PRIMAGE, una arquitectura para repositorio de imágenes y datos relacionados que proporciona recursos de almacenamiento y procesamiento (incluidas infraestructura y herramientas de software) y que permite cumplir los criterios de calidad recomendados en las pautas de desarrollo de biomarcadores de imagen de la *European Society of Radiology* [88].

3.4. Plataforma

La plataforma PRIMAGE implementa un modelo federado para sincronizar datos entre diferentes núcleos de almacenamiento vinculados a distintos entornos de procesamiento, incluidos recursos en la nube. La arquitectura permite

computación de alto rendimiento para proporcionar un marco simplificado para desarrollar aplicaciones [30].

Los datos se almacenaron en una interfaz web basada en QUIBIM Precision®, una solución web respaldada por el programa SME Instrument Phase II (pequeñas y medianas empresas) de la Comisión Europea [77].

La plataforma se alojó en Microsoft Azure como aplicación web. Esta se ejecuta en un plan de servicio P1v2 basado en Linux con 1 núcleo, 3.5 GB de RAM y 250 GB de almacenamiento, con la plataforma funcionando en un contenedor Docker.

La plataforma consta de diferentes capas. La primera es una capa de interfaz de usuario visible (*Front-end*) y está directamente expuesta al usuario final para interactuar con el software mediante la interfaz de usuario web (<https://primage.quibim.com/>). La segunda es la capa del servidor (*Back-end*), y está construida por diferentes servicios que procesan solicitudes desde la interfaz de usuario, aplicaciones externas y la interacción entre los servicios. Presenta cuatro componentes principales:

- Servidor de plataforma: componente que opera todas las solicitudes relacionadas con la interfaz de usuario.
- MIUC (Conector Universal de Imágenes Médicas): componente que realiza la conexión entre la plataforma y el PACS en la nube utilizando el protocolo DICOM.
- Convertidor de ROI: permite convertir archivos de segmentación entre formatos. Se utiliza en el visor web DICOM incluido en la plataforma y permite importar y exportar archivos de segmentación en formato NIfTI (Iniciativa de Tecnología Informática de Neuroimagen).
- Programador de trabajos: permite programar la ejecución de módulos de análisis.

La tercera y última capa es la de persistencia, que permite la permanencia de la información no volátil del software. Esta capa está compuesta por:

- PACS en la nube: las imágenes médicas que no se cargan manualmente a través de la interfaz web se almacenan aquí.

- MongoDB: base de datos donde la aplicación almacena la información semiestructurada.
- Almacenamiento en la nube: la aplicación almacena todos los archivos (estudios de imágenes, resultados, archivos de configuración...) en la nube.

Se proporciona un esquema de las capas de la plataforma en la Figura 7.

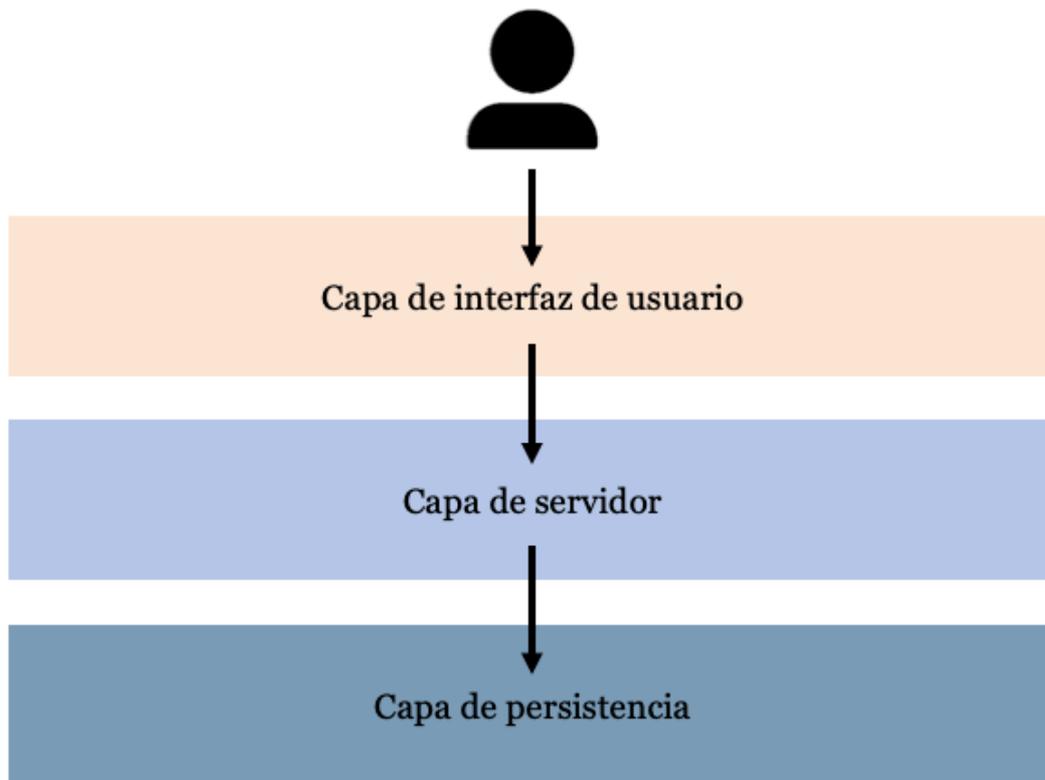


Figura 7. Esquema del flujo de datos en la plataforma.

La plataforma permite el acceso con credenciales (Figura 8A). En ella se llevó a cabo el almacenamiento de imágenes en DICOM y la incorporación de datos clínicos. Para el proceso de ingesta de datos, la plataforma proporciona un formulario electrónico específico (Figura 8B).

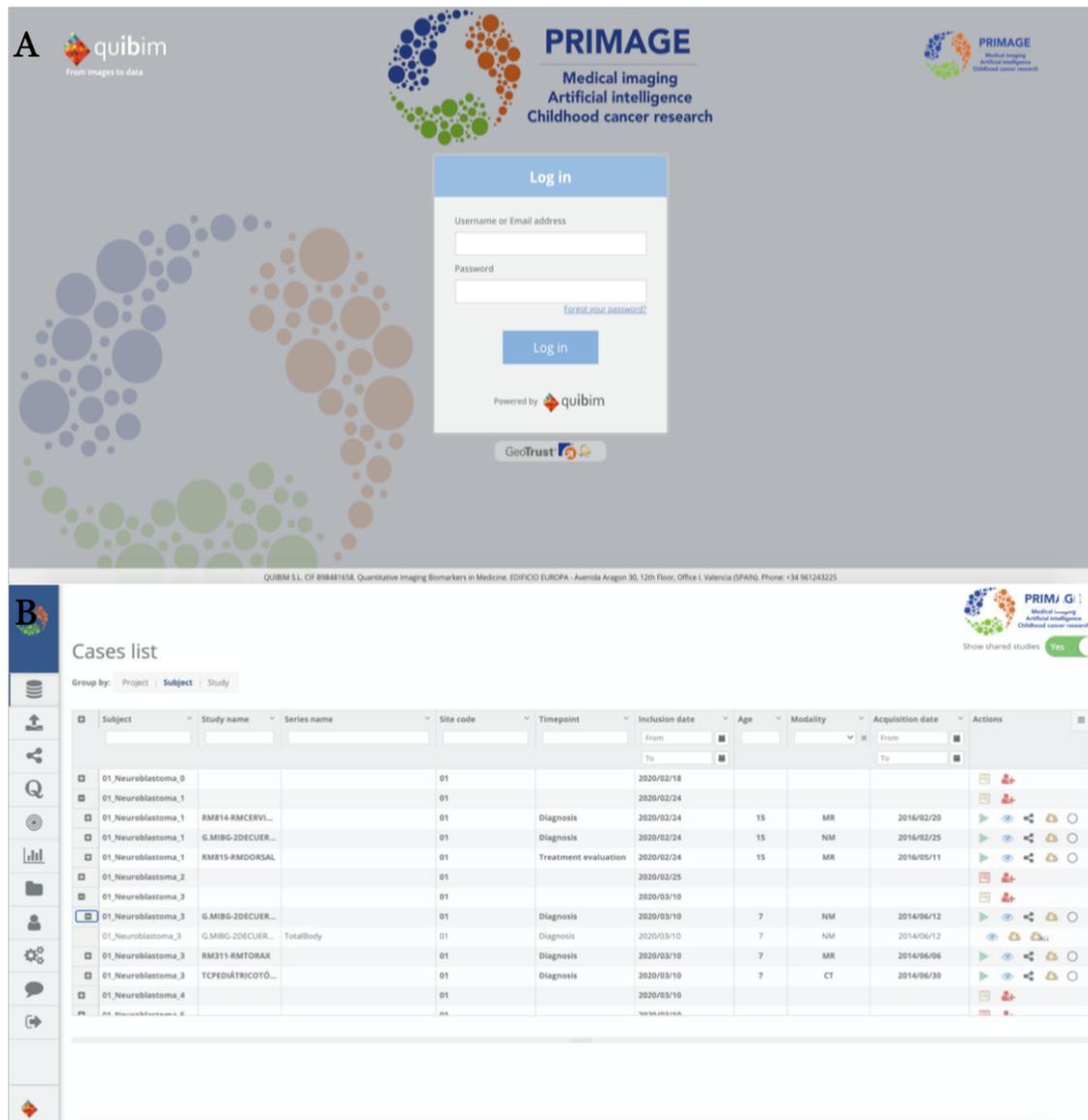


Figura 8. Ejemplo de la vista de la plataforma PRIMAGE. A. Vista de acceso. B. Visor de pacientes.

Por otro lado, la plataforma permite la visualización y anotación de las imágenes (etiquetas DICOM, selección de regiones de interés para segmentación) (Figura 9A), y la creación de módulos de análisis en contenedores Docker (Figura 9B).

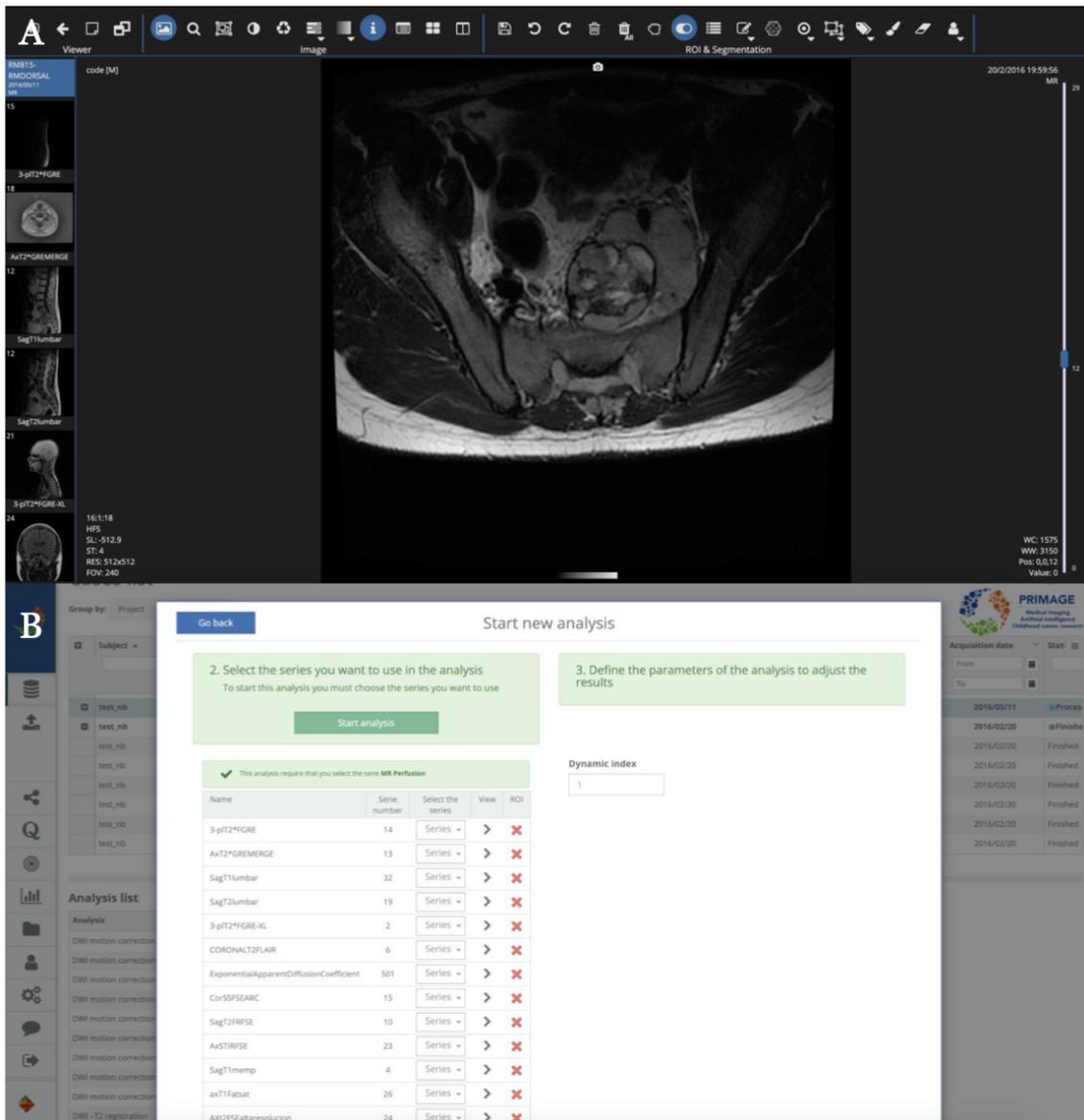


Figura 9. Ejemplo de la vista de la plataforma PRIMAGE. A. Visor de imagen. B. Vista de lanzamiento de los análisis.

3.5. Evaluación clínica y biomarcadores

Se recogieron diversas variables clínicas, analíticas y de imagen para la completa caracterización de la población. La recogida de datos se llevó a cabo de forma retrospectiva de las historias clínicas de los pacientes en cada uno de los centros de procedencia y se almacenó en la plataforma PRIMAGE. Las variables se recogieron atendiendo a las recomendaciones del Grupo Internacional de Riesgo del Neuroblastoma para la estratificación de riesgo de pacientes con

neuroblastoma [15]. Se estableció un cuestionario que se integró en la plataforma del proyecto para la recogida estandarizada de variables.

3.5.1. Variables clínicas

Se recogieron las siguientes variables:

- Datos de pacientes: Sexo
- Información al diagnóstico
 - o Fecha de diagnóstico (fecha de biopsia), edad al diagnóstico (en meses), hallazgo incidental (sí/no), fecha de los primeros síntomas.
 - o Hallazgos clínicos: datos que no se han empleado para esta tesis doctoral y que se emplearán para publicaciones futuras.

3.5.2. Variables analíticas

Se obtuvieron los siguientes datos analíticos de las muestras de sangre, biopsia tumoral, estudio de médula ósea, biopsia líquida y estudios moleculares:

- Anatomía patológica tumoral: evaluada/no evaluada, fecha de biopsia, tipo histológico (neuroblastoma, ganglioneuroblastoma, glanglioneuroblastoma nodular, glanglioneuroblastoma entremezclado, ganglioneuroma, ganglioneuroma maduro, ganglioneuroma en maduración, no disponible), grado de diferenciación (diferenciado, poco diferenciado, indiferenciado), índice de mitosis/kariorrexis (alto, intermedio, bajo, no evaluado), tipo de muestra (tumor primario, ganglios linfáticos, médula ósea, otras metástasis), tipo de biopsia (biopsia aspiración con aguja fina, biopsia con aguja gruesa, biopsia excisional, resección quirúrgica, desconocido).
- Estudio de médula ósea: evaluado/no evaluado, fecha de estudio, aspirado (evaluado/no evaluado; resultados).
- Biopsia líquida: evaluado/no evaluado (método; sí/no).
- Estudio molecular: evaluado/no evaluado, tipo de muestra (tumor primario, ganglios linfáticos, médula ósea, otras metástasis), estatus MYCN (evaluado/no evaluado; método de evaluación; estatus: amplificado/no amplificado/sin resultado).

- Estatus 11q (evaluado/no evaluado; estatus (aberración/no aberración/sin resultado)).
- Ploidía del ADN (evaluado/no evaluado; método de evaluación; estatus diploide-tetraploide / triploide-hexaploide / pentaploide / sin resultado).

3.5.3. Variables de imagen

El estudio se realizó con las imágenes de RM. Las variables recogidas se evaluaron de manera independiente en las RM al diagnóstico y las RM realizadas durante el tratamiento:

- RM: realizada/no realizada, proveedor, modelo, fecha de realización.

Se evaluaron los IDRF a partir de las RM, con las siguientes variables (sí/no/desconocido):

- Extensión tumoral ipsilateral dentro del cuello-tórax; extensión tumoral ipsilateral dentro del tórax-abdomen; extensión tumoral ipsilateral dentro del abdomen-pelvis
- Cuello: tumor que engloba la arteria carótida y/o vertebral y/o la vena yugular interna; tumor que se extiende hasta la base del cráneo; tumor que comprime la tráquea
- Unión cervicotorácica: tumor que engloba las raíces del plexo braquial, tumor que engloba los vasos subclavios y/o la arteria vertebral y/o carótida; tumor que comprime la tráquea
- Tórax: tumor que engloba la aorta y/o sus ramas principales; tumor que comprime la tráquea y/o los bronquios principales; tumor mediastínico inferior que infiltra la unión costovertebral T9 y T12
- Toracoabdominal: tumor que engloba la aorta y/o la vena cava
- Abdomen/pelvis: tumor que infiltra la porta hepática y/o el ligamento hepatoduodenal; tumor que envuelve ramas de la arteria mesentérica superior en la raíz mesentérica; tumor que engloba el origen del eje celíaco y/o de la arteria mesentérica superior; tumor que invade uno o ambos pedículos renales; tumor que engloba la aorta y/o la vena cava; tumor que engloba los vasos ilíacos; tumor pélvico que cruza la escotadura ciática

- Extensión del tumor intraespinal cualquiera que sea su ubicación, siempre que más de un tercio del canal espinal en el plano axial esté invadido y/o los espacios leptomeníngeos perimedulares no sean visibles y/o la señal de la médula espinal sea anormal
- Infiltración de órganos/estructuras adyacentes: pericardio, diafragma, riñón, hígado, bloque duodenopancreático y mesenterio.

Se evaluó la localización tumoral al diagnóstico:

- Suprarrenal, abdominal no suprarrenal, cervical, pélvico, torácico.
- Metástasis: presencia (sí/no), y localización.

3.5.4. Variables de estadificación

Tras la extracción de las variables anteriormente descritas, se obtuvieron las siguientes variables para la estadificación tumoral:

- INRGSS (a partir de los IDRF): L1 / L2 / M / MS
- INRG (considerando la edad, la categoría histológica, el grado de diferenciación tumoral, la amplificación de N-MYC, la aberración 11q y la ploidía del ADN): muy bajo/bajo/intermedio/alto

3.5.5. Variables clínicas establecidas que definen la agresividad

La agresividad tumoral se refiere en esta Tesis Doctoral a la capacidad de un tumor para crecer, invadir los tejidos circundantes y diseminarse a distancia. Esto puede evaluarse a través de distintos marcadores biológicos, como la tasa de proliferación celular, el grado histológico, la presencia de ciertas mutaciones genéticas, la invasión vascular o la presencia de metástasis, entre otros. También puede estimarse considerando la supervivencia que presentan los pacientes, puesto que existe una relación indirecta entre agresividad tumoral y supervivencia global: en general, tumores más agresivos tienden a ser más difíciles de tratar y tienen un pronóstico menos favorable, lo que puede llevar a una menor supervivencia global en comparación con tumores menos agresivos. No obstante, hay excepciones y factores individuales que pueden influir en el resultado.

La supervivencia global es el patrón de referencia tradicional en los ensayos clínicos aleatorizados para pacientes oncológicos [89], y se refiere al tiempo transcurrido desde el inicio (diagnóstico, tras intervención) hasta el fallecimiento por cualquier causa. En nuestro estudio se empleó la supervivencia global como marcador clínico para definir la agresividad tumoral global. La fecha de inicio para la medición del tiempo se estableció en la fecha de diagnóstico por imagen, dado que en 75 casos no se disponía de anatomía patológica definitiva/biopsia al diagnóstico. En el resto de casos, se estudió el tiempo entre la fecha al diagnóstico por imagen y la biopsia, y se vio que en todos los casos era inferior a 7 días.

Se evaluó el tiempo a evento, el tiempo de supervivencia en el caso de aquellos que han sufrido el evento (fallecimiento), o el tiempo hasta el último momento de seguimiento en el caso de aquellos pacientes para los que el evento no ha sucedido, así como el estado del paciente en ese punto temporal (vivo/fallecido).

3.6. Imagen de RM

3.6.1. Protocolos de imagen de RM

Dado que se trata de un estudio multicéntrico, los pacientes presentaban gran variabilidad en los equipos y en los parámetros de adquisición de las imágenes, con una gran heterogeneidad en los proveedores, la intensidad de campo magnético, los protocolos y la región anatómica incluida.

Se realizaron exploraciones con RM de General Electric Healthcare (Signa Excite HDxt, Signa Explorer, Discovery, Genesis; n = 80); Siemens Medical (Aera, Skyra, Symphony, Avanto, Magneto Espree, TrioTim, Sonata, Sprecetra, Verio, Harmony; n = 298); Philips Healthcare (Intera, Achieva, Ingenia, Panorama; n = 129) y Canon (Titan; n = 1). La intensidad de campo fue de 1.5 Tesla (n = 397) o 3 Tesla (n = 111).

El protocolo básico del que disponían todas las series analizadas incluyó, como mínimo, una secuencia T2 spin echo (T2wSE) o T2* eco de gradiente (T2*wGE), con o sin supresión grasa (FS) o recuperación de inversión (IR). Las secuencias potenciadas en T2 incluidas para cada subanálisis se describen en los correspondientes apartados de la sección de Resultados (secciones 4.1., 4.2.,

4.3.2, 4.3.3., 4.4.1., 4.4.3.). Asimismo, los pacientes disponían como mínimo de una secuencia potenciada en T1, usualmente Eco de Gradiente. El estudio de difusión se realizó en 188 pacientes, con valores $b=0$ y $b=1000$. Del total de la población estudiada, 70 pacientes disponían de estudios de perfusión con contraste y al menos 3 adquisiciones temporales. Las imágenes torácicas se obtuvieron con sincronización respiratoria. En caso de requerir sedación, ésta se realizó con Midazolam por vía intravenosa u oral en aquellos pacientes sin vía intravenosa.

La media de tiempo total de adquisición fue de 20 minutos (± 2 minutos). El campo de visión (*field of view*, FOV) medio adquirido fue de 410 mm (mediana 440, rango 225-500 mm). Las dimensiones en plano de los vóxeles fue variable, como se comentará posteriormente.

3.6.2. Selección de secuencias

Se seleccionaron las secuencias sobre las que realizar el análisis. El estudio se llevó a cabo en secuencias T2-T2*w convencionales de RM. Se escogieron imágenes con orientación en el plano transversal obtenidas con secuencias potenciadas en T2wSE FS o T2 STIR (*Short Time Inversion Recovery*), ya que son secuencias adquiridas en la práctica clínica habitual que proporcionan mayor contraste entre diferentes tejidos, permitiendo una óptima visualización del tumor. En caso de no encontrarse disponibles o presentar artefactos significativos de movimiento que impidiesen la correcta visualización tumoral, se usaron secuencias sin supresión grasa, T2wSE o T2*w GE.

3.7. Preparación de la imagen RM

Al tratarse de un estudio multicéntrico, y como se ha comentado anteriormente, existía una amplia variabilidad en la adquisición de las imágenes. Con la finalidad de reducir esta heterogeneidad se realizó un proceso estandarizado de preparación de las imágenes de RM, con distintos pasos preparatorios que se detallan a continuación. Esta preparación de la imagen permite realizar la

segmentación y la posterior extracción de características radiómicas sobre imágenes de RM con menor heterogeneidad y artefactos pese a las diferencias en los parámetros de adquisición.

Esta fase preparatoria consistió en aplicar un Filtro de Difusión Anisotrópica (ADF) para reducir el ruido de las imágenes, corrección del sesgo por campo magnético N4 para corregir la heterogeneidad en la distribución de la señal, normalización de señal mediante Z-score para la estandarización de las intensidades, y re-muestreo para la armonización espacial del tamaño de los vóxeles (Figura 10). El flujo seguido se detalla a continuación.

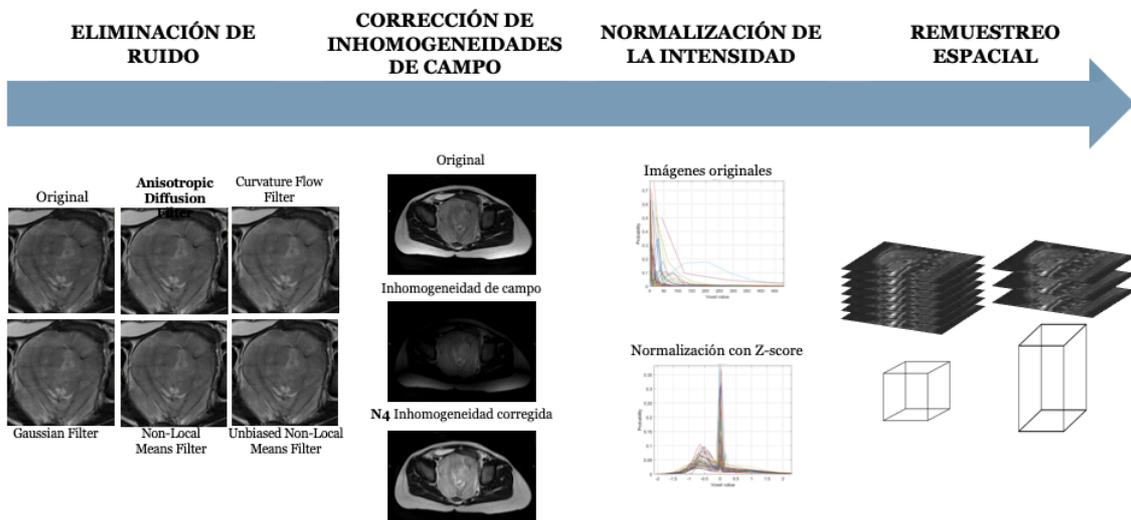


Figura 10. Flujo de preparación de la imagen.

3.7.1. Filtros de ruido (*denoising*)

El filtrado del ruido es un proceso fundamental en el procesamiento de las imágenes que pretende reducir o eliminar el ruido presente en una imagen con la finalidad de mejorar su calidad. El "ruido" en una imagen se refiere a las variaciones aleatorias o no deseadas en el brillo o color de los píxeles.

En la práctica clínica habitual, las imágenes de RM se obtienen y reconstruyen con escáneres, secuencias y protocolos diversos, con lo que presentan variaciones en los niveles de ruido que afectan la calidad de las imágenes, impactan la relación

señal-ruido y comprometen la reproducibilidad de los biomarcadores de imagen. Para garantizar la usabilidad de las imágenes de RM se debe mejorar su calidad mediante la aplicación de técnicas de filtrado de ruido que maximicen su eliminación y preserven los detalles morfológicos [90].

Existen varios métodos para realizar la eliminación de ruido de imágenes, que pueden incluir técnicas de filtrado espacial, transformadas de dominio frecuencial, y algoritmos basados en aprendizaje profundo, como las redes neuronales convolucionales (CNN). En el seno del proyecto PRIMAGE se realizó un subanálisis con el objetivo de medir el impacto de los diferentes filtros de ruido en la calidad final de la imagen y en la estabilidad de las métricas radiómicas, para seleccionar los parámetros óptimos del filtro tanto en fantomas como en imágenes del mundo real de pacientes con tumores neuroblásticos [91].

En dicho estudio se utilizaron cinco de los principales filtros de ruido: el filtro gaussiano (GF) [92], el filtro de flujo de curvatura (CFF) que preserva la definición de bordes mientras suaviza perpendicularmente las curvas de isointensidad [100], el filtro de difusión anisotrópica (ADF) que elimina el ruido de alta frecuencia mientras conserva los bordes [94], el filtro de medias no locales (NLMF) [13] y el filtro de medias no locales sin sesgo (UNLMF) [102]. La representación visual del efecto de estos filtros se observa en la Figura 11. Todos ellos pertenecen al grupo de filtros "preservadores de bordes", con la excepción del filtro gaussiano, que pretende eliminar el ruido mientras preserva las estructuras morfológicas. La comparación de estas 5 familias de filtros en las imágenes de RM potenciadas en T2wSE, tanto en fantomas como en pacientes pediátricos con neuroblastoma, demostró que el filtro ADF debería utilizarse para una extracción más reproducible de las características radiómicas y así facilitar la determinación de los mejores biomarcadores de imagen en oncología [91]. Por ello, se seleccionó el filtro ADF para aplicar a las imágenes de RM como parte de la preparación de las imágenes en esta Tesis Doctoral. Los filtros descritos están integrados en la biblioteca SimpleITK [103].

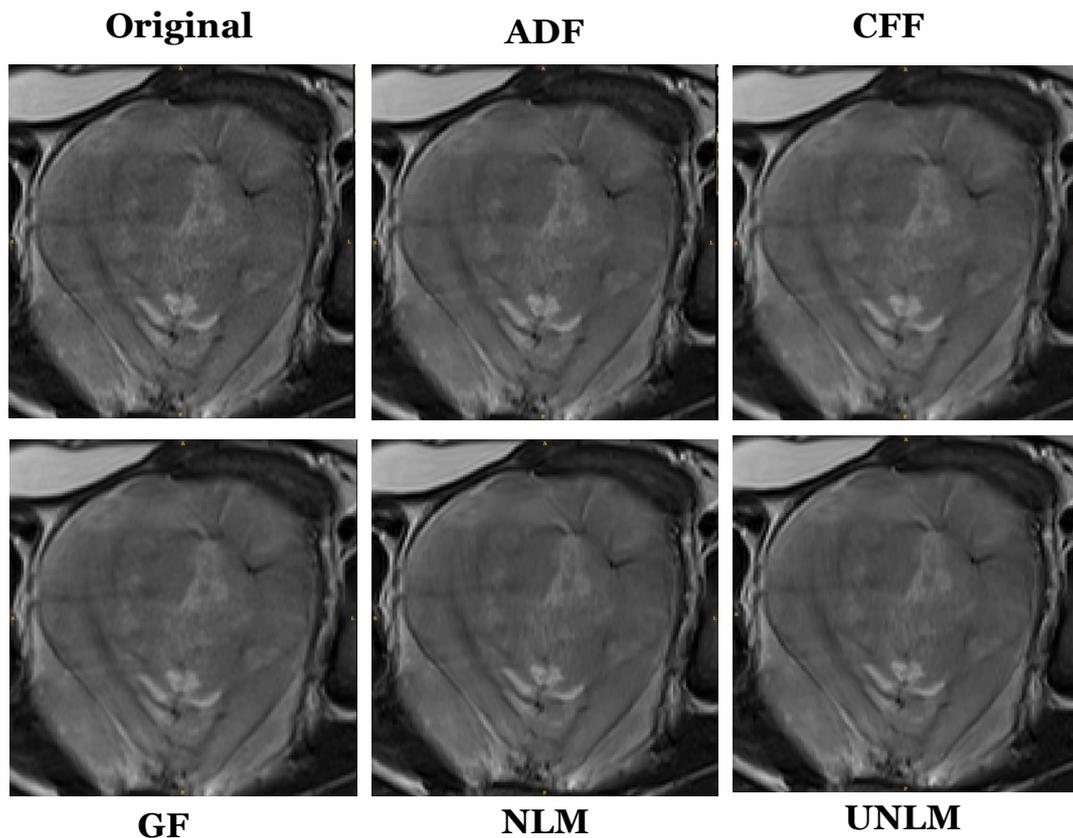


Figura 11. Imagen original y tras la aplicación de diversos filtros de denoising. RM potenciada en T2wSE de paciente con neuroblastoma pélvico. ADF=Anisotropic Diffusion Filter. CFF=Curvature Flow Filter. GF= Gaussian Filter. NLM=Non-Local Means Filter. UNLM=Unbiased Non-Local Means Filter.

3.7.2. Eliminación de inhomogeneidades

El sesgo por campo magnético (*bias-field*) en imágenes de RM se refiere a una variación de intensidad que ocurre de manera gradual en la imagen debido a factores como la inhomogeneidad del campo magnético, imperfecciones en la bobina de radiofrecuencia y artefactos en la adquisición. Este sesgo puede afectar la precisión de las mediciones cuantitativas en la imagen y, por lo tanto, es importante corregirlo para obtener resultados precisos. Su corrección se realiza mediante la aplicación de técnicas de procesamiento y filtros que permiten eliminar o reducir esta variación de la intensidad de la imagen, permitiendo así una representación más homogénea de las estructuras anatómicas [97].

Con la finalidad de corregir este sesgo en las imágenes de RM, se utilizó la corrección por filtro N4, un método de actualidad para corregir inhomogeneidades en RM que ha sido utilizado ampliamente en la literatura [98, 99]. El efecto de su corrección puede observarse en la Figura 12. El filtro se aplicó mediante el software ANTS (*Advanced Normalization Tools*) [100].

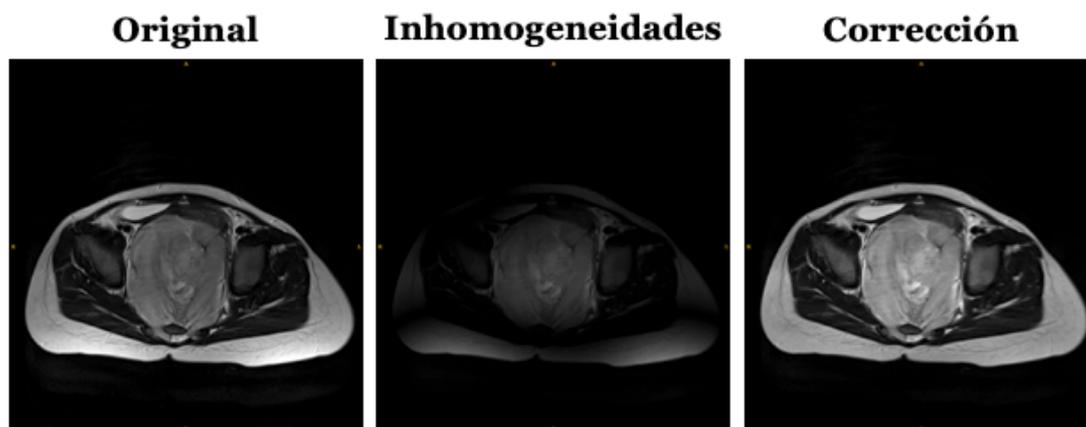


Figura 12. Efecto de corrección del sesgo por campo magnético. Imagen original, extracción de las inhomogeneidades e imagen corregida tras la eliminación de estas inhomogeneidades con corrección N4.

3.7.3. Normalización de intensidad

La normalización de las intensidades en las imágenes de RM se refiere al proceso de ajustar los niveles de brillo y contraste de los píxeles con el fin de facilitar la comparación de imágenes diferentes. Este proceso se lleva a cabo mediante la aplicación de transformaciones matemáticas que redistribuyen los valores de los píxeles en la escala de grises. En nuestro estudio se aplicó la metodología Z-score como proceso de normalización para reducir variaciones sistemáticas debidas a la adquisición de imágenes, reconstrucción y procesado [41, 101]. Este filtro se utiliza ampliamente para estandarizar los datos entre imágenes y ha demostrado previamente unos resultados sólidos para el análisis radiómico [102]. La representación de la distribución de los valores de las intensidades de cada vóxel en las imágenes originales y tras la aplicación de Z-score puede observarse en la Figura 13, observando la distribución de los valores alrededor del valor 0 tras

normalizar las intensidades. La normalización se realizó en el entorno de Pyradiomics [103].

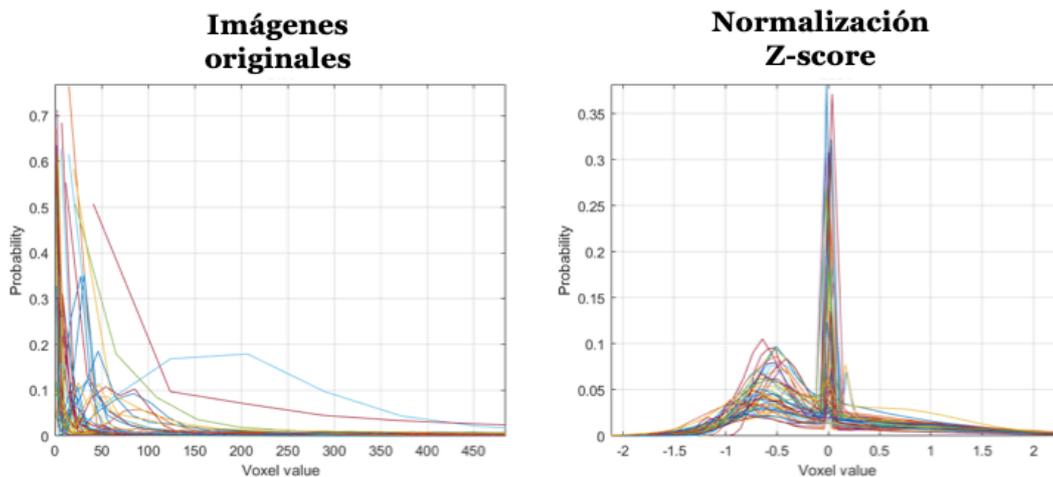


Figura 13. Distribución de los valores de las intensidades de cada vóxel en las imágenes originales y tras la aplicación de Z-score.

3.7.4. Cambio de tamaño y registro (remuestreo espacial)

El remuestreo de la imagen implica la modificación de su resolución espacial, de modo que los píxeles originales de la imagen se reorganizan para ajustarla a una nueva cuadrícula o escala de píxeles. Para permitir la comparación entre las diversas imágenes se llevó a cabo un remuestreo para homogeneizar el tamaño de los vóxeles de todas las imágenes al tamaño estandarizado de 1 x 1 x 6 mm, siendo 6 mm la dimensión del espesor de corte. El remuestreo se realizó en el entorno de Pyradiomics [103].

3.8. Segmentación manual en RM

3.8.1. Metodología de segmentación

La segmentación consiste en la delimitación y etiquetado de los vóxeles que forman parte de una región de interés (en este caso, los tumores neuroblásticos). Como se ha expuesto anteriormente, la segmentación se llevó a cabo en

secuencias convencionales de RM. Se segmentó en imágenes con orientación transversal en secuencias potenciadas en T2wSE FS o T2 STIR. En caso de no encontrarse disponibles o presentar artefactos de movimiento que impidiesen la correcta delimitación tumoral, se usaron secuencias T2wSE o T2*wGE.

Se segmentaron los tumores en 3D incluyendo la totalidad del tumor. Esta segmentación volumétrica se realizó incluyendo todos aquellos vóxeles que formaban parte del tumor neuroblástico en todas las imágenes de la secuencia analizada (Figura 14).

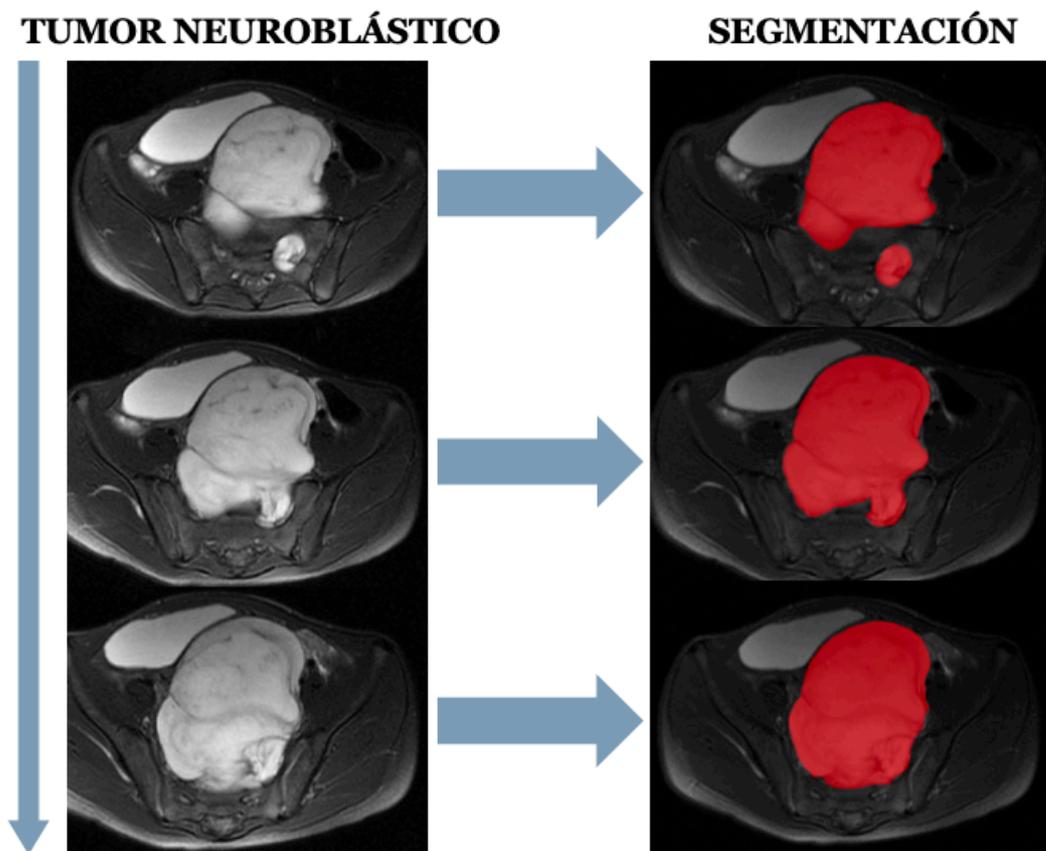


Figura 14. Ejemplo de segmentación manual de un neuroblastoma pélvico en tres imágenes de Resonancia Magnética T2w STIR transversal (de craneal a caudal).

En esta Tesis Doctoral, la segmentación se realizó empleando dos herramientas diferentes: ITK-SNAP y la plataforma PRIMAGE. En un primer momento, para la segmentación manual y para el entrenamiento de la red de segmentación automática, se utilizó el programa de acceso abierto ITK-SNAP (versión 3.8.0) [45], herramienta que permite a los usuarios navegar por imágenes

tridimensionales y delinear manualmente regiones anatómicas de interés. Tras el entrenamiento de la red, la validación de la herramienta de segmentación y los posteriores análisis se realizaron en el entorno de la plataforma PRIMAGE, cuyas características se han descrito en el apartado 3.4. de Material y Métodos. En ambos entornos las imágenes se obtuvieron en formato DICOM. Las máscaras de segmentación manual obtenidas se exportaron en formato NifTI y se consideraron como las regiones de interés “patrón de referencia” o “verdad fundamental” (*ground truth*, conjunto de datos verificados con respecto a los cuales se pueden comparar otros datos).

Durante el proceso de segmentación manual con la herramienta ITK-SNAP®, las secuencias transversales escogidas para realizar la segmentación se evaluaron conjuntamente con las imágenes potenciadas en T2wSE en plano coronal o sagital para identificar los límites del tumor en al menos dos planos. Asimismo, se revisaron secuencias DW y DCE (en caso de encontrarse disponibles) para ayudar a definir los márgenes tumorales que resultasen dudosos en secuencias T2-T2*w (Figura 15).

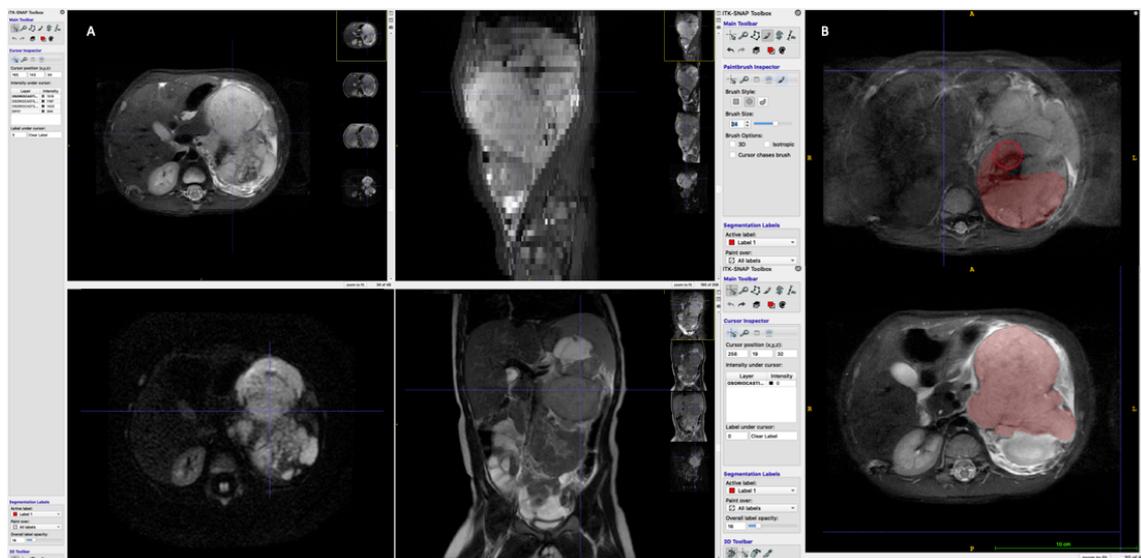


Figura 15. Entorno de segmentación ITK-SNAP®. A. Herramienta para segmentación manual, correlación de imágenes STIR transversal y sagital, DW transversal y T2wSE coronal de un tumor neuroblástico suprarrenal izquierdo. B. Proceso de segmentación manual en distintos cortes transversales.

3.8.2. Criterios restrictivos de segmentación

Los tumores neuroblásticos pueden ser lesiones de gran volumen y morfología irregular. Su ubicación suele ser cercana a otras estructuras anatómicas muy relevantes como los vasos, órganos sólidos, asas intestinales o vías respiratorias. Estas estructuras pueden estar en contacto con el tumor, comprimidas, englobadas o incluso infiltradas por él. Los ganglios linfáticos locorreionales pueden también contactar con los tumores, resultado en ocasiones indistinguibles de los mismos.

Previamente a la realización de las segmentaciones, se elaboró una metodología en la que se establecieron criterios para delimitar los vóxeles que con seguridad formaran parte del tumor, empleando un enfoque restrictivo (Figura 16):

- Si el tumor se encontraba bien delimitado respecto a las estructuras adyacentes, se realizó la segmentación siguiendo los márgenes del tumor.
- En aquellos casos en los que el tumor se encontraba en contacto con un vaso o lo englobaba, el vaso se excluyó de la segmentación.
- Si un tumor infiltraba órganos vecinos, estando los márgenes mal definidos entre ellos, se revisaron las secuencias DW y DCE para distinguir la porción infiltrante del tumor de la estructura infiltrada, excluyendo esta de la segmentación.
- Los ganglios linfáticos en contacto estrecho con el tumor en los que la separación de ambas estructuras no fuese posible fueron incluidos en la segmentación.
- Los ganglios a distancia y las metástasis fueron excluidos de la segmentación y posterior análisis.

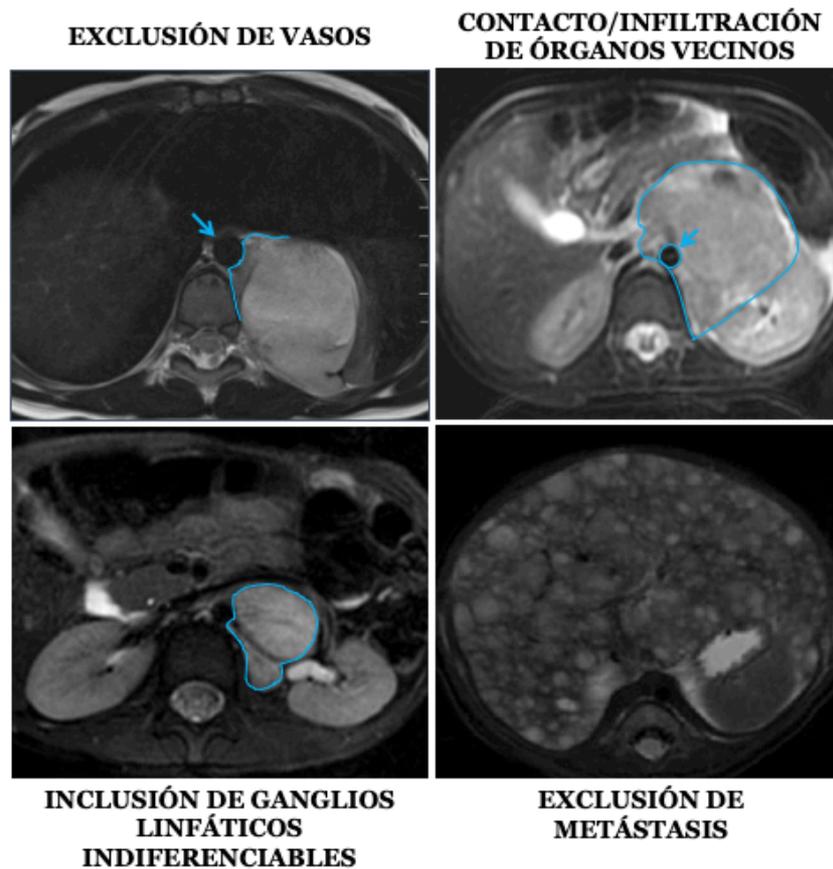


Figura 16. Criterios restrictivos de segmentación manual.

3.9. Metodología estadística

Las variables continuas se expresaron como la media \pm desviación estándar (DE), o mediana y Rango Intercuartílico (IQR) (en caso de distribución no normal). Las variables categóricas se expresaron como frecuencia absoluta y su porcentaje (%). La normalidad de las variables [104] se analizó mediante la prueba de Shapiro-Wilk.

3.9.1. Segmentación

La comparación de los resultados de segmentación es un problema abordado frecuentemente en la literatura, para el que se han descrito distintas métricas. La verdad fundamental o segmentación de referencia consistió en la segmentación manual realizada por una radióloga pediátrica (Radióloga 1, autora de esta Tesis Doctoral), tanto para el cálculo de la variabilidad interobservador como para la comparación de la segmentación manual con la automática.

La principal métrica utilizada en este estudio para la evaluación de resultados de segmentación fue el Coeficiente de Similitud de Dice (CSD) [104], siendo este un índice de superposición espacial y una métrica de validación de reproducibilidad [105]. Su valor puede oscilar entre 0 (lo cual indica ausencia de superposición espacial entre las dos máscaras de segmentación) y 1 (indicando una superposición completa). El índice CSD se ha utilizado ampliamente para calcular la métrica de superposición entre resultados de la segmentación y la verdad de referencia, y se define como [106, 107]:

$$CSD = \frac{2VP}{2VP + FP + FN}$$

Fórmula: FP, Falsos Positivos; VP, Verdaderos Positivos; FN, Falsos Negativos.

Siendo su representación visual (Figura 17):

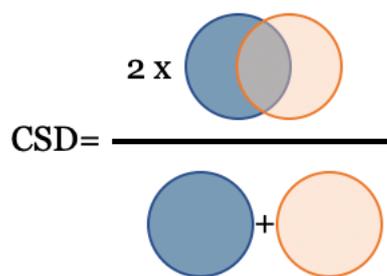


Figura 17. Representación visual del Coeficiente de Similitud de Dice, en el que cada círculo representa una máscara de segmentación.

El índice de Jaccard (JAC) es una métrica de superposición espacial definida como la intersección entre las máscaras de los dos conjuntos de datos dividida por su unión [108]:

$$JAC = \frac{VP}{VP + FP + FN}$$

Fórmula: FP, Falsos Positivos; VP, Verdaderos Positivos; FN, Falsos Negativos.

Siendo su representación visual (Figura 18):

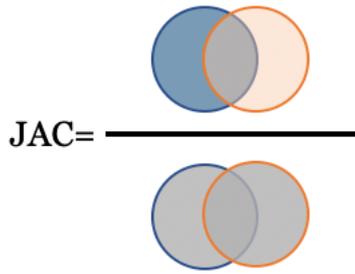


Figura 18. Representación visual del índice de Jaccard, en el que cada círculo representa una máscara de segmentación.

La distancia de Hausdorff (HD) es una métrica de distancia espacial basada en el cálculo de las distancias entre todos los pares de vóxeles de cada máscara (A o B), midiendo la distancia máxima de una máscara de segmentación (A) al punto más lejano en la otra máscara (B) (Figura 19).

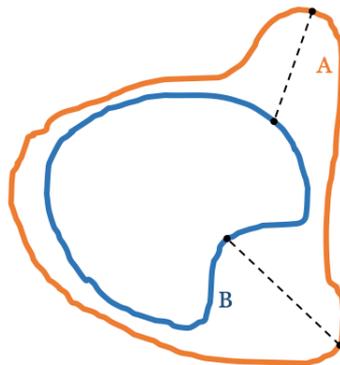


Figura 19. Representación visual de la distancia de Hausdorff, en el que se considera la distancia máxima entre pares de vóxeles de dos máscaras de segmentación.

La HD mide el máximo de las distancias entre cada punto y está definida por [109]:

$$HD(A, B) = \max(h(A, B), h(B, A)),$$

$$\text{donde } h(A, B) = \max \min ||a - b||,$$

$$\text{y } a \in A, b \in B.$$

Donde $h(A, B)$ representa la distancia máxima desde cualquier punto en el conjunto A hasta su punto más cercano en el conjunto B. $h(B, A)$ se refiere a la distancia máxima

desde cualquier punto en B hasta su punto más cercano en A. $\|a-b\|$ denota la distancia entre dos puntos a y b, en este caso, a pertenece al conjunto A y b pertenece al conjunto B. La función $\max_{a \in A} \min_{b \in B} \|a-b\|$ encuentra el máximo de todas las distancias mínimas entre cada punto en el conjunto A y todos los puntos en el conjunto B. La expresión $a \in A$ implica que a es un elemento del conjunto A, y $b \in B$ que b es un elemento del conjunto B: en el contexto de la definición de la distancia de Hausdorff, a recorre todos los puntos en el conjunto A y b recorre todos los puntos en el conjunto B, evaluando la distancia entre cada par de puntos (a,b) para calcular la distancia máxima desde cualquier punto en A hasta su punto más cercano en B y viceversa.

Se calculó la métrica AUC ROC (Área bajo la curva de Característica Operativa del Receptor). La curva ROC, como una representación de sensibilidad frente a 1-especificidad, normalmente supone más de una medición. Para el caso en el que una segmentación se compara con la segmentación o verdad de referencia, consideramos AUC como [109]:

$$AUC = 1 - \frac{1}{2} \left(\frac{FP}{FP + VN} + \frac{FN}{FN + VP} \right)$$

Fórmula: FP, Falsos Positivos; VP, Verdaderos Positivos; FN, Falsos Negativos; VN, Verdaderos Negativos.

Se diseñaron específicamente dos métricas basadas en la superposición espacial para obtener una comprensión más profunda de la dirección de los errores encontrados: la proporción de falsos positivos (FP) y la de falsos negativos (FN). Ambas métricas se calcularon de manera independiente con respecto a la segmentación de referencia. La segmentación empleada como verdad fundamental, realizada por la Radióloga 1, incluye aquellos vóxeles que son verdaderos positivos (VP) (pertenecen al tumor y han sido incluidos en la segmentación) y falsos negativos (FN) (pertenecen al tumor y no se han segmentado). Esta máscara fue comparada, en primer lugar, con la segmentación manual realizada por la Radióloga 2 y posteriormente con la segmentación automática obtenida por el modelo de segmentación automática (máscaras no

correspondientes con la verdad fundamental), que incluyen vóxeles VP y FP (vóxeles segmentados por la Radióloga 2 o la segmentación automática no pertenecientes al tumor) (Figura 20).

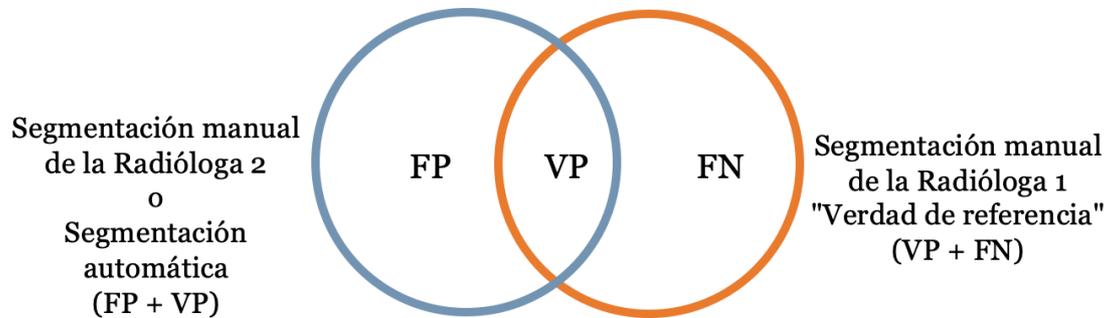


Figura 20. Representación visual de falsos positivos, verdaderos positivos y falsos negativos de la máscara de referencia (círculo naranja) y la no correspondiente con la segmentación de referencia (círculo azul). Abreviaturas: FP, Falsos Positivos; VP, Verdaderos Positivos; FN, Falsos Negativos.

La Tasa de Falsos Positivos modificada (TFPm) consideró aquellos vóxeles que fueron identificados por la red o la Radióloga 2 como tumor, pero correspondían a otras estructuras, divididos por los vóxeles que realmente correspondían a la máscara de la verdad de referencia realizada por la Radióloga 1 (vóxeles VP + FN). Esta definición difiere de la ratio de FP utilizada en problemas estadísticos estándar al excluir el término de verdadero negativo (VN) de la expresión matemática, ya que los vóxeles de VN corresponden al fondo de la imagen en una tarea de segmentación y no a las máscaras de segmentación destinadas a ser comparadas.

$$TFPm = \frac{FP}{VP + FN}$$

La Tasa de Falsos Negativos (TFN) de la segmentación automática o realizada por la Radióloga 2 a la segmentación de referencia, midió los vóxeles pertenecientes al tumor que la red o la Radióloga 2 no incluyeron como tales, divididos entre los vóxeles de la segmentación de referencia realizada por la Radióloga 1.

$$TFN = \frac{FN}{VP + FN} = 1 - \text{Sensibilidad}$$

Por motivos de consistencia y para facilitar la comprensión de los resultados, las métricas TFPm y TFN se presentan como 1-métrica, lo que resulta en un máximo de 1 para un acuerdo vóxel a vóxel completo y un mínimo de 0 para una similitud nula.

En el análisis de validación independiente, con la finalidad de valorar la influencia de diferentes variables (localización, momento temporal, campo magnético y secuencia de RM) sobre el CSD, se realizó un test ANOVA.

3.9.2. Reproducibilidad de la radiómica

Según la terminología recomendada para el desarrollo de biomarcadores cuantitativos de imagen, la reproducibilidad consiste en la precisión de la medición con condiciones que varían entre las mediciones replicadas [34]. Para evaluar la reproducibilidad se utilizó el coeficiente de correlación de concordancia (CCC), método comúnmente utilizado para evaluar la relación entre datos apareados [110], y su cálculo es:

$$CCC = \frac{2\sigma_X\sigma_Y\rho}{\sigma_X^2 + \sigma_Y^2 + (\mu_X - \mu_Y)^2},$$

donde μ_X , μ_Y son los valores medios, $2\sigma_X$, $2\sigma_Y$ son las varianzas correspondientes, y ρ es el coeficiente de correlación evaluado para cada característica radiómica en todos los sujetos dentro de diferentes adquisiciones. Los valores del CCC cercanos a 1 indican una alta reproducibilidad. El CCC se clasificó como excelente (si ≥ 0.90); bueno (0.75–0.89); moderado (0.50–0.75); o bajo (<0.50) [42].

Dado que este cálculo depende de la variabilidad inherente de los datos subyacentes [41], también se calculó el coeficiente de variación (CoV) para cada parámetro de radiómica. Esta prueba de repetibilidad de precisión se utiliza con frecuencia para investigaciones de reproducibilidad basadas en la variabilidad intra-sujeto, definida como:

$$CoV = \frac{\sigma_X}{\mu_X} * 100$$

Donde σ_X es la desviación estándar y μ_X es la media de la diferencia absoluta de los valores de características entre dos exploraciones de RM. Un CoV bajo indica que los puntos de datos están agrupados alrededor de la media, sugiriendo menos variabilidad y mayor consistencia. La clasificación de los resultados del CoV fue la siguiente: excelente ($\text{CoV} \leq 10\%$); bueno (11–20%); moderado (21–30%); y bajo ($>30\%$) [33].

Además, para evaluar las diferencias entre las series, se realizó un análisis descriptivo exploratorio inicial y un gráfico de cajas y bigotes, y para confirmar los hallazgos, se llevó a cabo un ANOVA.

3.9.3. Relación de la radiómica y la radiómica profunda con variables analíticas y de estadificación

Se llevó a cabo un análisis de la relación entre diferentes variables clínicas para tumores neuroblásticos en 508 pacientes con los parámetros de radiómica y radiómica profunda extraídos. Por un lado, se realizó un análisis de la relación de estos factores de riesgo con los 107 parámetros de radiómica previamente obtenidos. Por otro lado, se evaluaron las relaciones de estos factores de riesgo con las 50 variables de radiómica profunda.

Las variables seleccionadas para el análisis incluyeron sistemas de estadificación (el INRGSS y el INRGSS) y la variable analítica MYCN. La distribución de los pacientes entre los diferentes grupos se detalla en la Tabla 6.

Tabla 6. Tamaño muestral en las diferentes variables analizadas para el análisis de radiómica y radiómica profunda.

Variable analítica/estadificación		Pacientes
MYCN	Amplificado	121
	No amplificado	351
	Ganancia	27
	Total	499
INRGSS	M	215
	MS	42
	L1	72
	L2	179
	Total	508

INRG	Alto	275
	Intermedio	23
	Bajo	33
	Muy bajo	44
	Total	375

Los pacientes se estratificaron en diferentes grupos en base a criterios clínicos.

Estatus MYCN: se diferenció a los pacientes con amplificación del MYCN de aquellos sin amplificación.

- Grupo 1: amplificado
- Grupo 2: no amplificado + ganancia

Clasificación INRG: se realizó una división de los pacientes agrupando a aquéllos de riesgo bajo o muy bajo.

- Grupo 1: alto
- Grupo 2: intermedio
- Grupo 3: bajo y muy bajo

Clasificación INRGSS: se dividió a los pacientes en 3 grupos, los dos primeros implicando diseminación tumoral, y el tercer grupo representando los tumores localizados.

- Grupo 1: M
- Grupo 2: MS
- Grupo 3: L1+L2

Asimismo, los grupos se subdividieron por edad en menores de 18 meses, o iguales o mayores a 18 meses (Tabla 7). Por consenso se establece el punto de corte en 18 meses [8], ya que los pacientes menores de 18 meses pueden presentar mejor pronóstico que los niños mayores, especialmente el estadio MS.

Tabla 7. Distribución de pacientes en función de las diferentes variables analíticas y de estadificación analizadas para el análisis de radiómica y radiómica profunda y de la edad.

Variable analítica/estadificación		Pacientes	Edad < 18 meses	Edad ≥18 meses
MYCN	Amplificado	121	45	76
	No amplificado + Ganancia	378	186	192
INRGSS	M	215	52	163
	MS	42	42	0
	L1 + L2	251	147	104
INRG	Alto	275	54	221
	Intermedio	23	5	18
	Bajo + Muy bajo	77	4	73

Se analizaron las diferencias entre grupos para las variables de radiómica y radiómica profunda. Para comparar dos grupos, para cada variable se realizó un test de Shapiro-Wilk con la finalidad de comprobar si seguían una distribución normal, y un test de Levene para evaluar la varianza. Para las variables con solo dos grupos, se utilizó la prueba T-test (o la prueba Mann-Whitney en caso de no cumplir el supuesto de normalidad), ambas con un nivel de significancia de 0.05, según el esquema aportado en la Figura 21.

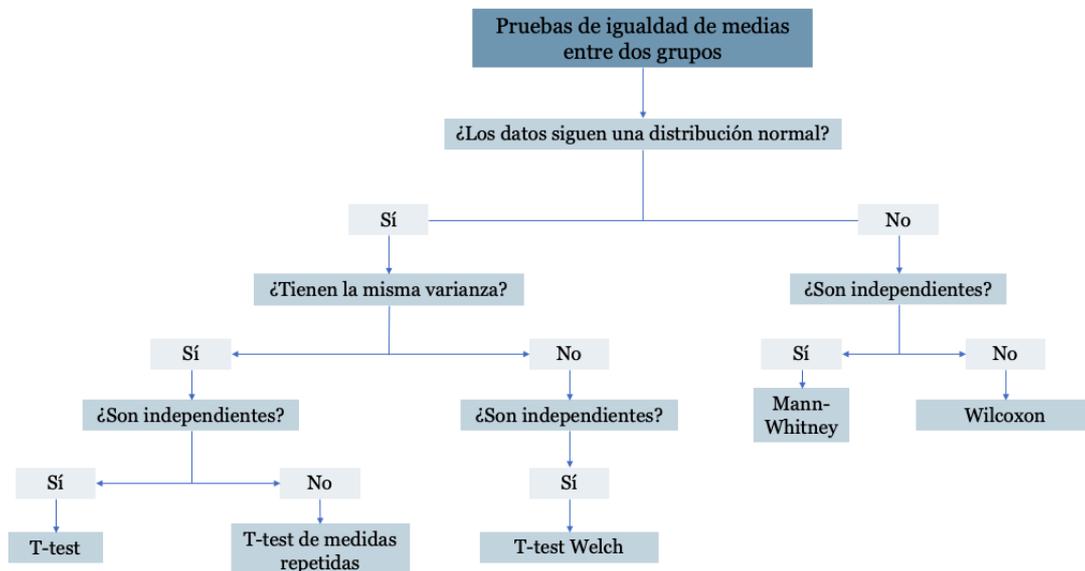


Figura 21. Pruebas de igualdad de medias entre dos grupos.

Para comparar medias entre más de dos grupos se evaluó la distribución de los datos (test Shapiro-Wilk) y la heterogeneidad de la varianza. En caso de seguir una distribución normal, se empleó un test ANOVA. En caso de que la

distribución no fuese normal, se realizó un test de Kruskal Wallis (KW) para grupos independientes, ambas con un nivel de significancia de 0.05. Posteriormente, se realizó una prueba post hoc entre grupos con la prueba de Bonferroni, también con un nivel de significancia de 0.05 (Figura 22).

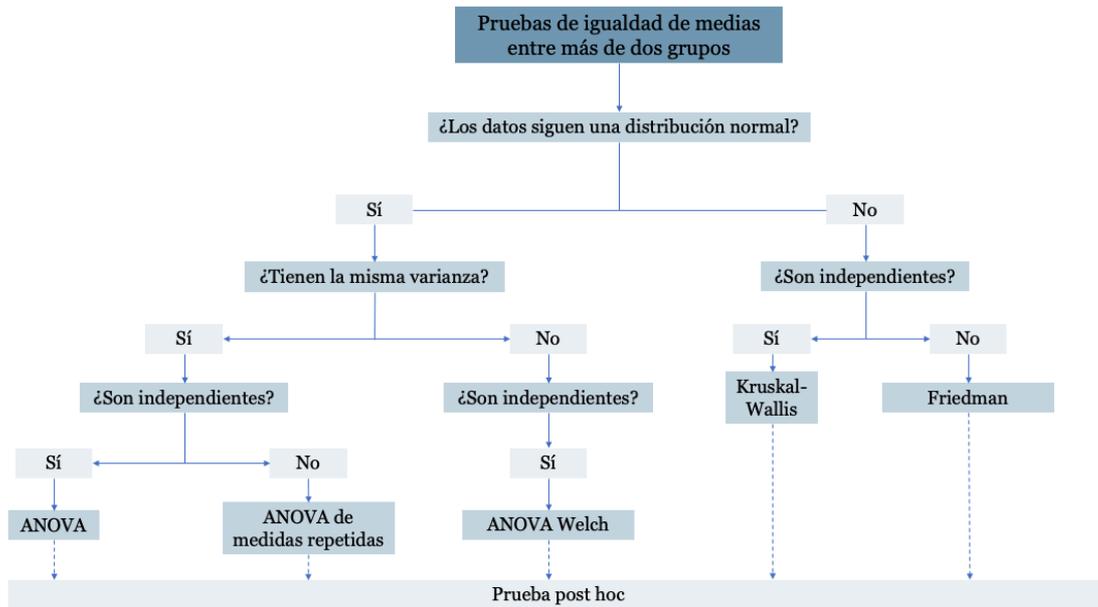


Figura 22. Pruebas de igualdad de medias entre más de dos grupos.

Los análisis se realizaron empleando el software Python, empleando las librerías *scipy.stats* y *pingouin* [95, 96].

3.9.4. Modelos predictivos de supervivencia

La evaluación de modelos de supervivencia implica métricas específicas diseñadas para medir la capacidad del modelo para predecir el tiempo hasta el evento de interés. Se emplearon:

- C-index: índice de concordancia, mide la proporción de pares de eventos y no eventos que se ordenan correctamente de acuerdo con sus tiempos de supervivencia predichos. La fórmula del C-Index en el contexto de supervivencia es la siguiente:

$$C - \text{index} = \frac{\text{Número de pares concordantes}}{\text{Número total de pares evaluados}}$$

Los "pares concordantes" son aquellos en los que el individuo que experimenta el evento tiene un riesgo más alto según las predicciones del modelo que el individuo censurado o que no ha experimentado el evento.

El "número total de pares evaluados" incluye todos los pares de individuos en los cuales uno ha experimentado el evento y el otro no. Esto también tiene en cuenta los casos en los que los tiempos de supervivencia están empatados.

El C-Index varía entre 0.5 y 1.0. Un valor de 0.5 indica que el modelo no tiene capacidad de discriminación mejor que al azar, mientras que un valor de 1.0 indica una discriminación perfecta.

- AUC-ROC: se emplea para evaluar la capacidad de discriminación de un modelo entre clases, donde la clase puede ser definida por diferentes eventos, como la supervivencia o la no supervivencia. Mide la capacidad del modelo para clasificar correctamente los eventos de supervivencia frente a los eventos de no supervivencia a lo largo de diferentes umbrales de probabilidad. Un valor de 0 indica que el modelo clasifica incorrectamente todas las instancias, mientras que un valor de 1 indica que el modelo clasifica todas las instancias correctamente.
- IPW (Ponderación de probabilidad inversa): medida de correlación de rangos entre los puntajes de riesgo predichos y los puntos de tiempo observados, asigna pesos a las observaciones en función de la inversa de la probabilidad de selección de esas observaciones en la muestra. Estos pesos se utilizan para calcular estimaciones potenciadas que tienen en cuenta la estructura de muestreo o la presencia de datos faltantes.

Las arquitecturas empleadas para la elaboración de los diferentes modelos de supervivencia se expondrán en el apartado 4.7. de Resultados.

4. RESULTADOS

RESULTADOS

En el siguiente apartado se detallan, en primer lugar, las características de la población estudiada. A continuación, se expone el proceso de segmentación llevado a cabo tras realizar las modificaciones para la preparación y homogeneización de las imágenes de RM expuestas en el apartado 3.7. La segmentación se realizó, en un primer momento, de forma manual por dos radiólogas, y se llevó a cabo un análisis de la variabilidad interobservador para establecer la máxima variabilidad tolerable al comparar dos máscaras de segmentación. Tras este análisis, se detalla cómo se desarrolló un modelo de detección y segmentación automática tumoral en RM empleando una red neuronal convolucional, que posteriormente se validó de manera independiente. Consecutivamente se explica cómo, una vez segmentados los tumores de manera automática, se extrajeron diferentes parámetros de radiómica y se realizó un estudio para evaluar la estabilidad y reproducibilidad de la misma tras someter las imágenes a diferentes modificaciones. A continuación, se expone el análisis llevado a cabo para relacionar la radiómica con diferentes variables analíticas y de estadificación.

Tras esto, se detalla el estudio realizado con radiómica profunda: la extracción de estos parámetros y la reducción de su dimensionalidad; para finalmente exponer el análisis llevado a cabo para relacionar la radiómica profunda con diferentes variables analíticas y de estadificación.

Seguidamente, se define cómo se llevó a cabo el entrenamiento y validación de un modelo predictivo de supervivencia global, y la comparación con otros modelos realizados con diferentes arquitecturas para su comparación. Finalmente, se expone la explicabilidad de los modelos y se detallan las variables con mayor impacto que podrían constituir biomarcadores de imagen para la predicción de supervivencia global.

4.1. Características de la población

Se incluyeron 508 pacientes con tumores neuroblásticos. La serie de pacientes fue en aumento conforme al avance del proyecto. Las edades de los pacientes estuvieron comprendidas entre los 0 (al nacimiento) y los 18 años (media de edad

de 29±34 meses), con 243 niñas y 265 niños, y con histología de neuroblastoma (n=455), ganglioneuroma (n=19) y ganglioneuroblastoma (n=34). Dado el tamaño muestral reducido de pacientes con ganglioneuroma y ganglioneuroblastoma y sus características de imagen similares al neuroblastoma, estos fueron incluidos en los análisis de imagen junto con los neuroblastomas. La denominación durante esta Tesis Doctoral se realizará indistintamente como neuroblastoma y tumor neuroblástico.

La localización del tumor fue abdominopélvica (415 casos, de ellos 250 se situaban en glándula suprarrenal, 129 abdominales y 36 con localización pélvica) o cervicotorácica (93 casos, de los que 62 eran torácicos y 31 cervicales). En la Tabla 8 se describe la población total incluida en esta Tesis Doctoral.

Tabla 8. Características basales de la población estudiada. Los datos se expresan como media ± desviación estándar cuando la distribución es normal.

Descripción de las características basales de la población estudiada		Número de pacientes
Pacientes	Número	508
Edad al diagnóstico	Media± desviación estándar (meses)	29±34
Sexo	Hombre	265
	Mujer	243
Histología	Neuroblastoma	455
	Ganglioneuroblastoma	34
	Ganglioneuroma	19
Localización	Abdominopélvica	415
	Cervicotorácica	93

Respecto a la procedencia de los pacientes en este estudio multicéntrico, como se ha indicado en el apartado 3.1. de Material y Métodos, la mayoría de los niños estaban incluidos en 2 ensayos clínicos llevados a cabo por SIOOPEN: el Estudio de Neuroblastoma de Alto Riesgo de la Sociedad Europea de Oncología Pediátrica (SIOOPEN HR-NBL1/SIOOPEN) que incluye a más de 3000 pacientes de 12 países, dirigido por el Instituto de Investigación del Cáncer Infantil St. Anna (Viena, Austria) (n=147), y el ensayo clínico del Protocolo Europeo de Neuroblastoma de Riesgo Bajo e Intermedio de SIOOPEN (LINES/SIOOPEN), dirigido por el Hospital Universitario y Politécnico La Fe (Valencia, España) (n=116). Se reclutaron también pacientes que no estaban incluidos en ninguno de los estudios clínicos mencionados anteriormente, procedentes de diferentes hospitales europeos o institutos de investigación que colaboran en el proyecto PRIMAGE. De estos, se reclutaron pacientes del Registro Alemán de Neuroblastoma NB2016 de la Sociedad de Oncología y Hematología Pediátrica, dirigido por el Departamento de Oncología Pediátrica del Hospital Universitario de Niños de Colonia (n=236), casos reclutados de hospitales españoles (Hospital La Paz y Hospital Niño Jesús, Madrid, España), Hospital Son Espases (Palma de Mallorca, España), Hospital Reina Sofía (Córdoba, España)) (n=6), y pacientes procedentes del Hospital Universitario de Pisa, Instituto Gaslini (Italia (n=3).

Las características basales de la población estudiada para cada subanálisis serán descritas en los apartados correspondientes (apartados 4.2., 4.3.2, 4.3.3., 4.4.1., 4.4.3.).

4.2. Segmentación manual: análisis de la variabilidad interobservador

Tras la preparación de las imágenes de RM descrita en el apartado 3.7. de Material y Métodos, se llevó a cabo la segmentación de los tumores y se realizó un primer análisis de la variabilidad interobservador para la segmentación manual de los tumores neuroblásticos (Figura 23).

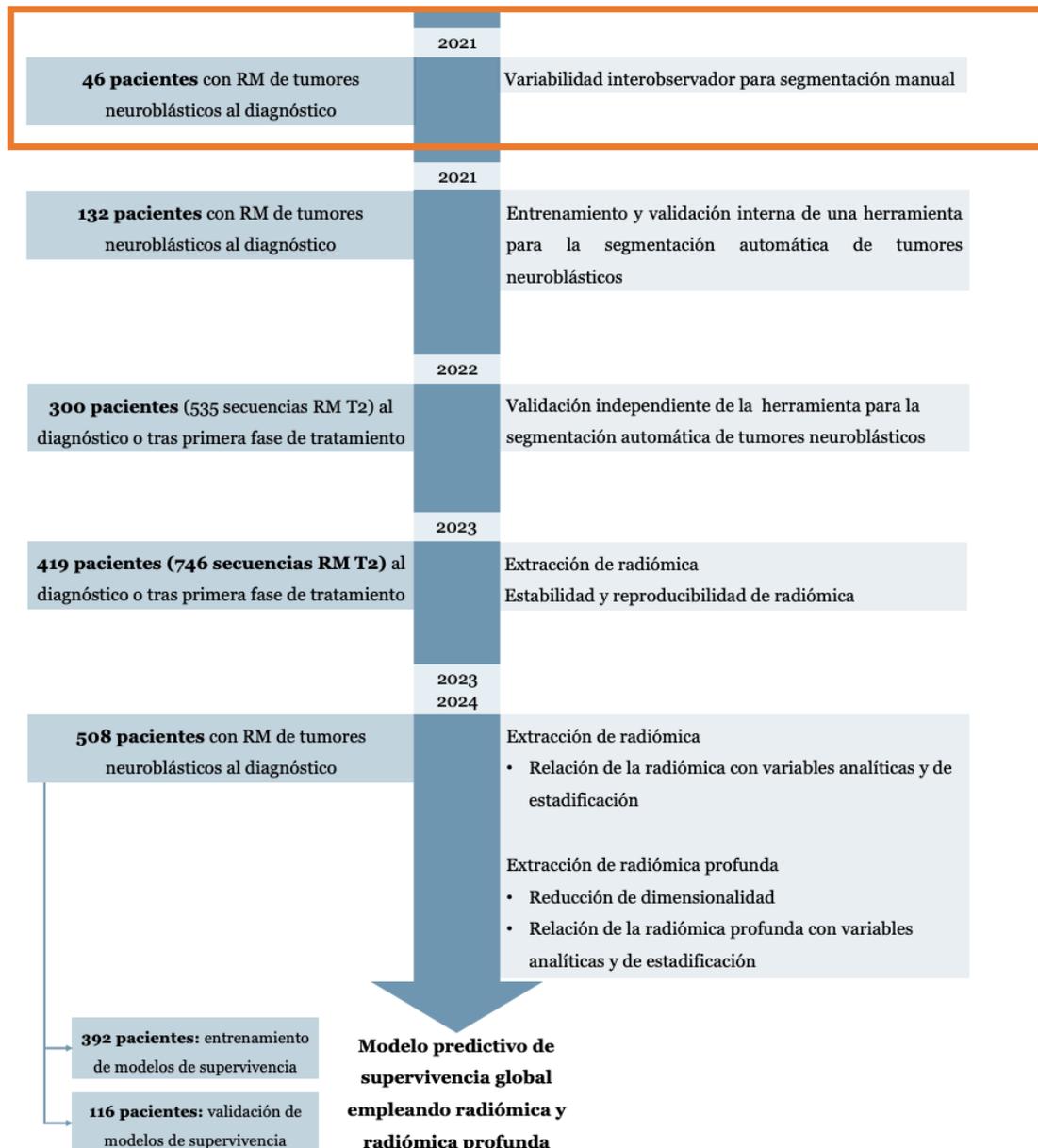


Figura 23. Flujo de trabajo: análisis de la variabilidad interobservador.

Para establecer la máxima variabilidad tolerable que cabría esperar entre dos segmentaciones diferentes, se realizó un estudio comparativo de la segmentación manual llevada a cabo por dos radiólogas (Radióloga 1, con 6 años de experiencia, y Radióloga 2, con 30 años de experiencia, ambas con experiencia previa en la segmentación manual de tumores). Ambas radiólogas llevaron a cabo de forma independiente la segmentación de los mismos 46 casos de pacientes con neuroblastoma. Las segmentaciones se realizaron de forma manual siguiendo los

criterios restrictivos de segmentación expuestos anteriormente. Se extrajeron las diversas métricas expuestas en el apartado 3.9.1., estableciendo como segmentación de referencia la segmentación manual realizada por la Radióloga 1, autora de esta Tesis Doctoral (Figura 24).

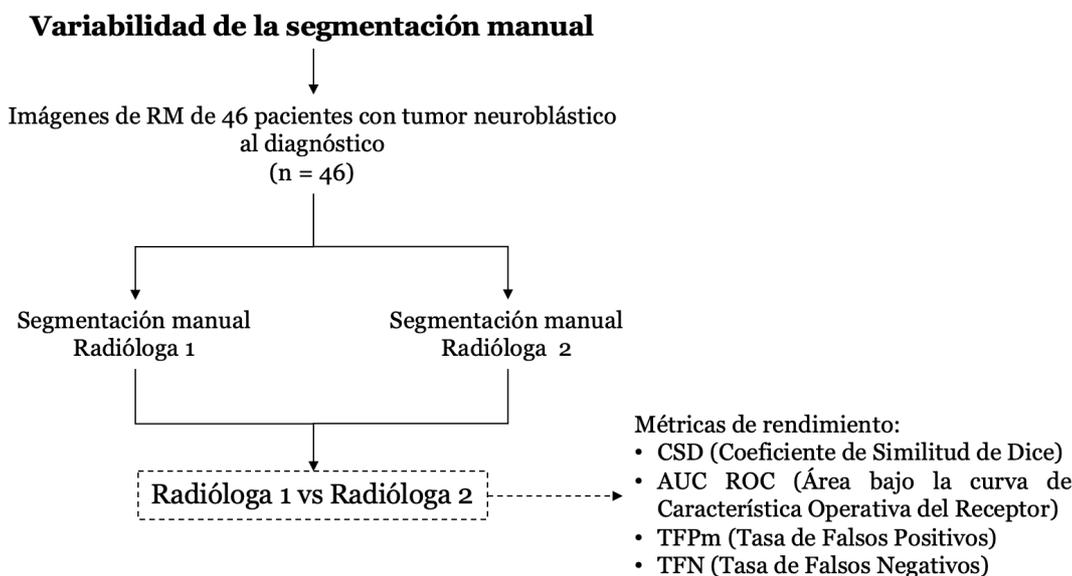


Figura 24. Flujo de comparación de variabilidad interobservador en la segmentación manual de tumores neuroblásticos.

Los 46 pacientes con tumor neuroblástico incluidos en este estudio habían sido diagnosticados o referidos al Hospital Universitari i Politècnic La Fe desde otro centro ante sospecha clínica de tumor neuroblástico para su diagnóstico y/o tratamiento. Todas las RM fueron realizadas al momento del diagnóstico inicial, sin haberse realizado ningún tratamiento ni intervención. La edad media fue de 36.5 ± 30.1 meses (media y DE), con discreto predominio de niñas (26 niñas, 20 niños). La histología tumoral en todos los casos fue de neuroblastoma. La localización se clasificó como abdominopélvica (32 casos: 20 localizados en glándulas suprarrenales, 8 de localización abdominal no adrenal y 4 de localización pélvica), o cervicotorácica (14 casos: 13 torácicos y 1 exclusivamente cervical). De estos 46 pacientes, 40 presentaban una RM adquirida en un equipo 1.5T, y 6 en una RM de 3T, siendo los proveedores General Electric Healthcare (Signa Excite HDxt, Signa Explorer) (n = 32), Siemens Medical (Aera, Avanto) (n = 10) y Philips Healthcare (Achieva) (n = 4). La segmentación del tumor se realizó

preferiblemente en las imágenes transversales potenciadas en T2wSE FS (28 casos). Se utilizaron imágenes potenciadas en T2 sin supresión grasa en los casos en los que no se encontraban disponibles las imágenes T2wSE FS (18 casos) (Tabla 9).

Tabla 9. Características basales de la muestra empleada para la comparación de segmentaciones manuales entre dos observadoras. Los datos se expresan como media \pm desviación estándar cuando la distribución es normal.

Descripción de la muestra empleada para el análisis de variabilidad interobservador		Número de pacientes
Pacientes	Número	46
Edad al diagnóstico	Media \pm desviación estándar (meses)	36.5 \pm 30.1
Sexo	Hombre	26
	Mujer	20
Histología	Neuroblastoma	46
	Ganglioneuroblastoma	0
	Ganglioneuroma	0
Localización	Abdominopélvico	32
	Cervicotorácico	14
Proveedor	GE Healthcare	32
	Siemens	10
	Philips	4
Intensidad del campo magnético	1.5T	40
	3T	6
Secuencia segmentada	T2wSE	18
	T2wSE FS	28

Los resultados de la segmentación manual obtenidos por la Radióloga 1 (segmentación de referencia) se compararon con los de la Radióloga 2 para medir la variabilidad entre observadoras. Se evidenció que la mediana del CSD era 0.969 (± 0.032 IQR), y la del JAC de 0.939 (± 0.060 IQR). La mediana de TFPm fue de 0.939 (± 0.063 IQR), lo que resultó en una alta concordancia entre ambas radiólogas en cuanto a los vóxeles no tumorales incluidos en la segmentación. La mediana de TFN fue 0.998 (± 0.008 IQR). El AUC ROC fue de 0.998 (Tabla 10) (Figura 25).

Tabla 10. Variabilidad interobservador para la segmentación manual de tumores neuroblásticos considerando diferentes métricas de rendimiento. Abreviaturas: CSD, Coeficiente de Similitud Dice; JAC, índice Jaccard; AUC ROC, Área Bajo la Curva; HD, Hausdorff Distance; 1-TFPm, Tasa de Falsos Positivos modificada; 1-TFN, Tasa de Falsos Negativos.

	CSD	JAC	AUC ROC	HD	1-TFPm	1-TFN
Mediana	0.969	0.939	0.998	1.619	0.939	0.998
IQR	0.032	0.060	0.004	2.057	0.063	0.008
CI	0.042	0.043	0.021	4.338	0.044	0.042

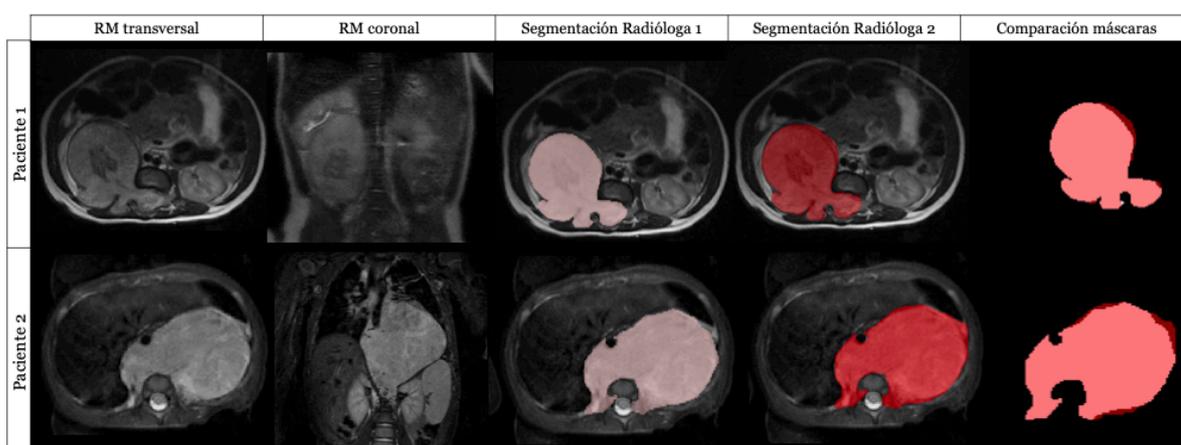


Figura 25. Comparación de dos casos segmentados por Radióloga 1 (etiqueta rosa) y Radióloga 2 (etiqueta roja) y superposición y comparación de máscaras. El caso 1 se segmentó en T2wSE, el caso 2 se segmentó en T2wSE FS. En ambos casos, el Coeficiente de Similitud de Dice fue 0.957.

4.3. Segmentación automática en RM

Se llevó a cabo un segundo estudio con el objetivo de entrenar y validar internamente una herramienta de segmentación automática (Figura 26) para facilitar la tarea de localización y delimitación tumoral.

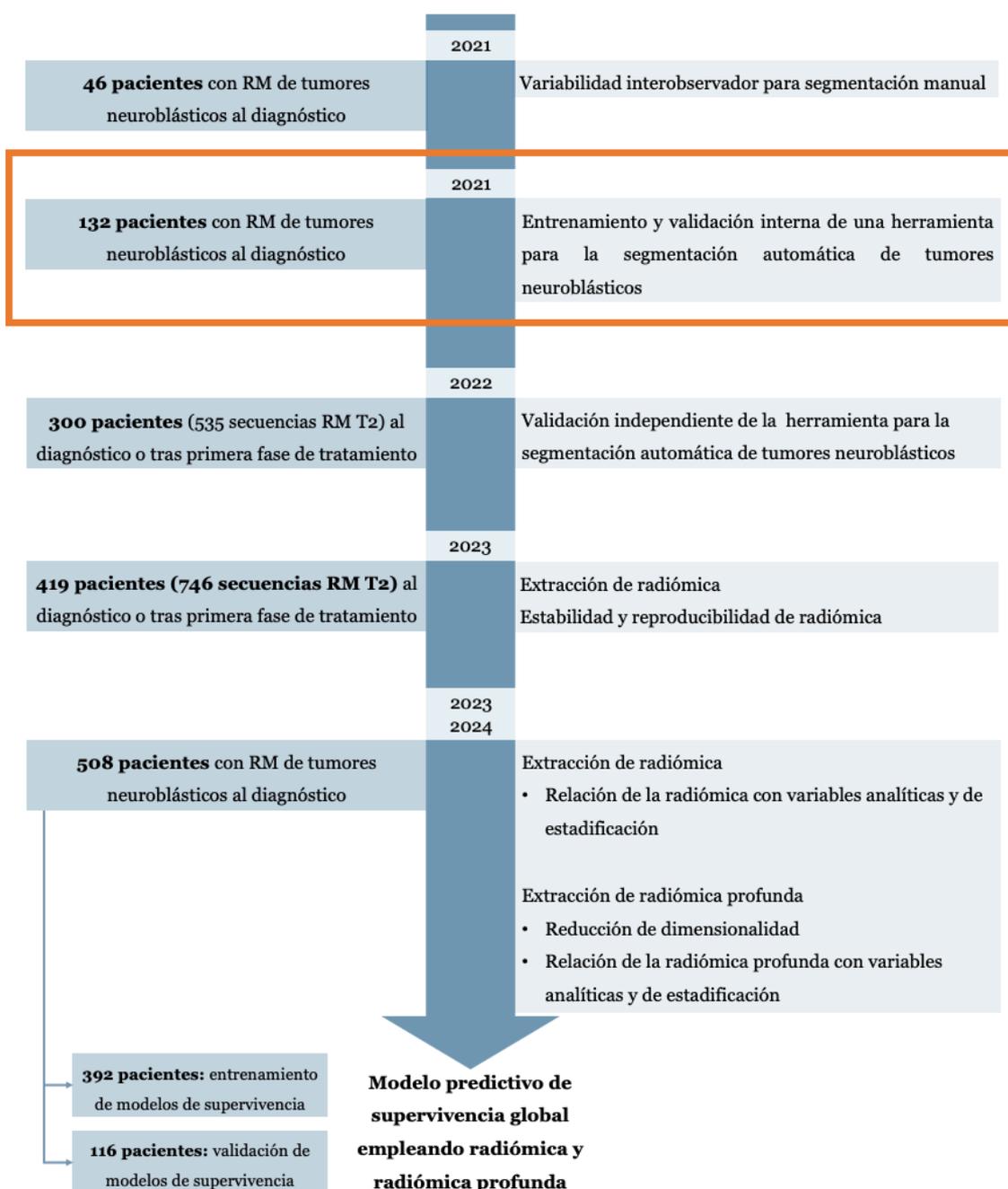


Figura 26. Flujo de trabajo: entrenamiento y validación interna de una herramienta de segmentación automática.

4.3.1. Herramienta para la segmentación: nnU-Net

El modelo de segmentación automática se desarrolló utilizando una red neuronal convolucional basada en el aprendizaje profundo para segmentación médica. Se empleó una CNN recientemente descrita, denominada nnU-Net [53], que ha constituido un nuevo estado del arte. Esta red se caracteriza por llevar a cabo un proceso de autoconfiguración (Figura 27): ante una nueva tarea de segmentación se procede a extraer las propiedades del conjunto de datos en forma de una “huella digital del conjunto de datos”. Una serie de reglas heurísticas modela las interdependencias entre los parámetros y opera en esta huella digital para deducir los “parámetros basados en reglas” que están condicionados por los datos del proceso. Estos se complementan con los “parámetros fijos”, que están predefinidos y no requieren adaptación. Se lleva a cabo el entrenamiento de hasta tres configuraciones en una validación cruzada de cinco grupos. Finalmente, nnU-Net realiza de manera automática la selección empírica del conjunto óptimo de estos modelos y determina si es necesario aplicar un postprocesamiento (“parámetros empíricos”).

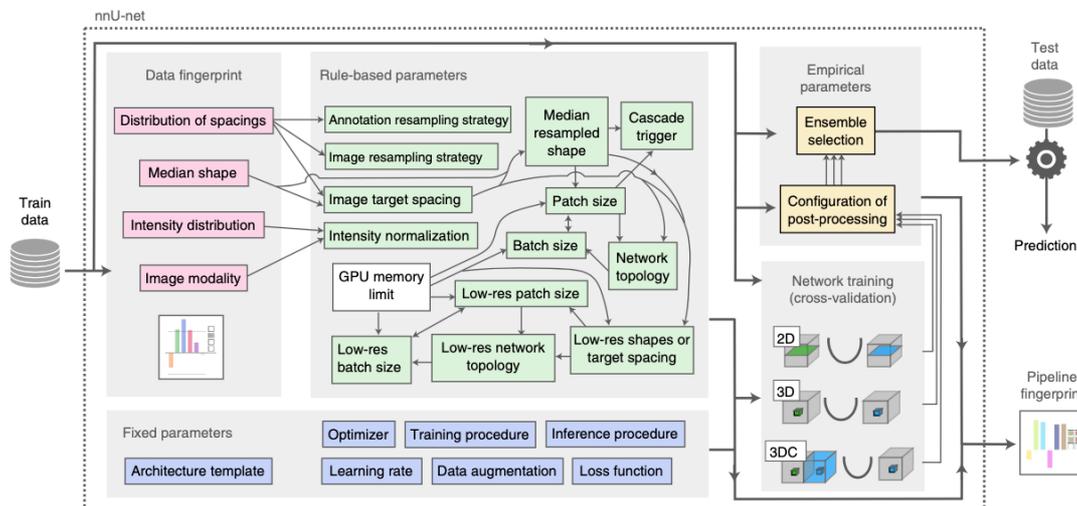


Figura 27. Configuración del método automatizado nnU-Net para la segmentación de imágenes biomédicas basada en aprendizaje profundo. Huella digital del conjunto de datos (en color rosa), reglas heurísticas que modelan las dependencias entre parámetros (flechas), parámetros basados en reglas (en color verde), parámetros fijos (en color azul), parámetros empíricos (en amarillo). Extraído del estudio de Isensee [53].

La arquitectura de la red empleada fue en 3D, y el modelo fue entrenado con una estrategia de validación cruzada, que es una técnica estadística comúnmente utilizada en aprendizaje automático para estimar el rendimiento de un modelo predictivo y que consiste en dividir los datos en varios conjuntos, entrenar el modelo en algunos de ellos y posteriormente evaluar su rendimiento en el conjunto restante. Esto se repite varias veces, cambiando qué datos se utilizan para entrenar y cuáles para evaluar, lo que permite una estimación más precisa del rendimiento del modelo en datos no vistos [113].

Todas las imágenes fueron remuestreadas a un nuevo espaciado de píxeles: $[z, x, y] = [8, 0.695, 0.695]$, correspondiente a los valores promedio dentro del conjunto de datos de entrenamiento.

En el contexto del aprendizaje automático y CNN, los datos de entrenamiento se dividen en diferentes subconjuntos (lote o *batch*). Una iteración es una actualización de los parámetros del modelo durante el proceso de entrenamiento para un lote de los datos de entrenamiento. Durante una iteración, el modelo realiza predicciones sobre los datos, calcula el error y actualiza los parámetros del modelo utilizando un algoritmo de optimización. Una vez que todos los lotes de datos de entrenamiento se han procesado, decimos que ha finalizado un *epoch*. El número total de iteraciones necesarias para completar un *epoch* depende del tamaño del lote (*batch size*) y del tamaño del conjunto de datos de entrenamiento. Los *epoch* son una medida del número de veces que el algoritmo de entrenamiento ha pasado por todo el conjunto de datos de entrenamiento. En cada *epoch*, el algoritmo ajusta los pesos de la CNN en función de los errores de predicción con el objetivo de mejorar el rendimiento del modelo.

El entrenamiento del modelo desarrollado en esta Tesis Doctoral se realizó a lo largo de 1000 epochs con 250 iteraciones cada una y un tamaño de lote de 2.

La función de pérdida (el cálculo de la discrepancia entre las predicciones del modelo y los valores reales del conjunto de datos para optimizar cada iteración) se basó en el CSD. En la Figura 28 se representa el entrenamiento de cada conjunto de datos, representando el eje X el número de *epochs* a través del conjunto de datos de entrenamiento, y el eje Y las pérdidas. La curva azul

representa la función de pérdida de entrenamiento, mide la discrepancia entre las predicciones de la red y las etiquetas reales del conjunto de datos de entrenamiento (idealmente, esta curva debería descender a medida que la red aprende de los datos de entrenamiento, minimizándose la discrepancia). La curva roja representa la función de pérdida de test, la discrepancia en un conjunto de datos de prueba independiente que la red nunca ha visto durante el entrenamiento. Esta curva es fundamental para evaluar la capacidad de generalización del modelo: si aumenta mientras la curva de entrenamiento sigue disminuyendo, podría indicar sobreajuste. La curva verde representa el CSD, el aumento en esta curva indica una mejora en la precisión de la segmentación.

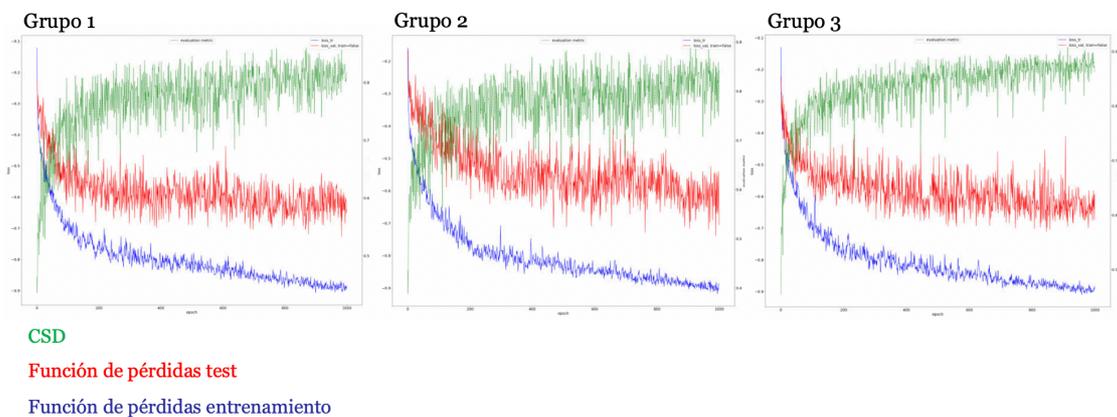


Figura 28. Entrenamiento de la red neuronal convolucional en 3 de los subconjuntos empleados para el entrenamiento. Eje X: número de iteraciones completas en el conjunto de datos de entrenamiento. Eje Y: pérdidas. Curva azul: función de pérdida de entrenamiento. Curva roja: función de pérdida de test. Curva verde: Coeficiente de Similitud de Dice.

Las curvas obtenidas en el entrenamiento de la red para segmentación de neuroblastoma indicaron convergencia: la función de pérdida de entrenamiento y la función de pérdida de test disminuyeron y se estabilizaron, indicando que la red aprendió bien y generalizó a nuevos datos. La métrica CSD aumentó, por lo que la red mejoró sus resultados en la tarea de segmentación.

4.3.2. Desarrollo de una herramienta de detección y segmentación automática

Se llevó a cabo un estudio con el objetivo de entrenar y validar internamente una herramienta de segmentación automática, con la hipótesis de que la arquitectura nnU-Net puede utilizarse para detectar y segmentar automáticamente tumores neuroblásticos en imágenes de RM, proporcionando una solución más robusta, universal y libre de errores que la obtenida mediante el proceso de segmentación manual.

Se realizó un estudio retrospectivo multicéntrico e internacional de 132 pacientes pediátricos con tumores neuroblásticos que habían sido sometidos a una RM al diagnóstico (incluyendo las segmentaciones manuales llevadas a cabo por la Radióloga 1 de los 46 pacientes descritos anteriormente para el cálculo de la variabilidad interobservador en segmentación manual). Los pacientes y los datos de RM se obtuvieron retrospectivamente de tres centros, siendo estos de España (n = 73, Hospital Universitari i Politènic La Fe, incluyendo 21 casos del ensayo clínico del Protocolo Europeo de Neuroblastoma de Riesgo Bajo e Intermedio (LINES/SIOPEN)), Austria (n = 57, Instituto de Investigación del Cáncer Infantil del Estudio de Neuroblastoma de Alto Riesgo SIOPEN (HR-NBL1/SIOPEN)) e Italia (n = 4, Hospital Universitario de Pisa).

La edad al diagnóstico fue de $37,6 \pm 39,3$ meses (media \pm DE, rango de 0 a 252 meses, mediana de 24,5 meses \pm IQR de 54), con discreto predominio de mujeres (70 mujeres, 62 hombres). La histología del tumor fue neuroblastoma (104 casos), ganglioneuroblastoma (18 casos) y ganglioneuroma (10 casos). La ubicación del tumor se clasificó como abdominopélvica (105 casos, 59 de ellos en la glándula suprarrenal, 32 con ubicación abdominal no suprarrenal y 14 con ubicación pélvica) o cervicotorácica (27 casos, 18 de ellos torácicos, 2 con una ubicación cervical exclusiva y 7 afectando tanto la región torácica como cervical).

Las imágenes de RM presentaban una alta variabilidad en la adquisición de datos, incluyendo diferentes escáneres, proveedores y protocolos, al tratarse de estudios

provenientes de distintas instituciones. Las imágenes de RM se adquirieron con un escáner de 1.5T (n = 116) o 3T (n = 16), fabricados por General Electric Healthcare (Signa Excite HDxt, Signa Explorer) (n = 51), Siemens Medical (Aera, Skyra, Symphony, Avanto) (n = 54) o Philips Healthcare (Intera, Achieva, Ingenia) (n = 27).

Todos los tumores (132 casos) fueron segmentados manualmente por la Radióloga 1, que fue considerada la verdad fundamental o segmentación de referencia. La segmentación del tumor se realizó en las imágenes transversales potenciadas en T2wSE FS (48 casos). Se utilizaron imágenes potenciadas en T2wSE cuando no estaban disponibles las imágenes T2wSE FS (84 casos).

Las características basales de la muestra empleada para el desarrollo de la herramienta de detección y segmentación automática se describen en la Tabla 11.

Tabla 11. Características basales de la muestra empleada para el desarrollo de la herramienta de detección y segmentación automática. Los datos se expresan como media \pm desviación estándar cuando la distribución es normal.

Descripción de la muestra empleada para el desarrollo de la herramienta de detección y segmentación automática		Número de pacientes
Pacientes	Número	132
Edad al diagnóstico	Media \pm desviación estándar (meses)	37,6 \pm 39,3
Sexo	Hombre	62
	Mujer	70
Histología	Neuroblastoma	104
	Ganglioneuroblastoma	18
	Ganglioneuroma	10
Localización	Abdominopélvico	105

	Cervicotorácico	27
Proveedor	GE Healthcare	51
	Siemens	54
	Philips	27
Intensidad del campo magnético	1.5T	116
	3T	16
Secuencia segmentada	T2wSE	84
	T2wSE FS	48

4.3.2.1. Distribución de pacientes para el desarrollo de una herramienta de segmentación automática

Se llevó a cabo el entrenamiento y validación interna del modelo de segmentación automática basado en la arquitectura nnU-Net a partir de las segmentaciones manuales de referencia, dividiendo el conjunto de datos en dos cohortes. Se seleccionó una primera cohorte (80% de los casos, $n = 106$) para entrenar y ajustar el modelo mediante un enfoque de validación cruzada de 5 grupos. Una segunda cohorte (20% de los pacientes, $n = 26$) se utilizó para la validación interna.

La cohorte de datos de entrenamiento-ajuste ($n = 106$) se dividió en 5 subconjuntos de 21 o 22 casos no superpuestos. Con el fin de construir un modelo reproducible y universal, y para eliminar el sesgo de muestreo y garantizar la heterogeneidad de todos los conjuntos de datos, se implementó una división equilibrada y estratificada de los casos de cada una de las cohortes tanto en los subconjuntos de entrenamiento como en el de validación interna. Se utilizó un muestreo estratificado empleando la biblioteca *Scikit-learn* [114], para elaborar grupos heterogéneos considerando cuatro variables: proveedor (Siemens/Philips/GE), intensidad de campo magnético (1.5 T/3 T), ubicación del tumor (abdominopélvica/cervicotorácica) y secuencia segmentada (T2wSE/T2wSE FS) (Tabla 12).

Tabla 12. División equilibrada y estratificada de los casos para cada uno de los subconjuntos para el entrenamiento y validación interna de la herramienta de

segmentación automática. Se emplea como ejemplo la composición del conjunto de datos de validación interna (20% de los casos, n = 26), considerando cuatro variables para garantizar la heterogeneidad de los subconjuntos: secuencia segmentada, proveedor del equipamiento, intensidad del campo magnético y localización tumoral.

Grupo de validación interna (n=26)			
Secuencia segmentada	Proveedor del equipamiento	Intensidad del campo magnético	Localización del tumor
T2wSE	Philips	1.5	Abdominopélvico
T2wSE	Siemens	1.5	Abdominopélvico
T2wSE	Philips	1.5	Abdominopélvico
T2wSE	GE	1.5	Abdominopélvico
T2wSE	GE	1.5	Cervicotorácico
T2wSE	Philips	1.5	Abdominopélvico
T2wSE	Siemens	1.5	Abdominopélvico
T2wSE	Philips	1.5	Cervicotorácico
T2wSE	Siemens	1.5	Abdominopélvico
T2wSE FS	GE	1.5	Abdominopélvico
T2wSE	Siemens	3	Abdominopélvico
T2wSE	GE	1.5	Abdominopélvico
T2wSE	GE	1.5	Cervicotorácico
T2wSE	Philips	1.5	Abdominopélvico
T2wSE FS	Philips	1.5	Abdominopélvico
T2wSE	Siemens	3	Abdominopélvico
T2wSE	Siemens	1.5	Abdominopélvico
T2wSE FS	Siemens	1.5	Abdominopélvico
T2wSE FS	GE	1.5	Cervicotorácico
T2wSE FS	GE	1.5	Abdominopélvico
T2wSE FS	GE	1.5	Abdominopélvico
T2wSE FS	Siemens	1.5	Abdominopélvico
T2wSE	GE	1.5	Cervicotorácico
T2wSE FS	Siemens	1.5	Abdominopélvico
T2wSE FS	Siemens	3	Abdominopélvico
T2wSE FS	GE	1.5	Cervicotorácico

Cada uno de los 5 subconjuntos de entrenamiento-ajuste ($n = 21$ o 22 casos no superpuestos) se empleó como set de ajuste (conjunto retenido), mientras que los demás subconjuntos se utilizaron colectivamente como un conjunto de datos de entrenamiento. Por ello, se ajustaron y evaluaron los 5 conjuntos de datos, obteniéndose así 5 modelos de segmentación.

Los 5 modelos de segmentación resultantes obtenidos mediante el método de validación cruzada se utilizaron como una solución de conjunto que se aplicó a todos los casos, tanto a los del conjunto de datos de entrenamiento-ajuste ($n = 106$), como a los del conjunto de datos de validación ($n = 26$) para medir el rendimiento de entrenamiento y validación de manera independiente. La Figura 29 representa la distribución de pacientes y los procesos realizados para el entrenamiento y validación interna de la herramienta de segmentación automática.

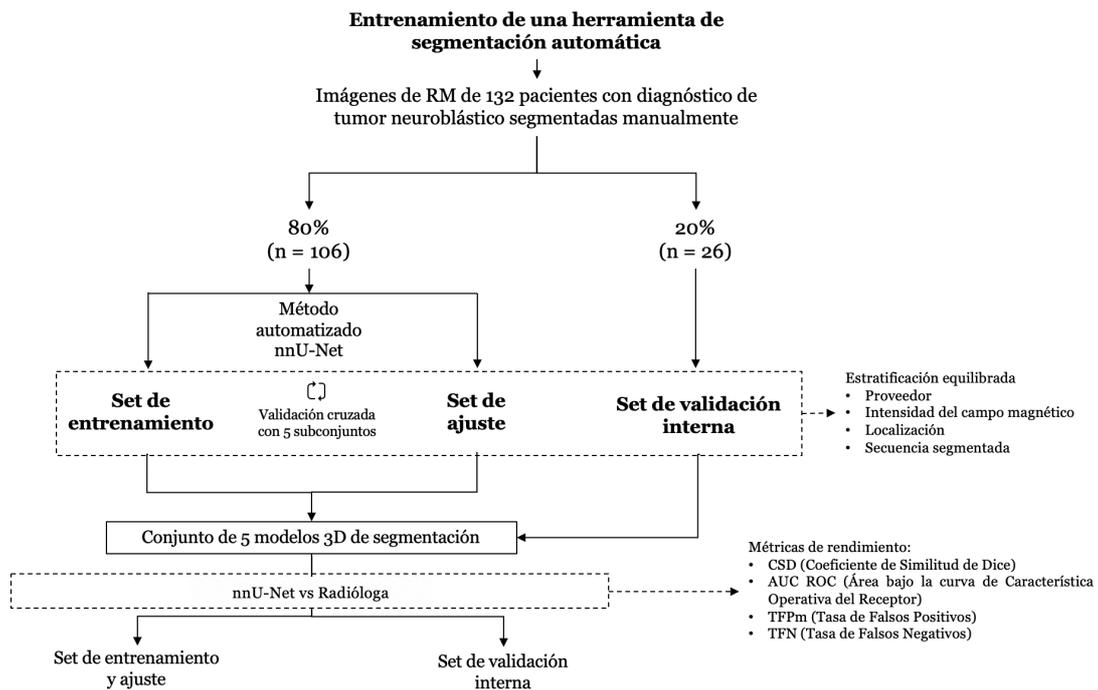


Figura 29. Diagrama del diseño del estudio para el entrenamiento con validación cruzada y ajuste del modelo de segmentación basado en nnU-Net utilizando 132 casos segmentados manualmente.

4.3.2.2. Entrenamiento del modelo de segmentación automática

Los resultados de este primer estudio se reflejan en un artículo publicado en 2022 (Veiga-Canuto et al. Cancers 2022) [115].

El volumen del tumor se obtuvo a partir de las 132 máscaras realizadas por la Radióloga 1. El volumen mediano de todas las máscaras fue de 116,518 mm³ ($\pm 219,084$ IQR).

Como los 106 casos de la cohorte de entrenamiento se dividieron en cinco subconjuntos de 21 o 22 casos para realizar una validación cruzada, cada subconjunto logró diferentes resultados en las métricas de rendimiento, con una mediana de CSD de entre 0.873 (subconjunto 1) y 0.901 (subconjunto 3) (Tabla 13).

Tabla 13. Métricas de rendimiento para los 5 subconjuntos de entrenamiento mediante validación cruzada. Abreviaturas: CSD, Coeficiente de Similitud Dice; JAC, índice Jaccard; AUC ROC, Área Bajo la Curva; HD, Hausdorff Distance; 1-TFPm, Tasa de Falsos Positivos modificada; 1-TFN, Tasa de Falsos Negativos.

Sub conjunto	Métrica	CSD	JAC	AUC ROC	HD	1-TFPm	1-TFN
1	Mediana	0.895	0.809	0.940	3.966	0.922	0.882
	IQR	0.121	0.186	0.116	3.773	0.082	0.233
	IC	0.146	0.147	0.117	9.984	0.074	0.148
2	Mediana	0.873	0.775	0.926	5.539	0.944	0.856
	IQR	0.110	0.168	0.100	6.169	0.100	0.100
	IC	0.127	0.122	0.066	8.982	0.088	0.132
3	Mediana	0.899	0.817	0.936	5.823	0.935	0.875
	IQR	0.131	0.201	0.064	9,347	0.133	0.133
	IC	0.123	0.121	0.062	9.124	0.125	0.124
4	Mediana	0.901	0.821	0.948	4.831	0.949	0.897
	IQR	0.122	0.193	0.062	4.020	0.088	0.124
	IC	0.046	0.066	0.030	1,506	0.090	0.061
5	Mediana	0.874	0.777	0.927	5.315	0.958	0.856
	IQR	0.134	0.208	0.110	5.709	0.033	0.221
	IC	0.141	0.133	0.071	3.495	0.032	0.142

En la Figura 30 se muestran ejemplos de segmentaciones realizadas automáticamente por nnU-Net y las segmentaciones de los mismos casos realizadas manualmente por la Radióloga 1 (y empleadas para el entrenamiento de la red), con superposición de ambas máscaras para su comparación.

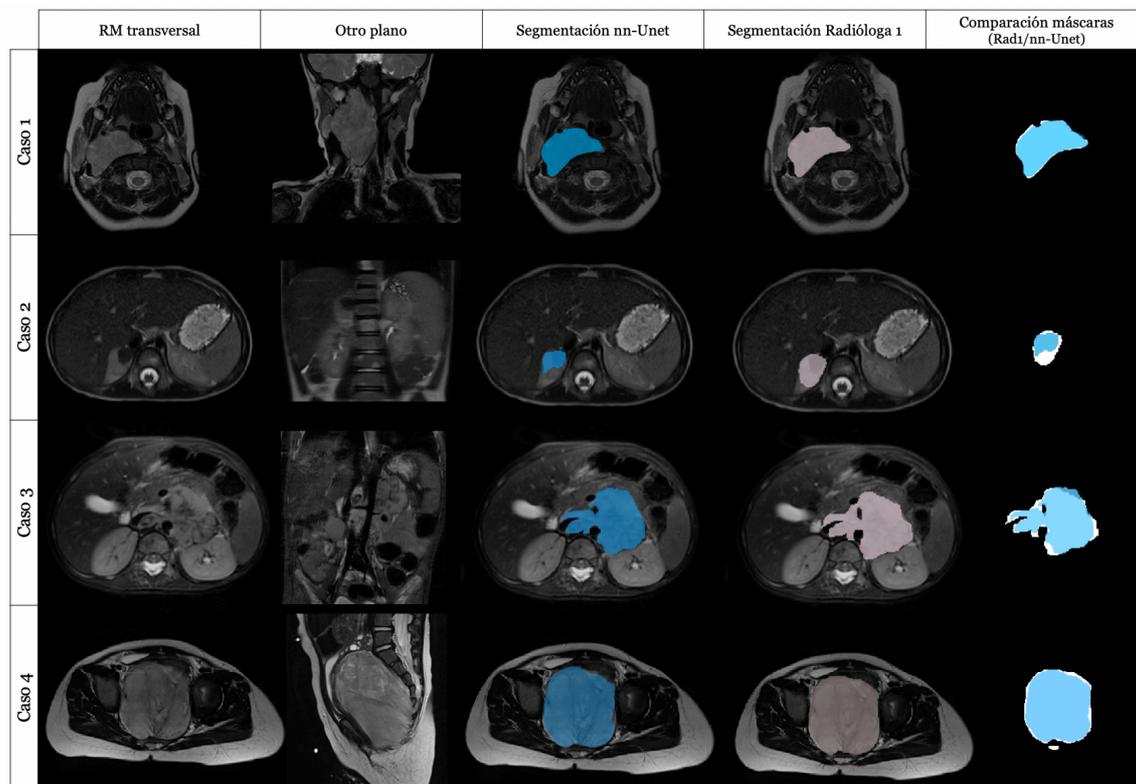


Figura 30. Ejemplos de segmentaciones realizadas automáticamente por nnU-Net (en azul) y comparación con segmentaciones realizadas manualmente por la Radióloga 1 (en rosa). Imágenes originales de RM transversal, plano coronal (casos 1 a 3), y sagital (caso 4). Caso 1: neuroblastoma cervical derecho, secuencia T2wSE; Coeficiente de Similitud de Dice de 0.954. Caso 2: neuroblastoma suprarenal derecho, secuencia T2wSE; Coeficiente de Similitud de Dice de 0.617. Caso 3: neuroblastoma abdominal, secuencia T2wSE FS; Coeficiente de Similitud de Dice de 0.869. Caso 4: neuroblastoma pélvico, secuencia T2wSE; Coeficiente de Similitud de Dice 0.949.

De los 106 casos, 79 obtuvieron un CSD ≥ 0.8 , mientras que 27 tuvieron un valor de CSD < 0.8 . De estos 27 casos, 9 tuvieron un CSD de 0 a 0.19, siendo en todos los casos < 0.01 ; 4 casos tuvieron un CSD ≥ 0.2 a 0.39; 2 casos tuvieron un CSD ≥ 0.4 a 0.59; y 12 casos tuvieron un CSD ≥ 0.6 a 0.8 (Figura 31).

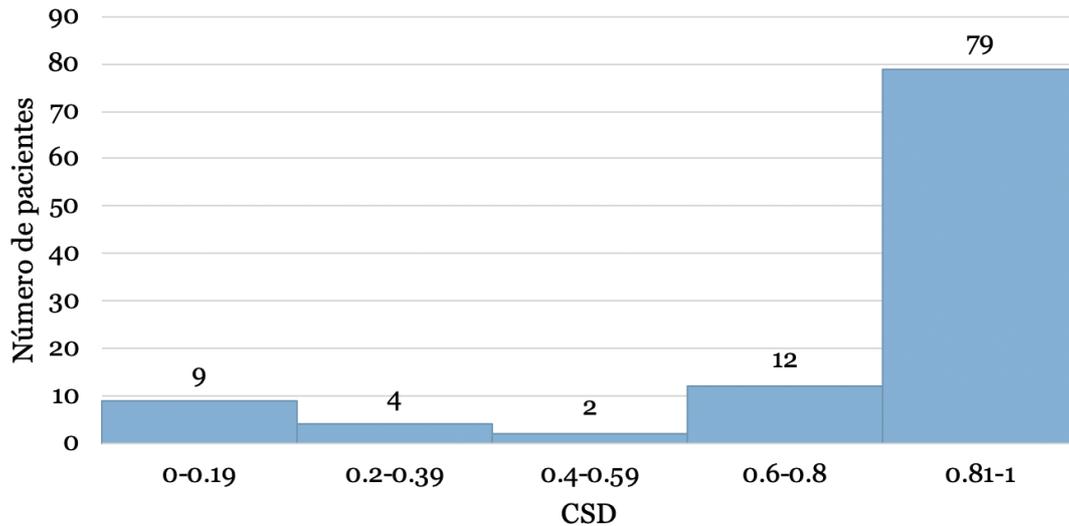


Figura 31. Distribución de los 106 pacientes del grupo de entrenamiento en función del valor de Coeficiente de Similitud de Dice.

Los casos con un CSD <0.8 tuvieron una edad media de 32.7 ± 30.3 meses y mediana de 19.8 meses. Mostraron un volumen mediano de 75.733 mm^3 (± 42.882 IQR), siendo este menor al volumen mediano de todas las máscaras. Se observó la siguiente distribución de casos con CSD <0.8 entre subconjuntos de entrenamiento: los subconjuntos 1, 3 y 5 tuvieron 5 casos cada uno, mientras que los subconjuntos 2 y 4 tuvieron 6 casos cada uno. Estas máscaras fueron analizadas por la Radióloga 1 para identificar las razones de la baja concordancia. Con respecto a los 9 casos con CSD <0.01 , la red había segmentado ganglios linfáticos extensos en lugar del tumor primario en 4 casos (4% del total de casos del grupo de entrenamiento). En otros 2 casos, la red segmentó otras estructuras en lugar del tumor (2% de casos totales, segmentando la vesícula biliar o el riñón izquierdo, Figura 32). En otros 3 casos (3% del total de casos), la red no identificó como tumor ninguna estructura y, por lo tanto, no realizó ninguna máscara.

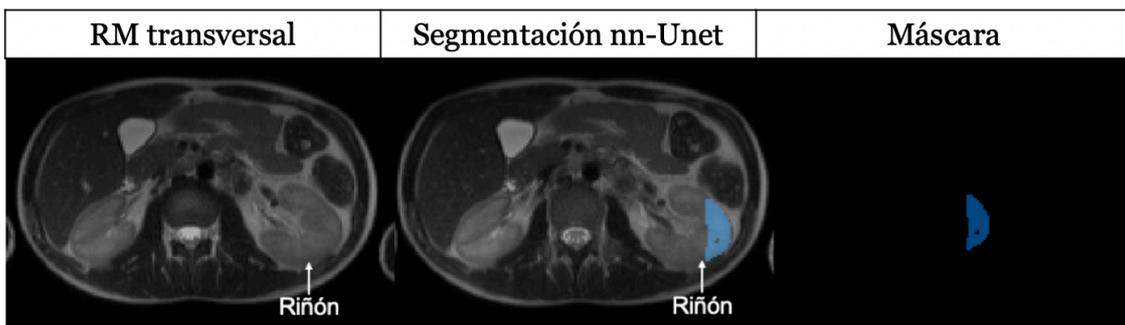


Figura 32. Ejemplo de segmentación con Coeficiente de Similitud de Dice <0.01 , en el que la red segmentó el riñón izquierdo en lugar del tumor.

De los 18 casos restantes con un CSD <0.8 , en 17 casos la red localizó bien el tumor, pero no lo segmentó completamente o presentó variabilidad en los bordes, especialmente en los casos con ganglios linfáticos circundantes. Un caso presentaba tumores bilaterales y la red solo detectó uno de ellos.

Posteriormente, los cinco modelos de segmentación resultantes obtenidos mediante el método de validación cruzada se utilizaron como solución de conjunto para probar todos los casos del entrenamiento-ajuste ($n = 106$). Se obtuvo una mediana de CSD de $0.965 (\pm 0.018 \text{ IQR})$ y una mediana de AUC ROC de $0.981 (\pm 0.010 \text{ IQR})$. Al observar la dirección de los errores entre ambos conjuntos (verdad fundamental frente a segmentación automática), la TFPm mediana fue $0.968 (\pm 0.015 \text{ IQR})$, de modo que la máscara automática incluyó como tumor el 3.2% de los vóxeles que no habían sido incluidos en la segmentación manual de referencia. La TFN mediana fue de $0.963 (\pm 0.021 \text{ IQR})$, por lo que la herramienta automática no incluyó el 3.7% de los vóxeles incluidos en la máscara de la segmentación de referencia. Los resultados se estratificaron en función de su ubicación (abdominopélvica o cervicotorácica) y la intensidad del campo magnético (1,5 T o 3 T) (Tabla 14).

Tabla 14. Solución de conjunto del modelo de segmentación a partir de los 5 modelos de segmentación resultantes obtenidos mediante el método de validación cruzada, aplicada a todos los casos de entrenamiento-ajuste. Se describen las métricas de desempeño para los resultados finales. Los resultados se detallan según la ubicación (abdominopélvica o cervicotorácica) y la intensidad del campo magnético (1,5 T o 3 T).

	CSD	JAC	AUC ROC	HD	1-TFPm	1-TFN
Mediana	0.965	0.931	0.981	1.062	0.968	0.963
IQR	0.018	0.033	0.010	0.673	0.015	0.021
IC	0.031	0.031	0.015	1.762	0.025	0.031

Cervicotorácico (n = 21)						
Mediana	0.956	9.159	0.975	1.190	0.962	0.950
IQR	0.024	0.044	0.012	0.758	0.015	0.024
IC	0.036	0.053	0.018	1.960	0.037	0.036
Abdominopélvico (n = 85)						
Mediana	0.966	9.339	0.982	1.024	0.969	0.645
IQR	0.015	0.029	0.009	0.658	0.014	0.019
IC	0.037	0.037	0.018	2.148	0.030	0.038
1.5 T (n = 93)						
Mediana	0.965	0.932	0.981	1.059	0.969	0.963
IQR	0.018	0.033	0.011	0.709	0.016	0.021
IC	0.029	0.021	0.014	1.672	0.021	0.029
3 T (n = 13)						
Mediana	0.964	0.930	0.982	1.066	0.967	0.964
IQR	0.013	0.025	0.005	0.383	0.007	0.010
IC	0.145	0.141	0.073	8.149	0.138	0.145

Para comparar las medias de CSD atendiendo a los efectos de la ubicación (abdominopélvica o cervicotorácica) y la fuerza del campo magnético (1.5 o 3 T), se realizó una prueba de ANOVA sobre la variable respuesta CSD considerando como factores categóricos las variables localización (abdominopélvico sí/no) y campo magnético (1.5/3), y como covariables la edad y volumen tumoral. No se observó un efecto significativo en el CSD de ninguna de las variables. Al introducir la edad y el volumen como factores correctivos en la evaluación de los efectos de la ubicación y el campo magnético en el CSD se evidenció que la edad y el volumen no presentaron un efecto significativo y no mostraron ninguna tendencia en el CSD (valor $p = 0.052$ para la edad y 0.169 para el volumen). Por lo tanto, los efectos de la ubicación y el campo magnético y su interacción siguen siendo no significativos cuando se introduce la corrección por edad y volumen (Tabla 15).

Tabla 15. Modelo lineal generalizado sobre la variable respuesta Coeficiente de Similitud Dice (CSD) considerando como factores categóricos las variables localización (abdominopélvico sí/no) y campo magnético (1.5/3) y como covariables la edad y volumen tumoral. La edad y el volumen no presentaron un efecto significativo y no mostraron ninguna tendencia en el CSD.

Variable dependiente: CSD final		
Variable	F	P valor
Modelo corregido	1.613	0.163
Intersección	641.238	0.000
Localización	1.549	0.216
Campo magnético	0.141	0.708
Localización x Campo magnético	1.469	0.228
Edad	3.874	0.052
Volumen	1.923	0.168

4.3.2.3. Validación interna del modelo de segmentación automática

Como se ha expuesto previamente, se llevó a cabo una validación interna con 26 casos para tras el entrenamiento del modelo de segmentación automática, con la finalidad de valorar la presencia de sobreajuste, lo que podría resultar en una sobreestimación del rendimiento del modelo. El resultado de la mediana del CSD para la validación al comparar la segmentación automática con la llevada a cabo por la Radióloga 1 fue de 0.918 (± 0.067 IQR) (Tabla 16).

Para visualizar de manera gráfica la distribución del CSD en los diferentes subconjuntos, se obtuvieron diagramas de cajas y bigotes del CSD de cada uno de los subconjuntos de entrenamiento y el conjunto de validación interna (Figura 33, los resultados del CSD representados se describen en la Tabla 13 y en la Tabla 16).

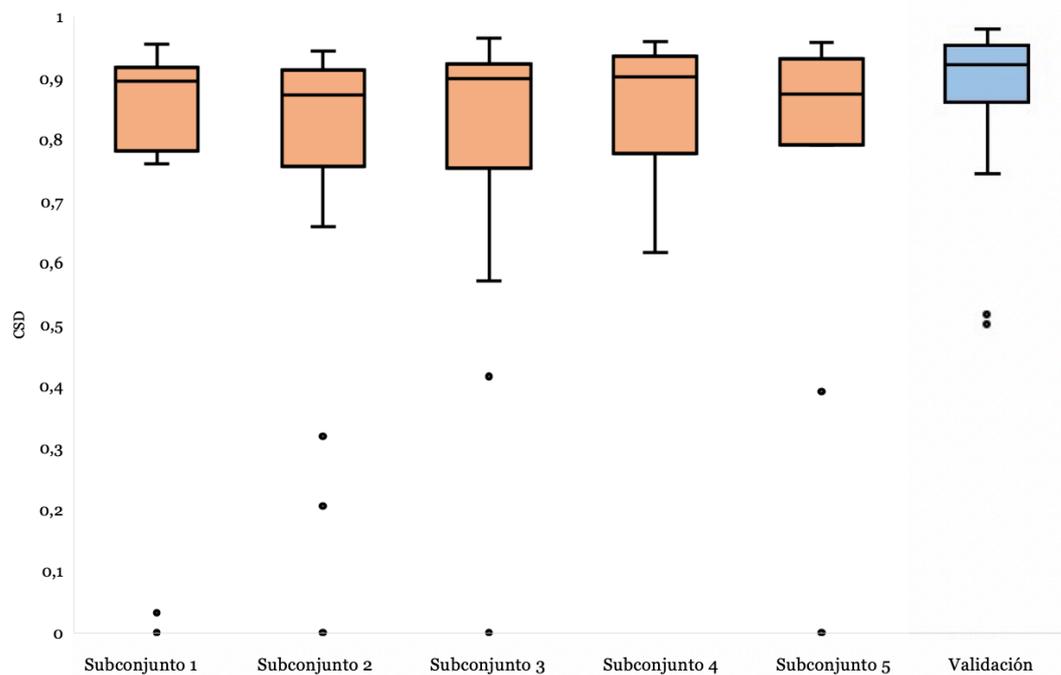


Figura 33. Diagramas de cajas y bigotes representando el Coeficiente de Similitud de Dice de cada uno de los subconjuntos de entrenamiento (naranja) y del conjunto de validación (azul). Los resultados del CSD representados se describen en la Tabla 13 y en la Tabla 16.

De los 26 casos en el conjunto de datos de validación, 4 (15%) tuvieron un valor de CSD < 0.8 : 3 casos tuvieron un CSD ≥ 0.4 a 0.59 ; y 1 caso tuvo un CSD ≥ 0.6 a 0.8 . Estos casos fueron analizados por la Radióloga 1 para identificar las razones del bajo nivel de acuerdo. Respecto a los 3 casos con CSD < 0.6 , la red había segmentado ganglios linfáticos extensos además del tumor primario en 2 casos, e identificado solo una parte del tumor en 1 caso. En el caso con un CSD ≥ 0.6 a 0.8 , la red segmentó ganglios linfáticos además del tumor primario.

Además de comparar los resultados obtenidos mediante la herramienta de segmentación automática para la cohorte de validación con las segmentaciones manuales realizadas por la Radióloga 1, se evaluó el acuerdo entre radiólogas y el acuerdo entre la herramienta automática y la Radióloga 2. Para ello, la Radióloga 2 segmentó manualmente los 26 casos del conjunto de datos de validación. Se compararon la segmentación de la Radióloga 2 con las segmentaciones de la

Radióloga 1 y del modelo automático (Tabla 16), obteniendo una mediana de DSC entre 0.915 y 0.920.

Tabla 16. Métricas de rendimiento para el conjunto de validación interna del modelo de segmentación automática. Se muestran los resultados de la comparación de la Radióloga 1 frente al modelo automático y a la Radióloga 2, y de la Radióloga 2 frente al modelo automático. Abreviaturas: CSD, Coeficiente de Similitud Dice; JAC, índice Jaccard; AUC ROC, Área Bajo la Curva; HD, Hausdorff Distance; 1-TFPm, Tasa de Falsos Positivos modificada; 1-TFN, Tasa de Falsos Negativos.

	CSD	JAC	AUC ROC	HD	1-TFPm	1-TFN
Radióloga 1 vs. Modelo automático						
Mediana	0.918	0.848	0.968	3.490	0.943	0.938
IQR	0.080	0.135	0.051	7.756	0.088	0.104
IC	0.059	0.074	0.473	3.006	0.134	0.063
Radióloga 1 vs. Radióloga 2						
Mediana	0.920	0.686	0.950	7.104	0.929	0.930
IQR	0.090	0.524	0.192	14.864	0.015	0.024
IC	0.038	0.116	0.053	5.709	0.166	0.058
Radióloga 2 vs. Modelo automático						
Mediana	0.915	0.972	0.950	2.626	0.915	0.912
IQR	0.443	0.517	0.122	17.377	0.436	0.189
IC	0.114	0.134	0.054	6.443	0.161	0.104

Impacto en el ahorro de tiempo

Para evaluar el ahorro de tiempo que supone la herramienta automática y poder valorar su impacto clínico, se realizó una comparación del tiempo medio necesario para segmentar manualmente 20 casos (418 cortes) desde el principio, con el tiempo medio requerido por el modelo automático para segmentarlos. Los casos segmentados manualmente requirieron un tiempo medio de 56 minutos por caso, mientras que el tiempo medio necesario para obtener cada máscara con el modelo automático fue de 10 segundos (0.167 minutos), lo que resultó en una reducción de tiempo del 99.7%. Dado que puede existir cierta variabilidad en las máscaras automáticas finales, se llevó a cabo una validación visual de las

máscaras. Todas las segmentaciones fueron validadas visualmente, y se realizaron ediciones y ajustes manuales de las máscaras automáticas cuando fue necesario (12 casos fueron editados, incluyendo 92 cortes). El tiempo medio para realizar estos procesos fue de 4.08 minutos (± 2.35 DE) y el tiempo mediano fue de 4 minutos. Esto se comparó con el tiempo necesario para segmentar manualmente las máscaras desde el principio. La segmentación automática asociada a una validación visual y edición manual (en caso de ser necesaria) reduce el tiempo necesario para la tarea de segmentación manual (sin intervención de la herramienta) en un 92.8%.

4.3.3. Validación de la herramienta de detección y segmentación automática en RM

La validación independiente es un paso fundamental para que los modelos de IA sean aplicables a la práctica clínica, ya que permite confirmar su reproducibilidad y generalización. Se realizó un tercer subanálisis con la finalidad de llevar a cabo una validación independiente de la herramienta de detección y segmentación automática previamente desarrollada e internamente validada, proponiendo una metodología de segmentación semiautomática (Figura 34).

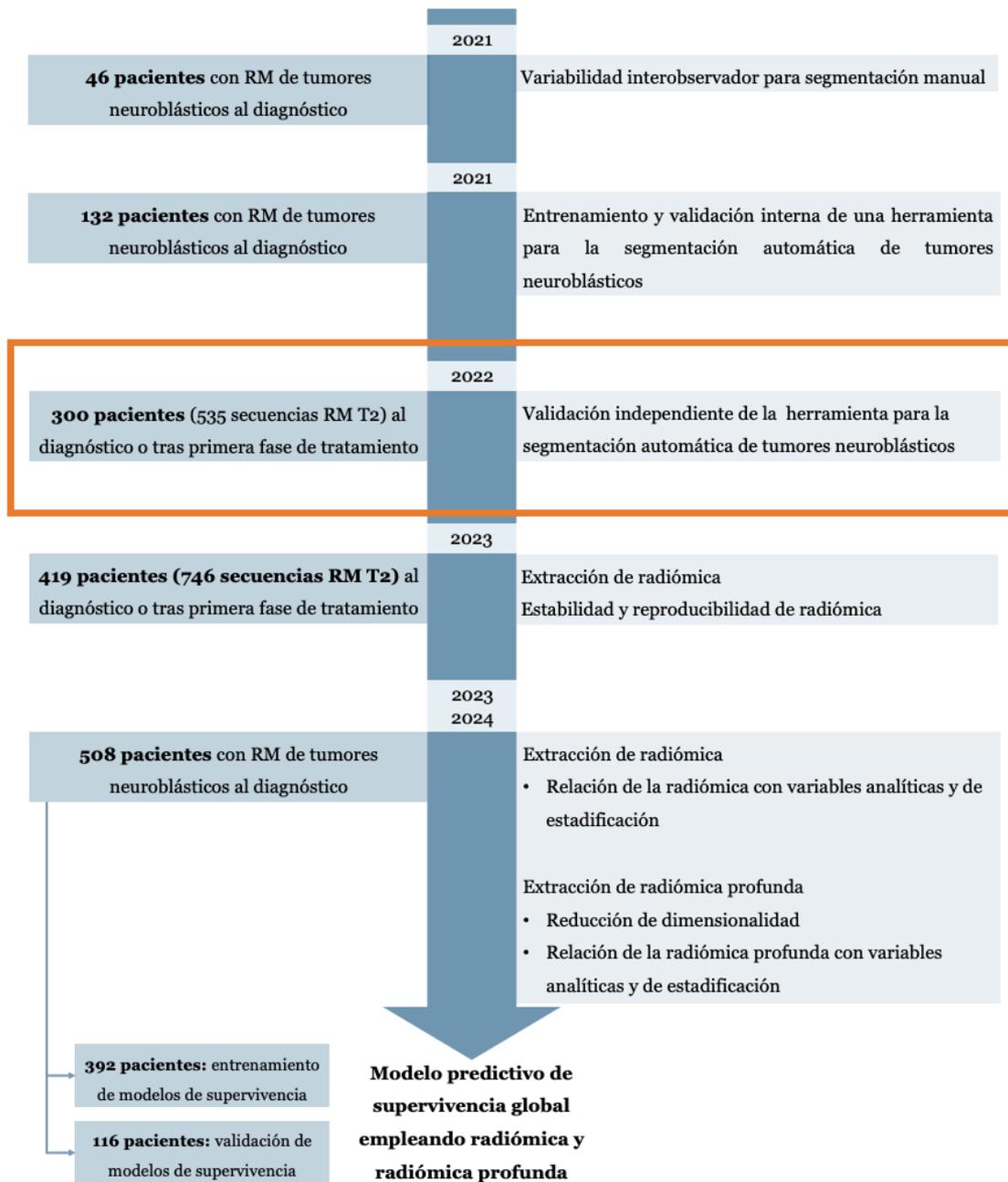


Figura 34. Flujo de trabajo: validación externa de la herramienta de detección y segmentación automática.

La hipótesis de este estudio fue que la herramienta de detección y segmentación automática basada en la arquitectura nnU-Net y previamente desarrollada tiene un excelente rendimiento para detectar y segmentar automáticamente tumores neuroblásticos en imágenes de RM en una población y equipamiento de RM diferente al utilizado para entrenar la herramienta.

Para llevar a cabo este proceso de validación externa, se reclutaron un total de 300 pacientes en la plataforma del proyecto PRIMAGE. Para incrementar la heterogeneidad de la muestra, se incluyeron todas las secuencias de RM disponibles potenciadas en T2wSE o T2*wGE con o sin supresión grasa. Se revisaron visualmente un total de 810 secuencias potenciadas T2wSE, T2*wGE, T2wSE FS y T2 STIR de estos 300 pacientes, tanto al diagnóstico como tras la primera pauta de quimioterapia. Los pacientes incluidos en este análisis se habían sometido a una RM de la región anatómica del tumor en diferentes momentos del tiempo: únicamente al momento del diagnóstico (n = 264 pacientes), únicamente después del tratamiento inicial con quimioterapia (n = 10), o tanto al diagnóstico como tras la primera línea de tratamiento (n = 26). Se excluyeron aquellas imágenes previamente empleadas para el entrenamiento y validación interna de la herramienta de segmentación (n = 132), aquellas que no tenían tumor objetivable (estudios de extensión de cabeza y cuello en pacientes con tumor abdominal (n = 50); exámenes realizados después del tratamiento quimioterápico sin restos de tumor (n = 15); RM realizadas tras cirugía (n = 50)), secuencias no orientadas en el plano transversal (n = 2) y secuencias con artefactos importantes (n = 26). Finalmente, se emplearon 535 secuencias de RM de 300 niños para la validación externa independiente de la herramienta de segmentación, de las cuales 486 correspondieron a secuencias realizadas al diagnóstico y 49 tras primera línea de tratamiento (Figura 35).

Validación independiente de la herramienta de detección y segmentación automática

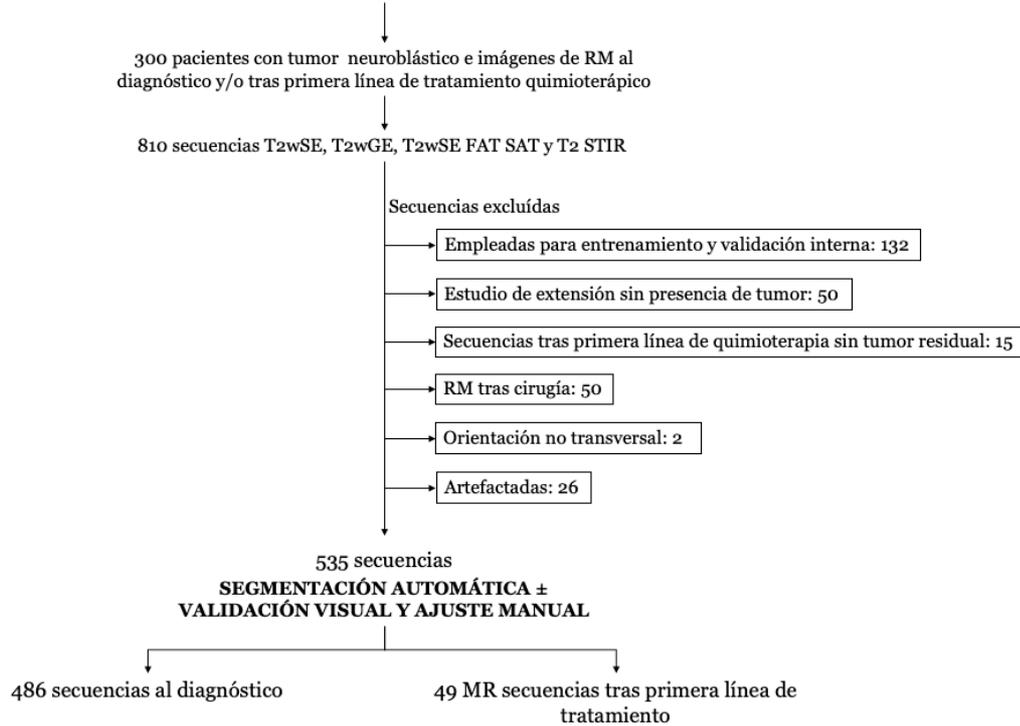


Figura 35. Diseño del estudio para la validación independiente de la herramienta de detección y segmentación automática de tumores neuroblásticos.

Los pacientes y los datos de RM se obtuvieron principalmente de dos ensayos clínicos europeos, llevados a cabo por SIOPEN: LINES/SIOPEN (n = 107) y HR-NBL1/SIOPEN (n = 119). Adicionalmente, se incluyeron pacientes que no formaban parte de los ensayos clínicos mencionados, reclutados en diferentes instituciones europeas: pacientes del Registro Alemán de Neuroblastoma NB2016 de la Sociedad de Oncología y Hematología Pediátrica, dirigido por el Departamento de Oncología Pediátrica del Hospital Universitario de Niños de Colonia (n = 62), pacientes del Hospital Universitario de Pisa, Instituto Gaslini (Pisa, Italia) (n = 3), así como pacientes del Hospital La Paz y Hospital Niño Jesús (Madrid, España), Hospital Son Espases (Palma de Mallorca, España), Hospital Reina Sofía (Córdoba, España) y del Hospital La Fe que no formaban parte del estudio LINES/SIOPEN (n = 9).

La edad al primer diagnóstico fue de 18 ± 32 meses (media \pm DE, rango de 0 a 212 meses), con distribución por sexos de 155 niñas y 145 niños. La histología del

tumor fue neuroblastoma (263 casos), ganglioneuroblastoma (27 casos) y ganglioneuroma (10 casos).

La ubicación del tumor se clasificó como abdominopélvica (430 secuencias de las 535 incluidas estando localizados 269 de los tumores en la glándula suprarrenal, 143 con ubicación abdominal no suprarrenal y 18 con ubicación pélvica) o cervicotorácica (105 secuencias, 95 de los tumores torácicos y 10 cervicales).

Las imágenes de RM presentaban una alta variabilidad en la adquisición de datos, incluyendo diferentes escáneres, proveedores y protocolos, provenientes de distintas instituciones. Las imágenes de RM se adquirieron con un escáner de 1.5 T (n = 435 secuencias) o 3 T (n = 100), fabricados por General Electric Healthcare (Signa Excite HDxt, Signa Explorer, Discovery, Genesis; n = 105); Siemens Medical (Aera, Skyra, Symphony, Avanto, Magneto Espree, TrioTim, Sonata, Sprecetra, Verio, Harmony; n = 318); Philips Healthcare (Intera, Achieva, Ingenia, Panorama; n = 109) o Canon (Titan; n = 3). De las 535 secuencias empleadas, 307 fueron T2wSE, 11 T2*wGE FS, 176 T2wSE FS y 41 T2wSE STIR. En la Tabla 17 se exponen las características de la muestra en el conjunto de datos de validación independiente y se comparan con el grupo de entrenamiento. La distribución de la población fue similar, excepto por la edad en el momento del diagnóstico, que fue menor en el grupo de validación. En el grupo de validación se incluyeron RM de pacientes tras la primera línea de tratamiento con quimioterapia. Asimismo, se incluyó un proveedor nuevo (Canon), y dos nuevas secuencias para aumentar la heterogeneidad de la muestra: T2wSE STIR y T2*wGE FS.

Tabla 17. Comparación de las características de la muestra en el conjunto de datos de entrenamiento original y el de validación.

Comparación de las características de la muestra en el conjunto de datos de entrenamiento original y el de validación		Grupo de entrenamiento/ajuste y validación interna n (%)	Grupo de validación independiente (%)
Pacientes		132	300
Secuencias de RM	Número	132	535
Edad al diagnóstico	Media± DE (meses)	37±39	18±32

Sexo	Masculino	62 (47)	145 (48)
	Femenino	70 (53)	155 (52)
Histología	Neuroblastoma	104 (79)	263 (88)
	Ganglioneuroblastoma	18 (13)	27 (9)
	Ganglioneuroma	10 (8)	10 (3)
Momento de realización del estudio	Al diagnóstico	132 (100)	486 (91)
	Tras tratamiento	0 (0)	49 (9)
Localización	Abdominopélvico	105 (80)	430 (80)
	Cervicotorácico	27 (20)	105 (20)
Proveedor	GE	51 (39)	105 (20)
	Siemens	54 (41)	318 (59)
	Philips	27 (20)	109 (20)
	Canon	0 (0)	3 (1)
Intensidad del campo magnético	1,5T	116 (88)	435 (81)
	3T	16 (12)	100 (19)
Secuencia segmentada	T2wSE	84 (64)	307 (57)
	T2wSE FS	48 (36)	176 (33)
	T2wSE STIR	0 (0)	41 (8)
	T2*wGE FS	0 (0)	11 (2)

El algoritmo de segmentación automática previamente entrenado fue integrado en la plataforma PRIMAGE. Posteriormente, las segmentaciones se realizaron sobre las secuencias de RM mediante su ejecución en la plataforma de PRIMAGE. El resultado fue la máscara del tumor primario accesible desde el visor DICOM integrado en la plataforma PRIMAGE para su revisión final. La Radióloga 1 revisó cada secuencia de RM sin conocer el resultado de la segmentación y localizó visualmente el tumor y sus límites. A continuación, revisó las máscaras de segmentación en todos los cortes de RM que incluyesen lesión tumoral, validando visualmente el rendimiento de la red y asegurándose de que el tumor estuviera

bien segmentado. En caso de visualizar algún error de segmentación realizó un ajuste manual de la máscara. Finalmente, se compararon las máscaras automáticas con las editadas manualmente (Figura 36).

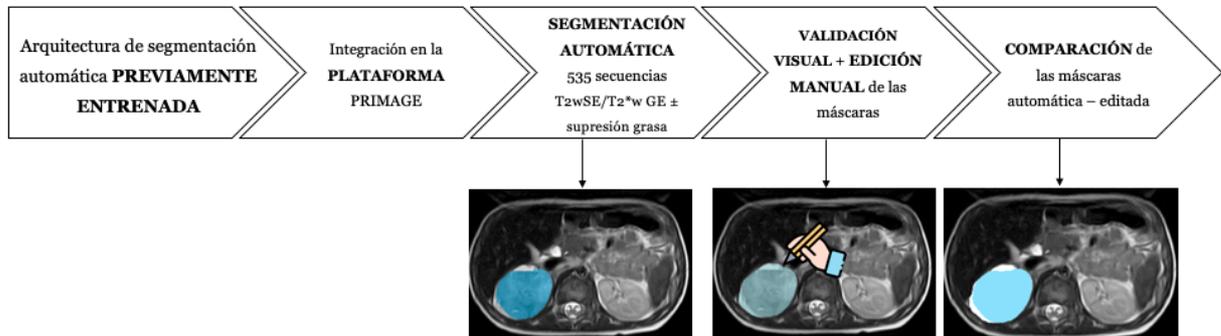


Figura 36. Esquema del proceso llevado a cabo para la validación independiente de la herramienta de segmentación.

De las 535 secuencias de RM potenciadas en T₂/T₂^{*}, la mediana del CSD fue de 0.997 (± 0.056 IQR) y la mediana del JAC fue de 0.996 (± 0.106 IQR), ambos resultados considerados como excelentes. La HD media fue de 7.081 (± 19.999 DE) mm, considerándose pequeña la distancia máxima de la máscara de segmentación automática al punto más cercano en la máscara corregida. El AUC ROC media fue de 0.930, mostrando un excelente acuerdo entre máscaras. El 1-TFPm medio fue de 0.847 (± 1.123 DE), lo que resulta en una alta concordancia entre la red y la corrección manual en cuanto a los vóxeles no tumorales incluidos. El 1-TFN medio fue de 0.917 (± 0.215 DE), lo cual implica que la red no omitió una cantidad importante de tumor durante la segmentación (Tabla 18).

Tabla 18. Métricas de rendimiento para el conjunto de validación independiente, comparando las máscaras obtenidas automáticamente con las corregidas manualmente. Abreviaturas: CSD, Coeficiente de Similitud Dice; JAC, índice Jaccard; AUC ROC, Área Bajo la Curva; HD, Hausdorff Distance; 1-TFPm, Tasa de Falsos Positivos modificada; 1-TFN, Tasa de Falsos Negativos.

	CSD	JAC	HD	AUC ROC	1-TFPm	1-TFN
Mediana	0.997	0.996	0.000	0.999	1.000	1.000
IQR	0.056	0.106	3.000	0.027	0.004	0.031
Media	0.887	0.862	7.081	0.930	0.847	0.917
DE	0.262	0.279	19.999	0.191	1.123	0.215

En la Figura 37 se muestran algunos ejemplos del rendimiento de la herramienta de segmentación automática y la comparación con las máscaras de segmentación tras corrección manual en diferentes casos, con variabilidad en sus ubicaciones y parámetros de adquisición.

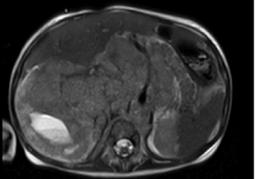
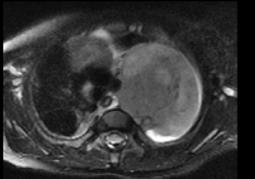
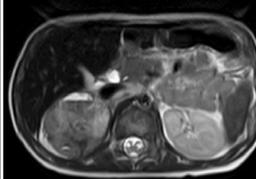
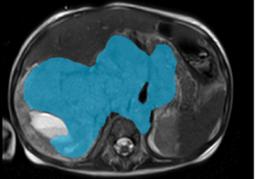
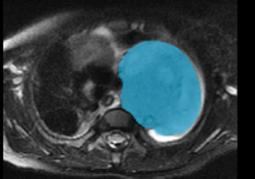
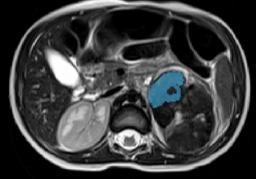
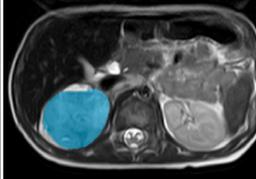
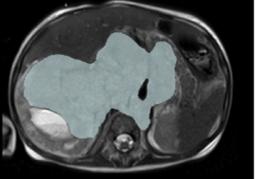
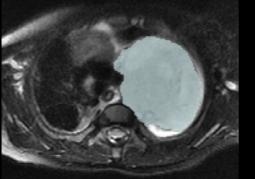
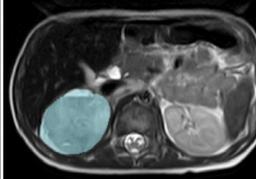
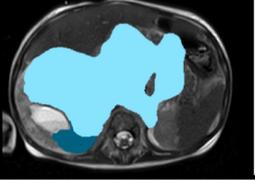
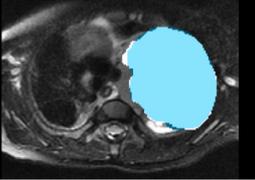
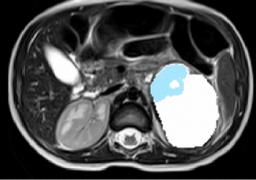
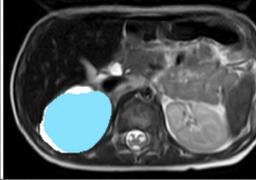
	Caso 1	Caso 2	Caso 3	Caso 4
Imagen fuente				
Segmentación automática				
Corrección manual				
Comparación máscaras				
Características y CSD	Localización: abdominopélvica Momento de RM: diagnóstico Equipamiento RM: Siemens Campo magnético: 1.5T Potenciación: T2 SE CSD: 0.957	Localización: cervicotorácica Momento de RM: diagnóstico Equipamiento RM: GE Campo magnético: 3T Potenciación: STIR CSD: 0.948	Localización: abdominopélvica Momento de RM: tratamiento Equipamiento RM: Siemens Campo magnético: 1.5T Potenciación: T2 SE CSD: 0.187	Localización: abdominopélvica Momento de RM: tratamiento Equipamiento RM: Philips Campo magnético: 3T Potenciación: T2 SE CSD: 0.980

Figura 37. Ejemplos de máscaras de segmentación automática antes y después de la edición manual en cuatro casos con variabilidad en la ubicación y parámetros de adquisición. Adaptado de Veiga-Canuto et al [116].

Se realizó un subanálisis descriptivo atendiendo a los valores atípicos obtenidos. El 74 de 535 de las secuencias de RM presentaron un valor de CSD < 0.8 (14%) y por tanto un bajo nivel de acuerdo. Estos casos con alta variabilidad fueron analizados visualmente por la Radióloga 1 para identificar las razones del bajo nivel de concordancia entre las máscaras. En 32 de los 74 casos (6% del total de 535 secuencias) se evidenció un CSD < 0.19 , lo que reflejó un rendimiento no exitoso de la red. En 18 de estos casos el CSD fue de 0 (puesto que la red no segmentó ninguna estructura), mientras que en 14 casos la red segmentó otras estructuras diferentes al tumor: ganglios linfáticos (10 casos) o grasa (4 casos). En los restantes 42 casos con valor de CSD < 0.8 (8% del total de 535 secuencias), el CSD fue de entre 0.2 y 0.8. En estos casos la red pudo identificar el tumor, pero la segmentación fue incompleta.

En el grupo de pacientes con bajo nivel de acuerdo < 0.8 ($n = 74$) el volumen medio del tumor fue de $121.140 \text{ mm}^3 (\pm 281.210 \text{ DE})$, siendo un volumen menor en comparación con el volumen medio de los tumores con CSD ≥ 0.8 ($n = 461$) ($194.660 \text{ mm}^3 \pm 222.750 \text{ DE}$). La proporción de casos después de la quimioterapia fue mayor en el grupo con bajo nivel de acuerdo, siendo de 16/74 casos en este grupo (22%) frente a 33/461 casos en el grupo con CSD ≥ 0.8 (7%). En cuanto a la ubicación del tumor, 6/74 casos presentaron una localización cervicotorácica (8%) y 68/74 (92%) fueron abdominopélvicos, mientras que en el grupo con CSD ≥ 0.8 la proporción fue de 99/461 cervicotorácicos (22%) frente a 362/461 abdominopélvicos (79%). En cuanto al campo magnético, 63/74 pacientes con bajo nivel de acuerdo se adquirieron en un equipo de 1.5T (85%) y 11 se realizaron en un equipo de 3T (15%), con proporciones similares a los pacientes con CSD ≥ 0.8 , siendo estas de 372/461 (80%) en un equipo de 1.5T y 89/461 (20%) en un equipo de 3T. En referencia a la potenciación, 40/74 secuencias estaban potenciadas en T2wSE (54%), 29/74 en T2wSE FS (40%), 4/74 en STIR (5%) y 1/74 T2 GE (1%), mostrando proporciones similares en

comparación con el grupo con CSD ≥ 0.8 , siendo estas de 267/461 T2wSE (58%), 147/461 en T2wSE FS (32%), 37/461 en STIR (8%) y 10/461 T2 GE (2%),

Por otro lado, se realizó un análisis descriptivo estratificando los resultados del CSD en las 535 secuencias, considerando la ubicación (abdominopélvica o cervicotorácica), el momento del estudio (diagnóstico o tratamiento), la fuerza del campo magnético (1.5T o 3T) y la ponderación de la secuencia (T2WSE, T2wSE FS, T2 GE, STIR) (Tabla 19). Se observaron unos resultados similares en el CSD para todos los subgrupos (de 0.995 a 0.999) excepto para los casos evaluados después del tratamiento, en los que el CSD fue de 0.902.

Tabla 19. Métricas de rendimiento para el conjunto de validación independiente del modelo de segmentación automática estratificadas en función del momento de realización de la RM, la ubicación, el campo magnético y la secuencia segmentada. Abreviaturas: CSD, Coeficiente de Similitud Dice; JAC, índice Jaccard; AUC ROC, Área Bajo la Curva; HD, Hausdorff Distance; 1-TFPm, Tasa de Falsos Positivos modificada; 1-TFN, Tasa de Falsos Negativos.

	CSD	JAC	HD	AUC ROC	1-TFPm	1-TFN
Tumor al diagnóstico (n = 486)						
Mediana	0.999	0.997	0.000	0.999	1.000	1.000
IQR	0.036	0.070	2.207	0.013	0.003	0.024
Media	0.901	0.879	7.115	0.931	0.853	0.923
DE	0.250	0.266	20.720	0.196	1.112	0.208
Tumor tras tratamiento (n = 49)						
Mediana	0.902	0.821	2.803	0.999	1.000	1.000
IQR	0.535	0.247	6.000	0.770	0.866	0.526
Media	0.752	0.691	6.737	0.926	0.785	0.854
DE	0.334	0.344	10.553	0.137	1.245	0.275
Cervicotorácico (n = 105)						
Mediana	0.999	0.999	0.000	0.999	1.000	1.000
IQR	0.025	0.049	2.000	0.002	0.021	0.001
Media	0.960	0.938	3.933	0.994	0.928	0.988
DE	0.109	0.145	11.712	0.023	0.212	0.046
Abdominopélvico (n = 430)						
Mediana	0.997	0.995	0.000	0.999	1.000	1.000
IQR	0.071	0.132	3.527	0.027	0.004	0.025
Media	0.869	0.843	7.849	0.923	0.814	0.916
DE	0.284	0.300	21.484	0.208	1.251	0.222

1.5T (n = 434)						
Mediana	0.998	0.996	0.000	1.000	1.000	1.000
IQR	0.055	0.104	2.979	0.032	0.004	0.039
Media	0.879	0.855	7.372	0.925	0.841	0.911
DE	0.275	0.291	20.622	0.197	1.174	0.223
3T (n = 101)						
Mediana	0.995	0.990	0.000	0.999	1.000	1.000
IQR	0.057	0.108	3.000	0.002	0.002	0.018
Media	0.918	0.888	5.944	0.951	0.871	0.942
DE	0.196	0.220	17.251	0.163	0.884	0.177
T2WSE (n = 307)						
Mediana	0.997	0.994	0.000	0.999	1.000	1.000
IQR	0.049	0.094	2.855	0.029	0.005	0.049
Media	0.904	0.877	6.078	0.938	0.873	0.907
DE	0.232	0.253	17.649	0.170	0.828	0.249
T2WSE FS (n = 176)						
Mediana	0.998	0.997	0.000	0.999	1.000	1.000
IQR	0.072	0.135	3.577	0.027	0.003	0.017
Media	0.849	0.827	9.737	0.907	0.769	0.917
DE	0.317	0.327	24.984	0.239	1.618	0.226
STIR (n = 41)						
Mediana	0.999	0.999	0.000	0.999	1.000	1.000
IQR	0.077	0.125	2.207	0.009	0.001	0.017
Media	0.911	0.881	3.393	0.976	0.991	0.952
DE	0.210	0.241	8.908	0.081	0.026	0.162
T2 GE FS (n = 11)						
Mediana	0.996	0.992	0.000	0.998	1.000	1.000
IQR	0.009	0.014	0.433	0.043	0.128	0.079
Media	0.929	0.901	7.499	0.943	0.792	0.888
DE	0.176	0.229	18.535	0.109	0.480	0.217

Para valorar la influencia del CSD en estos cuatro factores (ubicación, momento, campo magnético, ponderación de secuencia), se llevó a cabo un análisis simultáneo (ANOVA), mostrando que ninguno de los factores o sus interacciones tuvo una influencia significativa en el CSD para un nivel de significación de 0.05 (Tabla 20).

Tabla 20. Análisis de la varianza de la influencia sobre la variable dependiente Coeficiente de Similitud Dice del momento de realización del estudio, la localización, la fuerza del campo magnético y la secuencia, y de sus interacciones para el conjunto de validación independiente de la herramienta de segmentación automática.

Variable dependiente: CSD		
Variable a comparar	Prueba F	P valor
Momento del estudio	0.081	0.775
Localización	2.561	0.110
Campo magnético	2.374	0.124
Potenciación	0.107	0.956
Localización x Campo magnético	0.171	0.679
Momento del estudio x Localización	1.272	0.260
Localización x Potenciación	0.478	0.698
Momento del estudio x Campo magnético	3.117	0.078
Campo magnético x Potenciación	0.794	0.453
Momento del estudio x Potenciación	0.383	0.682

A pesar de no mostrar diferencias significativas, cabe destacar que la mediana del CSD en el diagnóstico fue de 0.999, y después del tratamiento fue de 0.902, y que el volumen medio \pm DE de los tumores en el diagnóstico fue mayor ($210.389 \text{ mm}^3 \pm 227.830$) que después del tratamiento ($43.467 \text{ mm}^3 \pm 49.745$). En la figura 38 se muestran ejemplos del rendimiento de la herramienta de segmentación automática y la edición manual en dos casos con diferente localización y con RM realizadas en diferentes momentos (al diagnóstico y tras tratamiento).

	Caso 1a	Caso 1b	Caso 1c	Caso 2a	Caso 2b
Imagen fuente					
Segmentación automática					
Corrección manual					
Comparación máscaras					
Características y CSD	Localización: abdominopélvica Momento de RM: diagnóstico Equipamiento RM: Siemens Campo magnético: 1.5T Potenciación: T2 SE CSD: 0.977	Localización: abdominopélvica Momento de RM: tratamiento Equipamiento RM: Siemens Campo magnético: 1.5T Potenciación: T2 SE CSD: 0.286	Localización: abdominopélvica Momento de RM: tratamiento Equipamiento RM: Siemens Campo magnético: 1.5T Potenciación: T2 SE CSD: 1	Localización: cervicotorácica Momento de RM: diagnóstico Equipamiento RM: Philips Campo magnético: 1.5T Potenciación: T2 SE CSD: 1	Localización: cervicotorácica Momento de RM: tratamiento Equipamiento RM: Siemens Campo magnético: 1.5T Potenciación: T2 SE CSD: 1

Figura 38. Ejemplos del rendimiento de la herramienta de segmentación automática y la edición manual en dos casos con variabilidad en el momento de adquisición, la ubicación y parámetros de adquisición. Adaptado de Veiga-Canuto et al [116].

Impacto en el ahorro de tiempo

Se registró el tiempo empleado por la Radióloga 1 (en segundos) para validar visualmente y editar manualmente la máscara. El tiempo medio necesario para la revisión visual de las máscaras generadas sin necesidad de edición manual fue de 7.8 ± 7.5 segundos (media \pm DE). 136 máscaras requirieron corrección manual adicional, que requirió un tiempo de 124 ± 120 segundos (media \pm DE). La mediana del CSD de las 136 máscaras editadas fue de 0.887 ± 0.499 (mediana; IQR).

4.4. Estudio de radiómica

4.4.1. Extracción de parámetros de radiómica

El análisis de radiómica implica la extracción matemática de características cuantitativas de regiones de interés de imágenes, proporcionando información sobre relaciones de forma, textura y vóxeles [41, 60]. La siguiente fase de esta Tesis Doctoral consistió en la extracción de características de radiómica y la realización de un análisis de reproducibilidad de la misma (Figura 39).

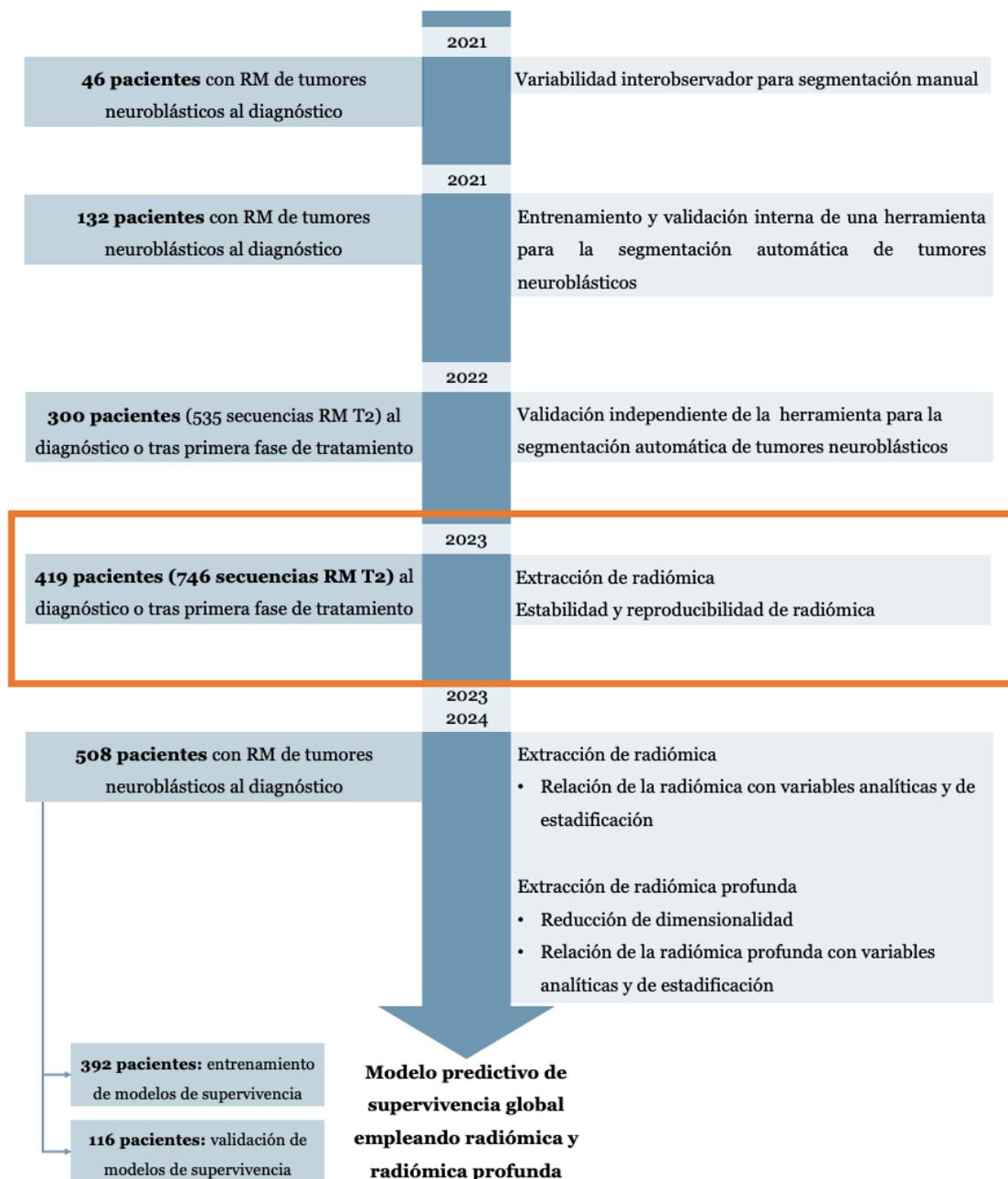


Figura 39. Flujo de trabajo: extracción de radiómica y análisis de su estabilidad y reproducibilidad.

Para este análisis se seleccionaron inicialmente 479 pacientes (incluyendo los 132 pacientes empleados para el entrenamiento de la red de segmentación, los 300 pacientes incluidos en la validación y 47 nuevos pacientes) con tumor neuroblástico con RM realizada en el momento del diagnóstico y/o tras la primera línea de tratamiento con quimioterapia. Se revisaron visualmente 810 secuencias de estos 479 pacientes, potenciadas en T2wSE, T2*wGE FS, T2wSE FS y T2wSE STIR, excluyendo aquellas secuencias que correspondían con estudios de extensión sin presencia de tumor (n = 17), aquellas secuencias sin tumor residual tras la primera línea de quimioterapia (n = 14), las secuencias realizadas tras cirugía (n = 10) y las secuencias que presentaban artefactos importantes (n = 23). En total, tras la exclusión de 64 secuencias, se incluyeron 746 secuencias pertenecientes 419 pacientes (Figura 40).

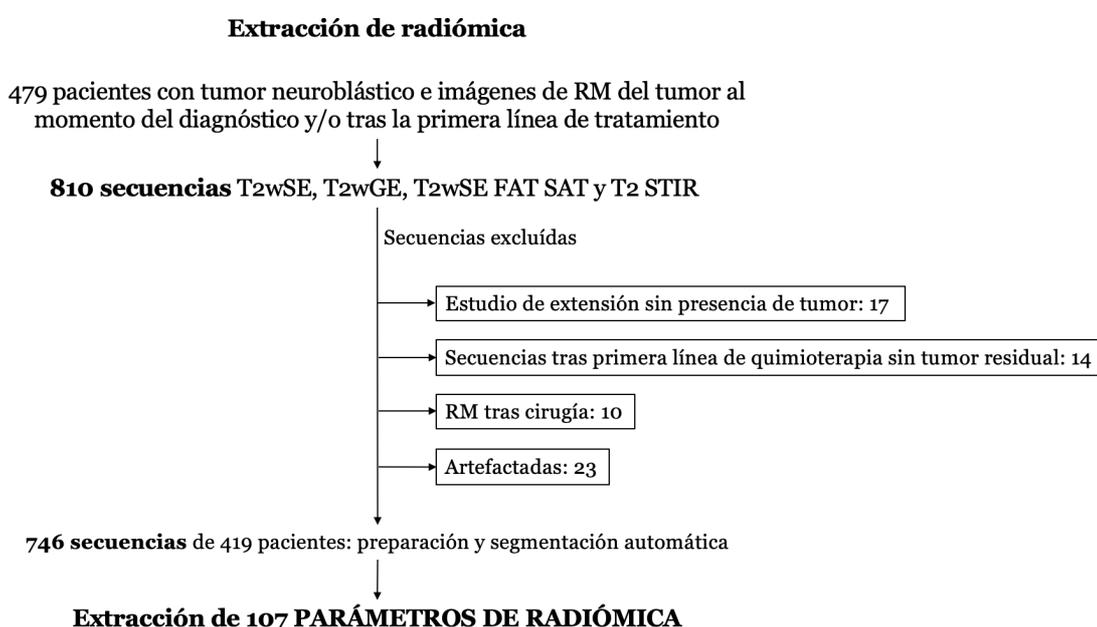


Figura 40. Flujo de pacientes seleccionados para la extracción de parámetros de radiómica.

En cuanto a la muestra de los 419 pacientes, se observó un discreto predominio de pacientes de sexo masculino (n = 220) frente a femenino (n = 199), en distintas

localizaciones (n = 328 con localización abdominopélvica; n = 91 con localización cervicotorácica).

En referencia a las secuencias, se incluyeron secuencias de RM realizadas en el momento del diagnóstico (n = 625 secuencias) y/o después del tratamiento inicial con quimioterapia (n = 121 secuencias).

Las imágenes de RM presentaban una alta variabilidad en cuanto al proveedor (General Electric Healthcare (Signa Excite HDxt, Signa Explorer, Discovery, Genesis; n = 160); Siemens Medical (Aera, Skyra, Symphony, Avanto, Magneto Espree, TrioTim, Sonata, Sprecetra, Verio, Harmony; n = 430); Philips Healthcare (Intera, Achieva, Ingenia, Panorama; n = 153) o Canon (Titan; n = 3)), y a la intensidad del campo magnético (1.5 T (n = 602) o 3 T (n = 144)). De las 535 secuencias empleadas, 307 fueron T2wSE, 381 T2wSE FS, 47 T2wSE STIR y 11 T2*wGE FS. Las características de la muestra para la extracción de radiómica detallada por pacientes y secuencias se expone en la Tabla 21.

Tabla 21. Características basales de la muestra empleada para para la extracción y posterior análisis de radiómica. Los datos se expresan como media \pm desviación estándar cuando la distribución es normal.

Descripción de la muestra para la extracción de radiómica		Número de pacientes
PACIENTES		
Pacientes	Número	419
Edad al diagnóstico	Media \pm desviación estándar (meses)	29 \pm 34
Sexo	Hombre	220
	Mujer	199
Histología	Neuroblastoma	366
	Ganglioneuroblastoma	34
	Ganglioneuroma	19
Localización	Abdominopélvico	328
	Cervicotorácico	91

SECUENCIAS		
Secuencias	Número	746
Momento de realización del estudio	Al diagnóstico	625
	Tras tratamiento	121
Proveedor	GE	160
	Siemens	430
	Philips	153
	Canon	3
Intensidad del campo magnético	1.5T	602
	3T	144
Secuencia segmentada	T2wSE	307
	T2wSE FS	381
	T2wSE STIR	47
	T2*wGE FS	11

Como se ha expuesto anteriormente, las imágenes fueron preparadas para asegurar un marco común antes de la segmentación del tumor y la extracción de radiómica. Esta fase preparatoria consistió en la aplicación de un filtro de difusión anisotrópica (ADF) con el fin de minimizar el ruido [91], corrección del sesgo por campo magnético N4 para corregir la inhomogeneidad de la señal [98, 99], normalización de la señal mediante Z-score [102] y re-muestreo para la armonización espacial. Estos pasos de preparación de imágenes se aplicaron y previamente a la segmentación y a la extracción de radiómica como verdad de referencia ya que representan un escenario del mundo real en el que las imágenes están armonizadas. Las imágenes preparadas con estas modificaciones se consideraron como Serie 1.

Posteriormente, se realizó la segmentación automática de aquellas secuencias que no dispusiesen previamente de segmentación (211 secuencias nuevas respecto al estudio expuesto en el apartado 4.3.3) empleando el flujo previamente descrito: realización de la segmentación automática mediante la herramienta previamente entrenada y validación visual con edición manual en los casos que requirieran edición (se editaron 18 secuencias). Tras la segmentación, se extrajeron las características de radiómica de las regiones de interés (diagrama representado en Figura 40).

Se extrajeron 107 características de radiómica de las 746 secuencias de RM. Las características extraídas se clasificaron como:

- Características de forma (n = 14): describen la forma de la región de interés y sus propiedades geométricas.
- Características de primer orden o de distribución del histograma de intensidades (n = 18): describen la distribución de los valores de los vóxeles sin tener en cuenta relaciones espaciales. Están basadas en el histograma de intensidades.
- Características de textura original o de segundo orden (n = 75): describen la relación de intensidades entre vóxeles vecinos, aportando una medida de la disposición espacial de las intensidades de los vóxeles. Dichas relaciones se miden a través de distintas matrices de grises, incluyendo la matriz de co-ocurrencia de niveles de gris (GLCM, n = 24), matriz de dependencia de niveles de gris (GLDM, n = 14), matriz de longitud de segmentos consecutivos de píxeles con el mismo nivel de gris (GLRLM, n = 16), matriz de tamaño de zona de niveles de gris (GLSZM, n = 16), y matriz de diferencia de tono de gris vecino (NGTDM, n = 5). Los análisis se llevaron a cabo utilizando Pyradiomics [103].

La lista completa de las características obtenidas se presenta en la Tabla 22.

Tabla 22. Lista de las 107 características radiómicas de forma, de primer orden y de segundo orden que fueron extraídas de tumores neuroblásticos en imágenes de resonancia magnética potenciada en T2/T2*.

Forma	Primer orden	Segundo orden				
		GLCM	GLDM	GLRLM	GLSZM	NGTDM
Elongation	10th percentile	Autocorrelation	Dependence entropy	Gray level nonuniformity	Gray level nonuniformity	Busyness
Flatness	90th percentile	Cluster prominence	Dependence nonuniformity	Gray level nonuniformity normalized	Gray level nonuniformity normalized	Coarseness
Least axis length	Energy	Cluster shade	Dependence nonuniformity normalized	Gray level variance	Gray level variance	Complexity
Major axis length	Entropy	Cluster tendency	Dependence variance	High gray level run emphasis	High gray level Zone emphasis	Contrast
Maximum 2D diameter (x-z plane, row)	Interquartile range	Contrast	Gray level nonuniformity	Long run emphasis	Large area emphasis	Strength
Maximum 2D diameter (y-z plane, column)	Kurtosis	Correlation	Gray level variance	Long run High gray level emphasis	Large area High gray level emphasis	
Maximum 2D diameter (x-y plane, slice)	Maximum	Difference Average	High gray level emphasis	Long run low gray level emphasis	Large area low gray level emphasis	
Maximum 3D diameter	Media absolute deviation	Difference Entropy	Large dependence emphasis	Low gray level run emphasis	Low gray level Zone emphasis	
Mesh Volume	Mean	Difference variance	Large dependence high gray level emphasis	Run entropy	Size zone nonuniformity	
Minor axis length	Median	Inverse difference	Large dependence low gray level emphasis	Run length nonuniformity	Size Zone nonuniformity normalized	
Sphericity	Minimum	Inverse difference moment	Low gray level emphasis	Run length nonuniformity normalized	Small area emphasis	
Area	Range	Inverse difference moment normalized	Small dependence emphasis	Run percentage	Small area High gray level emphasis	
Surface to volumen ratio	Robust mean absolute deviation	Inverse difference normalized	Small dependence High gray level emphasis	Run variance	Small area low gray level emphasis	
Voxel volume	Root means squared	Informational measure of correlation 1	Small dependence low gray level emphasis	Short run emphasis	Zone entropy	
	Skewness	Informational measure of correlation 2		Short run high gray level emphasis	Zone percentage	
	Totalenergy y	Maximum Correlation Coefficient (MCC)		Short run low gray level emphasis	Zone variance	
	Uniformity	Joint average				
	Variance	Joint energy				
		Joint entropy				
		Inverse variance				
		Maximum probability				
		Sum Average				
		Sum entropy				
		Sum squares				

4.4.2. Estabilidad de la radiómica

4.4.2.1. Modificaciones en las imágenes y máscaras

Con la finalidad de evaluar la reproducibilidad y estabilidad de las diferentes características de radiómica en función de diferentes modificaciones en la

imagen, se generaron ocho series adicionales en las que se llevaron a cabo diferentes modificaciones en la preparación (series 2 a 7) y la segmentación (series 8 y 9). Posteriormente, se realizó una comparación de cada una de estas nuevas series con la serie de referencia (serie 1). En la Figura 41 se aporta un esquema del flujo empleado.

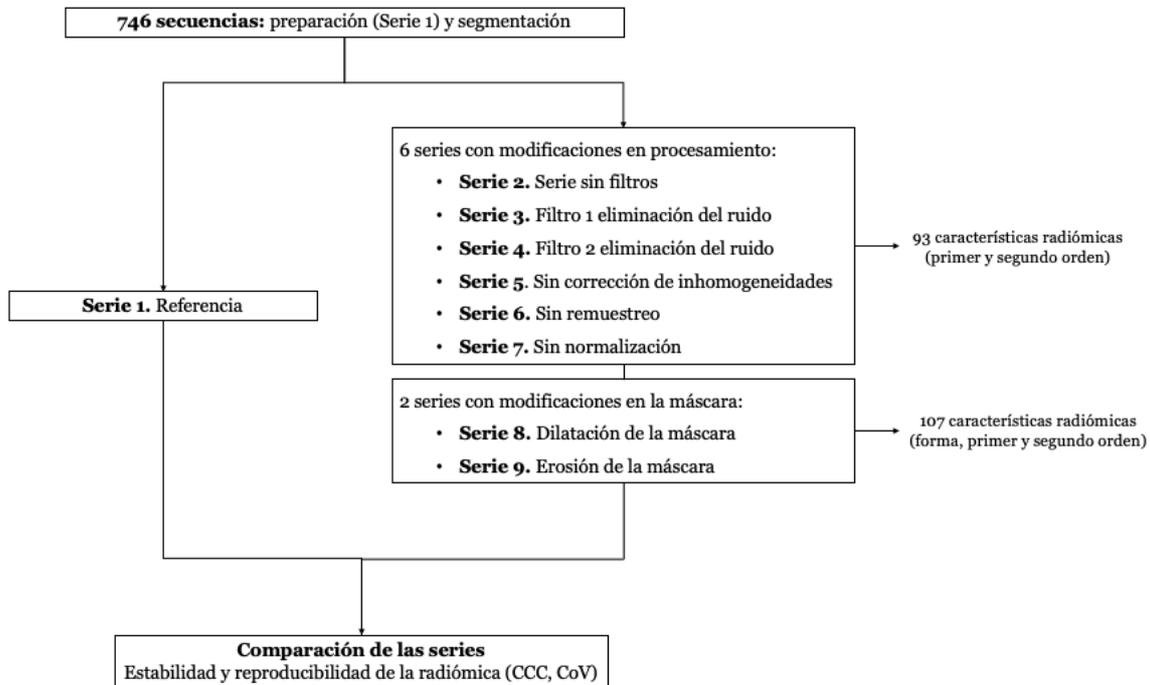


Figura 41: Esquema del flujo para evaluar la estabilidad y reproducibilidad de la radiómica. Modificación de diferentes parámetros de preparación de la imagen y edición de las máscaras, y comparación con la serie original.

Los diferentes filtros y modificaciones empleados (expuestos en el apartado 4.7.) fueron de código abierto con un coste computacional asequible y ampliamente utilizados en estudios previos en la literatura:

- Denoising (reducción de ruido): se aplicaron tres filtros preservadores de bordes comúnmente utilizados:
 - Filtro de difusión anisotrópica (ADF) [101]. Como se ha expuesto previamente, en un estudio previo realizado en el GIBI2³⁰, este filtro mostró los mejores resultados en cuanto a la reproducibilidad de las características radiómicas [98].

- Filtro de flujo de curvatura (CFF), que preserva la definición de los bordes y suaviza perpendicularmente a las curvas de isointensidad [93].
- Filtro bilateral, que reemplaza la intensidad de cada píxel con un promedio ponderado de los valores de intensidad de píxeles cercanos [115].
- Inhomogeneidades: se utilizó la corrección del sesgo por campo magnético N4 para corregir la no uniformidad de baja frecuencia presente en las imágenes [105, 106].
- Remuestreo: todas las series de RM fueron remuestreadas a un tamaño de voxel de 1 x 1 x 6 mm para la armonización espacial y para permitir la comparación de imágenes.
- Normalización: se aplicó la metodología Z-score como proceso de normalización para reducir variaciones sistemáticas debidas a la adquisición de imágenes, reconstrucción y postprocesamiento [40, 108].
- Edición de máscaras: se editó el tamaño de las máscaras realizando una erosión y dilatación, para imitar un escenario en el que los tumores hubieran sido segmentados por diferentes observadores. Se realizó una alteración en la estructura de las máscaras en forma de disco con un diámetro de 2 vóxeles. Se llevó a cabo una modificación con un grosor fijo para mantener una variabilidad consistente en las máscaras. Estas modificaciones se realizaron utilizando la biblioteca SimpleITK [103].

La Serie 1 de referencia fue la serie sometida al proceso de preparación de la imagen de la que inicialmente se extrajeron los parámetros de radiómica, descrita en el apartado 4.4.1. Se detallan a continuación las modificaciones realizadas en cada una de las series, se presenta un esquema de las modificaciones para cada serie en la Tabla 23 y se representan visualmente en la Figura 42.

En las Series 2, 3, 4 y 5 se aplicaron los procesos de normalización y remuestreo (al igual que en la Serie 1 de referencia), lo que permitió la evaluación de las modificaciones de los filtros de reducción de ruido e inhomogeneidades. En la Serie 2 no se aplicó ningún filtro de reducción de ruido ni filtro de inhomogeneidades. En las series 3 y 4 se evaluaron diferentes filtros de reducción

de ruido (bilateral y CFF). La Serie 5 evaluó el efecto de eliminar el filtro de inhomogeneidades, manteniendo el filtro de reducción de ruido ADF como la serie de referencia.

Las Series 6 y 7 permitieron la evaluación de remuestreo y normalización: se aplicaron los mismos filtros de reducción de ruido e inhomogeneidades que a la Serie 1 de referencia, pero se eliminaron las modificaciones relacionadas con el remuestreo (Serie 6) o la normalización (Serie 7).

Las Series 8 y 9 conservaron los mismos parámetros que la segmentación de referencia, pero se realizó la modificación de las máscaras con dilatación (Serie 8) o erosión (Serie 9).

Todas las modificaciones se aplicaron en la plataforma del proyecto PRIMAGE. Los módulos de análisis fueron probados e integrados en la plataforma. La segmentación se realizó en la serie de referencia y posteriormente se extrapoló al resto de las series (excepto en 8 y 9, donde se realizaron modificaciones de máscara).

Tabla 23. Esquema de las modificaciones de imagen aplicadas para el análisis de la estabilidad de la radiómica. Se obtuvieron nueve series diferentes: la serie original de referencia (serie 1) y las modificadas, que incluyeron la ausencia de filtros de reducción de ruido e inhomogeneidades (serie 2), la reducción de ruido (series 3 y 4), corrección de inhomogeneidades (serie 5), remuestreo (serie 6), normalización (serie 7) y edición de máscaras (series 8 y 9).

		Nombre de la serie								
		Serie 1 Referencia	Serie 2 Sin filtros	Serie 3 Filtro eliminación ruido	Serie 4 Filtro eliminación ruido	Serie 5 Inhomogenei- dades	Serie 6 Remuestreo	Serie 7 Normalización	Serie 8 Dilatación máscara	Serie 9 Erosión máscara
Modificaciones aplicadas	N = 746 secuencias (419 pacientes)									
	Eliminación de ruido	ADF	No	Bilateral	CFF	ADF	ADF	ADF	ADF	ADF
	Inhomogeneida des	Sí	No	Sí	Sí	No	Sí	Sí	Sí	Sí
	Normalización	Sí	Sí	Sí	Sí	Sí	Sí	No	Sí	Sí
	Remuestreo	Sí	Sí	Sí	Sí	Sí	No	Sí	Sí	Sí
	Máscara	No editada	No editada	No editada	No editada	No editada	No editada	No editada	Dilatación	Erosión

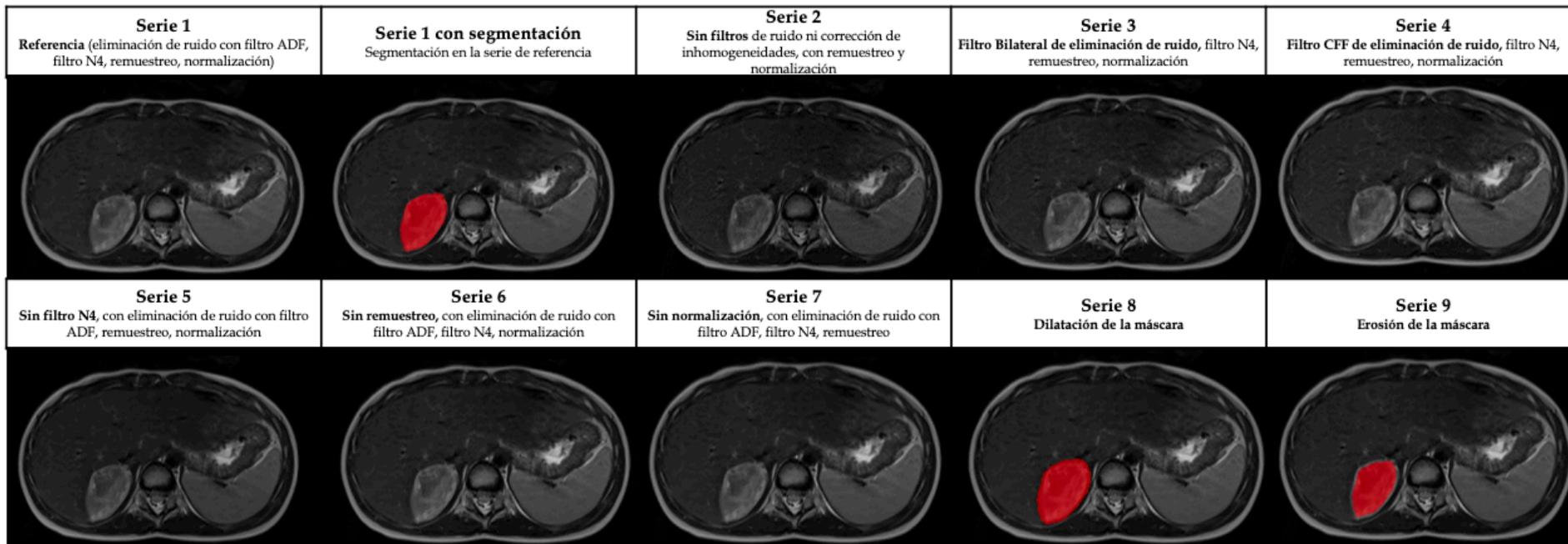


Figura 42. Representación gráfica del impacto de cada uno de los pasos de procesamiento en la imagen de resonancia magnética potenciada en T2wSE en un paciente con un tumor neuroblástico en la glándula suprarrenal derecha.

4.4.2.2. *Extracción de características de imagen*

De nuevo, se extrajeron un total de 107 características radiómicas de las 746 secuencias de RM potenciadas en T2/T2* como previamente se había realizado para la Serie 1, extrayendo las características de radiómica para las 8 series con modificaciones (Serie 2 a 9). Como se ha expuesto anteriormente, las características extraídas se clasificaron como radiómica de forma (n = 14), distribución del histograma de intensidades o de primer orden (n = 18) y características de textura original o de segundo orden (n = 75), incluyendo la matriz de co-ocurrencia de niveles de gris (GLCM, n = 24), matriz de dependencia de niveles de gris (GLDM, n = 14), matriz de longitud de segmentos consecutivos de píxeles con el mismo nivel de gris (GLRLM, n = 16), matriz de tamaño de zona de niveles de gris (GLSZM, n = 16) y matriz de diferencia de tono de gris vecino (NGTDM, n = 5). Los análisis se realizaron utilizando Pyradiomics [110].

En las series en las que la forma no fue alterada (series 2 a 7), las características de forma no se analizaron, ya que la forma del tumor y la máscara permanecieron estables en todas las transformaciones. En estos casos, se consideraron 93 características radiómicas de primer y segundo orden. Por el contrario, en las series donde se modificó la máscara, se evaluaron características radiómicas de forma, de primer orden y de segundo orden (107 características).

4.4.2.3. *Reproducibilidad radiómica entre series*

Se clasificaron las características de radiómica según el nivel de acuerdo (comentado previamente en el subapartado 3.9.2. Reproducibilidad de la radiómica del apartado Material y Métodos): el CCC se clasificó como excelente (≥ 0.90); bueno (0.75–0.89); moderado (0.50–0.75); o bajo (<0.50), y se presentaron los resultados en número y porcentaje de variables de cada una de las 8 series editadas (serie 2 a 9) según el nivel de acuerdo en comparación con la verdad de referencia (serie 1) [41] (Tabla 24).

Tabla 24. Número de características de radiómica y porcentaje de aquellas que presentaron un nivel de acuerdo excelente, bueno, moderado o pobre en cuanto al CCC para las 8 series con modificaciones (series 2 a 9) en comparación con la verdad fundamental (serie 1). E = excelente, B = bueno, M = moderado, P = pobre.

Modificación	Serie 2 Sin filtros				Serie 3 Filtro eliminación ruido bilateral				Serie 4 Filtro eliminación ruido CFF				Serie 5 Inhomogeneidades				Serie 6 Remuestreo				Serie 7 Normalización				Serie 8 Dilatación máscara				Serie 9 Erosión máscara				
	E	B	M	P	E	B	M	P	E	B	M	P	E	B	M	P	E	B	M	P	E	B	M	P	E	B	M	P	E	B	M	P	
Forma (n=14)	n																									14	0	0	0	14	0	0	0
	%																									100	0	0	0	100	0	0	0
Primer orden (n=18)	n	17	1	0	0	18	0	0	0	17	1	0	0	17	1	0	0	16	1	0	1	2	0	0	16	13	4	0	1	16	2	0	0
	%	94.44	5.56	0	0	100	0	0	0	94.44	5.56	0	0	94.44	5.56	0	0	88.89	5.56	0	5.56	11.11	0	0	88.89	72.22	22.22	0	5.56	88.89	11.11	0	0
GLCM (n=24)	n	21	2	1	0	19	5	0	0	19	4	1	0	22	1	1	0	13	7	4	0	3	0	1	20	18	6	0	0	21	3	0	0
	%	87.50	8.33	4.17	0	79.17	20.83	0	0	79.17	16.67	4.17	0	91.67	4.17	4.17	0	54.17	29.17	16.67	0	12.50	0	4.17	83.33	75.00	25.00	0	0	87.50	12.50	0	0
GLDM (n=14)	n	8	4	0	2	9	3	2	0	8	4	0	2	12	0	0	2	3	7	2	2	0	0	3	11	11	3	0	0	9	5	0	0
	%	57.14	28.57	0	14.29	64.29	21.43	14.29	0	57.14	28.57	0	14.29	85.71	0	0	14.29	21.43	50.00	14.29	14.29	0	0	21.43	78.57	78.57	21.43	0	0	64.29	35.71	0	0
GLRM (n=16)	n	7	5	0	4	9	4	3	0	7	5	0	4	11	1	0	4	3	7	4	2	0	2	3	11	11	5	0	0	11	4	1	0
	%	43.75	31.25	0	25.00	56.25	25.00	18.75	0	43.75	31.25	0	25.00	68.75	6.25	0	25.00	18.75	43.75	25.00	12.50	0	12.50	18.75	68.75	68.75	31.25	0	0	68.75	25.00	6.25	0
GLSZM (n=16)	n	9	2	2	3	9	3	4	0	8	3	2	3	11	1	1	3	2	6	4	4	0	0	1	15	8	7	1	0	9	5	2	0
	%	56.25	12.50	12.50	18.75	56.25	18.75	25.00	0	50.00	18.75	12.50	18.75	68.75	6.25	6.25	18.75	12.50	37.50	25.00	25.00	0	0	6.25	93.75	50.00	43.75	6.25	0	56.25	31.25	12.50	0
NGTDM (n=5)	n	5	0	0	0	4	0	0	1	5	0	0	0	5	0	0	0	1	1	2	1	0	0	0	5	4	1	0	0	4	0	1	0
	%	100	0	0	0	80	0	0	20	100	0	0	0	100	0	0	0	20.00	20.00	40.00	20.00	0	0	0	100.00	80.00	20.00	0	0	80.00	0	20.00	0
Total (n=93 series 2-7) (n=107 series 8,9)	n	67	14	3	9	68	15	9	1	64	17	3	9	78	4	2	9	38	29	16	10	5	2	8	78	65	26	1	1	84	19	4	0
	%	72.04	15.05	3.23	9.68	73.12	16.13	9.68	1.08	68.82	18.28	3.23	9.68	83.87	4.30	2.15	9.68	40.86	31.18	17.20	10.75	5.38	2.15	8.60	83.87	60.75	24.30	0.93	0.93	78.50	17.76	3.74	-

4.4.2.4. Reproducibilidad tras la modificación de filtros

Dado que el objetivo de nuestro estudio fue evaluar las características de radiómica que mostraron mayor reproducibilidad, para los posteriores análisis únicamente se seleccionaron aquellas variables estables para cada serie, con un nivel de acuerdo excelente ($CCC > 0.90$).

Tras el análisis de variables radiómicas con un $CCC > 0.90$ en las series con modificaciones de filtros (series 2 a 5), la serie 5 sin el filtro de inhomogeneidad N4 mostró el mayor número de variables reproducibles ($n=78$ de 93 características radiómicas, 84%) (Tabla 25). Esto indica que la eliminación de inhomogeneidades tuvo menor impacto en la radiómica que el resto de modificaciones. La serie 3 (filtro de reducción de ruido bilateral) presentó un $CCC > 0.90$ en 68 de 93 variables (73%), mientras que en la serie 4 (filtro de reducción de ruido CFF), el nivel de acuerdo se redujo a 64 (69%). La serie 2 (sin filtros de reducción de ruido ni inhomogeneidades) tuvo un total de 67 características reproducibles (72%).

Tabla 25. Número y porcentaje de características de radiómica de primer y segundo orden que presentaron un excelente nivel de acuerdo para las 6 series con modificaciones en el procesamiento de imágenes (series 2 a 7) en comparación con la serie de referencia. CFF: Filtro de Flujo de Curvatura.

Modificación		Serie 2 Sin filtros	Serie 3 Filtro eliminación ruido bilateral	Serie 4 Filtro eliminación ruido CFF	Serie 5 Inhomogenei- dades	Serie 6 Remuestreo	Serie 7 Normalización	
Acuerdo excelente								Total %
Primer orden (n=18)	n	17	18	17	17	16	2	
	%	94.44	100	94.44	94.44	88.89	11.11	80.55
GLCM (n=24)	n	21	19	19	22	13	3	
	%	87.5	79.17	79.17	91.67	54.17	12.5	67.36
GLDM (n=14)	n	8	9	8	12	3	0	
	%	57.14	64.29	57.14	85.71	21.43	0	47.62

GLRM (n=16)	n	7	9	7	11	3	0	
	%	43.75	56.25	43.75	68.75	18.75	0	38.54
GLSZM (n=16)	n	9	9	8	11	2	0	
	%	56.25	56.25	50	68.75	12.5	0	40.63
NGTDM (n=5)	n	5	4	5	5	1	0	
	%	100	80	100	100	20	0	66.67
Total (n=93)	n	67	68	64	78	38	5	
	%	72.04	73.12	68.82	83.87	40.86	5.38	57.35

Cuando se combinaron las series 2, 3, 4 y 5, y se analizó la influencia de los filtros, 57 de 93 (61%) características radiómicas demostraron una alta reproducibilidad ($CCC > 0.90$), cuyos nombres se detallan en la Tabla 26.

Tabla 26. Nombres de las características de radiómica que mostraron un $CCC > 0.90$ después de la modificación de filtros (series 2 a 5) (n = 57).

Primer orden	Segundo orden				
	GLCM	GLDM	GLRLM	GLSZM	NGTDM
10th percentile	Cluster prominence	Dependence entropy	Gray level nonuniformity	Gray level nonuniformity	Coarseness
90th percentile	Cluster shade	Dependence nonuniformity	Gray level variance	Gray level variance	Complexity
Energy	Cluster tendency	Gray level nonuniformity	Run Entropy	Large area High gray level emphasis	Contrast
Entropy	Contrast	Gray level variance	Run length nonuniformity	Size Zone nonuniformity	Strength
Interquartile range	Correlation	Small dependence emphasis	Run length nonuniformity normalized	Zone entropy	
Kurtosis	Difference average	Small dependence low gray level emphasis		Zone percentage	
Maximum	Difference entropy				
Mean absolute deviation	Difference variance				
Mean	Inverse difference				
Median	Inverse difference moment				
Minimum	Inverse difference moment normalized				
Range	Inverse difference normalized				
Robust mean absolute deviation	Inverse variance				
Root means quared	Joint average				
Skewness	Joint entropy				
Total energy	Maximum Correlation Coefficient				
Variance	Sum Average				
	Sum entropy				
	Sum squares				

4.4.2.5. *Reproducibilidad tras eliminar el remuestreo y la normalización*

La normalización y el remuestreo son parte de los algoritmos de curación y armonización aplicados a las imágenes en el escenario del mundo real antes de la extracción de características radiómicas, por lo que se evaluó su impacto en las diferentes series y se analizó su reproducibilidad. Tras eliminar el remuestreo, 38 de 93 características mantuvieron un excelente nivel de acuerdo (40%), mientras que solo 5 de 93 (5%) exhibieron un $CCC > 0.90$ si no se aplicaba la normalización Z-score (Tabla 25).

Cuando se combinaron las series 6 y 7, hubo 41/93 (44%) características radiómicas que presentaron una alta reproducibilidad, que se detallan en la Tabla 27. Se evidenció que la normalización tiene el mayor impacto en la radiómica de primer y segundo orden, con un menor número de características reproducibles, ya que el valor de las intensidades se modifica tras aplicar el z-score.

Tabla 27. Características radiómicas que mostraron un $CCC > 0.90$ después de la aplicación de remuestreo y normalización combinadas (series 6 y 7) (n= 41).

Primer orden	Segundo orden				
	GLCM	GLDM	GLCM	GLSZM	GLCM
Remuestreo					
10th percentile	Cluster prominence	Gray level variance	Gray level nonuniformity normalized	Gray level nonuniformity normalized	Complexity
90th percentile	Cluster shade	Large dependence low gray level emphasis	Gray level variance	Gray level variance	
Entropy	Cluster tendency	Low gray level emphasis	Low gray level run emphasis		
Interquartile range	Difference Average				
Kurtosis	Difference Entropy				
Maximum	Difference variance				
Mean absolute deviation	Joint average				
Mean	Joint energy				
Median	Joint entropy				
Range	Maximum probability				
Robust mean absolute deviation	Sum Average				
Root means quared	Sum entropy				
Skewness	Sum squares				
Total energy					
Uniformity					
Variance					
Normalización					
Kurtosis	Correlation				
Skewness	Inverse difference momento normalized				
	Inverse difference normalized				

4.4.2.6. Reproducibilidad tras modificación de la máscara

Para evaluar el impacto en la radiómica tras la modificación de las máscaras (series 8 y 9), se incluyeron las características de radiómica de forma, por lo que se evaluaron 107 características de radiómica.

En la serie 8 (dilatación de la máscara), 79/107 (74%) características mostraron un excelente acuerdo en comparación con la serie de referencia. En la serie 9 (erosión de la máscara), se observó un mayor número de variables reproducibles con un CCC>0.90 (n=84, 78%) (Tabla 28). Todas las variables de forma permanecieron estables y reproducibles en ambas series 8 y 9 (14/14 características radiómicas, 100% en cada serie).

Tabla 28. Número y porcentaje de características de radiómica de primer y segundo orden que presentaron un excelente nivel de acuerdo para las 2 series con modificaciones en la forma de la máscara (series 8 y 9) en comparación con la serie de referencia (serie 1).

Modificación		Serie 8 Dilatación máscara	Serie 9 Erosión máscara	
Acuerdo excelente				Total
Forma (n=14)	n	14	14	
	%	100	100	100
Primer orden (n=18)	n	13	16	
	%	72.22	88.89	80.56
GLCM (n=24)	n	18	21	
	%	75	87.5	81.25
GLDM (n=14)	n	11	9	
	%	78.57	64.29	71.43
GLRM (n=16)	n	11	11	
	%	68.75	68.75	68.75
GLSZM (n=16)	n	8	9	
	%	50	56.25	53.13
NGTDM (n=5)	n	4	4	
	%	80	80	80
Total (n=107)	n	79	84	
	%	73.83	78.50	76.17

Cuando se combinaron las modificaciones de morfología, 75/107 características (70%) permanecieron con un CCC>0.90. Las características radiómicas con un excelente nivel de acuerdo con respecto a erosiones y dilataciones se enumeran en la Tabla 29.

Tabla 29. Características radiómicas que mostraron un CCC>0.90 después de la aplicación de dilataciones y erosiones en las máscaras (series 8 y 9) (n= 75).

Forma	Primer orden	Segundo orden				
		GLCM	GLDM	GLRLM	GLSZM	NGTDM
Dilation						
Elongation	90th percentile	Cluster prominence	Dependence entropy	Gray level nonuniformity	Gray level nonuniformity	Coarseness
Flatness	Energy	Cluster tendency	Dependence nonuniformity	Gray level nonuniformity normalized	Large area emphasis	Complexity
Least axis length	Entropy	Contrast	Dependence nonuniformity normalized	Gray level variance	Large area High gray level emphasis	Contrast
Major axis length	Interquartile range	Correlation	Dependence variance	Long run emphasis	Large area low gray level emphasis	Strength
Maximum 2D diameter (x-z plane)	Maximum	Difference average	Gray level nonuniformity	Long run low gray level emphasis	Size zone nonuniformity	
Maximum 2D diameter (y-z plane)	Mean absolute deviation	Difference entropy	Gray level variance	Low gray level run emphasis	Zone entropy	
Maximum 2D diameter (x-y plane)	Mean	Difference variance	Large dependence emphasis	Run length nonuniformity	Zone percentage	
Maximum 3D diameter	Median	Inverse difference	Large dependence low gray level emphasis	Run length nonuniformity normalized	Zone variance	
Volume	Robust mean absolute deviation	Inverse difference moment	Low gray level emphasis	Run percentage		
Minor axis length	Root mean squared	Inverse difference normalized	Small dependence emphasis	Run variance		
Sphericity	Total energy	Informational measure of correlation1	Small dependence low gray level emphasis	Short run emphasis		
Shape surface Area	Uniformity	Informational measure of correlation2				
Surface to volume ratio	Variance	Inverse variance				
Voxel volume		Joint energy				
		Joint entropy				
		Maximum probability				
		Sum entropy				
		Sum squares				

Erosión							
Elongation	10th percentile	Cluster prominence	Dependence entropy	Gray level nonuniformity	Gray level nonuniformity	Busyness	
Flatness	90th percentile	Cluster shade	Dependence nonuniformity	Gray level nonuniformity normalized	Gray level variance	Complexity	
Least axis length	Energy	Cluster tendency	Dependence nonuniformity normalized	Gray level variance	Large area emphasis	Contrast	
Major axis length	Entropy	Contrast	Dependence variance	Long run emphasis	Large area High gray level emphasis	Strength	
Maximum 2D diameter (x-z plane)	Interquartile range	Correlation	Gray level nonuniformity	Long run low gray level emphasis	Large area low gray level emphasis		
Maximum 2D diameter (y-z plane)	Maximum	Difference average	Gray level variance	Run entropy	Size zone nonuniformity		
Maximum 2D diameter (x-y plane)	Mean absolute deviation	Difference entropy	Large dependence emphasis	Run length nonuniformity	Zone entropy		
Maximum 3D diameter	Mean	Difference variance	Large dependence low gray level emphasis	Run length nonuniformity normalized	Zone percentage		
Volume	Median	Inverse difference	Small dependence emphasis	Run percentage	Zone variance		
Minor axis length	Range	Inverse difference moment		Run variance			
Sphericity	Robust mean absolute deviation	Inverse difference moment normalized		Short run emphasis			
Area	Root mean squared	Inverse difference normalized					
Surface to volume ratio	Skewness	Informational measure of correlation1					
Voxel volume	Total energy	Informational measure of correlation2					
	Uniformity	Inverse variance					
	Variance	Joint energy					
		Joint entropy					
		Maximum correlation coefficient					
		Maximum probability					
		Sum entropy					
		Sum squares					

4.4.2.7. Menor reproducibilidad después de perturbaciones en la imagen

Las características radiómicas de segundo orden GLSZM y GLRM fueron los grupos de parámetros que mostraron el menor porcentaje medio de acuerdo: GLRM mostró un acuerdo del 39% y GLSZM del 41% después de las modificaciones de preparación de imagen (series 2 a 7) (Tabla 25). GLRM mostró un 69% de acuerdo y GLSZM mostró un 53% de acuerdo después de las modificaciones de la máscara (series 8 y 9) (Tabla 28).

Hubo 15 características radiómicas de segundo orden que mostraron un nivel pobre de acuerdo en todas las series, como se muestra en la Tabla 30.

Tabla 30. Características radiómicas que mostraron un nivel pobre de acuerdo en todas las series en comparación con la verdad fundamental.

Forma	Primer orden	Segundo orden				
		GLCM	GLDM	GLRLM	GLSZM	NGTDM
		Autocorrelation	High gray level emphasis	High gray level run emphasis	High gray level zone emphasis	Busyness
			Large dependence high gray level emphasis	Long run high gray level emphasis	Lowgray level zone emphasis	
			Small dependence high gray level emphasis	Short run high gray level emphasis	Size zone nonuniformity normalized	
				Short run low gray level emphasis	Small area emphasis	
					Small area high gray level emphasis	
					Small area low gray level emphasis	

4.4.2.8. Análisis estadístico de reproducibilidad de radiómica

Para evaluar las diferencias entre las series se llevaron a cabo dos análisis exploratorios para las series 2 a 7 y para las series 8 y 9 por separado, y se realizaron gráficos de cajas y bigotes.

El análisis para las series 2 a 7 reveló diferencias significativas entre las series. Las series 6 y 7 mostraron un rendimiento notablemente inferior en comparación a las demás. Los gráficos de cajas y bigotes y *raincloud* para el CCC de las series 2 a 7 mostraron diferencias visuales entre las series. Las Series 6 y 7 mostraron un rendimiento más pobre, con valores de media y mediana más bajos, y una mayor proporción de casos por debajo del umbral de 0.9 para CCC, junto con valores de dispersión más altos. Las Series 2, 3, 4 y 5 mostraron resultados comparables entre sí (Figura 43).

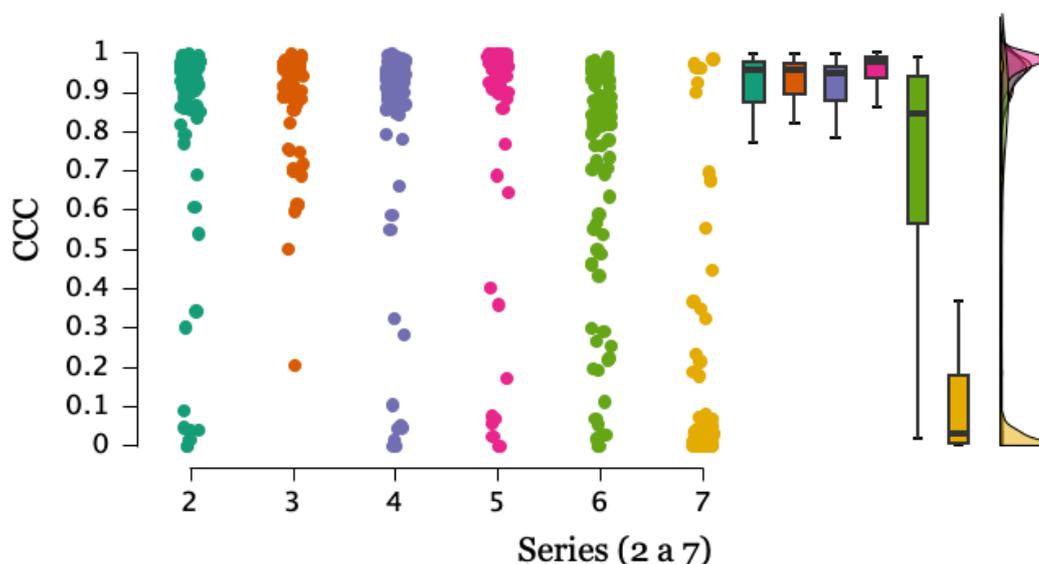


Figura 43. Gráficos de cajas y bigotes y *raincloud* para el Coeficiente de Correlación de Concordancia en las series 2 a 7, que muestran diferencias visuales entre las series. Las series 6 y 7 mostraron un rendimiento más pobre.

Para confirmar estos hallazgos se realizó un ANOVA para el grupo de series 2 a 7, tratando las series como un factor y considerando la estructura de los datos en bloques debido al cálculo de 93 parámetros para el primer grupo de series. Los resultados del ANOVA confirmaron la existencia de diferencias estadísticamente significativas entre las series para un nivel de significación de 0.05 (Tabla 31, Figura 44).

Tabla 31. Análisis de la varianza para las series 2 a 7. Los resultados del ANOVA confirmaron la existencia de diferencias estadísticamente significativas entre las series ($P < 0.001$).

ANOVA - CCC para series 2 a 7		
Casos	Prueba F	P valor
Serie_2_7	282.055	<0.001

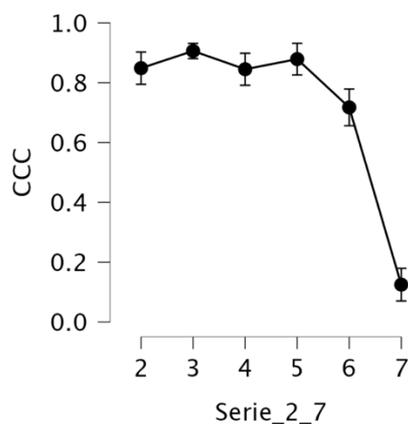


Figura 44. Medias marginales con intervalos de confianza para los valores de Coeficiente de Correlación de Concordancia obtenidos en las series 2 a 7, en los que se observan diferencias entre las series 6 y 7 y el resto de series.

Posteriormente, se aplicó una prueba de comparaciones múltiples post hoc utilizando el método de Bonferroni. El análisis reveló que no había diferencias significativas entre las Series 2, 3, 4 y 5. Sin embargo, las Series 6 y 7 mostraron diferencias significativas con el resto de las series y también entre ellas (Tabla 32).

Tabla 32. Comparaciones post hoc de las series 2 a 7. Se observa que las series 6 y 7 presentan diferencias significativas con el resto de series y también entre ellas.

Comparaciones Post Hoc - Series_2_7					
Series		Diferencia media	SE	t	p _{bonf}
2	3	-0.057	0.025	-2.286	0.341
	4	0.004	0.025	0.143	1.000
	5	-0.030	0.025	-1.204	1.000
	6	0.131	0.025	5.214	< 0.001
	7	0.723	0.025	28.770	< 0.001
3	4	0.061	0.025	2.429	0.233
	5	0.027	0.025	1.082	1.000
	6	0.189	0.025	7.500	< 0.001
	7	0.781	0.025	31.056	< 0.001

Comparaciones Post Hoc - Series_2_7					
Series		Diferencia media	SE	t	p _{bonf}
4	5	-0.034	0.025	-1.347	1.000
	6	0.128	0.025	5.071	< 0.001
	7	0.720	0.025	28.627	< 0.001
5	6	0.161	0.025	6.418	< 0.001
	7	0.754	0.025	29.973	< 0.001
6	7	0.592	0.025	23.556	< 0.001

Por su parte, las series 8 y 9 presentaron mejor rendimiento. En los gráficos de cajas y bigotes y *raincloud* para el CCC de las series 8 y 9 se observaron valores de CCC medios y medianos más altos y desviaciones estándar más bajas respecto al bloque de series 2 a 7 (Figura 45).

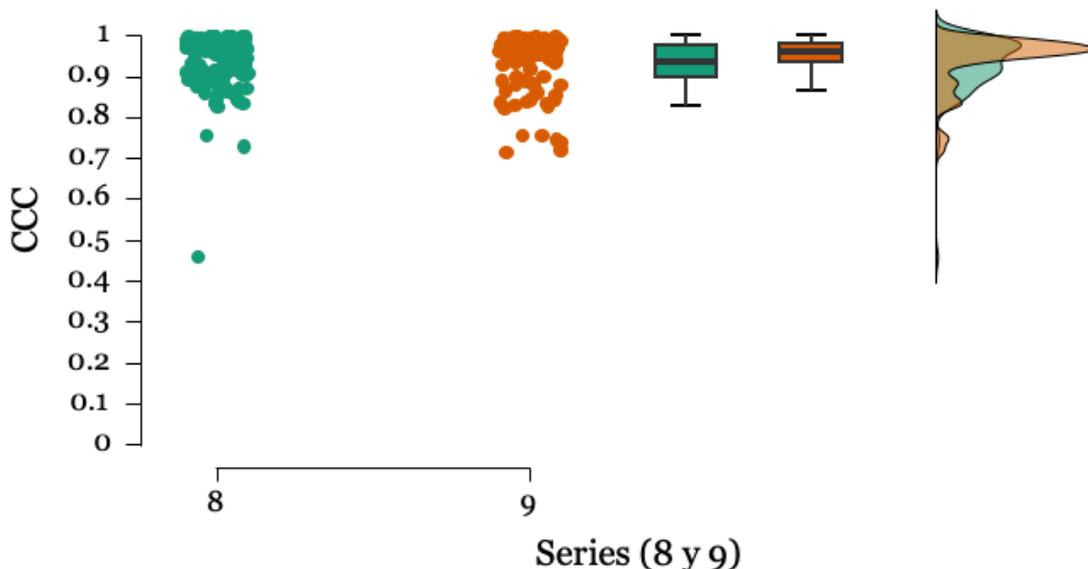


Figura 45. Gráficos de cajas y bigotes y *raincloud* para el Coeficiente de Correlación de Concordancia de las series 8 y 9.

Se realizó un ANOVA para las series 8 y 9, tratando las series como un factor y considerando la estructura organizativa de los datos en bloques debido al cálculo de 107 parámetros. Las Series 8 y 9 mostraron resultados más favorables para lograr valores altos de CCC. Estas dos series se diferenciaron en que se emplearon

o bien dilataciones o bien erosiones de la máscara. El análisis centrado en estas dos series, considerando sus datos emparejados, concluyó que no había diferencias significativas entre ellas (Tabla 33).

Tabla 33. Análisis de la varianza para las series 8 y 9. Los resultados confirmaron la ausencia de diferencias estadísticamente significativas entre las series ($P = 0.054$).

ANOVA - CCC para series 8 y 9		
Casos	Prueba F	P valor
Serie_2_7	3.806	0.054

4.4.2.9. Variabilidad intrasujeto

Al analizar la variabilidad intrasujeto a través del CoV, únicamente 8/107 características radiómicas, siendo todas de segundo orden, mostraron un resultado excelente (<0.10) en al menos dos series (7%), lo que indica una excelente repetibilidad (Tabla 34).

Tabla 34. Características radiómicas que mostraron un Coeficiente de Variación <0.10 (en negrita).

		Modificación								
		Serie 1 Referencia	Serie 2 Sin filtros	Serie 3 Eliminación ruido bilateral	Serie 4 Eliminación ruido CFF	Serie 5 Inhomogeneidades	Serie 6 Remuestreo	Serie 7 Normalización	Serie 8 Dilatación máscara	Serie 9 Erosión máscara
Características de radiómica	GLCM: Inverse difference moment normalized	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	GLCM: Inverse difference normalized	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
	GLDM: Dependence entropy	0.08	0.07	0.07	0.07	0.07	0.07	0.10	0.07	0.08
	GLRLM: Short run emphasis	0.09	0.07	0.08	0.07	0.09	0.08	0.07	0.08	0.10
	GLRLM: Run Entropy	0.10	0.10	0.09	0.09	0.10	0.10	0.17	0.09	0.11

GLRLM: Run percentage	0.12	0.10	0.11	0.10	0.12	0.11	0.09	0.11	0.13
GLSZM: Zone entropy	0.10	0.09	0.09	0.09	0.10	0.09	0.11	0.08	0.12
GLSZM: Small area emphasis	0.12	0.10	0.11	0.11	0.12	0.16	0.13	0.10	0.15

4.4.3. Relación de la radiómica con variables analíticas y de estadificación

Para llevar a cabo las fases finales de esta Tesis Doctoral, se añadieron 89 pacientes respecto al análisis de radiómica previo. Se dispuso, por tanto, de imágenes de RM de 508 pacientes incluidos en la plataforma PRIMAGE. Todas las RM seleccionadas para este estudio fueron en el momento del diagnóstico. Se empleó una secuencia de RM por paciente: se revisaron aquellos casos en los que un paciente presentase dos o más secuencias (previamente empleadas para el análisis de radiómica), y se seleccionó una secuencia por paciente. Para la selección, se priorizaron las secuencias T2wSE FS o STIR por presentar mejor contraste entre estructuras, y, en caso de presentar más de una secuencia T2wSE FS o STIR por paciente, se llevó a cabo una revisión visual y selección de la que presentase mayor resolución y menos artefactos (Figura 46).

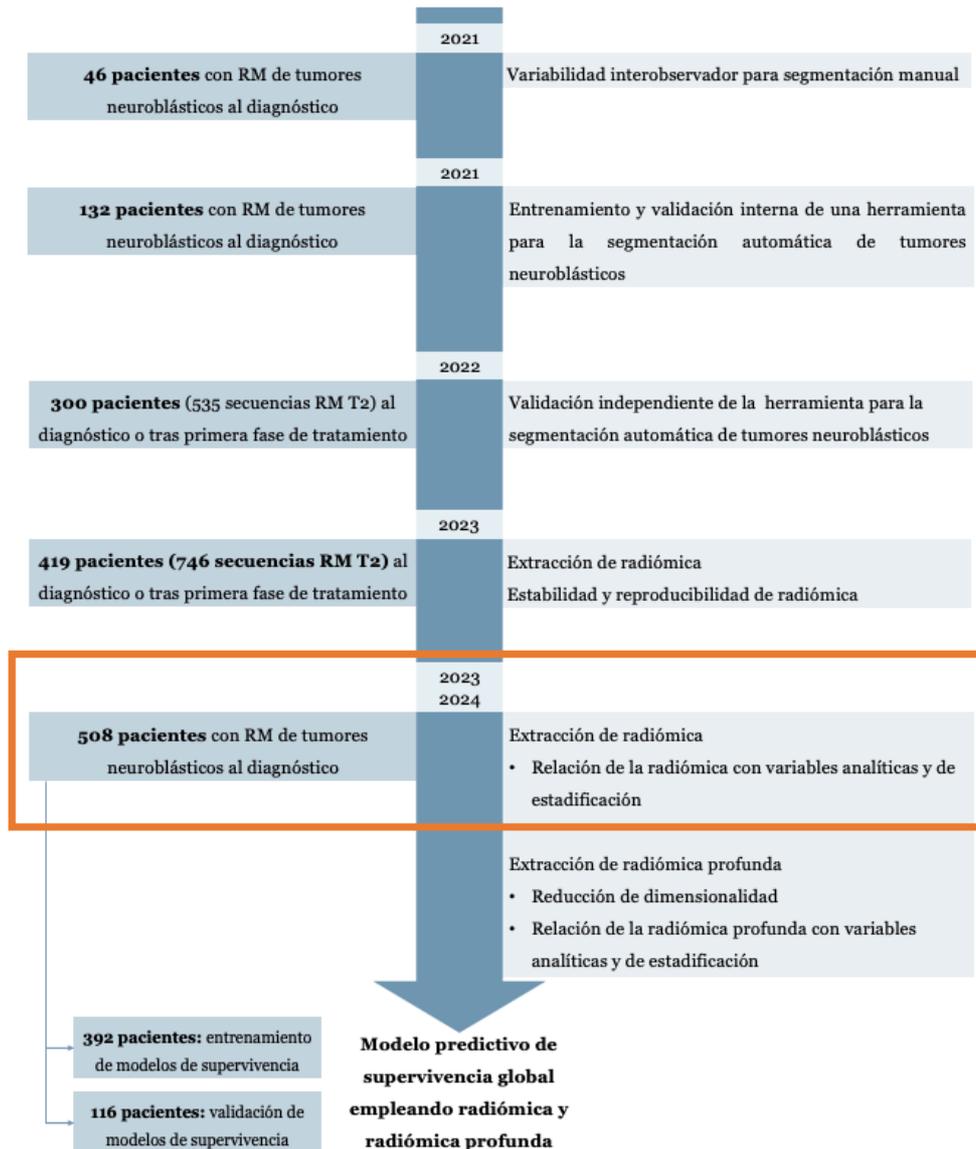


Figura 46. Flujo de trabajo: relación de radiómica con variables analíticas y de estadificación.

Las características de los pacientes se detallan en la Tabla 35. La distribución de proveedores de equipamiento de RM fue la siguiente: General Electric Healthcare (Signa Excite HDxt, Signa Explorer, Discovery, Genesis) n = 111; Siemens Medical (Aera, Skyra, Symphony, Avanto, Magneto Espree, TrioTim, Sonata, Sprecra, Verio, Harmony), n = 279); Philips Healthcare (Intera, Achieva, Ingenia, Panorama; (n = 115; o Canon (Titan; n = 3), y a la intensidad del campo magnético

(1.5 T (n = 408) o 3 T (n = 100)). De las 508 secuencias empleadas, 218 fueron T2wSE, 269 T2wSE FS y 21 T2wSE STIR.

Tabla 35. Características basales de la población estudiada en los análisis finales de esta Tesis Doctoral. Los datos se expresan como media \pm desviación estándar cuando la distribución es normal.

Descripción de la muestra		Número de pacientes
Pacientes	Número	508
Edad al diagnóstico	Media \pm desviación estándar (meses)	29 \pm 34
Sexo	Hombre	265
	Mujer	243
Histología	Neuroblastoma	455
	Ganglioneuroblastoma	34
	Ganglioneuroma	19
Localización	Abdominopélvico	415
	Cervicotorácico	93
Momento de realización del estudio	Al diagnóstico	508
	Tras tratamiento	0
Proveedor	GE	111
	Siemens	279
	Philips	115
	Canon	3
Intensidad del campo magnético	1.5T	408
	3T	100
Secuencia segmentada	T2wSE	218
	T2wSE FS	269
	T2wSE STIR	21
	T2*wGE FS	0

A continuación, se llevó a cabo un análisis de la relación entre una de las variables analíticas (parámetro molecular: estatus del protooncogén MYCN) y variables de estadificación (INRGSS, INRG) para tumores neuroblásticos con los 107 parámetros de radiómica previamente obtenidos.

4.4.3.1.MYCN

De los 508 pacientes incluidos en este análisis, 499 habían sido sometidos a análisis del estado del protooncogén MYCN (Tabla 6). Se dividió en dos grupos a los pacientes en función del MYCN: amplificado o no amplificado (Tabla 36).

Tabla 36. Distribución de la muestra en función del MYCN.

Variable analítica		Pacientes	Edad < 18 meses	Edad ≥18 meses
MYCN	Grupo 1: Amplificado	121	45	76
	Grupo 2: No amplificado + Ganancia	378	186	192

Tras comprobar la normalidad de los datos, se llevó a cabo la comparación entre dos grupos mediante t-test para grupos independientes, y U de Mann-Whitney en caso de no cumplir el supuesto de normalidad, ambas con un nivel de significancia de 0.05.

En el grupo de pacientes <18 meses se obtuvieron diferencias significativas en 6 variables de radiómica entre el grupo de MYCN amplificado frente al grupo no amplificado (detalladas en la Tabla 37). En el grupo de pacientes ≥18 meses no se obtuvieron diferencias significativas entre los grupos.

Tabla 37. Variables de radiómica con diferencias significativas en la amplificación de MYCN en <18 meses.

Variable radiómica	P valor	Prueba
Primer orden: Interquartile Range	0.044	Mann-Whitney
GLCM: Maximum Probability	0.049	Mann-Whitney
GLRLM: Run Length Non Uniformity	0.022	Mann-Whitney
GLRLM: Run Length Non Uniformity Normalized	0.047	T-test
GLSZM: Gray Level Non Uniformity	0.013	Mann-Whitney
GLSZM: Size Zone Non Uniformity	0.017	Mann-Whitney

4.4.3.2. INRGSS

Se realizó la estadificación INRGSS a todos los pacientes incluidos en este análisis (n = 508). Se dividió en tres grupos a los pacientes en función de la estadificación INRGSS (Tabla 38).

Tabla 38. Distribución de la muestra en función del INRGSS.

Variable estadificación		Pacientes	Edad < 18 meses	Edad ≥18 meses
INRGSS	Grupo 1: M	215	52	163
	Grupo 2: MS	42	42	0
	Grupo 3: L1 + L2	251	147	104

Tras evaluar la distribución de los datos y la heterogeneidad de la varianza, se compararon las diferencias entre los tres grupos. Si los datos seguían una distribución normal, se aplicó un test ANOVA. Si la distribución no era normal, se llevó a cabo un test de KW para grupos independientes, ambos con un nivel de significancia de 0.05. Finalmente se efectuó una prueba post hoc entre grupos siguiendo la prueba de Bonferroni, también con un nivel de significancia de 0.05.

En el grupo de pacientes <18 meses se obtuvieron diferencias significativas en 43 variables de radiómica (Tabla 39), todas ellas entre el grupo 1 (M) frente a los grupos 2 y 3 (Ms y L1 + L2), sin observar diferencias significativas entre los grupos 2 y 3 (Ms frente a L1 + L2).

Tabla 39. Variables de radiómica con diferencias significativas en función de INRGSS en pacientes <18 meses. Abreviaturas: G: grupo; P: valor de p.

Variable radiómica	G	G	p	Prueba	G	G	p	Prueba
Forma: Major Axis Length	1	2	0.032	ANOVA	1	3	0.016	ANOVA
Forma: Maximum 2D Diameter Column	1	2	0.013	KW	1	3	0.005	KW
Forma: Maximum 2D Diameter Row	1	2	0.019	KW	1	3	0.011	KW
Forma: Maximum 2D Diameter Slice	1	2	0.030	KW	1	3	0.006	KW
Forma: Maximum 3D Diameter	1	2	0.010	KW	1	3	0.004	KW
Forma: Mesh Volume	1	2	0.035	KW	1	3	0.002	KW

Forma: Minor Axis Length	1	2	0.046	KW	1	3	0.007	KW
Forma: Surface Area	1	2	0.019	KW	1	3	0.004	KW
Forma: Voxel Volume	1	2	0.035	KW	1	3	0.002	KW
Primer orden: Energy	1	2	0.039	KW	1	3	0.001	KW
Primer orden: Total Energy	1	2	0.039	KW	1	3	0.001	KW
GLDM: Dependence Non Uniformity	1	2	0.027	KW	1	3	0.002	KW
GLRLM: Run Length Non Uniformity	1	2	0.017	KW	1	3	0.002	KW
GLSZM: Gray Level Non Uniformity	1	2	0.012	KW	1	3	0.006	KW
GLSZM: Size Zone Non Uniformity	1	2	0.015	KW	1	3	0.031	KW
GLSZM: Zone Entropy	1	2	0.034	KW	1	3	0.018	KW
GLCM: Correlation	1	2	0.043	ANOVA				
GLDM: Dependence Entropy	1	2	0.020	KW				
Forma: Least Axis Length					1	3	0.011	ANOVA
Forma: Surface Volume Ratio					1	3	0.002	KW
GLDM: Gray Level Non Uniformity					1	3	0.009	KW
GLDM: Small Dependence Low Gray Level Emphasis					1	3	0.016	KW
GLRLM: Gray Level Non Uniformity					1	3	0.007	KW
GLSZM: Large Area High Gray Level Emphasis					1	3	0.011	KW
NGTDM: Busyness					1	3	0.017	KW
NGTDM: Coarseness					1	3	0.002	KW
NGTDM: Strength					1	3	0.023	KW

Para los pacientes ≥ 18 meses se obtuvieron diferencias significativas en una variable de radiómica entre los grupos 1 y 3. No se realizó este análisis de correlaciones respecto al grupo 2 al no encontrarse ningún paciente mayor de 18 meses en este grupo, puesto que, por definición, el grupo Ms incluye pacientes menores de 18 meses (Tabla 40):

Tabla 40. Variables de radiómica profunda con diferencias significativas en función de INRGSS en pacientes ≥ 18 meses.

Variable radiómica	Grupo	Grupo	p-valor	Prueba
Primer orden: Skewness	1	3	0.016	KW

4.4.3.3. INRG

Tras evaluar la distribución de los datos y la heterogeneidad de la varianza, se compararon las diferencias entre los tres grupos en función del riesgo estratificado por el sistema INRG, como se detalla en la Tabla 41. De los 508 pacientes incluidos, se obtuvo la estadificación INRG en 375 (Tabla 6) (en los casos restantes no se disponía de todos los datos necesarios para poder establecer la estadificación).

Tabla 41. Distribución de la muestra en función del INRGSS.

Variable estadificación		Pacientes	Edad < 18 meses	Edad ≥18 meses
INRG	Grupo 1: alto	275	54	221
	Grupo 2: intermedio	23	5	18
	Grupo 3: bajo + muy bajo	77	4	73

De nuevo, se compararon medias entre más de dos grupos.

En los pacientes <18 meses no se obtuvieron diferencias significativas entre grupos. Para los pacientes ≥18 meses se obtuvieron diferencias significativas en 5 variables de radiómica entre los grupos 1 y 3 (riesgo alto frente a riesgo bajo/muy bajo) (Tabla 42).

Tabla 42. Variables de radiómica con diferencias significativas en función de INRG en pacientes ≥18 meses.

Variable radiómica	Grupo	Grupo	P-valor	Prueba
GLCM: Inverse difference moment normalized	1	3	0.044	KW
GLCM: Inverse difference normalized	1	3	0.023	KW
GLDM: Dependence Variance	1	3	0.027	KW
GLDM: Large Dependence High Gray Level Emphasis	1	3	0.039	KW
GLDM: Small Dependence Low Gray Level Emphasis	1	3	0.031	KW

4.4.3.4. Análisis de los resultados

El análisis de la relación de las 107 variables de radiómica con las variables analíticas y de estadificación mostró relación estadísticamente significativa en 55 casos. El mayor número de variables significativas se observó en el INRGSS, donde se identificaron diferencias significativas en 43 parámetros de radiómica del grupo M frente al resto de grupos en menores de 18 meses. Por ello, la radiómica podría constituir una herramienta de diferenciación de pacientes con metástasis en pacientes con neuroblastoma menores de 18 meses.

5 variables radiómicas mostraron relación significativa para el INRG en mayores de 18 meses, con diferencias entre los grupos de alto y bajo riesgo, por lo que se evidencian diferencias en características de las imágenes en tumores de alto frente a los de bajo riesgo.

El MYCN no mostró relación significativa con la radiómica en mayores de 18 meses (Tabla 43).

Tabla 43. Resumen de los resultados de las relaciones estadísticamente significativas en cada subgrupo según la edad y las variables analíticas y de estadificación con las variables de radiómica.

Variable analítica o de estadificación	Número de variables de radiómica con relación estadísticamente significativa	
	Edad <18 meses	Edad ≥18 meses
MYCN	6	0
INRGSS	43	1
INRG	0	5
Total	49	6

Por tanto, 55 parámetros de radiómica mostraron relaciones significativas en los distintos análisis llevados a cabo. Algunas características de radiómica mostraron relación significativa en más de uno de los análisis.

En cuanto a la radiómica de forma, se observó relación estadísticamente significativa en 20 ocasiones, correspondiendo con 11 de las 14 características radiómicas de forma analizadas (79%). De ellas, distintas variables se repitieron

en algunos de los análisis (entre paréntesis se indica el número de repeticiones): Major Axis Length (2), Maximum 2D Diameter Column (2), Maximum 2D Diameter Row (2), Maximum 2D Diameter Slice (2), Maximum 3D Diameter (2), Mesh Volume (2), Minor Axis Length (2), Surface Area (2), Voxel Volume (2).

En el análisis de la radiómica de primer orden se observaron 6 ocasiones en las que se observaron diferencias estadísticamente significativas, correspondientes con 4 de las 18 variables de radiómica de primer orden analizadas (22%). Mostraron repetición en diversos análisis las variables: energy (2), total energy (2).

Respecto a la radiómica de segundo orden, se observaron relaciones significativas en 4 variables del grupo GLCM (de 24 variables analizadas, 16%). En el grupo GLDM se observaron relaciones significativas en 8 ocasiones, de las cuales mostraron repeticiones las siguientes variables: Dependence Non Uniformity (2), Small Dependence Low Gray Level Emphasis (2), por lo que 6/14 variables incluidas fueron estadísticamente significativas (43%). Se observaron resultados significativos en 5 de las características GLRLM, con repetición de la variable Run Length Non Uniformity (3), por lo que 3/16 de las variables incluidas fueron significativas (18%). En cuanto a GLSZM, se observaron 9 variables, mostrando repetición Gray Level Non Uniformity (3), Size Zone Non Uniformity (3), Zone Entropy (2), por lo que 6/16 variables fueron representativas (38%). Finalmente, 3 de las 5 variables NGTDM mostraron relaciones significativas (60%).

4.5. Estudio de radiómica profunda

4.5.1. Extracción de radiómica profunda

Como paso siguiente en esta Tesis Doctoral, se llevó a cabo la extracción de características de radiómica profunda (Figura 47), características que capturan información relevante y discriminativa de las imágenes, como patrones, texturas y estructuras importantes, que pueden ser cruciales para tareas de clasificación, detección, segmentación e identificación de biomarcadores.

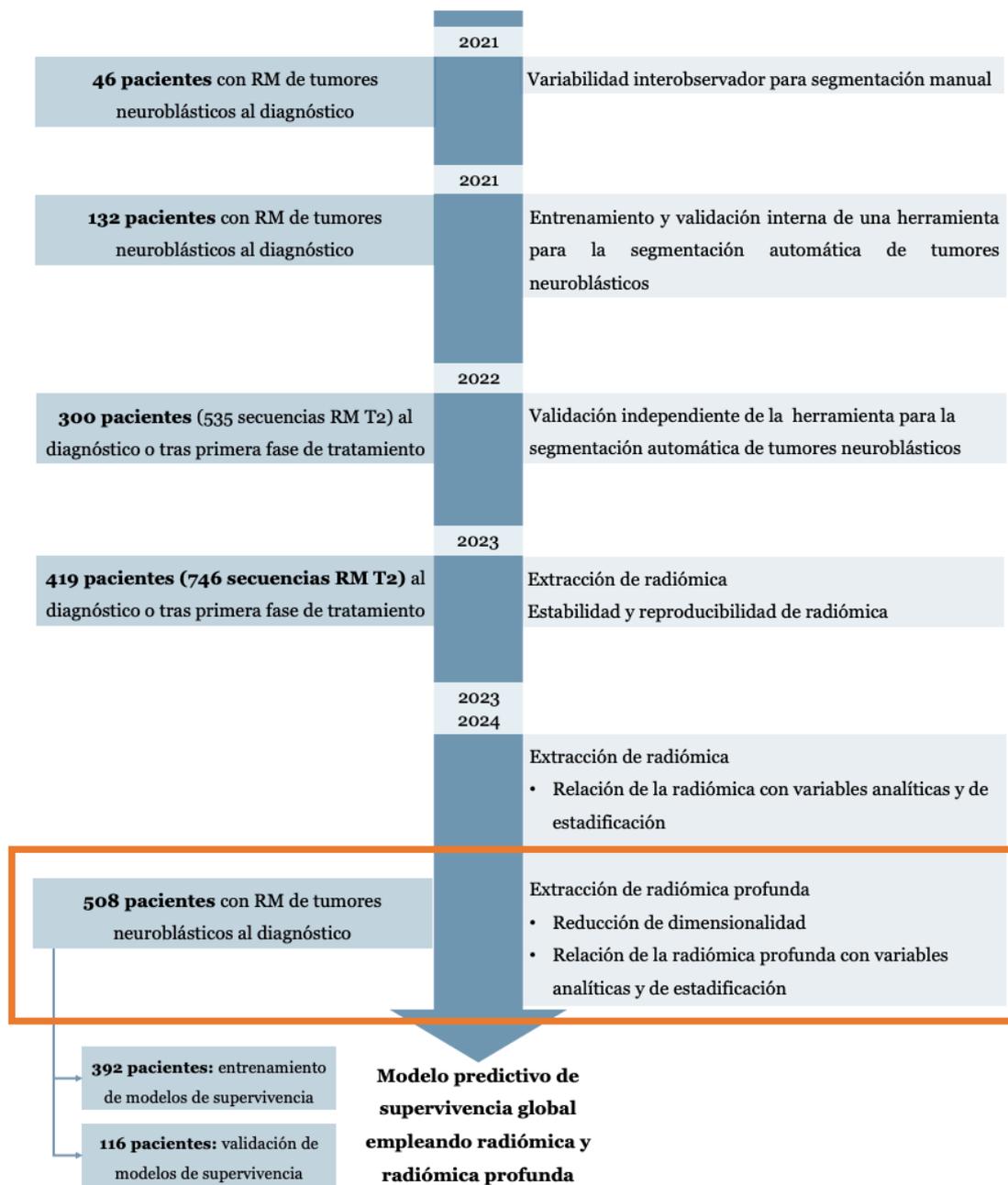


Figura 47. Flujo de trabajo: extracción de radiómica profunda y relación con variables analíticas y de estadificación.

La radiómica se extrajo a partir de imágenes de RM de 508 pacientes incluidos en la plataforma PRIMAGE. Todas las RM seleccionadas para este estudio fueron en el momento del diagnóstico. Se empleó una secuencia de RM por paciente.

Las características de los pacientes se han detallado previamente en la Tabla 35.

Para la extracción de características se empleó una CNN preentrenada con una amplia gama de imágenes que permite captar los patrones de las imágenes de manera más eficaz y robusta que si se hubiese entrenado una red neuronal *ad hoc* con las imágenes médicas disponibles en este estudio. La red empleada, llamada ResNet50, es una CNN con 50 capas de profundidad. Esta red ha sido entrenada con más de 1 millón de imágenes de la base de datos ImageNet [118], incluyendo imágenes médicas en 2D como radiografías, TC o RM, y puede clasificar imágenes en 1000 categorías de objetos.

La extracción de radiómica profunda en esta Tesis Doctoral se realizó utilizando el módulo de análisis correspondiente, usando Python como lenguaje de programación y la librería ResNet50. El proceso se implementó en la plataforma de PRIMAGE. Las características profundas se extrajeron de cada paciente en el corte de RM donde el tumor presentase un tamaño mayor.

Durante el proceso de extracción de estas características, la imagen de entrada en la CNN se somete a una serie de convoluciones produciendo un conjunto de mapas de características. Estos mapas se reducen de tamaño mediante una capa de agrupamiento máximo y se someten nuevamente a otro conjunto de convoluciones. La capa previa a la última fase de la red contiene representaciones intermedias de los datos que han sido aprendidas a medida que se procesa la información, y proporciona una representación detallada de las características extraídas por las capas convolucionales y de agrupamiento. Finalmente, estos mapas de características obtenidos se transforman en un vector unidimensional para realizar la tarea final de clasificación o regresión. En este caso, los datos se extrajeron de la capa previa al clasificador final de la red, ya que contiene representaciones intermedias de los datos que han sido aprendidas por la red a medida que procesa la información, y proporcionan una representación de los datos en lugar de predicciones finales (Figura 48).

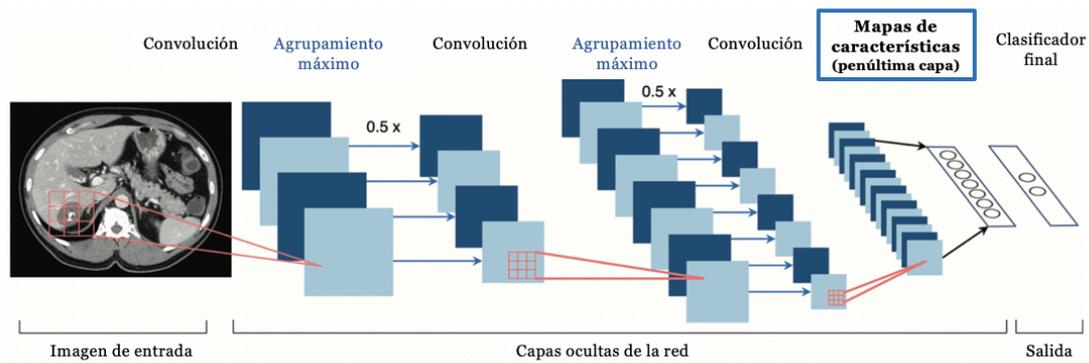


Figura 48. Esquema de funcionamiento de una red neuronal convolucional. Adaptado de Chartrand et al [40].

Finalmente, las características de radiómica profunda extraídas a través de la red neuronal fueron introducidas en un modelo de aprendizaje automático para su posterior análisis y obtención de resultados. Esto representa una estrategia eficaz para combinar las capacidades de extracción de características de una red neuronal con la capacidad de modelado de un algoritmo de aprendizaje automático y permite su integración con otros tipos de variables. Se extrajeron un total de 2048 características de radiómica profunda.

4.5.2. Reducción de variables de radiómica profunda

Se llevó a cabo una reducción de la dimensionalidad de las variables de radiómica profunda con la finalidad de eliminar características redundantes e inestables, justificada por la “maldición de la dimensionalidad”, atribuida a Richard E. Bellman [118], que expone que la alta dimensionalidad representa un desafío significativo para el análisis y la interpretación de los datos al aumentar el número de variables, por lo que es fundamental comprender su impacto en los datos.

En primer lugar, se extrajo la información mutua entre las 2048 variables originales. La información mutua entre dos variables aleatorias es un valor no-negativo que mide la dependencia entre variables. Es igual a 0 únicamente si dos variables aleatorias son independientes entre sí, y los valores mayores implican mayor dependencia. Se eliminó el 75.6% de las variables con mayor información mutua, para mantener las 500 variables con mayor independencia entre sí.

A continuación, se realizó un estudio de varianzas para reducir más la dimensionalidad. Se extrajo la varianza de cada una de las características de radiómica profunda, y se obtuvo la distribución de las varianzas del conjunto de datos, descartando aquellas cuya varianza se situase en el percentil 5 (varianza pequeña), considerando que las variables cuya varianza es pequeña presentarán un bajo poder predictivo al no observarse diferencias notables con respecto a la población a estudio. Posteriormente, se construyeron matrices de correlación que permitieron la extracción de correlaciones mediante el coeficiente de Pearson (para evaluar la fuerza y la dirección de la relación lineal entre dos variables continuas). Se revisaron los pares de mayor a menor correlación y se estableció un límite de 0.75 en valor absoluto, de modo que, si se evidenció una correlación mayor del 75% entre ambas variables, se descartó una de ellas. Para determinar qué variable entre ambas descartar, se aplicó una transformación a factores mediante un análisis factorial con rotación de varimax a todas las variables, en el cual se logró transformar las variables originales a unos nuevos factores de una dimensionalidad mucho menor, maximizando de manera simultánea las ponderaciones en cada uno para que las variables originales estuviesen representadas mayoritariamente en uno de estos factores. Tras esta transformación, se descartaron aquellas variables que tuvieran un peso menor para cada factor. Este proceso se realizó de forma iterativa para reducir la dimensionalidad y redundancias, hasta eliminar todos los pares de correlaciones tanto positivas como negativas con un valor mayor del 75%, seleccionando finalmente 50 factores.

4.5.3. Relación de la radiómica profunda con variables analíticas y de estadificación

Tras la extracción de variables de radiómica profunda y la reducción de su dimensionalidad, se llevó a cabo un análisis de la relación entre una de las variables analíticas (parámetro molecular: estatus del protooncogén MYCN) y variables de estadificación (INRGSS, INRG) para tumores neuroblásticos con los 50 parámetros de radiómica profunda reducida.

4.5.3.1. MYCN

Como se ha expuesto en el apartado 4.5.3.1., Se dividió en dos grupos a los pacientes en función del MYCN: amplificado o no amplificado (Tabla 36).

Tras comprobar la normalidad de los datos, se llevó a cabo la comparación entre dos grupos mediante t-test para grupos independientes, y U de Mann-Whitney en caso de no cumplir el supuesto de normalidad, ambas con un nivel de significancia de 0.05

En el grupo de pacientes <18 meses se obtuvieron diferencias significativas en 2 variables de radiómica profunda entre los grupos 1 y 2 respecto al MYCN (Tabla 44).

Tabla 44. Variables de radiómica profunda con diferencias significativas en la amplificación de MYCN en <18 meses.

Variable radiómica profunda	Valor p	Prueba
18	0.046	Mann-Whitney
33	0.000	Mann-Whitney

En el grupo de pacientes ≥ 18 meses se obtuvieron diferencias significativas en 9 variables de radiómica profunda (Tabla 45).

Tabla 45. Variables de radiómica profunda con diferencias significativas en la amplificación de MYCN en ≥ 18 meses.

Variable radiómica profunda	P valor	Prueba
5	0.045	Mann-Whitney
8	0.002	Mann-Whitney
9	0.047	Mann-Whitney
10	0.016	Mann-Whitney
14	0.016	T-test
25	0.022	Mann-Whitney
36	0.006	Mann-Whitney
48	0.033	Mann-Whitney
50	0.032	Mann-Whitney

4.5.3.2. INRGSS

Como se ha expuesto en el apartado 4.5.3.2., se dividió en tres grupos a los pacientes en función de la estadificación INRGSS (Tabla 38).

Tras evaluar la distribución de los datos y la heterogeneidad de la varianza, se compararon las diferencias entre los tres grupos. En caso de seguir una distribución normal se empleó un test ANOVA. En caso de que la distribución no fuese normal, se realizó un test de KW para grupos independientes, ambas con un nivel de significancia de 0.05. Posteriormente, se realizó una prueba post hoc entre grupos con la prueba de Bonferroni, también con un nivel de significancia de 0.05.

En el grupo de pacientes <18 meses se obtuvieron diferencias significativas en 4 variables de radiómica profunda (Tabla 46), todas ellas entre el grupo 3 (L1 + L2) frente a los grupos 1 y 2 (pacientes con metástasis), sin observar diferencias significativas entre los grupos 1 (M) y 2 (Ms).

Tabla 46. Variables de radiómica profunda con diferencias significativas en función de INRGSS en pacientes <18 meses.

Variable	Grupo	Grupo	p-valor	Prueba
7	1	3	0.006	KW
9	1	3	0.005	KW
10	2	3	0.035	ANOVA
16	2	3	0.011	KW

Para los pacientes ≥ 18 meses se obtuvieron diferencias significativas en 5 variables de radiómica profunda entre los grupos 1 y 3. No se pudo realizar este análisis de correlaciones respecto al grupo 2 al no encontrarse ningún paciente mayor de 18 meses en este grupo, puesto que, por definición, el grupo Ms incluye pacientes menores de 18 meses) (Tabla 47).

Tabla 47. Variables de radiómica profunda con diferencias significativas en función de INRGSS en pacientes ≥ 18 meses.

Variable	Grupo	Grupo	p-valor	Prueba
12	1	3	0.014	KW
16	1	3	0.003	KW

39	1	3	0.012	KW
4	1	3	0.003	KW
48	1	3	0.006	KW

4.5.3.3. INRG

Tras evaluar la distribución de los datos y la heterogeneidad de la varianza, se compararon las diferencias entre los tres grupos en función del riesgo estratificado por el sistema INRG, como se ha expuesto en la sección 4.5.3.3. (Tabla 41).

En los pacientes <18 meses se obtuvieron diferencias significativas entre grupos para 4 variables de radiómica profunda (Tabla 48). Cabe destacar que el tamaño muestral de los grupos 2 y 3 en pacientes <18 meses fue pequeño (5 y 4 pacientes, respectivamente) por lo que los resultados podrían no ser confiables, ya que la muestra más pequeña puede introducir más variabilidad y aumentar la posibilidad de obtener resultados no representativos de la verdadera relación entre las variables.

Tabla 48. Variables de radiómica profunda con diferencias significativas en función de INRG en pacientes <18 meses.

Variable	Grupo	Grupo	P valor	Prueba
9	1	3	0.024	KW
28	1	3	0.017	KW
33	1	3	0.009	KW
49	1	3	0.011	KW

Para los pacientes ≥ 18 meses se obtuvieron diferencias significativas en 5 variables de radiómica profunda. En ambos casos las diferencias fueron entre los grupos 1 y 3 (riesgo alto y riesgo bajo/muy bajo) (Tabla 49).

Tabla 49. Variables de radiómica profunda con diferencias significativas en función de INRG en pacientes ≥ 18 meses.

Variable	Grupo	Grupo	P valor	Prueba
3	1	3	0.005	KW
4	1	3	0.005	KW

14	1	3	0.004	ANOVA
16	1	3	0.028	KW
45	1	3	0.045	KW

4.5.3.4. Análisis de los resultados

El análisis de la relación de las 50 variables de radiómica profunda con las variables analíticas y de estadificación mostró relación estadísticamente significativa en 29 casos (Tabla 50). El mayor número de variables significativas se observó en el MYCN en pacientes ≥ 18 meses.

Tabla 50. Resumen de los resultados de las relaciones estadísticamente significativas en cada subgrupo según la edad y las variables analíticas y de estadificación con la radiómica profunda.

Variable analítica o de estadificación	Número de variables de radiómica profunda con relación estadísticamente significativa	
	Edad <18 meses	Edad ≥ 18 meses
MYCN	2	9
INRGSS	4	5
INRG	4	5
Total	10	19

Por tanto, 29 parámetros de radiómica profunda mostraron relaciones significativas en los distintos análisis llevados a cabo. De ellas, las variables 4, 10, 14, 33 y 48 se repitieron en dos ocasiones, mientras que las variables 9 y 16 se repitieron en 3 ocasiones.

Por consiguiente, la correlación entre la radiómica y la radiómica profunda con una variedad de parámetros analíticos y de estadificación permite la identificación de patrones subyacentes y fenotipos tumorales. Estos análisis subrayan la utilidad de la correlación de imágenes con parámetros como la expresión génica, abriendo la puerta al campo emergente de la radiogenómica, que posee un potencial considerable para la extracción de la heterogeneidad tumoral.

4.6. Modelo predictivo de supervivencia global.

Tras realizar el análisis de relaciones entre parámetros analíticos y de estadificación de manera univariable, se procedió a explorar las capacidades de las características de imagen de manera multivariable, construyendo modelos predictivos de tiempo hasta evento para predecir la supervivencia global (Figura 49).

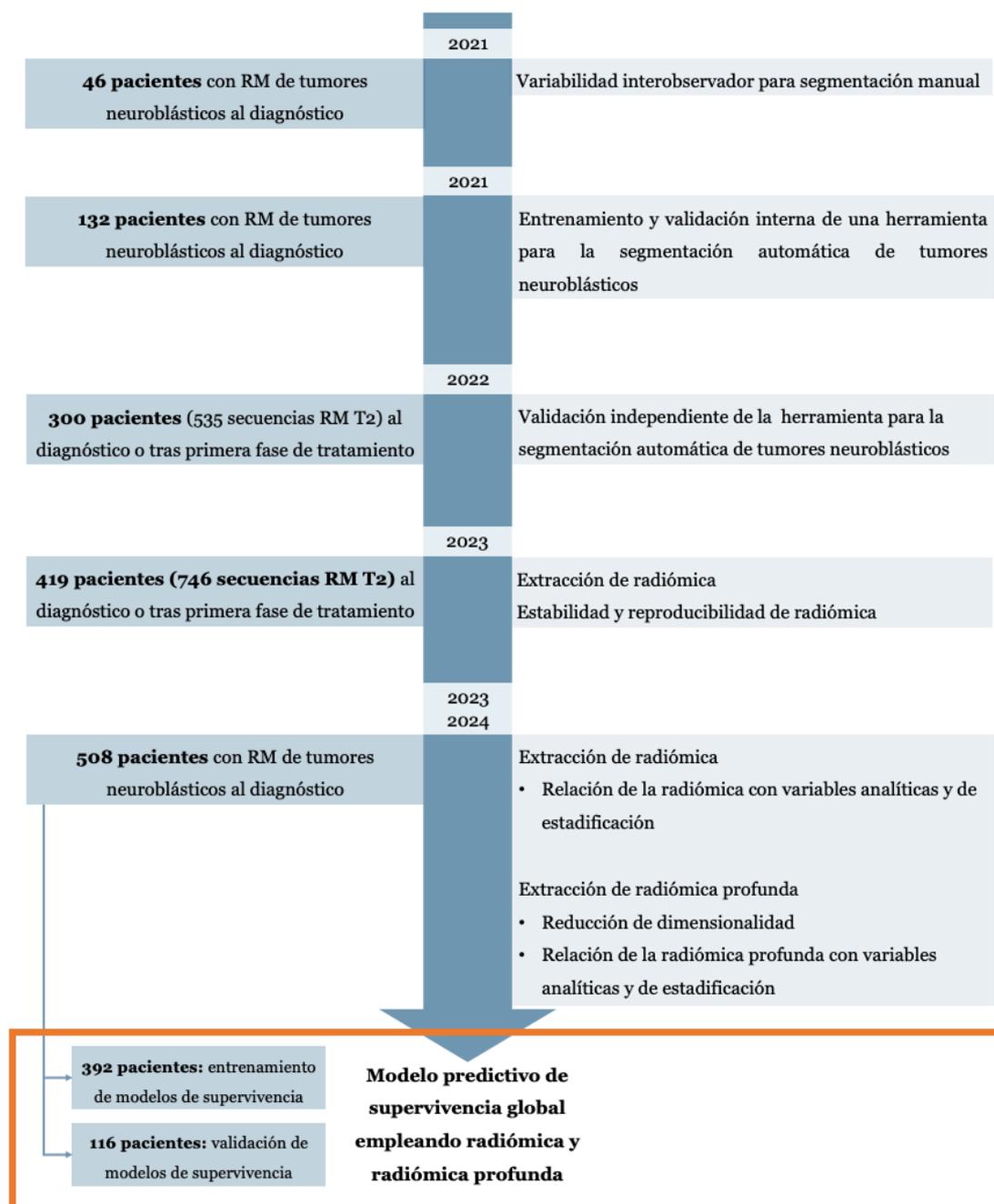


Figura 49. Flujo de trabajo: construcción de un modelo predictivo de supervivencia global.

Los modelos de tiempo hasta evento se abordaron como modelos de supervivencia o análisis de supervivencia, un tipo de problema de regresión donde se debe predecir un valor continuo, pero difiere de la regresión tradicional en el hecho de que parte de los datos solo pueden ser observados parcialmente y, por lo tanto, están censurados. El análisis de supervivencia tiene como objetivo establecer una conexión entre las covariables y el tiempo de un evento, estableciendo la probabilidad de que suceda un evento en un tiempo determinado.

El modelo se realizó con las 107 características radiómicas y las 50 características de radiómica profunda (factores de radiómica profunda seleccionados mediante técnicas de reducción de dimensionalidad detalladas en la sección 4.6.2), todas ellas previamente normalizadas mediante normalización z-score con la finalidad de evitar que el modelo únicamente seleccionase aquellas con mayores valores absolutos.

Para garantizar la robustez de los resultados, se llevó a cabo una partición de los casos. De los 508 pacientes disponibles para la realización de modelos finales, se dividió el conjunto de datos en un 77% para el entrenamiento ($n = 392$) y un 23% para la validación ($n = 116$). Esta distribución se realizó en función de la procedencia de los casos, empleando para la validación externa los casos procedentes del Hospital Universitari i Politècnic La Fe, y para el entrenamiento los casos procedentes del resto de instituciones: HR-NBL1/SIOPEN (NCT00030719), liderado por St. Anna Children's Hospital y CCRI (Viena, Austria), pacientes del Registro Alemán de Neuroblastoma NB2016 de la Sociedad de Oncología y Hematología Pediátrica, dirigido por el Departamento de Oncología Pediátrica del Hospital Universitario de Niños de Colonia, pacientes del Hospital Universitario de Pisa, Instituto Gaslini (Pisa, Italia), así como pacientes del Hospital La Paz y Hospital Niño Jesús (Madrid, España), Hospital Son Espases (Palma de Mallorca, España), y Hospital Reina Sofía (Córdoba, España).

La evaluación de modelos de supervivencia implica métricas específicas diseñadas para medir la capacidad del modelo para predecir el tiempo hasta el evento de interés. Se emplearon (como se detalla en el apartado 3.9.4), el C-index, el AUC-ROC y la IPW.

Para construir el modelo de supervivencia global, se escogió la arquitectura Cox (*Cox Proportional Hazards*): el modelo de riesgos proporcionales de Cox es uno de los modelos estadísticos más utilizados en análisis de supervivencia. Se emplea para desarrollar un modelo predictivo destinado a analizar datos de tiempo hasta la ocurrencia de un evento, generando una función de supervivencia que estima la probabilidad de que ocurra el evento de interés en un momento específico, considerando los valores de las variables predictoras pertinentes. La formulación de la función de supervivencia y la determinación de los coeficientes de regresión asociados a los predictores se logran mediante la utilización de los sujetos observados en el estudio. Posteriormente, el modelo puede ser aplicado a nuevos casos que dispongan de mediciones para las variables predictoras, facilitando así la predicción del tiempo hasta el evento de interés para estos casos. La información proveniente de los sujetos censurados, es decir, aquellos que no han experimentado el evento de interés durante el período de observación, también contribuye de manera significativa a la estimación precisa del modelo. El modelo asume que la relación entre las covariables y la función de riesgo es constante a lo largo del tiempo [120].

4.6.1. Entrenamiento de un modelo predictivo de supervivencia global

El entrenamiento del modelo se realizó con 392 pacientes. Para ello se empleó una metodología de validación cruzada, dividiendo a la muestra ($n = 392$) en 5 subconjuntos o *folds* con el objetivo de aumentar la robustez y generalizabilidad del modelo, así como de observar posibles sesgos en los datos de entrenamiento. Tres de los subconjuntos constaron de 78 pacientes, y dos de los subconjuntos estuvieron conformados por 79 pacientes.

Para el grupo de entrenamiento del modelo de predicción de supervivencia global desarrollado con 107 características radiómicas y 50 características de radiómica

profunda se obtuvo un índice AUC de 0.60 y un C-index de 0.63. Los resultados muestran la media y la desviación estándar de los 5 subconjuntos empleados mediante validación cruzada (Tabla 51).

Tabla 51. Resultados del modelo de supervivencia Cox para el conjunto de datos de entrenamiento.

	C-index media +/- DE	AUC media +/- DE	IPW media +/- DE
Modelo COX	0.63 +/- 0.07	0.60 +/- 0.10	0.61 +/- 0.13

Se realizó un entrenamiento del modelo y se evaluó el AUC-ROC en función del tiempo desde el momento del diagnóstico del paciente para el conjunto de datos de entrenamiento (para cada uno de los 5 subconjuntos). Se observan las diferentes curvas de AUC ROC correspondientes a los 5 subconjuntos empleados mediante técnicas de validación cruzada (Figura 50).

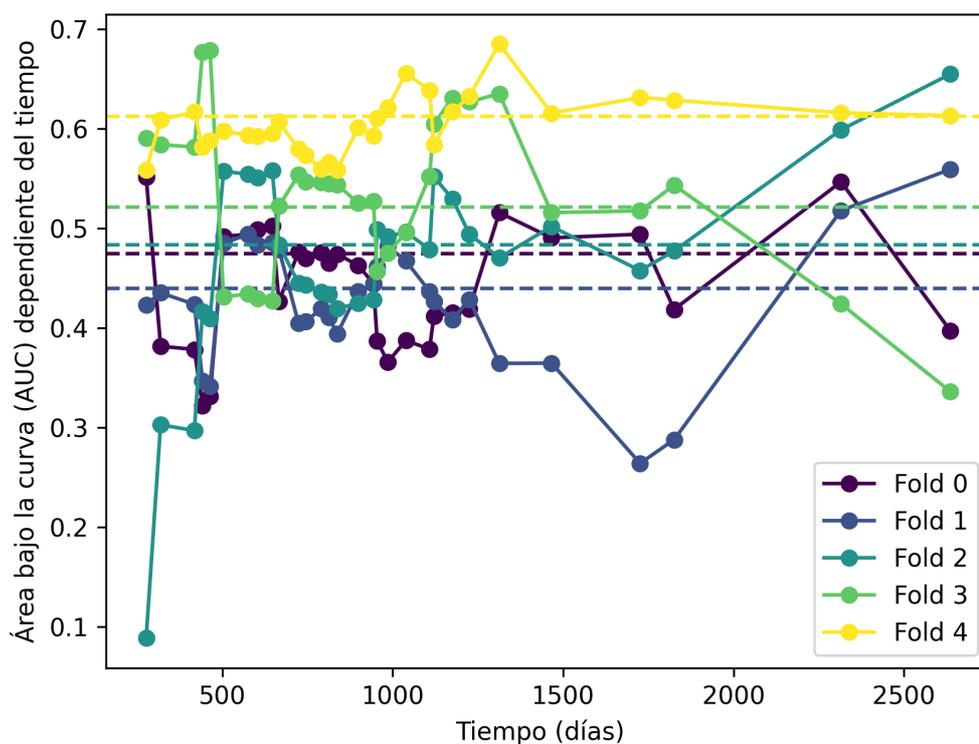


Figura 50. Área bajo la curva en función del tiempo de los grupos de entrenamiento de validación cruzada para el modelo de supervivencia global. Las líneas discontinuas representan la media de tiempo de cada subconjunto.

Se obtuvieron curvas de supervivencia para cada subconjunto evidenciando una menor similitud de las predicciones del modelo respecto a los datos reales en todos los subconjuntos conforme aumentó el tiempo transcurrido desde el momento del diagnóstico de cada paciente. Se observó una tendencia en las curvas de supervivencia predichas para cada subconjunto, con una pendiente más abrupta respecto a los datos reales y menor probabilidad de supervivencia en función del tiempo. Se representaron las curvas con el intervalo de confianza del 95% para la estimación de supervivencia en cada punto específico en el tiempo (bandas adyacentes a la curva), para mostrar la incertidumbre asociada con la estimación de la tasa de supervivencia en cada punto del tiempo (Figura 51).

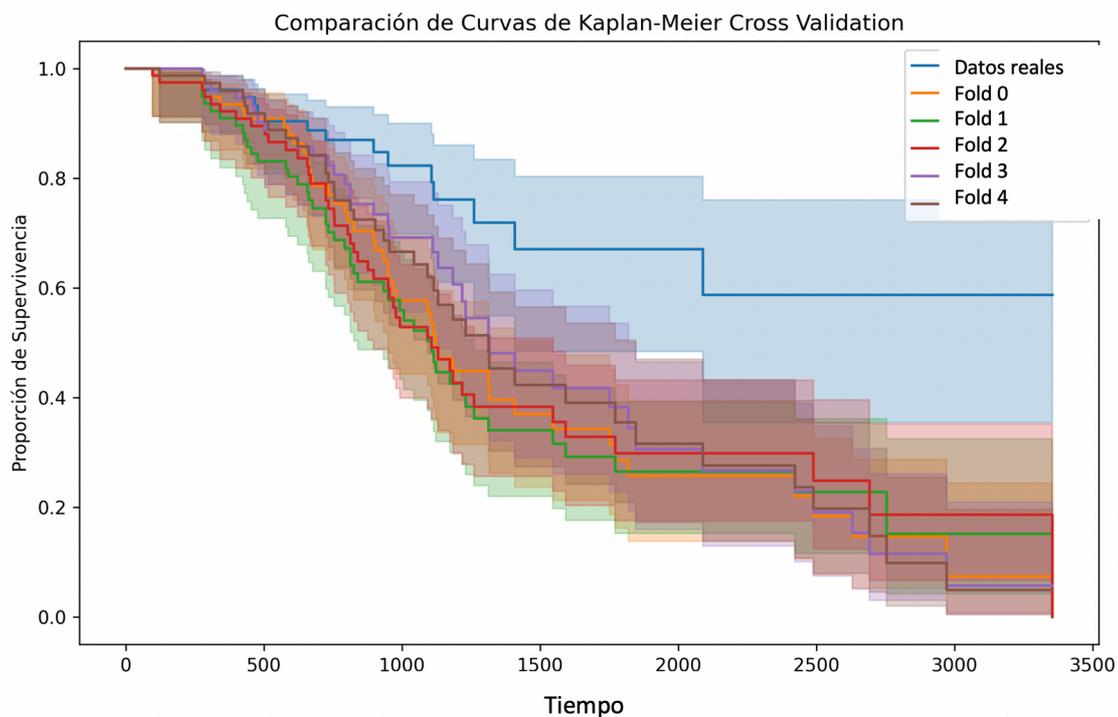


Figura 51. Curvas Kaplan-Meier del modelo en función del tiempo para los 392 pacientes del grupo de entrenamiento, divididos en 5 subconjuntos de datos. La curva azul representa los datos reales. Curvas con intervalo de confianza del 95%.

4.6.2. Validación de un modelo de supervivencia global

Con el objetivo de evaluar la universalidad y capacidad de generalización del modelo para facilitar una posible adopción del modelo a diferentes entornos

hospitalarios, se llevó a cabo una validación externa, es decir, una validación llevada a cabo con un conjunto de datos perteneciente a un centro nuevo del que no se hubiesen empleado datos para el entrenamiento. Para esta validación externa se emplearon 116 pacientes reservados de la base de datos. La media de C-index fue de 0.64, mientras que la AUC media fue de 0.63 (Tabla 52).

Tabla 52. Resultados de las métricas empleadas para evaluar el modelo de supervivencia de Cox en el conjunto de datos de validación.

	C-index media	AUC media	IPW media
Modelo COX	0.64	0.63	0.63

Se obtuvieron las curvas de supervivencia de Kaplan-Meier para el conjunto de datos de validación. Se compararon los datos reales con las predicciones realizadas por el modelo de Cox, evidenciando menor similitud de los datos predichos respecto a los datos reales conforme aumenta el tiempo transcurrido desde el momento de diagnóstico del paciente (Figura 52).

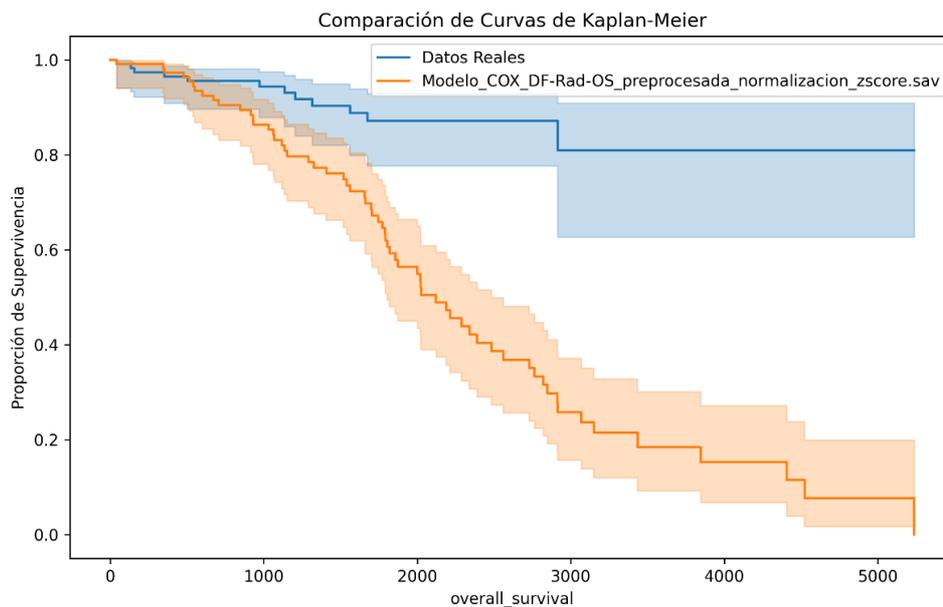


Figura 52. Curva Kaplan-Meier del modelo en función del tiempo para los 116 pacientes del grupo de validación. La curva azul representa los datos reales. Curvas con intervalo de confianza del 95%.

Se empleó un diagrama de barras con los datos de tiempo de supervivencia de los pacientes agrupados en subgrupos de 742 días para observar la frecuencia absoluta de dicha variable, y así poder comprender la discrepancia en las curvas Kaplan-Meier de datos reales frente a la predicha por el modelo a partir de los 2000 días de supervivencia, momento a partir del cual el modelo presenta mayor discrepancia en las predicciones respecto a la curva de datos reales. Se observó una disminución significativa en el número de pacientes con supervivencia mayor a 2270 días, lo cual permite explicar discrepancias de la curva de datos reales respecto a la curva predicha por el modelo de Cox (Figura 53).

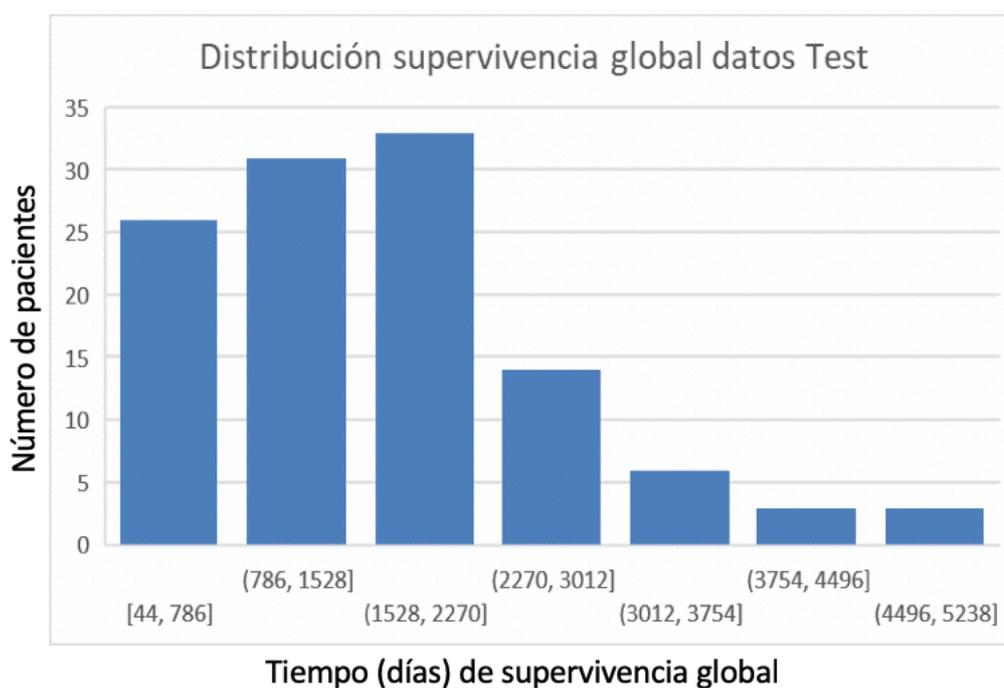


Figura 53. Diagrama de barras representando la distribución de la variable tiempo de supervivencia de los pacientes desde el momento de diagnóstico (agrupados en subgrupos de 742 días).

4.6.3. Entrenamiento de otros modelos de supervivencia global

De manera simultánea, se emplearon distintas arquitecturas y métodos de clasificación y regresión para construir modelos de supervivencia diferentes al previamente descrito, para realizar una comparación y analizar el impacto de emplear distintas estrategias de modelado con el mismo conjunto de datos de entrenamiento y validación. Las arquitecturas empleadas han sido previamente utilizadas de manera generalizada para problemas genéricos de clasificación y regresión, y en este caso están adaptadas para emplearse en problemas de supervivencia (considerando los eventos censurados).

- Gradient Boosting: es una técnica de aprendizaje automático que construye un modelo predictivo en forma de un conjunto de modelos débiles (generalmente árboles de decisión) de manera secuencial. Cada nuevo modelo se ajusta a los errores residuales del modelo anterior[121].
- FSVM (*Fuzzy Support Vector Machines*, Máquina de Vectores de Soporte Difusos): modelos diseñados para predecir el tiempo hasta un evento (supervivencia), que, en lugar de clasificar los datos en dos clases distintas (como en un modelo estándar), permiten una mayor flexibilidad al asignar una membresía difusa a cada punto de datos, lo que significa que un punto puede tener cierto grado de pertenencia a múltiples clases simultáneamente [122].
- NSVM (*Naive version of a linear Survival Support Vector Machine*, Máquina de Vectores de Soporte Naive), un enfoque de generación de características utilizando la diferencia entre pares de vectores de características en los datos originales y aplicando un clasificador de vectores de soporte lineal estándar, que permite que el modelo pueda distinguir entre las características de dos muestras diferentes[123].

Para el entrenamiento de estos modelos se mantuvieron los mismos conjuntos y subconjuntos de datos en la validación cruzada que los empleados para el entrenamiento del modelo de Cox original.

Los resultados obtenidos para los distintos modelos en el conjunto de datos de entrenamiento fueron inferiores a los resultados del modelo Cox, mostrando un valor de C-index medio inferior en todos modelos frente al modelo de Cox (un 14% inferior en el modelo gradient boosting, 24% inferior en el modelo FSVM y 25% inferior en el modelo NSVM en la métrica C-index). Asimismo, mostraron valores inferiores en el AUC media y la IPW media. En la Tabla 53 se muestran los valores medios de C-index, AUC media e IPW media junto con la desviación estándar obtenida tras la realización de validación cruzada en los diferentes modelos de predicción de supervivencia, comparándolos con el modelo original Cox.

Tabla 53. Resultados de los modelos de supervivencia Gradient Boosting, Support Vector Machines for Survival Analysis y Neural Networks for Survival Analysis para el conjunto de datos de entrenamiento y comparación con el modelo Cox. Valores medios de C-index, AUC media e IPW media y desviación estándar obtenida tras la realización de validación cruzada en los diferentes modelos de predicción de supervivencia.

	C-index media +/- DE	AUC media +/- DE	IPW media +/- DE
Modelo COX	0.63 +/-0.07	0.60 +/- 0.10	0.61 +/- 0.13
Gradient Boosting	0.54 +/- 0.06	0.58 +/- 0.10	0.53 +/- 0.09
FSVM	0.48 +/-0.08	0.53 +/-0.12	0.60 +/-0.11
NSVM	0.47 +/-0.08	0.51 +/-0.12	0.56 +/-0.11

4.6.4. Validación de otros modelos de supervivencia global

Para la validación externa de estos modelos se mantuvo el mismo conjunto de datos que el empleado para la validación externa del modelo de Cox original.

Los resultados obtenidos para los distintos modelos en el conjunto de datos de validación fueron inferiores a los resultados del modelo Cox, mostrando un valor de C-index medio inferior en todos modelos frente al modelo de Cox (un 3%

inferior en el modelo gradient boosting, 30% inferior en el modelo FSVM y 11% inferior en el modelo NSVM en la métrica C-index). Asimismo, mostraron valores inferiores en el AUC media y la IPW media. En la Tabla 54 se muestran los valores medios de C-index, AUC media e IPW media para la validación externa en los diferentes modelos de predicción de supervivencia, comparándolos con el modelo original Cox.

Tabla 54. Resultados de los modelos de supervivencia *Gradient Boosting*, *Support Vector Machines for Survival Analysis* y *Neural Networks for Survival Analysis* para el conjunto de datos de validación y comparación con el modelo Cox. Valores medios de C-index, AUC media e IPW media.

	C-index media	AUC media	IPW media
Modelo COX	0.64	0.63	0.63
Gradient Boosting	0.62	0.56	0.37
FSVM	0.44	0.48	0.64
NSVM	0.57	0.61	0.67

4.7. Explicabilidad de los modelos de supervivencia global desarrollados

4.7.1. Posibles biomarcadores de imagen para predicción de supervivencia global en tumores neuroblásticos

La aplicación de modelos de aprendizaje automático para la predicción de supervivencia en pacientes con tumores neuroblásticos emerge como una herramienta prometedora en la práctica clínica. La capacidad de estos modelos para evaluar el riesgo de supervivencia global puede influir significativamente en la toma de decisiones médicas, como la selección de tratamientos y la planificación del seguimiento de manera personalizada. Sin embargo, la adopción exitosa de estos modelos en entornos clínicos requiere una comprensión clara de su funcionamiento interno, así como la capacidad de evaluar posibles errores o sesgos que puedan surgir en su implementación.

Una estrategia fundamental para mejorar la explicabilidad de los modelos de supervivencia global es el uso de SHAP (*SHapley Additive exPlanations*) *values*. Los SHAP *values* proporcionan una forma de medir la contribución relativa de

cada variable de entrada a la predicción de supervivencia de un paciente. Se basan en la teoría de juegos cooperativos y se calculan evaluando cómo cambia la predicción del modelo al considerar diferentes combinaciones de variables, permitiendo así una comprensión más profunda de cómo las características de los pacientes influyen en su pronóstico. Los SHAP *values* ofrecen una explicación individualizada de la importancia de cada variable en el modelo de predicción de supervivencia. Esto permite entender las predicciones del modelo y discernir qué características clínicas o biomarcadores de imagen tienen el mayor impacto en la estimación del riesgo de supervivencia de un paciente [124].

En un gráfico de *SHAP values*, la interpretación se puede hacer de la siguiente manera:

- Eje Y: indica los nombres de las características en orden de importancia, desde arriba hacia abajo. Las características en la parte superior del gráfico son las que más contribuyen a la predicción del modelo, mientras que las características en la parte inferior tienen menos impacto.
- Eje X: representa el valor de SHAP y su grado de cambio.
- Valores: un impacto positivo de la variable se asocia con un mayor riesgo de que suceda el evento (en este caso, muerte) para los pacientes, mientras que un impacto negativo se asocia con un menor riesgo.
- Color de los puntos: el color de cada punto en el gráfico representa el valor de la característica correspondiente en el conjunto de datos original. Los puntos rojos indican valores altos, mientras que los puntos azules indican valores bajos. Esto proporciona una perspectiva visual de cómo varían los valores de las características en relación con su impacto en la predicción del modelo.

En este estudio se obtuvieron las variables más importantes empleadas en el modelo de Cox a partir del conjunto de datos de validación. Este enfoque proporciona una visión detallada de qué características del tumor o del paciente están más fuertemente asociadas con la supervivencia global, lo que puede servir

como base para la identificación de biomarcadores de imagen potencialmente útiles en la predicción del pronóstico de los pacientes con tumores neuroblásticos.

De entre las variables más relevantes obtenidas en el proceso de validación del modelo Cox basado en radiómica y radiómica profunda (Figura 54) se observó que ninguna de las variables de radiómica profunda se encontraba entre las 20 más predictivas seleccionadas por el modelo de Cox para realizar las predicciones o inferencias, por lo que las variables de radiómica resultaron de mayor impacto para la predicción de supervivencia.

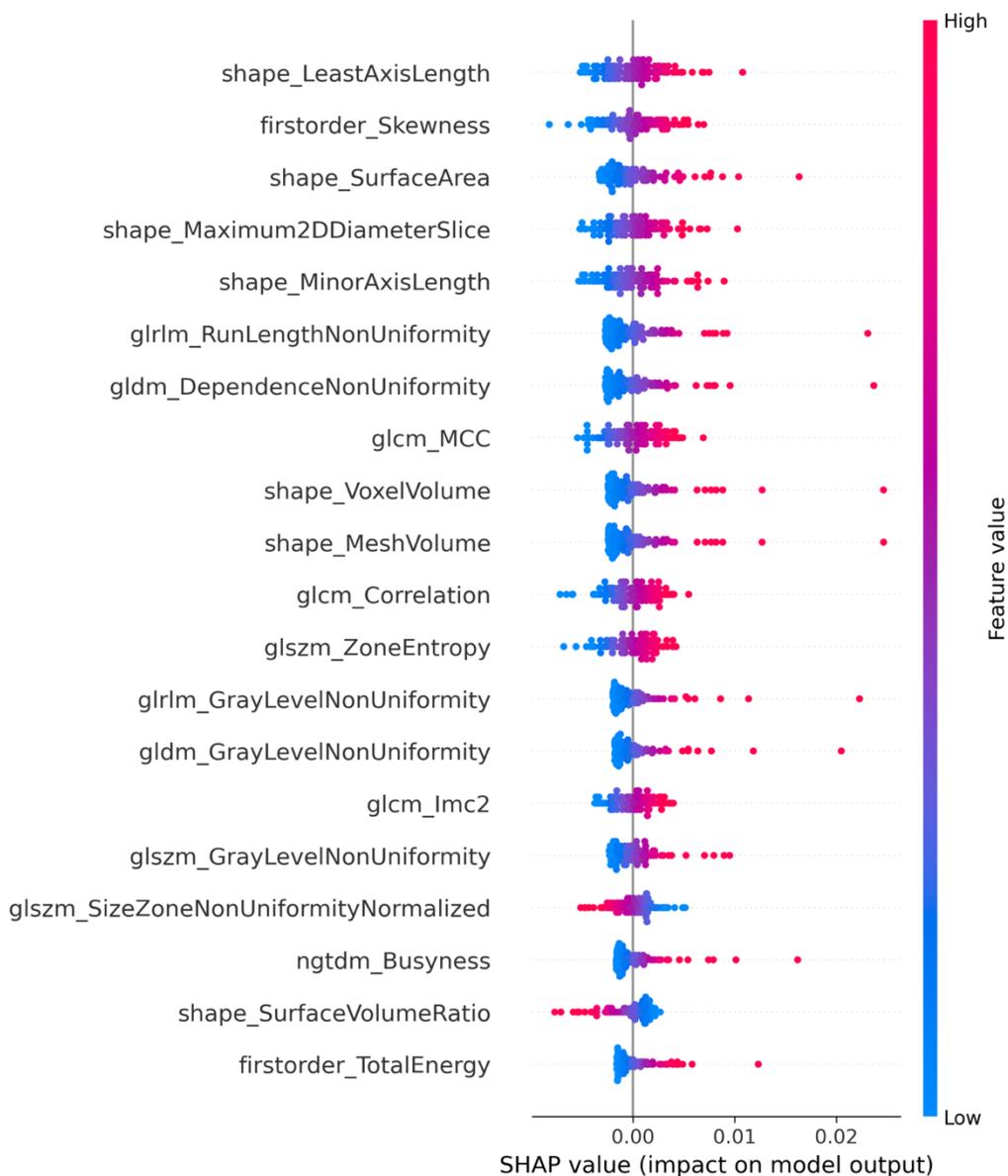


Figura 54. Representación mediante *SHAP values* del impacto de cada una de las variables más representativas del modelo Cox basado en radiómica y radiómica profunda

para el conjunto de datos de validación. El color rojo indica valores altos de la característica, mientras que el color azul indica valores bajos. Un impacto positivo de la variable se asocia con un mayor riesgo para los pacientes, mientras que un impacto negativo se asocia con un menor riesgo para los pacientes.

De las 20 variables más representativas, 7 fueron características de radiómica de forma, dos de primer orden y 11 de segundo orden.

Se puede observar el impacto y la direccionalidad de las variables con mayor representación. Por ejemplo, la variable de radiómica de forma *Least axis length*, que mide las variaciones en la longitud del eje más corto del tumor en 2D presenta un impacto positivo de la variable, lo cual implica que un mayor diámetro del eje más corto del tumor se asocia con un mayor riesgo de que suceda el evento (muerte).

Ante un valor mayor de la variable de radiómica de primer orden *Skewness*, referida a la asimetría de la distribución de las intensidades de los vóxeles pertenecientes a la región tumoral en torno al valor medio de la distribución (los tejidos altamente heterogéneos muestran una asimetría absoluta más alta que los homogéneos), existe mayor riesgo de fallecimiento del paciente.

La variable de radiómica de segundo orden *GLRLM Run length non uniformity* (No uniformidad de longitud de ejecución de GLRLM) mide la uniformidad en la distribución de las longitudes de las ejecuciones en la imagen. Un valor más alto de esta característica indica una mayor variabilidad en las longitudes de ejecución, lo que sugiere una textura más heterogénea en la imagen, y asocia a mayor riesgo de evento para el paciente.

4.7.2. Variables más predictivas del modelo de supervivencia

A continuación, se discuten los resultados obtenidos de las variables más predictivas en el modelo de supervivencia de Cox (Figura 54) explicando su impacto en base a los resultados hallados en los análisis de los estudios previos (4.4.2. Estabilidad de la radiómica; 4.4.3. Relación de la radiómica con variables

analíticas y de estadificación; 4.5.3. Relación de la radiómica profunda con variables analíticas y de estadificación).

El impacto en los análisis previos de las 10 variables más representativas se describe a continuación:

- *Shape least axis length*: longitud del eje más corto del tumor, en los análisis realizados previamente, esta variable mostró un elevado grado de reproducibilidad y estabilidad tras las modificaciones de erosión y dilatación de máscaras. Asimismo, mostró diferencias significativas entre los grupos M y L1+L2 para la predicción del INRGSS en <18 meses. En el modelo de supervivencia se observa una relación positiva el aumento de longitud del eje más corto del tumor y el riesgo de evento para el paciente.
- *Skewness*: esta variable, referida a la asimetría en la distribución de vóxeles, mostró una alta estabilidad tras modificaciones de remuestreo, normalización y erosión. Presentó diferencias significativas entre los grupos M y L1+L2 para INRGSS en >18 meses. En el modelo de supervivencia se observa una relación positiva entre el valor tomado por esta variable y el riesgo de evento para el paciente.
- *Surface area*: esta variable, referida a la superficie tumoral, mostró una alta estabilidad tras modificaciones de erosión. Se obtuvieron diferencias significativas entre los grupos M y L1+L2 y los grupos M y Ms para INRGSS en <18 meses. En el modelo de supervivencia se observa una relación positiva entre la superficie tumoral y el riesgo de evento para el paciente.
- *Maximum 2D diameter slice*: se obtuvo un elevado grado de reproducibilidad y estabilidad tras las modificaciones de erosión y dilatación de máscaras, se obtuvieron diferencias significativas entre los grupos M y L1+L2 y los grupos M y Ms para INRGSS en <18 meses. En el modelo de supervivencia se observa una relación positiva entre el valor máximo del diámetro en 2D y el riesgo de muerte.
- *Minor axis length*: esta variable se refiere a la longitud del eje menor de una elipse que mejor se ajusta a la forma de la región de interés (el tumor). Se obtuvo un elevado grado de reproducibilidad y estabilidad tras las modificaciones de erosión y dilatación de máscaras, se obtuvieron

diferencias significativas entre los grupos M y L1+L2 y los grupos M y Ms para INRGSS en <18 meses. En el modelo de supervivencia se observa una relación positiva entre la longitud del eje menor del tumor y el riesgo de evento.

- *GLRLM Run length non uniformity*: como se ha expuesto, esta variable mide la uniformidad en la distribución de las longitudes de las ejecuciones en la imagen y se asocia a la heterogeneidad. Esta característica de radiómica de segundo orden presentó estabilidad en tras la modificación de filtros, dilatación y erosión. Mostró diferencias significativas para predecir la amplificación del MYCN en menores de 18 meses. Se obtuvieron diferencias significativas entre los grupos M y L1+L2 y los grupos M y Ms para INRGSS en <18 meses. En el modelo de supervivencia se observa una relación positiva, una textura más heterogénea en la imagen y asocia a mayor riesgo de evento para el paciente.
- *GLDM Dependence non uniformity* (No Uniformidad de Dependencia): esta variable mide la variabilidad en las dependencias de los pares de valores de vóxeles en la imagen. Un valor mayor de este parámetro indica que hay una amplia variabilidad en estas relaciones, lo que sugiere una textura más heterogénea en la imagen. En los análisis previos, la variable presentó estabilidad tras la modificación de filtros, dilatación y erosión. Se obtuvieron diferencias significativas entre los grupos M y L1+L2 y los grupos M y Ms para INRGSS en <18 meses. En el modelo de supervivencia se observa una relación positiva entre el valor tomado por esta variable y el riesgo de evento para el paciente.
- *GLCM Maximum Correlation Coefficient*: Coeficiente de Correlación Máxima de GLCM, es una característica de segundo orden que representa la máxima correlación lineal entre los valores de intensidad de los vóxeles en diferentes direcciones en la imagen, y cuantifica la máxima correlación encontrada entre los vóxeles vecinos en una dirección específica. En los análisis previos mostró estabilidad tras modificación de filtros y erosión de la máscara. Sin embargo, no se observaron correlaciones significativas entre esta variable y las variables clínicas y analíticas estudiadas. En el modelo de supervivencia se observa una relación positiva entre la correlación entre los vóxeles y el riesgo de evento para el paciente.

- *Shape Voxel Volume*: esta característica de radiómica de forma representa el volumen total ocupado por todos los vóxeles dentro de la ROI (el tumor). En los análisis previos mostró estabilidad tras dilatación y erosión de la máscara. Se obtuvieron diferencias significativas entre los grupos M y L1+L2 y los grupos M y Ms para INRGSS en <18 meses. En el modelo de supervivencia se observa una relación positiva entre el volumen total ocupado por todos los vóxeles dentro del tumor y el riesgo de muerte.
- *Shape Mesh Volume*: esta variable se refiere al volumen de la malla, es decir, una representación tridimensional del volumen que está contenido dentro de la malla que representa la forma de la ROI en la imagen. Esta característica mostró estabilidad tras dilatación y erosión de la máscara. Se obtuvieron diferencias significativas entre los grupos M y L1+L2 y los grupos M y Ms para INRGSS en <18 meses. En el modelo de supervivencia se observa una relación positiva entre la correlación entre el volumen del tumor y el riesgo de evento.

Por todo ello, las diez variables expuestas muestran un valor predictivo alto en el modelo de supervivencia global desarrollado, todas ellas con relación positiva entre el valor de la variable y el riesgo de evento. Asimismo, las variables expuestas presentaron diferentes grados de reproducibilidad y estabilidad en los análisis realizados previamente, y se relacionaron con algunas de las variables clínicas y de estadificación. Por tanto, las diez variables de radiómica expuestas podrían considerarse como variables candidatas a biomarcadores de imagen para la predicción de supervivencia global en pacientes pediátricos con tumores neuroblásticos.

5. DISCUSIÓN

DISCUSIÓN

En el estudio de los tumores neuroblásticos, la imagen es fundamental para la detección, caracterización y predicción de respuesta al tratamiento. Sin embargo, al tratarse de tumores con una gran heterogeneidad, existe una necesidad clínica de mejorar la caracterización por imagen y la gradación del riesgo.

Esta Tesis Doctoral describe una metodología detallada para un análisis computacional preciso de la imagen y la obtención de parámetros cuantitativos. La detección y segmentación tumoral y la extracción de parámetros cuantitativos de imagen a partir de la RM permiten llevar a cabo un análisis avanzado de los tumores, extrayendo características que van más allá de las identificadas por el ojo humano, y que aportan información complementaria a los sistemas de estadificación actual, y que puede ser correlacionada con diferentes métodos de caracterización tumoral. Asimismo, el análisis de estos parámetros permite establecer modelos de predicción de supervivencia, que pueden complementar a los sistemas empleados actualmente en la práctica clínica.

Por tanto, el análisis cuantitativo de imagen supone un progreso en el diagnóstico de estos pacientes y aportan un valor añadido a la predicción de agresividad de los tumores neuroblásticos.

Esta Tesis Doctoral ha consistido en un estudio multicéntrico que ha incluido a un gran número de pacientes diagnosticados de tumor neuroblástico con gran heterogeneidad, incluyendo un gran espectro de tumores con diferentes localizaciones y de imágenes de RM con distintos proveedores y campos magnéticos.

Por otro lado, las herramientas desarrolladas en esta Tesis Doctoral, siendo la herramienta de segmentación automática y el modelo de predicción de supervivencia global basado en radiómica y radiómica profunda, han sido validadas en pacientes no empleados para el entrenamiento de las mismas para garantizar su reproducibilidad.

5.1. Resultados más importantes y comparación con otros trabajos

5.1.1. Segmentación automática

La variabilidad entre observadores al realizar la segmentación manual de los tumores neuroblásticos en imágenes de RM potenciadas en T2w indica que existe una alta concordancia entre observadores (mediana de CSD de 0.969 ± 0.032 , IQR). Las discrepancias entre observadoras pueden deberse a la naturaleza heterogénea de los tumores neuroblásticos y a la variabilidad intrínseca de la segmentación manual relacionada con las habilidades individuales y el nivel de atención al detalle. Por lo tanto, la máscara de referencia manual puede tener algunos errores asociados intrínsecamente a la metodología de segmentación realizada por humanos. Joskowicz et al [43] evaluaron la variabilidad en las segmentaciones manuales en TC para tumores hepáticos, tumores pulmonares, contornos renales y hematomas cerebrales entre 11 observadores, y concluyeron que la variabilidad entre observadores es amplia y que incluso dos o tres observadores pueden no ser suficientes para establecer el rango completo de variabilidad inter observador. El estudio realizado por Montagne et al [125] comparó la variabilidad entre observadores para la segmentación de la próstata en RM realizada por 7 observadores y se concluyó que la variabilidad está influenciada por los cambios en la morfología de la próstata. Por lo tanto, la variabilidad en la superposición de volumen y segmentación para diferentes estructuras y observadores es amplia [126].

En esta Tesis Doctoral se llevó a cabo la delineación experta de tumores como la mejor aproximación, aunque no perfecta, a la verdad. La evaluación de la similitud píxel a píxel entre la verdad fundamental y la máscara segmentada de forma automática demuestra que la arquitectura de aprendizaje profundo nnU-Net puede utilizarse para detectar y segmentar los tumores neuroblásticos en las imágenes de RM con un CSD mediano de $0.965 (\pm 0.018 \text{ IQR})$, logrando un rendimiento sólido y superando los métodos y resultados obtenidos en estudios previos que abordaron el problema de la segmentación de neuroblastomas [46 – 48]. Sin embargo, no se ha hallado literatura previa que haya demostrado el rendimiento de una solución basada en CNN en la segmentación de tumores neuroblásticos.

La arquitectura nnU-Net establece un nuevo estado del arte en diversos desafíos de segmentación y muestra fuertes características de generalización para otras estructuras [53, 54]. Nuestros resultados sugieren que esta herramienta de segmentación automática introduce una variabilidad equivalente a la observada en el proceso de segmentación manual en tumores neuroblásticos. Estudios previos relacionados con tumores de mama mostraron que los algoritmos de segmentación pueden mejorar la variabilidad manual [46 – 48].

Al analizar la dirección de los errores en un problema de segmentación tumoral, podría considerarse como recomendación adecuada el dar más relevancia a la TFPm, con el objetivo de minimizar los vóxeles FP incluidos con respecto a la segmentación de referencia, ya que esta métrica representa aquellos vóxeles que pertenecen a órganos o estructuras adyacentes, lo que podría introducir un fuerte sesgo en la extracción de características cuantitativas de imágenes para el desarrollo de modelos radiómicos. La influencia de los FN (vóxeles tumorales no incluidos en la segmentación) en los modelos de radiómica parece menos importante, ya que puede no tener un impacto significativo si se omiten algunos vóxeles periféricos del tumor. El efecto de la variabilidad de la segmentación manual entre observadores en la robustez de las características radiómicas basadas en RM ha sido descrito previamente en otros tumores como el cáncer de mama [127].

Al evaluar la TFPm y la TFN entre las segmentaciones manuales realizadas por las dos radiólogas, la mediana de la TFPm fue de 0.939 (± 0.063 IQR), lo que indica que un 6.1% de los vóxeles fueron clasificados erróneamente como tumor, mientras que la mediana de la TFN fue de 0.998 (± 0.008 IQR), por lo que la segmentación manual de la Radióloga 2 no incluyó un 0.2% de vóxeles incluidos en la máscara de verdad fundamental o patrón de referencia. En cuanto al patrón de referencia frente a la segmentación automática, observamos que la mediana de la TFPm fue de 0.968 (± 0.015 IQR). Por lo tanto, la herramienta de segmentación automática genera máscaras con un promedio del 3.2% de vóxeles no tumorales. La mediana de la TFN corresponde a un valor de 0.963 (± 0.021 IQR), por lo que la herramienta automática no logra incluir un 3.7% de vóxeles

tumorales. Los resultados obtenidos demuestran que el modelo de segmentación automática logra un mejor rendimiento en cuanto a la TFPm, lo que supone una gran ventaja en las tareas de segmentación para la extracción posterior de características cuantitativas de imágenes.

En cuanto al tiempo requerido para el proceso de segmentación, se obtuvo una reducción promedio del tiempo del 99.7% al comparar el modelo automático con la metodología de segmentación manual. Dado que puede existir cierta variabilidad en las máscaras automáticas finales, se recomienda una validación visual realizada por radiólogos para editar y ajustar las máscaras automáticas cuando sea necesario. Este proceso de validación visual de las máscaras automáticas redujo el tiempo requerido para la segmentación realizada *de novo* en un 92.8%. El tiempo de corrección se vio influenciado por la dificultad intrínseca de la segmentación de cada tumor, ya que se incluyeron tumores más fáciles de segmentar (por ejemplo, más homogéneos, con márgenes más definidos, sin ganglios linfáticos) que no necesitaron correcciones o requirieron ajustes leves, mientras que otros tumores presentaron mayor complejidad, con apariencia similar de las estructuras circundantes o cambios en la intensidad de los vóxeles periféricos, requiriendo más tiempo para ser corregidos.

Por todo ello, la aplicación del modelo automático resulta en una gran ventaja en el tiempo requerido para realizar el proceso de segmentación, facilitando el flujo de trabajo para los radiólogos.

Además, el análisis visual de las máscaras realizadas por la red es útil para corregir los posibles errores y sesgos/atípicos dentro del conjunto de datos, lo que podría utilizarse para volver a entrenar el modelo, lo que resultaría en un aumento de su precisión general y robustez.

Tras llevar a cabo el entrenamiento y validación interna de la red automática, y con la finalidad de validar esta metodología de segmentación, se llevó a cabo un estudio internacional multicéntrico que incluyó un conjunto de datos independiente y heterogéneo de imágenes de RM de niños con tumores neuroblásticos, y que permitió confirmar y validar el rendimiento de la

herramienta para identificar y segmentar las lesiones. Las máscaras de segmentación nnU-Net fueron validadas visualmente por una radióloga pediátrica experimentada (Radióloga 1), con un CSD mediano de 0.997. Entre las 535 secuencias de RM T2/T2*w, solo el 14% (74 casos con un CSD < 0.8) requirió edición manual, y únicamente en el 6% (18 secuencias) la herramienta no logró segmentar ninguna estructura. Las 136 máscaras ajustadas manualmente tuvieron una mediana del CSD elevada, de 0.887.

La automatización del proceso de segmentación tiene el valor añadido sobre el método manual de que es determinista, siempre proporcionando los mismos resultados dado el mismo conjunto de imágenes de entrada, lo que mejora la repetibilidad [128]. La validación visual y la edición manual de las máscaras automáticas mejoran la usabilidad del método. Este modelo automático con validación visual podría integrarse en el flujo de trabajo de radiología incluso en conjuntos de datos grandes [129, 130].

Se han desarrollado, descrito y publicado múltiples herramientas de procesamiento de inteligencia artificial, pero muy pocas, aproximadamente el 6% [131], han sido sometidas a una validación externa. La validación en un conjunto de datos del mundo real e independiente es un proceso fundamental para determinar la precisión de un modelo y estimar su reproducibilidad y generalización, lo cual es esencial para basar las decisiones clínicas en modelos de predicción correctos.

En nuestro conocimiento y en el momento de su defensa, esta Tesis Doctoral es el primer estudio que realiza una validación externa internacional, multicéntrica y con múltiples proveedores de un modelo de segmentación automática para la identificación y segmentación de tumores neuroblásticos con imágenes de RM en diferentes partes del cuerpo. Únicamente algunos trabajos anteriores han abordado el problema de la validación externa de algoritmos de segmentación automática en diferentes áreas, como glioblastoma [57], displasia cortical focal [58] o segmentos hepáticos [132, 133]. La limitación más frecuente en cuanto a la validación externa es la inclusión de un número pequeño de sujetos. Como ejemplo, un trabajo que aborda la segmentación de craneofaringioma en RM

utilizando una CNN profunda basada en U-Net realizó una validación independiente con 38 pacientes [134], mientras que en un estudio sobre el impacto de las segmentaciones manuales y automáticas en los modelos de radiómica para predecir la respuesta de los cánceres de recto localmente avanzados, el número de pacientes en el conjunto de datos de validación independiente fue de 28 casos [135]. En otro estudio, la validación externa para demostrar la robustez de nuevas herramientas radiómicas para predecir la recurrencia bioquímica en el cáncer de próstata fue realizada en una serie de 88 pacientes [136, 137]. Nuestra serie es una de las más grandes que utiliza casos del mundo real para una validación independiente.

En esta Tesis Doctoral se realizó un estudio multicéntrico y multivendedor para la translación a un entorno clínico de una solución computacional, y demostró la reproducibilidad de una herramienta de segmentación CNN, siguiendo las recomendaciones internacionales para la validación de biomarcadores [138]. Otros estudios diseñados para validar algoritmos de segmentación de aprendizaje profundo son de un solo centro, como un trabajo que tuvo como objetivo cuantificar el índice de musculoesquelético y la sarcopenia en el carcinoma renal metastásico [139].

Nuestro estudio demostró que la herramienta automática tiene un rendimiento inferior después del tratamiento con quimioterapia (CSD mediano de 0.902 después del tratamiento frente a 0.999 al diagnóstico), aunque la diferencia no fue estadísticamente significativa. Se observó una reducción en el volumen del tumor y en la delimitación de sus bordes después del tratamiento del neuroblastoma [26, 140] que pudo haber influido en el rendimiento de la red. Además, trabajos previos han demostrado cambios en las características radiómicas en otros tumores tras el tratamiento, como el cáncer de páncreas [141].

La arquitectura de aprendizaje profundo nnU-Net ha establecido un nuevo estado del arte en numerosas segmentaciones, y como se ha demostrado, es generalizable. Este algoritmo de segmentación robusto, repetible y automatizado con validación visual final mejora la consistencia de la extracción de datos y fortalece el flujo de trabajo de la segmentación de tumores [53, 142, 143].

5.1.2. Reproducibilidad de la radiómica

Para lograr una efectiva transición de la radiómica a la práctica clínica es fundamental evaluar tanto la robustez como la reproducibilidad de las características radiómicas. La repetibilidad de estas características puede ser evaluada mediante la aplicación del enfoque de escaneo y reescaneo para probar la consistencia del rendimiento. Sin embargo, nuestro estudio basado en datos del mundo real no permite llevar a cabo este análisis.

La reproducibilidad, por otro lado, puede ser evaluada mediante la introducción de perturbaciones a la imagen, como filtros o normalización, así como mediante la utilización de distintas máscaras de segmentación, tal como se ha realizado en esta Tesis Doctoral. En nuestro conocimiento, este estudio representa la primera aproximación al análisis de la reproducibilidad de la radiómica en tumores pediátricos, aunque investigaciones previas han evaluado la precisión de las características radiómicas en otros tipos de tumores, como el carcinoma hepatocelular [33] o el glioblastoma [69, 70]. Una de las principales fortalezas y novedades de nuestro estudio radica en la extensa cohorte multicéntrica de pacientes con tumores neuroblásticos reclutada a lo largo de un extenso periodo de tiempo (recopilación de datos durante 18 años), que presentan heterogeneidad tanto en su ubicación como en su comportamiento.

En el análisis llevado a cabo para evaluar la reproducibilidad de la radiómica se aplicaron diversos cambios a una verdad de referencia que incluyó el algoritmo ADF para la reducción de ruido, la corrección del sesgo por campo magnético N4 para eliminar las inhomogeneidades, la normalización mediante z-score y el remuestreo. Estas modificaciones mejoraron la imagen para representar un escenario del mundo real en el que las imágenes están armonizadas.

Entre las distintas modificaciones de imagen, la eliminación de inhomogeneidades (filtro N4, serie 5) tuvo el menor impacto en las características radiómicas, mostrando mayor reproducibilidad en comparación con otras series. Un estudio previo que analizó la influencia de la corrección de inhomogeneidades y la reducción de ruido en la radiómica para el glioblastoma multiforme demostró

que la corrección de inhomogeneidades seguida de los filtros de ruido introdujo características más estables y reproducibles que los filtros de ruido seguidos de la corrección del sesgo por campo magnético [69].

El remuestreo y la normalización, etapas cruciales antes de la comparación, alteraron los resultados radiómicos presentando diferencias estadísticamente significativas entre la normalización y el remuestreo en comparación con el resto de las modificaciones. Al eliminar la normalización, solo 5 de las 93 (5%) características radiómicas permanecieron estables (2 características radiómicas de primer orden y 3 características GLCM de segundo orden). La normalización tiene un gran impacto en las características radiómicas de primer y segundo orden, principalmente debido a que el valor de las intensidades se modifica mediante la aplicación de un z-score, con un valor promedio de intensidad de 0. La normalización debe considerarse cuidadosamente para evitar sesgar los resultados y no evaluar el efecto del tumor, sino el efecto de las modificaciones en la preparación de la imagen. Nuestros resultados respaldan las conclusiones extraídas en trabajos previos [69]. Un estudio anterior analizó el efecto de la normalización en RM de la radiómica para el cáncer de próstata [68], y reveló que la normalización tuvo un impacto significativo en la mayoría de las características, lo que podría tener un efecto notable en los resultados de los modelos de radiómica. Estos resultados demuestran que la comparación de características radiómicas puede no ser confiable si la normalización no se ha realizado con los mismos métodos estandarizados.

La estabilidad de las características radiómicas de forma fue esperable en las series en las que la forma del tumor y la máscara no había sido modificada tras la aplicación de diferentes filtros, por lo que las características radiómicas de forma no se incluyeron en el análisis de las series 2-7. La alta robustez y reproducibilidad de las características de forma en la radiómica del glioblastoma multiforme han sido descritas previamente [69]. En las series en las que se modificó la forma de la máscara (erosión o dilatación), las variables de forma permanecieron estables y reproducibles, con un CCC por encima del punto de corte de 0.9. En estos casos, las modificaciones en la forma fueron relativamente pequeñas por lo que estas variables radiómicas tampoco experimentaron cambios importantes. Esto implica que las pequeñas modificaciones en la forma

de las segmentaciones, que pueden deberse a discrepancias en la segmentación manual y a la variabilidad entre observadores, no afectan la reproducibilidad de ese tipo de radiómica, pero sí afectan a la radiómica de primer y segundo orden. Al aplicar erosión en la máscara, las características radiómicas permanecieron más estables. Esto implica que la reducción del área del tumor aún engloba solo los vóxeles tumorales y la radiómica permanece más estable en comparación con la dilatación, donde se añaden a los análisis de radiómica vóxeles que no pertenecen al tumor. Estudios previos han descrito cambios relacionados con la máscara de segmentación. Un estudio que comparó los resultados de radiómica en segmentaciones realizadas por profesionales independientes de diferentes disciplinas mostró que la variabilidad en la segmentación afecta a la estabilidad de las características radiómicas en estudios basados en TC en el cáncer de páncreas [71]. Además, un estudio que incluyó segmentaciones manuales realizadas por diferentes individuos en tres tipos de tumores diferentes en TC mostró que la variabilidad en la segmentación entre observadores tiene una influencia relevante en el análisis de radiómica y está fuertemente influenciada por el tipo de tumor [144].

En cuanto a la reproducibilidad intrasujeto, muchas de las características radiómicas no mostraron una variabilidad significativa en las múltiples mediciones realizadas. Sin embargo, algunas características radiómicas de segundo orden presentaron una excelente reproducibilidad en todas las series, lo que podría considerarse como biomarcadores estables y tener un valor clínico potencial. Estudios previos que abordaron el CoV tras aplicar remuestreo de niveles de gris han demostrado que esta modificación mejora las características radiómicas de segundo orden [145].

5.1.3. Relación de la radiómica y radiómica profunda con variables analíticas y de estadificación

Se encontró una asociación estadísticamente significativa entre la radiómica y las variables analíticas y de estadificación en 55 casos. El mayor número de variables significativas se encontró en el INRGSS, donde se observaron diferencias en 43 parámetros radiómicos entre el grupo M y los demás grupos en pacientes

menores de 18 meses. Por tanto, las variables de radiómica podrían constituir biomarcadores no invasivos de relevancia para la diferenciación de neuroblastoma metastásico (M) en menores de 18 meses. Las características de radiómica de forma analizadas presentaron mayor proporción en las relaciones estadísticamente significativas con los distintos grupos de estadificación (79%) respecto a las de primer orden (22%) o segundo orden (16%).

Estudios previos han demostrado la relación de diversas características de radiómica con las propiedades del microambiente tumoral en diferentes tumores. Un estudio mostró correlación entre las características radiómicas extraídas de imágenes de TC del cáncer de pulmón de células no pequeñas y las firmas de células Th2, pudiendo ayudar a definir categorías de actividad inmune tumoral [146]. En otro estudio reciente se observaron diferencias estadísticamente significativas en parámetros de radiómica de primer orden y orden superior en TC con contraste en tumores pulmonares neuroendocrinos con el valor del índice Ki-67 [32]. Otro estudio demostró la relación de la radiómica en RM en cáncer de mama triple negativo con el estado de los linfocitos infiltrantes de tumor [147].

Por su parte, el análisis de la relación de la radiómica profunda con las variables analíticas y de estadificación mostró relación estadísticamente significativa en 29 casos. La radiómica profunda presenta mayor relación con la predicción de la mutación MYCN en pacientes mayores de 18 meses. La relación de la radiómica con el estado de mutación del MYCN ha sido analizado en estudios previos, como un estudio en el que se evaluó la radiómica extraída por RM y se desarrollaron clasificadores basados en aprendizaje automático para predecir la amplificación de MYCN en neuroblastomas, ofreciendo evidencia retrospectiva preliminar que sugiere la viabilidad de la radiómica de RM en la predicción de la amplificación de MYCN en neuroblastomas [21]. En una cohorte de 133 pacientes se demostró que los fenotipos de imagen de los neuroblastomas se correlacionan con el perfil genómico del tumor y el pronóstico del paciente. La combinación de datos anatómicos y patrón morfológico (volumen, forma, número de IDRFs) puede representar criterios pronósticos relevantes [148].

Por todo ello, la correlación entre la radiómica y radiómica profunda con diversos parámetros analíticos y de estadificación permite identificar patrones subyacentes y fenotipos tumorales. Estos análisis demuestran la utilidad del análisis correlacional de la imagen con parámetros como la expresión génica, abriendo la puerta al campo de la radiogenómica, campo que presenta un potencial considerable, ya que estudios futuros podrían lograr la extracción de la heterogeneidad tumoral en imágenes tridimensionales, pudiendo realizar biopsias virtuales del tumor completo, representando un avance significativo hacia la medicina de precisión al permitir la extracción de información genética y molecular para aplicar tratamientos más específicos y optimizados.

5.1.4. Modelos de predicción de supervivencia global

En los modelos elaborados en esta Tesis Doctoral, las variables radiómicas demostraron ser las características de imagen más relevantes para la predicción de la supervivencia global, superando a las características de radiómica profunda. El rendimiento del modelo Cox fue mayor que el resto de modelos.

Las variables que mostraron mayor representatividad en el modelo fueron todas variables de radiómica, sin incluir variables de radiómica profunda, por lo que el impacto de la radiómica resulta mayor para la predicción de la supervivencia global. La ausencia de variables de radiómica profunda predictoras de supervivencia global plantea la consideración del bajo impacto de las redes entrenadas (como ResNet) predominantemente con imágenes del mundo real en entornos de salud. Estas redes solo consideran información en un plano, y no capturan toda la heterogeneidad tumoral que una imagen 3D puede aportar.

En la literatura se ha demostrado la aplicación y beneficio del aprendizaje por transferencia desde ImageNet en imágenes médicas [149], sin embargo, las características de las imágenes médicas plantean desafíos distintos en comparación con las imágenes incluidas en el conjunto de datos de ImageNet, que son predominantemente imágenes no médicas. Las imágenes médicas suelen estar en escala de grises y presentan niveles más altos de ruido debido a su proceso de adquisición, y representan formas corporales complejas. Por lo tanto,

el aprendizaje por transferencia desde ImageNet puede introducir sesgos de los conjuntos de datos, con imágenes no relacionadas con la patología estudiada, comprometiendo así la robustez de las redes de aprendizaje profundo. Trabajos previos reflejan que el aprendizaje profundo necesita un gran volumen de datos para aumentar su rendimiento, pero en la práctica es difícil obtener conjuntos de datos que contengan imágenes médicas. Una posible solución a este problema es transferir el aprendizaje a través de modelos previamente entrenados, para hacerlos más apropiados para una tarea específica[150].

En nuestro conocimiento, no se han elaborado previamente estudios que permitan la predicción de supervivencia global en neuroblastoma. Sin embargo, en un estudio previo, realizado a partir de imágenes de RM potenciadas en T2w, analizó el potencial del análisis de radiómica en tumores neuroblásticos y permitió diferenciar el neuroblastoma del ganglioneuroblastoma y ganglioneuroma en niños. Los resultados indicaron que las características de radiómica, combinadas con la edad del paciente en el diagnóstico inicial, pueden proporcionar un método cuantitativo para distinguir estos tipos de tumores, lo que facilita la diferenciación patológica de tumores neuroblásticos periféricos en niños [151]. De manera similar, un estudio demostró la utilidad de la radiómica en secuencias T2w y DW de RM para establecer modelos que permitan diferenciar entre tumores renales benignos y malignos [152].

Otros estudios han permitido el desarrollo de modelos de aprendizaje profundo basados en radiómica en imágenes de RM potenciadas en T2w para predecir la supervivencia global en pacientes oncológicos [153], como un estudio desarrollado en pacientes con glioma en el que se llevó a cabo un análisis multivariante estándar, Cox-LASSO, para seleccionar características de radiómica para construir un modelo de predicción [76].

Cabe destacar que en los trabajos expuestos se han empleado secuencias de RM potenciadas en T2, puesto que esta es superior para identificar el límite del tumor y la diferenciación con las estructuras circundantes[76, 154].

Las 10 variables más representativas en el modelo de supervivencia fueron: *Shape least axis length*, *Skewness*, *Surface área*, *Maximum 2D diameter slice*, *Minor axis length*, *GLRLM Run length non uniformity*, *GLDM Dependence non uniformity*, *GLCM Maximum Correlation Coefficient*, *Shape Voxel Volume* y *Shape Mesh Volume*. Todas estas variables presentaron una relación positiva en el modelo de supervivencia, de modo que una mayor longitud del eje más corto del tumor, una mayor asimetría en la distribución de los vóxeles, una mayor superficie tumoral, un aumento del valor máximo del diámetro tumoral en 2D, una mayor longitud del eje menor del tumor, una textura más heterogénea o un mayor volumen tumoral demostraron asociar un mayor riesgo de evento para el paciente, aumentando el riesgo de muerte al aumentar estos parámetros.

Estas 10 variables presentaron diferentes grados de reproducibilidad y estabilidad en los análisis realizados previamente, y se relacionaron con algunas de las variables clínicas y de estadificación. Por tanto, las diez variables de radiómica expuestas podrían considerarse como variables candidatas a biomarcadores de imagen para la predicción de supervivencia global en pacientes pediátricos con tumores neuroblásticos.

La variable que mostró una mayor estabilidad en los análisis realizados previamente para evaluar la estabilidad y reproducibilidad de la radiómica fue *GLRLM Run length non uniformity* (no uniformidad de la longitud de ejecución), una característica de textura calculada a partir de la matriz GLRLM, matriz que mide la distribución de ejecuciones, secuencias de píxeles consecutivos con la misma intensidad de nivel de gris en una imagen. La *Run length non uniformity* mide la variabilidad de las longitudes de ejecución en la imagen, es decir, cuantifica cuánto se desvían las longitudes de ejecución consecutivas entre sí. Un valor más alto indica una mayor variabilidad en las longitudes de las ejecuciones, lo que sugiere una textura más compleja o heterogénea en la imagen [155]. Por tanto, esta heterogeneidad en la textura de la imagen se erige como biomarcador no invasivo para la predicción de supervivencia global en pacientes con neuroblastoma.

Por otro lado, no se han descrito hasta la fecha modelos de supervivencia basados en radiómica profunda.

5.2. Limitaciones y líneas de futuro

Nuestro estudio presentó diversas limitaciones. Se llevó a cabo en un conjunto de datos heterogéneo con parámetros de adquisición variables, momentos temporales (en el momento del diagnóstico o después del tratamiento) y ubicaciones diversas, lo que podría haber influido en los resultados. Sin embargo, todas estas variaciones reflejan la naturaleza del mundo real de los grandes conjuntos de datos retrospectivos utilizados para la tipificación fenotípica tumoral mediante radiómica.

Otra limitación potencial es que este estudio se realizó únicamente en tumores neuroblásticos, y oportunidades futuras podrían incluir el estudio de diferentes tipos de cáncer.

Además, el análisis solo incluyó imágenes potenciadas en T2, por lo que trabajos futuros podrían explorar el impacto de los análisis realizados en otras secuencias de RM, así como en otras pruebas de imagen como TC o SPECT.

En cuanto al entrenamiento de la herramienta de segmentación automática, las segmentaciones se realizaron solo por dos observadores, por lo que representa una pequeña fracción de la variabilidad real entre observadores y podría no ser suficiente para establecer un estándar de referencia. Además, ambas eran radiólogas con experiencia previa en radiología pediátrica y en problemas de segmentación. Por lo tanto, las segmentaciones manuales realizadas por otros usuarios (radiólogos menos experimentados, otros usuarios clínicos, personal no médico) podrían presentar una mayor variabilidad interobservador. Además, como se ha señalado previamente, la máscara considerada como patrón de referencia o verdad fundamental puede tener algunos errores asociados intrínsecamente al proceso de segmentación manual. Por otro lado, los tumores pueden asociarse con ganglios linfáticos extensos o pueden presentar contacto con ellos, siendo difícil en algunos casos su diferenciación, lo que, como se ha comprobado en esta Tesis Doctoral, puede conducir a errores en la segmentación

realizada por la herramienta automática de segmentación. Otra limitación es que la CNN fue entrenada con imágenes de RM en el plano transversal, y la validación solo se realizó en las imágenes transversales. Trabajos futuros podrían incluir el ajuste fino de la herramienta nnU-Net para realizar segmentaciones en planos sagitales y coronales.

En referencia a las limitaciones del estudio de validación externa de la herramienta de segmentación, únicamente una radióloga experimentada revisó las segmentaciones automáticas, por lo que no se evaluó la variabilidad interobservador. Sin embargo, esta limitación no es clínicamente relevante ya que en esta Tesis Doctoral se ha descrito una alta concordancia entre observadores para la segmentación manual en neuroblastoma (con un índice de superposición CSD mediano de 0.969).

En cuanto a los modelos de supervivencia y a la correlación de variables analíticas y de estadificación, cabe mencionar la construcción de modelos pronósticos que integren más variables clínicas y otros eventos clínicos finales, como el tiempo libre de evento o el tiempo a progresión para evaluar el impacto de la radiómica en los resultados clínicos de interés como líneas de desarrollo futuro.

6. CONCLUSIONES

CONCLUSIONES

- El análisis de las imágenes de Resonancia Magnética mediante algoritmos avanzados de procesamiento permite extraer características y descriptores de imagen cuantitativos con los que es posible establecer predicciones sobre la agresividad biológica, entendida como supervivencia global, de los tumores de estirpe neuroblástica en pacientes de edad pediátrica.
- La precisión de la segmentación manual de tumores neuroblásticos en imágenes de Resonancia Magnética potenciadas en T2w en el plano transversal llevada a cabo por radiólogos es comparable a la segmentación obtenida con la arquitectura de aprendizaje profundo nnU-Net. El modelo original de segmentación automática desarrollado en esta Tesis Doctoral supone una gran ventaja en tareas de segmentación, con una importante reducción del tiempo y de la participación del profesional de radiología en esta tarea.
- La herramienta de segmentación automática previamente entrenada nnU-Net, desarrollada en esta Tesis Doctoral, ha sido validada en un estudio independiente internacional, multicéntrico y multivendedor, mostrando su capacidad para localizar y segmentar los tumores neuroblásticos en las imágenes de Resonancia Magnética potenciadas en T2/T2* con gran porcentaje de éxito.
- Para garantizar la extracción reproducible de las características radiómicas, resulta esencial reportar cada paso del procesamiento de la imagen. Cualquier modificación de los pasos de preparación puede provocar cambios importantes en las características radiómicas. La normalización de la imagen es el factor más influyente en la estabilidad de las características radiómicas.
- Las variables radiómicas podrían constituir biomarcadores de imagen de relevancia para la diferenciación de neuroblastoma metastásico en

pacientes menores de 18 meses, mientras que la radiómica profunda presenta una mayor relación con la predicción de la mutación MYCN en pacientes mayores de 18 meses.

- El modelo de supervivencia global para neuroblastoma basado en la arquitectura Cox presenta una capacidad predictiva de supervivencia aceptable, siendo superior a otras arquitecturas. Las variables de radiómica son más relevantes para la predicción de la supervivencia global que las de radiómica profunda, pudiendo constituir biomarcadores de imagen para la predicción de supervivencia en tumores neuroblásticos. La construcción de modelos pronósticos que integren más variables clínicas y otros eventos clínicos finales se erige como línea de investigación futura.

7. BIBLIOGRAFÍA

BIBLIOGRAFÍA

- [1] S. N. Bhatnagar, «AN AUDIT OF MALIGNANT SOLID TUMORS IN INFANTS AND NEONATES», vol. 1, 2012.
- [2] PDQ Pediatric Treatment Editorial Board. Neuroblastoma Treatment (PDQ®): Health Professional Version. 2024 Feb 13. In: PDQ Cancer Information Summaries [Internet]. Bethesda (MD): National Cancer Institute (US); 2002-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK65747/>.
- [3] G. J. Lonergan, C. M. Schwab, E. S. Suarez, y C. L. Carlson, «From the Archives of the AFIP: Neuroblastoma, Ganglioneuroblastoma, and Ganglioneuroma: Radiologic-Pathologic Correlation», *RadioGraphics*, vol. 22, n.º 4, pp. 911-934, jul. 2002, doi: 10.1148/radiographics.22.4.g02j115911.
- [4] M. S. Irwin y J. R. Park, «Neuroblastoma», *Pediatr. Clin. North Am.*, vol. 62, n.º 1, pp. 225-256, feb. 2015, doi: 10.1016/j.pcl.2014.09.015.
- [5] American Cancer Society, «Key Statistics About Neuroblastoma». [En línea]. Disponible en: <https://www.cancer.org/cancer/types/neuroblastoma/about/key-statistics.html>
- [6] H. J. Brisse *et al.*, «Guidelines for Imaging and Staging of Neuroblastic Tumors: Consensus Report from the International Neuroblastoma Risk Group Project», *Radiology*, vol. 261, n.º 1, pp. 243-257, oct. 2011, doi: 10.1148/radiol.11101352.
- [7] B. De Bernardi *et al.*, «Neuroblastoma With Symptomatic Spinal Cord Compression at Diagnosis: Treatment and Results With 76 Cases», *J. Clin. Oncol.*, vol. 19, n.º 1, pp. 183-190, ene. 2001, doi: 10.1200/JCO.2001.19.1.183.
- [8] S. L. Cohn *et al.*, «The International Neuroblastoma Risk Group (INRG) Classification System: An INRG Task Force Report», *J. Clin. Oncol.*, vol. 27, n.º 2, pp. 289-297, ene. 2009, doi: 10.1200/JCO.2008.16.6785.
- [9] T. Monclair, V. Mosseri, G. Cecchetto, B. De Bernardi, J. Michon, y K. Holmes, «Influence of image-defined risk factors on the outcome of patients with localised neuroblastoma. A report from the LNESG1 study of the European International Society of Paediatric Oncology Neuroblastoma Group», *Pediatr. Blood Cancer*, vol. 62, n.º 9, pp. 1536-1542, sep. 2015, doi: 10.1002/pbc.25460.
- [10] J. Fischer *et al.*, «Complete surgical resection improves outcome in INRG high-risk patients with localized neuroblastoma older than 18 months», *BMC Cancer*, vol. 17, n.º 1, p. 520, dic. 2017, doi: 10.1186/s12885-017-3493-0.

- [11] K. Holmes *et al.*, «Influence of Surgical Excision on the Survival of Patients With Stage 4 High-Risk Neuroblastoma: A Report From the HR-NBL1/SIOPEN Study», *J. Clin. Oncol.*, vol. 38, n.º 25, pp. 2902-2915, sep. 2020, doi: 10.1200/JCO.19.03117.
- [12] J. F. Schäfer *et al.*, «Whole-body magnetic resonance imaging in pediatric oncology — recommendations by the Oncology Task Force of the ESPR», *Pediatr. Radiol.*, vol. 50, n.º 8, pp. 1162-1174, jul. 2020, doi: 10.1007/s00247-020-04683-4.
- [13] Matthay KK, Maris JM, Schleiermacher G, Nakagawara A, Mackall CL, Diller L, Weiss WA., «Neuroblastoma», *Nat Rev Primer*, vol. 10;2, p. 16078, nov. 2016.
- [14] P. Olivier *et al.*, «Guidelines for radioiodinated MIBG scintigraphy in children», *Eur. J. Nucl. Med. Mol. Imaging*, vol. 30, n.º 5, pp. B45-B50, may 2003, doi: 10.1007/s00259-003-1138-9.
- [15] K. K. Matthay *et al.*, «Criteria for evaluation of disease extent by 123I-metaiodobenzylguanidine scans in neuroblastoma: a report for the International Neuroblastoma Risk Group (INRG) Task Force», *Br. J. Cancer*, vol. 102, n.º 9, pp. 1319-1326, abr. 2010, doi: 10.1038/sj.bjc.6605621.
- [16] A. S. Littooi y B. De Keizer, «Imaging in neuroblastoma», *Pediatr. Radiol.*, vol. 53, n.º 4, pp. 783-787, sep. 2022, doi: 10.1007/s00247-022-05489-2.
- [17] R. Ladenstein *et al.*, «Validation of the mIBG skeletal SIOPEN scoring method in two independent high-risk neuroblastoma populations: the SIOPEN/HR-NBL1 and COG-A3973 trials», *Eur. J. Nucl. Med. Mol. Imaging*, vol. 45, n.º 2, pp. 292-305, feb. 2018, doi: 10.1007/s00259-017-3829-7.
- [18] V. Lewington *et al.*, «123I-mIBG scintigraphy in neuroblastoma: development of a SIOPEN semi-quantitative reporting method by an international panel», *Eur. J. Nucl. Med. Mol. Imaging*, vol. 44, n.º 2, pp. 234-241, feb. 2017, doi: 10.1007/s00259-016-3516-0.
- [19] C. Li *et al.*, «Prognostic value of metabolic indices and bone marrow uptake pattern on preoperative 18F-FDG PET/CT in pediatric patients with neuroblastoma», *Eur. J. Nucl. Med. Mol. Imaging*, vol. 45, n.º 2, pp. 306-315, feb. 2018, doi: 10.1007/s00259-017-3851-9.
- [20] D. A. Morgenstern *et al.*, «Risk stratification of high-risk metastatic neuroblastoma: A report from the HR-NBL-1/SIOPEN study», *Pediatr. Blood Cancer*, vol. 65, n.º 11, p. e27363, nov. 2018, doi: 10.1002/pbc.27363.
- [21] A. Ghosh, E. Yekeler, S. R. Teixeira, D. Dalal, y L. States, «Role of MRI radiomics for the prediction of MYCN amplification in neuroblastomas», *Eur. Radiol.*,

vol. 33, n.º 10, pp. 6726-6735, may 2023, doi: 10.1007/s00330-023-09628-7.

[22] E. F. Attiyeh, Q. Wang, P. W. McGrady, H. Shimada, y K. K. Matthay, «Chromosome 1p and 11q Deletions and Outcome in Neuroblastoma», *N. Engl. J. Med.*, 2005.

[23] K. T. Vo *et al.*, «Clinical, Biologic, and Prognostic Differences on the Basis of Primary Tumor Site in Neuroblastoma: A Report From the International Neuroblastoma Risk Group Project», *J. Clin. Oncol.*, vol. 32, n.º 28, pp. 3169-3176, oct. 2014, doi: 10.1200/JCO.2014.56.1621.

[24] G. M. Brodeur *et al.*, «International criteria for diagnosis, staging, and response to treatment in patients with neuroblastoma.».

[25] T. Monclair *et al.*, «The International Neuroblastoma Risk Group (INRG) Staging System: An INRG Task Force Report», *J. Clin. Oncol.*, vol. 27, n.º 2, pp. 298-303, ene. 2009, doi: 10.1200/JCO.2008.16.6876.

[26] J. R. Park *et al.*, «Revisions to the International Neuroblastoma Response Criteria: A Consensus Statement From the National Cancer Institute Clinical Trials Planning Meeting», *J. Clin. Oncol.*, vol. 35, n.º 22, pp. 2580-2587, ago. 2017, doi: 10.1200/JCO.2016.72.0177.

[27] Ø. E. Olsen, «Why measure tumours?», *Pediatr. Radiol.*, vol. 45, n.º 1, pp. 35-41, ene. 2015, doi: 10.1007/s00247-014-3148-0.

[28] M. S. Irwin *et al.*, «Revised Neuroblastoma Risk Classification System: A Report From the Children's Oncology Group», *J. Clin. Oncol.*, vol. 39, n.º 29, pp. 3229-3241, oct. 2021, doi: 10.1200/JCO.21.00278.

[29] Luis Martí-Bonmatí, Ángel Alverich Bayarri, *Imaging Biomarkers: Development and Clinical Integration*. Springer, 2016. [En línea]. Disponible en: <http://dx.doi.org/10.1007/978-3-319-43504-6>.

[30] J. Damián Segrelles Quilis, S. López-Huguet, P. Lozano, y I. Blanquer, «A federated cloud architecture for processing of cancer images on a distributed storage», *Future Gener. Comput. Syst.*, vol. 139, pp. 38-52, feb. 2023, doi: 10.1016/j.future.2022.09.019.

[31] L. Martí-Bonmatí *et al.*, «PRIMAGE project: predictive in silico multiscale analytics to support childhood cancer personalised evaluation empowered by imaging biomarkers», *Eur. Radiol. Exp.*, vol. 4, n.º 1, p. 22, dic. 2020, doi: 10.1186/s41747-020-00150-9.

[32] D. Cozzi *et al.*, «Radiomics in pulmonary neuroendocrine tumours (NETs)»,

Radiol. Med. (Torino), vol. 127, n.º 6, pp. 609-615, may 2022, doi: 10.1007/s11547-022-01494-5.

[33] G. Carbonell *et al.*, «Precision of MRI radiomics features in the liver and hepatocellular carcinoma», *Eur. Radiol.*, vol. 32, n.º 3, pp. 2030-2040, mar. 2022, doi: 10.1007/s00330-021-08282-1.

[34] D. C. Sullivan *et al.*, «Metrology Standards for Quantitative Imaging Biomarkers», *Radiology*, vol. 277, n.º 3, pp. 813-825, dic. 2015, doi: 10.1148/radiol.2015142202.

[35] H. Neubauer, M. Li, V. Müller, T. Pabst, y M. Beer, «Diagnostic Value of Diffusion-Weighted MRI for Tumor Characterization, Differentiation and Monitoring in Pediatric Patients with Neuroblastic Tumors», *RöFo - Fortschritte Auf Dem Geb. Röntgenstrahlen Bildgeb. Verfahr.*, vol. 189, n.º 07, pp. 640-650, jul. 2017, doi: 10.1055/s-0043-108993.

[36] A.-L. Peschmann *et al.*, «Quantitative DWI predicts event-free survival in children with neuroblastic tumours: preliminary findings from a retrospective cohort study», *Eur. Radiol. Exp.*, vol. 3, n.º 1, p. 6, dic. 2019, doi: 10.1186/s41747-019-0087-4.

[37] L. Cerdá Alberich *et al.*, «A Confidence Habitats Methodology in MR Quantitative Diffusion for the Classification of Neuroblastic Tumors», *Cancers*, vol. 12, n.º 12, p. 3858, dic. 2020, doi: 10.3390/cancers12123858.

[38] H. I. Serin *et al.*, «Diffusion weighted imaging in differentiating malignant and benign neuroblastic tumors», *Jpn. J. Radiol.*, vol. 34, n.º 9, pp. 620-624, sep. 2016, doi: 10.1007/s11604-016-0565-z.

[39] C. Li *et al.*, «Metabolic Evaluation of MYCN-Amplified Neuroblastoma by 4-[18F]FGln PET Imaging», *Mol. Imaging Biol.*, vol. 21, n.º 6, pp. 1117-1126, dic. 2019, doi: 10.1007/s11307-019-01330-9.

[40] G. Chartrand *et al.*, «Deep Learning: A Primer for Radiologists», *RadioGraphics*, vol. 37, n.º 7, pp. 2113-2131, nov. 2017, doi: 10.1148/rg.2017170077.

[41] J. E. Van Timmeren, D. Cester, S. Tanadini-Lang, H. Alkadhi, y B. Baessler, «Radiomics in medical imaging—“how-to” guide and critical reflection», *Insights Imaging*, vol. 11, n.º 1, p. 91, dic. 2020, doi: 10.1186/s13244-020-00887-2.

[42] M. Sun *et al.*, «Robustness and reproducibility of radiomics in T2 weighted images from magnetic resonance image guided linear accelerator in a phantom study», *Phys. Med.*, vol. 96, pp. 130-139, abr. 2022, doi: 10.1016/j.ejmp.2022.03.002.

[43] L. Joskowicz, D. Cohen, N. Caplan, y J. Sosna, «Inter-observer variability of

manual contour delineation of structures in CT», *Eur. Radiol.*, vol. 29, n.º 3, pp. 1391-1399, mar. 2019, doi: 10.1007/s00330-018-5695-5.

[44] H. K. Bø, O. Solheim, A. S. Jakola, K.-A. Kvistad, I. Reinertsen, y E. M. Berntsen, «Intra-rater variability in low-grade glioma segmentation», *J. Neurooncol.*, vol. 131, n.º 2, pp. 393-402, ene. 2017, doi: 10.1007/s11060-016-2312-9.

[45] P. A. Yushkevich, Y. Gao, y G. Gerig, «ITK-SNAP: An interactive tool for semi-automatic segmentation of multi-modality biomedical images», en *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Orlando, FL, USA: IEEE, ago. 2016, pp. 3342-3345. doi: 10.1109/EMBC.2016.7591443.

[46] H. J. Deglint, R. M. Rangayyan, F. J. Ayres, G. S. Boag, y M. K. Zuffo, «Three-Dimensional Segmentation of the Tumor in Computed Tomographic Images of Neuroblastoma», *J. Digit. Imaging*, vol. 20, n.º 1, pp. 72-87, mar. 2007, doi: 10.1007/10278-006-0769-3.

[47] S. Gassenmaier *et al.*, «Feasibility and possible value of quantitative semi-automated diffusion weighted imaging volumetry of neuroblastic tumors», *Cancer Imaging*, vol. 20, n.º 1, p. 89, dic. 2020, doi: 10.1186/s40644-020-00366-3.

[48] R. M. Rangayyan, S. Banik, y G. S. Boag, «Landmarking and segmentation of computed tomographic images of pediatric patients with neuroblastoma», *Int. J. Comput. Assist. Radiol. Surg.*, vol. 4, n.º 3, pp. 245-262, may 2009, doi: 10.1007/s11548-009-0289-y.

[49] S. M. McKinney *et al.*, «International evaluation of an AI system for breast cancer screening», *Nature*, vol. 577, n.º 7788, pp. 89-94, ene. 2020, doi: 10.1038/s41586-019-1799-6.

[50] K. Yasaka, H. Akai, O. Abe, y S. Kiryu, «Deep Learning with Convolutional Neural Network for Differentiation of Liver Masses at Dynamic Contrast-enhanced CT: A Preliminary Study», *Radiology*, vol. 286, n.º 3, pp. 887-896, mar. 2018, doi: 10.1148/radiol.2017170706.

[51] F. Feng, J. A. Ashton-Miller, J. O. L. DeLancey, y J. Luo, «Convolutional neural network-based pelvic floor structure segmentation using magnetic resonance imaging in pelvic organ prolapse», *Med. Phys.*, vol. 47, n.º 9, pp. 4281-4293, sep. 2020, doi: 10.1002/mp.14377.

[52] O. Ronneberger, P. Fischer, y T. Brox, «U-Net: Convolutional Networks for Biomedical Image Segmentation», en *Medical Image Computing and Computer-Assisted*

Intervention – MICCAI 2015, N. Navab, J. Hornegger, W. M. Wells, y A. F. Frangi, Eds., Cham: Springer International Publishing, 2015, pp. 234-241.

[53] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, y K. H. Maier-Hein, «nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation», *Nat. Methods*, vol. 18, n.º 2, pp. 203-211, feb. 2021, doi: 10.1038/s41592-020-01008-z.

[54] N. Alves, M. Schuurmans, G. Litjens, J. S. Bosma, J. Hermans, y H. Huisman, «Fully Automatic Deep Learning Framework for Pancreatic Ductal Adenocarcinoma Detection on Computed Tomography», *Cancers*, vol. 14, n.º 2, p. 376, ene. 2022, doi: 10.3390/cancers14020376.

[55] C. L. Ramspek, K. J. Jager, F. W. Dekker, C. Zoccali, y M. van Diepen, «External validation of prognostic models: what, why, how, when and where?», *Clin. Kidney J.*, vol. 14, n.º 1, pp. 49-58, feb. 2021, doi: 10.1093/ckj/sfaa188.

[56] G. C. M. Siontis, I. Tzoulaki, P. J. Castaldi, y J. P. A. Ioannidis, «External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination», *J. Clin. Epidemiol.*, vol. 68, n.º 1, pp. 25-34, ene. 2015, doi: 10.1016/j.jclinepi.2014.09.007.

[57] D. Bouget *et al.*, «Glioblastoma Surgery Imaging–Reporting and Data System: Validation and Performance of the Automated Segmentation Task», *Cancers*, vol. 13, n.º 18, p. 4674, sep. 2021, doi: 10.3390/cancers13184674.

[58] B. David *et al.*, «External validation of automated focal cortical dysplasia detection using morphometric analysis», *Epilepsia*, vol. 62, n.º 4, pp. 1005-1021, abr. 2021, doi: 10.1111/epi.16853.

[59] A. C. Yu, B. Mohajer, y J. Eng, «External Validation of Deep Learning Algorithms for Radiologic Diagnosis: A Systematic Review», *Radiol. Artif. Intell.*, vol. 4, n.º 3, p. e210064, 2022, doi: 10.1148/ryai.210064.

[60] X. Zhang *et al.*, «Deep Learning With Radiomics for Disease Diagnosis and Treatment: Challenges and Potential», *Front. Oncol.*, vol. 12, p. 773840, feb. 2022, doi: 10.3389/fonc.2022.773840.

[61] M. R. Tomaszewski y R. J. Gillies, «The Biological Meaning of Radiomic Features», *Radiology*, vol. 298, n.º 3, pp. 505-516, mar. 2021, doi: 10.1148/radiol.2021202553.

[62] P. Lambin *et al.*, «Radiomics: the bridge between medical imaging and personalized medicine», *Nat. Rev. Clin. Oncol.*, vol. 14, n.º 12, pp. 749-762, dic. 2017, doi: 10.1038/nrclinonc.2017.141.

- [63] A. Zwanenburg *et al.*, «The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping», *Radiology*, vol. 295, n.º 2, pp. 328-338, may 2020, doi: 10.1148/radiol.2020191145.
- [64] J. E. Park *et al.*, «Quality of science and reporting of radiomics in oncologic studies: room for improvement according to radiomics quality score and TRIPOD statement», *Eur. Radiol.*, vol. 30, n.º 1, pp. 523-536, ene. 2020, doi: 10.1007/s00330-019-06360-z.
- [65] H. Wang *et al.*, «Reproducibility and Repeatability of CBCT-Derived Radiomics Features», *Front. Oncol.*, vol. 11, p. 773512, nov. 2021, doi: 10.3389/fonc.2021.773512.
- [66] R. Berenguer *et al.*, «Radiomics of CT Features May Be Nonreproducible and Redundant: Influence of CT Acquisition Parameters», *Radiology*, vol. 288, n.º 2, pp. 407-415, ago. 2018, doi: 10.1148/radiol.2018172361.
- [67] A. Euler *et al.*, «Virtual Monoenergetic Images of Dual-Energy CT—Impact on Repeatability, Reproducibility, and Classification in Radiomics», *Cancers*, vol. 13, n.º 18, p. 4710, sep. 2021, doi: 10.3390/cancers13184710.
- [68] L. J. Isaksson *et al.*, «Effects of MRI image normalization techniques in prostate cancer radiomics», *Phys. Med.*, vol. 71, pp. 7-13, mar. 2020, doi: 10.1016/j.ejmp.2020.02.007.
- [69] H. Moradmand, S. M. R. Aghamiri, y R. Ghaderi, «Impact of image preprocessing methods on reproducibility of radiomic features in multimodal magnetic resonance imaging in glioblastoma», *J. Appl. Clin. Med. Phys.*, vol. 21, n.º 1, pp. 179-190, ene. 2020, doi: 10.1002/acm2.12795.
- [70] K. V. Hoebel *et al.*, «Radiomics Repeatability Pitfalls in a Scan-Rescan MRI Study of Glioblastoma», *Radiol. Artif. Intell.*, vol. 3, n.º 1, p. e190199, ene. 2021, doi: 10.1148/ryai.2020190199.
- [71] J. Wong *et al.*, «Effects of interobserver and interdisciplinary segmentation variabilities on CT-based radiomics for pancreatic cancer», *Sci. Rep.*, vol. 11, n.º 1, p. 16328, ago. 2021, doi: 10.1038/s41598-021-95152-x.
- [72] L. J. Jensen, D. Kim, T. Elgeti, I. G. Steffen, B. Hamm, y S. N. Nagel, «Stability of Liver Radiomics across Different 3D ROI Sizes—An MRI In Vivo Study», *Tomography*, vol. 7, n.º 4, pp. 866-876, dic. 2021, doi: 10.3390/tomography7040073.
- [73] Z. Ramli *et al.*, «Stability and Reproducibility of Radiomic Features Based on Various Segmentation Techniques on Cervical Cancer DWI-MRI», *Diagnostics*, vol. 12, n.º 12, p. 3125, dic. 2022, doi: 10.3390/diagnostics12123125.

- [74] R. Paul *et al.*, «Representation of Deep Features using Radiologist defined Semantic Features», en *2018 International Joint Conference on Neural Networks (IJCNN)*, Rio de Janeiro: IEEE, jul. 2018, pp. 1-7. doi: 10.1109/IJCNN.2018.8489440.
- [75] R. Paul *et al.*, «Explaining Deep Features Using Radiologist-Defined Semantic Features and Traditional Quantitative Features», *Tomography*, vol. 5, n.º 1, pp. 192-200, mar. 2019, doi: 10.18383/j.tom.2018.00034.
- [76] G. Li *et al.*, «An MRI radiomics approach to predict survival and tumour-infiltrating macrophages in gliomas», *Brain*, vol. 145, n.º 3, pp. 1151-1161, abr. 2022, doi: 10.1093/brain/awab340.
- [77] QUIBIM Precision, «<https://cordis.europa.eu/project/id/778064>».
- [78] «Organización Panamericana de la Salud y Consejo de Organizaciones Internacionales de las Ciencias Médica. Pautas éticas internacionales para la investigación relacionada con la salud con seres humanos, Cuarta Edición. Ginebra: Consejo de Organizaciones Internacionales de las Ciencias Médicas (CIOMS); 2016.»
- [79] J. Mongan, L. Moy, y C. E. Kahn, «Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers», *Radiol. Artif. Intell.*, vol. 2, n.º 2, p. e200029, mar. 2020, doi: 10.1148/ryai.2020200029.
- [80] L. Cerdá-Alberich *et al.*, «MAIC–10 brief quality checklist for publications using artificial intelligence and medical images», *Insights Imaging*, vol. 14, n.º 1, p. 11, ene. 2023, doi: 10.1186/s13244-022-01355-9.
- [81] C. Baeza-Delgado *et al.*, «A practical solution to estimate the sample size required for clinical prediction models generated from observational research on data», *Eur. Radiol. Exp.*, vol. 6, n.º 1, p. 22, dic. 2022, doi: 10.1186/s41747-022-00276-y.
- [82] R. D. Riley *et al.*, «Calculating the sample size required for developing a clinical prediction model», *BMJ*, p. m441, mar. 2020, doi: 10.1136/bmj.m441.
- [83] W. B. London *et al.*, «Clinical and Biologic Features Predictive of Survival After Relapse of Neuroblastoma: A Report From the International Neuroblastoma Risk Group Project», *J. Clin. Oncol.*, vol. 29, n.º 24, pp. 3286-3292, ago. 2011, doi: 10.1200/JCO.2010.34.3392.
- [84] Y. Al-Tonbary *et al.*, «Clinico-epidemiology of neuroblastoma in north east Egypt: A 5-year multicenter study», *Oncol. Lett.*, vol. 10, n.º 2, pp. 1054-1062, ago. 2015, doi: 10.3892/ol.2015.3335.
- [85] G. S. Collins, E. O. Ogundimu, y D. G. Altman, «Sample size considerations for the external validation of a multivariable prognostic model: a resampling study», *Stat.*

Med., vol. 35, n.º 2, pp. 214-226, ene. 2016, doi: 10.1002/sim.6787.

[86] «REGLAMENTO (UE) 2016/ 679 DEL PARLAMENTO EUROPEO Y DEL CONSEJO - de 27 de abril de 2016 - relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/ 46/ CE (Reglamento general de protección de datos)».

[87] Ebner H, Hayn D, Falgenhauer M, Nitzlader M, Schleiermacher G, Haupt R, Erminio G, Defferrari R, Mazzocco K, Kohler J, Tonini GP, Ladenstein R, Schreier G., «Piloting the European Unified Patient Identity Management (EUPID) Concept to Facilitate Secondary Use of Neuroblastoma Data from Clinical Trials and Biobanking», *Stud Health Technol Inf.*, vol. 31, n.º 8, p. 223, 2016, doi: 10.1055/s-0038-1625339.

[88] European Society of Radiology (ESR), «ESR statement on the stepwise development of imaging biomarkers», *Insights Imaging*, vol. 4, n.º 2, pp. 147-152, abr. 2013, doi: 10.1007/s13244-013-0220-5.

[89] IQVIA Institute for Human Data Science, «Evolving Oncology Endpoints. A New Horizon for Oncology Outcomes». julio de 2021.

[90] C. Lakshmi Devasena y M. Hemalatha, «Noise Removal in Magnetic Resonance Images using Hybrid KSL Filtering Technique», *Int. J. Comput. Appl.*, vol. 27, n.º 8, pp. 1-4, ago. 2011, doi: 10.5120/3324-4571.

[91] M. Fernández Patón *et al.*, «MR Denoising Increases Radiomic Biomarker Precision and Reproducibility in Oncologic Imaging», *J. Digit. Imaging*, vol. 34, n.º 5, pp. 1134-1145, oct. 2021, doi: 10.1007/s10278-021-00512-8.

[92] L. Alvarez, P.-L. Lions, y J.-M. Morel, «Image Selective Smoothing and Edge Detection by Nonlinear Diffusion. II», *SIAM J. Numer. Anal.*, vol. 29, n.º 3, pp. 845-866, jun. 1992, doi: 10.1137/0729052.

[93] Sethian, J. a., *Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science*. 1999.

[94] F. A. M. Cappabianco y P. P. C. E. da Silva, «Non-local Operational Anisotropic Diffusion Filter». arXiv, 11 de diciembre de 2018. Accedido: 11 de octubre de 2023. [En línea]. Disponible en: <http://arxiv.org/abs/1812.04708>

[95] A. Buades, B. Coll, y J.-M. Morel, «A Non-Local Algorithm for Image Denoising», en *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, CA, USA: IEEE, 2005, pp. 60-65. doi:

10.1109/CVPR.2005.38.

[96] B. C. Lowekamp, D. T. Chen, L. Ibáñez, y D. Blezek, «The Design of SimpleITK», *Front. Neuroinformatics*, vol. 7, 2013, doi: 10.3389/fninf.2013.00045.

[97] S. Song, Y. Zheng, y Y. He, «A review of Methods for Bias Correction in Medical Images», *Biomed. Eng. Rev.*, vol. 3, n.º 1, 2017, doi: 10.18103/bme.v3i1.1550.

[98] N. J. Tustison *et al.*, «N4ITK: Improved N3 Bias Correction», *IEEE Trans. Med. Imaging*, vol. 29, n.º 6, pp. 1310-1320, jun. 2010, doi: 10.1109/TMI.2010.2046908.

[99] A. Dovrou *et al.*, «A segmentation-based method improving the performance of N4 bias field correction on T2weighted MR imaging data of the prostate», *Magn. Reson. Imaging*, vol. 101, pp. 1-12, sep. 2023, doi: 10.1016/j.mri.2023.03.012.

[100] B. B. Avants, N. Tustison, y H. Johnson, «Advanced Normalization Tools (ANTS)».

[101] L. Hadjiiski *et al.*, «AAPM task group report 273: Recommendations on best practices for AI and machine learning for computer-aided diagnosis in medical imaging», *Med. Phys.*, vol. 50, n.º 2, feb. 2023, doi: 10.1002/mp.16188.

[102] Akihiro Haga, Wataru Takahashi, Shuri Aoki, Kanabu Nawa , Hideomi Yamashita, Osamu Abe, Keiichi Nakagawa, «Standardization of imaging features for radiomics analysis», *J Med Invest*, vol. 66, pp. 35-37, 2019, doi: 10.2152/jmi.66.35.

[103] J. J. M. Van Griethuysen *et al.*, «Computational Radiomics System to Decode the Radiographic Phenotype», *Cancer Res.*, vol. 77, n.º 21, pp. e104-e107, nov. 2017, doi: 10.1158/0008-5472.CAN-17-0339.

[104] L. R. Dice, «Measures of the amount of ecologic association between species», *Ecology*, vol. 26, pp. 297-302., 1945.

[105] K. H. Zou *et al.*, «Statistical validation of image segmentation quality based on a spatial overlap index1», *Acad. Radiol.*, vol. 11, n.º 2, pp. 178-189, feb. 2004, doi: 10.1016/S1076-6332(03)00671-8.

[106] C. Luo, C. Shi, X. Li, y D. Gao, «Cardiac MR segmentation based on sequence propagation by deep learning», *PLOS ONE*, vol. 15, n.º 4, p. e0230415, abr. 2020, doi: 10.1371/journal.pone.0230415.

[107] G. Chlebus *et al.*, «Reducing inter-observer variability and interaction time of MR liver volumetry by combining automatic CNN-based liver segmentation and manual corrections», *PLOS ONE*, vol. 14, n.º 5, p. e0217228, may 2019, doi: 10.1371/journal.pone.0217228.

[108] P. Jaccard, «THE DISTRIBUTION OF THE FLORA IN THE ALPINE

ZONE.1», *New Phytol.*, vol. 11, n.º 2, pp. 37-50, feb. 1912, doi: 10.1111/j.1469-8137.1912.tb05611.x.

[109] A. A. Taha y A. Hanbury, «Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool», *BMC Med. Imaging*, vol. 15, n.º 1, p. 29, dic. 2015, doi: 10.1186/s12880-015-0068-x.

[110] Lawrence I-Kuei Lin, «A Concordance Correlation Coefficient to Evaluate Reproducibility», *Biometrics*, vol. 45, n.º 1, pp. 255-268, 1989, doi: 10.2307/2532051.

[111] P. Virtanen *et al.*, «SciPy 1.0: fundamental algorithms for scientific computing in Python», *Nat. Methods*, vol. 17, n.º 3, pp. 261-272, mar. 2020, doi: 10.1038/s41592-019-0686-2.

[112] R. Vallat, «Pingouin: statistics in Python», *J. Open Source Softw.*, vol. 3, n.º 31, p. 1026, nov. 2018, doi: 10.21105/joss.01026.

[113] R. Bey, R. Goussault, F. Grolleau, M. Benchoufi, y R. Porcher, «Fold-stratified cross-validation for unbiased and privacy-preserving federated learning», *J. Am. Med. Inform. Assoc.*, vol. 27, n.º 8, pp. 1244-1251, ago. 2020, doi: 10.1093/jamia/ocaa096.

[114] F. Pedregosa *et al.*, «Scikit-learn: Machine Learning in Python», *Mach. Learn. PYTHON*, p. 6.

[115] D. Veiga-Canuto *et al.*, «Comparative Multicentric Evaluation of Inter-Observer Variability in Manual and Automatic Segmentation of Neuroblastic Tumors in Magnetic Resonance Images», *Cancers*, vol. 14, n.º 15, p. 3648, jul. 2022, doi: 10.3390/cancers14153648.

[116] D. Veiga-Canuto *et al.*, «Independent Validation of a Deep Learning nnU-Net Tool for Neuroblastoma Detection and Segmentation in MR Images», *Cancers*, vol. 15, n.º 5, p. 1622, mar. 2023, doi: 10.3390/cancers15051622.

[117] F. Banterle, M. Corsini, P. Cignoni, y R. Scopigno, «A Low-Memory, Straightforward and Fast Bilateral Filter Through Subsampling in Spatial Domain», *Comput. Graph. Forum*, vol. 31, n.º 1, pp. 19-32, feb. 2012, doi: 10.1111/j.1467-8659.2011.02078.x.

[118] «ImageNet. <http://www.image-net.org>».

[119] R. Bellman, R. Corporation, y K. M. R. Collection, *Dynamic Programming*. en Rand Corporation research study. Princeton University Press, 1957. [En línea]. Disponible en: <https://books.google.es/books?id=wtdoPwAACAAJ>

[120] Cox D. R., «Regression Models and Life-Tables.», *Ournal R. Stat. Soc. Ser. B Methodol.*, n.º 34, pp. 187-202, 1972, doi: <https://doi.org/10.1111/j.2517->

6161.1972.tb00899.x.

- [121] Boehmke, Bradley; Greenwell, Brandon, *Hands-On Machine Learning with R*, vol. 221–245. Chapman & Hall, 2019.
- [122] Chun-Fu Lin y Sheng-De Wang, «Fuzzy support vector machines», *IEEE Trans. Neural Netw.*, vol. 13, n.º 2, pp. 464-471, mar. 2002, doi: 10.1109/72.991432.
- [123] «Van Belle, V., Pelckmans, K., Suykens, J., & Van Huffel, S. (2007). Support vector machines for survival analysis. In Proc. of the Third International Conference on Computational Intelligence in Medicine and Healthcare (CIMED2007).».
- [124] S. M. Lundberg y S.-I. Lee, «A Unified Approach to Interpreting Model Predictions».
- [125] S. Montagne *et al.*, «Challenge of prostate MRI segmentation on T2-weighted images: inter-observer variability and impact of prostate morphology», *Insights Imaging*, vol. 12, n.º 1, p. 71, dic. 2021, doi: 10.1186/s13244-021-01010-9.
- [126] C. R. Meyer *et al.*, «Evaluation of Lung MDCT Nodule Annotation Across Radiologists and Methods», *Acad. Radiol.*, vol. 13, n.º 10, pp. 1254-1265, oct. 2006, doi: 10.1016/j.acra.2006.07.012.
- [127] R. W. Y. Granzier *et al.*, «MRI-based radiomics in breast cancer: feature robustness with respect to inter-observer segmentation variability», *Sci. Rep.*, vol. 10, n.º 1, p. 14163, dic. 2020, doi: 10.1038/s41598-020-70940-z.
- [128] N. M. deSouza *et al.*, «Standardised lesion segmentation for imaging biomarker quantitation: a consensus recommendation from ESR and EORTC», *Insights Imaging*, vol. 13, n.º 1, p. 159, oct. 2022, doi: 10.1186/s13244-022-01287-4.
- [129] S. S. F. Yip *et al.*, «Application of the 3D slicer chest imaging platform segmentation algorithm for large lung nodule delineation», *PLOS ONE*, vol. 12, n.º 6, p. e0178944, jun. 2017, doi: 10.1371/journal.pone.0178944.
- [130] C. Parmar *et al.*, «Robust Radiomics Feature Quantification Using Semiautomatic Volumetric Segmentation», *PLoS ONE*, vol. 9, n.º 7, p. e102107, jul. 2014, doi: 10.1371/journal.pone.0102107.
- [131] D. W. Kim, H. Y. Jang, K. W. Kim, Y. Shin, y S. H. Park, «Design Characteristics of Studies Reporting the Performance of Artificial Intelligence Algorithms for Diagnostic Analysis of Medical Images: Results from Recently Published Papers», *Korean J. Radiol.*, vol. 20, n.º 3, p. 405, 2019, doi: 10.3348/kjr.2019.0025.
- [132] A. Jimenez-Pastor *et al.*, «Precise whole liver automatic segmentation and quantification of PDFF and R2* on MR images», *Eur. Radiol.*, vol. 31, n.º 10, pp. 7876-

7887, oct. 2021, doi: 10.1007/s00330-021-07838-5.

[133] X. Han *et al.*, «Automated segmentation of liver segment on portal venous phase MR images using a 3D convolutional neural network», *Insights Imaging*, vol. 13, n.º 1, p. 26, dic. 2022, doi: 10.1186/s13244-022-01163-1.

[134] C. Chen *et al.*, «Automated segmentation of craniopharyngioma on MR images using U-Net-based deep convolutional neural network», *Eur. Radiol.*, nov. 2022, doi: 10.1007/s00330-022-09216-1.

[135] A. Defeudis *et al.*, «MRI-based radiomics to predict response in locally advanced rectal cancer: comparison of manual and automatic segmentation on external validation in a multicentre study», *Eur. Radiol. Exp.*, vol. 6, n.º 1, p. 19, dic. 2022, doi: 10.1186/s41747-022-00272-2.

[136] V. Bourbonne *et al.*, «External Validation of an MRI-Derived Radiomics Model to Predict Biochemical Recurrence after Surgery for High-Risk Prostate Cancer», *Cancers*, vol. 12, n.º 4, p. 814, mar. 2020, doi: 10.3390/cancers12040814.

[137] A. Bordron *et al.*, «External Validation of a Radiomics Model for the Prediction of Complete Response to Neoadjuvant Chemoradiotherapy in Rectal Cancer», *Cancers*, vol. 14, n.º 4, p. 1079, feb. 2022, doi: 10.3390/cancers14041079.

[138] J. P. B. O'Connor *et al.*, «Imaging biomarker roadmap for cancer studies», *Nat. Rev. Clin. Oncol.*, vol. 14, n.º 3, pp. 169-186, mar. 2017, doi: 10.1038/nrclinonc.2016.162.

[139] V. Roblot *et al.*, «Validation of a deep learning segmentation algorithm to quantify the skeletal muscle index and sarcopenia in metastatic renal carcinoma», *Eur. Radiol.*, vol. 32, n.º 7, pp. 4728-4737, jul. 2022, doi: 10.1007/s00330-022-08579-9.

[140] R. Bagatell *et al.*, «Assessment of Primary Site Response in Children With High-Risk Neuroblastoma: An International Multicenter Study», *J. Clin. Oncol.*, vol. 34, n.º 7, pp. 740-746, mar. 2016, doi: 10.1200/JCO.2015.63.2042.

[141] H. Nasief *et al.*, «A machine learning based delta-radiomics process for early prediction of treatment response of pancreatic cancer», *Npj Precis. Oncol.*, vol. 3, n.º 1, p. 25, dic. 2019, doi: 10.1038/s41698-019-0096-z.

[142] H. Kim *et al.*, «Can deep learning reduce the time and effort required for manual segmentation in 3D reconstruction of MRI in rotator cuff tears?», *PLOS ONE*, vol. 17, n.º 10, p. e0274075, oct. 2022, doi: 10.1371/journal.pone.0274075.

[143] J. D. Rudie *et al.*, «Longitudinal Assessment of Posttreatment Diffuse Glioma Tissue Volumes with Three-dimensional Convolutional Neural Networks», *Radiol. Artif.*

- Intell.*, vol. 4, n.º 5, p. e210243, sep. 2022, doi: 10.1148/ryai.210243.
- [144] M. Pavic *et al.*, «Influence of inter-observer delineation variability on radiomics stability in different tumor sites», *Acta Oncol.*, vol. 57, n.º 8, pp. 1070-1074, ago. 2018, doi: 10.1080/0284186X.2018.1445283.
- [145] M. Shafiq-ul-Hassan *et al.*, «Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels», *Med. Phys.*, vol. 44, n.º 3, pp. 1050-1062, mar. 2017, doi: 10.1002/mp.12123.
- [146] H. J. Yoon, J. Kang, H. Park, I. Sohn, S.-H. Lee, y H. Y. Lee, «Deciphering the tumor microenvironment through radiomics in non-small cell lung cancer: Correlation with immune profiles», *PLOS ONE*, vol. 15, n.º 4, p. e0231227, abr. 2020, doi: 10.1371/journal.pone.0231227.
- [147] G.-H. Su *et al.*, «Radiomics features for assessing tumor-infiltrating lymphocytes correlate with molecular traits of triple-negative breast cancer», *J. Transl. Med.*, vol. 20, n.º 1, p. 471, oct. 2022, doi: 10.1186/s12967-022-03688-x.
- [148] H. J. Brisse *et al.*, «Radiogenomics of neuroblastomas: Relationships between imaging phenotypes, tumor genomic profile and survival», *PLOS ONE*, vol. 12, n.º 9, p. e0185190, sep. 2017, doi: 10.1371/journal.pone.0185190.
- [149] A. Cadrin-Chênevert, «Moving from ImageNet to RadImageNet for Improved Transfer Learning and Generalizability», *Radiol. Artif. Intell.*, vol. 4, n.º 5, p. e220126, sep. 2022, doi: 10.1148/ryai.220126.
- [150] B. Sistaninejhad, H. Rasi, y P. Nayeri, «A Review Paper about Deep Learning for Medical Image Analysis», *Comput. Math. Methods Med.*, vol. 2023, pp. 1-10, may 2023, doi: 10.1155/2023/7091301.
- [151] H. Wang *et al.*, «Whole-tumor radiomics analysis of T2-weighted imaging in differentiating neuroblastoma from ganglioneuroblastoma/ganglioneuroma in children: an exploratory study», *Abdom. Radiol.*, vol. 48, n.º 4, pp. 1372-1382, abr. 2023, doi: 10.1007/s00261-023-03862-9.
- [152] Qing Xu, QingQiang Zhu ¹, Hao Liu ², LuFan Chang ², ShaoFeng Duan ³, WeiQiang Dou ³, SaiYang Li, Jing Ye ¹, «Differentiating Benign from Malignant Renal Tumors Using T2- and Diffusion-Weighted Images: A Comparison of Deep Learning and Radiomics Models Versus Assessment from Radiologists», *J Magn Reson Imaging*, vol. 55, n.º (4), pp. 1251-1259, 2022.
- [153] Z. Zhou, «Artificial intelligence on MRI for molecular subtyping of diffuse gliomas: feature comparison, visualization, and correlation between radiomics and deep

learning», *Eur. Radiol.*, vol. 32, n.º 2, pp. 745-746, feb. 2022, doi: 10.1007/s00330-021-08400-z.

[154] M. Kinoshita *et al.*, «Introduction of High Throughput Magnetic Resonance T2-Weighted Image Texture Analysis for WHO Grade 2 and 3 Gliomas», *PLOS ONE*, vol. 11, n.º 10, p. e0164268, oct. 2016, doi: 10.1371/journal.pone.0164268.

[155] «Radiomic Features - Pyradiomics». [En línea]. Disponible en: <https://pyradiomics.readthedocs.io/en/latest/features.html>

8. ANEXOS

ANEXOS

Anexo 1: Aprobaciones de los Comités Éticos de Investigación Clínica (CEIC) en las instituciones sanitarias que han participado en el desarrollo del estudio.

Anexo 2: Lista de comprobación de la guía CLAIM (Checklist for Artificial Intelligence in Medical Imaging). CLAIM sigue el modelo de la directriz STARD (Standards for Reporting Diagnostic accuracy studies) y se ha ampliado para abordar aplicaciones de IA en imágenes médicas que incluyen clasificación, reconstrucción de imágenes, análisis de texto y optimización del flujo de trabajo. Lista de comprobación de la escala MAIC-10, (Must Artificial Intelligence Criteria-10), guía para diseñar estudios y evaluar publicaciones relacionadas con la IA en el campo de las imágenes médicas.

Anexo 3: Artículos científicos derivados de la Tesis Doctoral hasta la fecha (no se incluyen las 9 comunicaciones a congresos, 3 ponencias ni el premio a mejor charla en el PhD Day del IIS La Fe 2021 por el trabajo Estudio comparativo multicéntrico de variabilidad interobservador en segmentación manual y automática de tumores neuroblásticos en imágenes de Resonancia Magnética).

- I. Veiga-Canuto D, Cerdà Alberich L, Fernández-Patón M, Jiménez Pastor A, Lozano-Montoya J, Miguel Blanco A, Martínez de Las Heras B, Sangüesa Nebot C; PRIMAGE Project consortium; Martí-Bonmatí L. Imaging biomarkers and radiomics in pediatric oncology: a view from the PRIMAGE (PRedictive In silico Multiscale Analytics to support cancer personalized diaGnosis and prognosis, Empowered by imaging biomarkers) project. *Pediatr Radiol.* 2023 Sep 25. doi: 10.1007/s00247-023-05770-y. PMID: 37747582.
- II. Veiga-Canuto D, Cerdà-Alberich L, Jiménez-Pastor A, Carot Sierra JM, Gomis-Maya A, Sangüesa-Nebot C, Fernández-Patón M, Martínez de Las Heras B, Taschner-Mandl S, Düster V, Pötschger U, Simon T, Neri E, Alberich-Bayarri Á, Cañete A, Hero B, Ladenstein R, Martí-Bonmatí L. Independent Validation of a Deep Learning nnU-Net Tool for Neuroblastoma Detection and Segmentation in MR Images. *Cancers (Basel).* 2023 Mar 6;15(5):1622. doi: 10.3390/cancers15051622. PMID: 36900410; PMCID: PMC10000775.
- III. Veiga-Canuto D, Cerdà-Alberich L, Sangüesa Nebot C, Martínez de Las Heras B, Pötschger U, Gabelloni M, Carot Sierra JM, Taschner-Mandl S, Düster V, Cañete

- A, Ladenstein R, Neri E, Martí-Bonmatí L. Comparative Multicentric Evaluation of Inter-Observer Variability in Manual and Automatic Segmentation of Neuroblastic Tumors in Magnetic Resonance Images. *Cancers (Basel)*. 2022 Jul 27;14(15):3648. doi: 10.3390/cancers14153648. PMID: 35954314; PMCID: PMC9367307.
- IV. Fernández Patón, M., Cerdá Alberich, L., Sangüesa Nebot, C., Martínez de Las Heras, B., Veiga Canuto, D., Cañete Nieto, A., & Martí-Bonmatí, L. (2021). MR Denoising Increases Radiomic Biomarker Precision and Reproducibility in Oncologic Imaging. *Journal of digital imaging*, 34(5), 1134–1145.
- V. Baeza-Delgado, C., Cerdá Alberich, L., Carot-Sierra, J. M., Veiga-Canuto, D., Martínez de Las Heras, B., Raza, B., & Martí-Bonmatí, L. (2022). A practical solution to estimate the sample size required for clinical prediction models generated from observational research on data. *European radiology experimental*, 6(1), 22.
- VI. Cerdá Alberich L, Sangüesa Nebot C, Alberich-Bayarri A, Carot Sierra JM, Martínez de Las Heras B, Veiga Canuto D, Cañete A, Martí-Bonmatí L. A Confidence Habitats Methodology in MR Quantitative Diffusion for the Classification of Neuroblastic Tumors. *Cancers (Basel)*. 2020 Dec 21;12(12):3858. doi: 10.3390/cancers12123858. PMID: 33371218; PMCID: PMC7767170.

Anexo 4: Proyectos de investigación concedidos relacionados con la Tesis Doctoral.

- I. Financiación
- II. Participación como Investigadora en el Proyecto PRIMAGE (PRedictive In silico Multiscale Analytics to support cancer personalized diaGnosis and prognosis, Empowered by imaging biomarker)
- III. Contrato de investigación para post-residentes del Instituto de Investigación Sanitaria La Fe como investigadora realizando el proyecto de esta Tesis Doctoral titulado Estudio de la agresividad biológica de tumores de estirpe neuroblástica en edad pediátrica mediante el análisis de las imágenes de Resonancia Magnética

ANEXO 1

DICTAMEN DEL COMITÉ DE ÉTICA DE LA INVESTIGACIÓN CON MEDICAMENTOS

MARÍA TORDERA BAVIERA, titular de la Secretaría Técnica del Comité de Ética de la Investigación con medicamentos del **CEIM - HOSPITAL UNIVERSITARIO Y POLITÉCNICO LA FE,**

CERTIFICA

Que este Comité ha evaluado en su sesión de fecha 22/09/2021, el Proyecto de Tesis:

Título: "**Estudio de la agresividad biológica de tumores de estirpe neuroblástica en edad pediátrica mediante el análisis de las imágenes de Resonancia Magnética.**"
Nº de registro: 2021-666-1

Documento	Versión - Fecha
Protocolo - TESIS	23 de agosto de 2021

Que dicho proyecto se ajusta a las normativas éticas sobre investigación biomédica con sujetos humanos y es viable en cuanto al planteamiento científico, objetivos, material y métodos, etc, descritos en la solicitud, así como la Hoja de Información al Paciente y el Consentimiento Informado.

En consecuencia, este Comité acuerda emitir **INFORME FAVORABLE** de dicho Proyecto de Tesis que será realizado en el HOSPITAL UNIVERSITARIO Y POLITÉCNICO LA FE, siendo el Director de la Tesis y Tutor del Hospital el **Dr. LUIS MARTÍ BONMATÍ,** del ÁREA DE IMAGEN MÉDICA, y el DOCTORANDO **Dña. DIANA VEIGA CANUTO.**

Que el CEIM - HOSPITAL UNIVERSITARIO Y POLITÉCNICO LA FE, tanto en su composición como en sus procedimientos, cumple con las normas de BPC (CPMP/ICH/135/95) y con la legislación vigente que regula su funcionamiento, y que la composición del CEIM - HOSPITAL UNIVERSITARIO Y POLITÉCNICO LA FE es la indicada en el anexo I, teniendo en cuenta que, en el caso de que algún miembro participe en el estudio o declare algún conflicto de interés, no habrá participado en la evaluación ni en el dictamen de la solicitud de autorización del estudio clínico.

Lo que firmo en Valencia, a 22/09/2021

Firmat per Maria Tordera Baviera el
27/09/2021 13:16:52



Fdo.: MARÍA TORDERA BAVIERA
Secretario/a Técnica del Comité de Ética de la Investigación con medicamentos

ANEXO I COMPOSICIÓN CEIm

Presidente:

ADELA CAÑETE NIETO - Facultativo Especialista en Pediatría. Jefe Sección de Unidad de Hematología y Oncología Pediátrica

Vicepresidente:

SALVADOR F ALIÑO PELLICER - Catedrático de Farmacología. Facultativo Especialista en Farmacología Clínica

Secretario:

MARÍA TORDERA BAVIERA - Farmacéutica Especialista en Farmacia Hospitalaria. Farmacéutica adjunta del Servicio de Farmacia

Vocales:

MARÍA VICTORIA PARICIO GÓMEZ - Diplomada Enfermería. Supervisora del Servicio de Hematología y Trasplante de Progenitores Hemáticos

JAVIER LLUNA GONZÁLEZ - Facultativo Especialista en Cirugía Pediátrica. Médico adjunto del Servicio de Cirugía Pediátrica

PAULA RAMÍREZ GALLEYMORE - Facultativo Especialista en Medicina Intensiva. Médico adjunto del Servicio de Medicina Intensiva

JOSÉ MARÍA CANELLES GAMIR - Farmacéutico de Atención Primaria del Departamento de Salud Valencia La Fe

SERAFÍN RODRÍGUEZ CAPELLÁN - Licenciado en Derecho. Técnico de Función Administrativa adscrito a la Dirección de Investigación

VICENTE INGLADA ALCAIDE - Miembro Lego, representante de los intereses de los pacientes

MIGUEL ÁNGEL CANO TORRES - Licenciado en Derecho. Técnico de Función Administrativa adscrito a la Dirección de Investigación

LUIS VICENTE MARTÍNEZ DOLZ - Facultativo Especialista en Cardiología. Jefe de Servicio de Cardiología

BONAVENTURA CASANOVA ESTRUCH - Facultativo Especialista en Neurología. Médico adjunto del Servicio de Neurología

SARA BRUGGER FRIGOLS - Facultativo Especialista en Radiodiagnóstico. Médico adjunto del Área de Imagen Médica

M^a ISABEL IZQUIERDO MACIÁN - Facultativo Especialista en Pediatría (Neonatología). Jefa de Servicio de Neonatología

MATTEO FRASSON - Facultativo Especialista en Cirugía. Médico adjunto del Servicio de Cirugía General y Digestiva

JOSÉ VICENTE SOLANAS PRATS - Facultativo Especialista en Medicina de Familia y Comunitaria. Médico del Centro de Salud Trinitat

ANTONIO ORDUÑA GALÁN - Ing. de aplicaciones y sistemas. Responsable Área de Seguridad y Calidad de Sistemas de inforamcion

OSCAR DÍAZ CAMBRONERO - Facultativo especialista en Anestesiología y Reanimación. Jefe Sección Anestesiología y Reanimación

LUIS MARTÍ BONMATÍ
AREA DE IMAGEN MEDICA

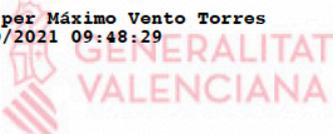
Valencia, 30 de septiembre de 2021.

Asunto: Autorización Inicio de estudio.

Adjunto le remito copia de los Informes Científico y Ético de Investigación, en el que se acuerda informar **favorablemente** sobre el Proyecto de Tesis titulado "*Estudio de la agresividad biológica de tumores de estirpe neuroblástica en edad pediátrica mediante el análisis de las imágenes de Resonancia Magnética.*", por usted presentado.

A la vista de los dictámenes emitidos, dicho Proyecto, puede iniciarse y llevarse a cabo.

Atentamente,

Firmat per **Máximo Vento Torres**
1'01/10/2021 09:48:29


Dr. Máximo Vento Torres
Director Científico

Valencia a 30 de septiembre de 2021

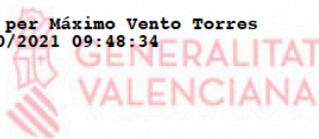
El **Dr. Máximo Vento Torres**, Presidente de la Comisión de Investigación del Instituto de Investigación Sanitaria La Fe,

INFORMA:

Que el Proyecto de tesis titulado: "*Estudio de la agresividad biológica de tumores de estirpe neuroblástica en edad pediátrica mediante el análisis de las imágenes de Resonancia Magnética.*" que será realizado en el HOSPITAL UNIVERSITARIO Y POLITÉCNICO LA FE, siendo el Director de la Tesis y Tutor del Hospital el **Dr. LUIS MARTÍ BONMATÍ**, del ÁREA DE IMAGEN MÉDICA, y el DOCTORANDO **Dña. DIANA VEIGA CANUTO**, contiene elementos objetivos suficientes en cuanto a la Hipótesis, Planteamientos y Plan de Trabajo que, a juicio de esta Comisión, permiten pronunciarse positivamente en cuanto a su viabilidad y aceptación para su realización.

ç

Firmat per Máximo Vento Torres
11/01/10/2021 09:48:34



Máximo Vento Torres
Presidente de la Comisión de Investigación

Votum:

EK Nr: 1637/2019

Projekttitle: PRIMAGE: Prädiktive In-Silico-Multiskalen-Analytik zur Unterstützung der personalisierten Diagnose und Prognose von Krebs mithilfe bildgebender Biomarker

Antragsteller/in: Frau Univ. Prof. Dr. Ruth Ladenstein

Institution: St. Anna Kinderkrebsforschung

Sponsor: HULAFE Fundación para la Investigación del Hospital La Fe de la Comunidad Valenciana

Teilnehmende Prüfzentren:

Ethik-Kommission	Prüfzentrum	Prüfärztin/arzt
Ethikkommission der Medizinischen Universität Wien	St. Anna Kinderkrebsforschung	Frau Univ. Prof. Dr. Ruth Ladenstein

Die Stellungnahme der Ethik-Kommission erfolgt aufgrund folgender eingereichter Unterlagen:

Conflict of Interest

Name	Version	Datum
Conflict_of_Interest_Antragsteller	V1.0	29.05.2019

Covering Letter

Name	Version	Datum
Cover Letter PRIMAGE	V1.0	29.05.2019
Cover Letter ergänzende Erklärungen	1.0	21.08.2019

Lebenslauf (CV)

Name	Version	Datum
CV WIEN Ruth Ladenstein_2019	1.0	12.03.2019



Sonstige

Name	Version	Datum
PRIMAGE_Ethics Evaluation	V1.0	20.07.2018
PRIMAGE_Consortium Agreement_draft_12_03_2019	V1.0	12.03.2019
PRIMAGE_Ethics Committee Dictum_HULAFE	V1.0	27.03.2019
WP11_PRIMAGE_D11.1	V1.0	31.05.2019
WP11_PRIMAGE_D11.3	V1.0	31.05.2019
WP11_PRIMAGE_D11.4	V1.0	31.05.2019
Schreier et al. 2016. EUPID. Health Informatics Meets e Health. Volume 223	V1.0	04.06.2019

Studienprotokoll (Prüfplan)

Name	Version	Datum
EU Projekt Teil 1 PRIMAGE_24-04-2018_final	V1.0	24.04.2018
EU Projekt Teil 2 PRIMAGE_24_04_2018_final	V1.0	24.04.2018
Engl RL PRIMAGE_Project description for Ethics Committee approval	V1.0	03.06.2019
Rollenbeschreibung CCRI PRIMAGE_310719_V1	V1	31.07.2019

Die Kommission fasst folgenden Beschluss (mit X markiert):

<input checked="" type="checkbox"/>	<p>Es besteht kein Einwand gegen die Durchführung der Studie.</p> <p>ACHTUNG: Unter Berücksichtigung der "ICH-Guideline for Good Clinical Practice" gilt dieser Beschluss ein Jahr ab Datum der Ausstellung. Gegebenenfalls hat der Antragsteller eine Verlängerung der Gültigkeit rechtzeitig zu beantragen.</p>
-------------------------------------	---

Ergänzende Kommentare der Sitzung am 02.07.2019:

Da die eingeladene Antragstellerin nicht an der Sitzung teilnehmen kann, wird der Tagesordnungspunkt unbehandelt vertagt.
Die Prüffärztin wird für die Sitzung am 30.07.2019 erneut eingeladen.

Ergänzende Kommentare der Sitzung am 30.07.2019:

Der Ethik-Kommission liegen derzeit die umfangreichen Unterlagen des Gesamtprojekts vor, aus denen nur schwer herauszulesen ist, wie genau die Beteiligung des lokalen Studienzentrums geplant ist. Aus diesem Grund ersucht die Ethik-Kommission um Vorlage eines Dokuments (Protokolls), welches die Beteiligung der St. Anna Kinderkrebsforschung klar und eindeutig darstellt.



Zur Versicherung:

Die Ethik-Kommission hält fest, dass der Abschluss einer Versicherung für diese Studie nicht erforderlich ist.

Andere: Frau Univ.Prof.Dr. Ruth Ladenstein stellt die Studie persönlich vor.

Die Ethik-Kommission ersucht die Antragsteller, bei der Wiedervorlage von geänderten Unterlagen ein Exemplar mit hervorgehobenen Änderungen beizulegen.

Ergänzende Kommentare:

Nachtrag vom 29. August 2019:

Die Antragsteller legen am 21.08.2019 überarbeitete Unterlagen vor, die von der Ethik-Kommission akzeptiert werden.

Die aktuelle Mitgliederliste der Ethik-Kommission ist unter folgender Adresse abrufbar:

<http://ethikkommission.meduniwien.ac.at/ethik-kommission/mitglieder/>

Mitglieder der Ethik-Kommission, die für diesen Tagesordnungspunkt als befangen anzusehen waren und daher laut Geschäftsordnung an der Entscheidungsfindung/Abstimmung nicht teilgenommen haben: **keine**

Dieses Dokument ist für berechnigte Benutzer/innen in digitaler Form unter folgender Adresse abrufbar:

<https://ekmeduniwien.at/vote/17571/download/>



Comitato Etico Regionale per la Sperimentazione Clinica della Regione Toscana
Sezione: COMITATO ETICO PEDIATRICO
Segreteria Tecnica Scientifica ubicata c/o: Meyer - Viale Pieraccini, 28 - 50139 Firenze
Telefono: 055-5662386 E-mail: comitato.etico@meyer.it

Firenze, il... 30/09/2019
Pag. 2 con la presente

Al promotore Ospedale Universitario Politecnico di La Fe

Allo sperimentatore principale Prof. Emanuele Neri
Sezione Dipartimentale Radiodiagnostica 3 - AOUPisana

Oggetto: Comunicazione del parere relativo alla richiesta di approvazione alla conduzione dello studio clinico dal titolo *PRedictive In-silico Multiscale Analytics to support cancer personalized diagnosis and prognosis, Empowered by imaging biomarkers business place (PRIMAGE)*
Codice Protocollo: PRIMAGE

In riferimento alla richiesta di cui all'oggetto, si trasmette il parere del Comitato Etico Regionale per la Sperimentazione Clinica della Toscana - sezione COMITATO ETICO PEDIATRICO riunitosi in data 24/09/2019.

Si ricorda che l'avvio della sperimentazione da parte del Promotore è subordinato al rilascio della disposizione autorizzativa della Direzione Generale dell'Azienda sanitaria.
Il Comitato si riserva la facoltà di monitorare nel corso del suo svolgimento, in accordo alle disposizioni normative vigenti, lo studio clinico approvato.

Il Responsabile Segreteria Tecnica Scientifica
Sezione Autorizzativa del Comitato Etico Pediatrico
per la Sperimentazione Clinica
c/o Azienda Ospedaliero Universitaria Meyer
Viale Pieraccini, 24 - 50139 FIRENZE

Il Comitato Etico in osservanza alle legislazioni vigenti in materia di
studi osservazionali
ha esaminato la richiesta di parere relativa allo studio in oggetto

Avendo valutato la seguente documentazione:

- ✓ Sintesi del protocollo in lingua italiana (versione 1 del 22/07/2019)
- ✓ Dichiarazione sulla natura osservazionale dello studio (versione -- del 01/08/2019)
- ✓ Protocollo di studio (versione 1 del 22/07/2019)
- ✓ Parere Unico del Centro Coordinatore (se applicabile) (versione -- del 27/03/2019)
- ✓ Elenco dei Centri partecipanti (se multicentrico) (versione 1 del 22/07/2019)
- ✓ Scheda di raccolta dati (versione 1 del 22/07/2019)
- ✓ Certificato marchio CE del dispositivo medico (versione -- del 11/01/2019)
- ✓ Scheda tecnica del dispositivo medico
- ✓ Scheda tecnico legale del dispositivo medico

- ✓ Manuale dispositivo
- ✓ Dichiarazione per l'accertamento della natura indipendente dello studio (versione -- del 01/08/2019)
- ✓ Lettera di intenti del promotore per il CE (versione - del 01/08/2019)
- ✓ Lettera di accettazione dello sperimentatore locale (versione - del 01/08/2019)
- ✓ Analisi d'impatto aziendale per la fattibilità locale (versione - del 01/08/2019)
- ✓ Curriculum vitae in formato UE aggiornato dello sperimentatore locale (versione - del 05/08/2019)
- ✓ Dichiarazione pubblica sul conflitto di interessi (versione - del 01/08/2019)
- ✓ Dichiarazione natura indipendente e no profit dello studio (versione - del 01/08/2019)
- ✓ Dichiarazione natura osservazionale (versione - del 01/08/2019)
- ✓ Medical Devide (DM) per indagini cliniche in valutazione B7 (versione - del 01/08/2019)
- ✓ Analisi impatto aziendale firmata dal DS (versione - del 06/09/2019)
- ✓ Foglio Informativo e consenso per genitori/tutore legale (versione - del 04/07/2019)
- ✓ Foglio Informativo e assenso per pazienti 14-18 anni (versione - del 22/07/2019)
- ✓ Foglio Informativo e assenso per pazienti 7-13 anni (versione - del 22/07/2019)

**Ha espresso:
PARERE FAVOREVOLE**

Numero registro pareri del Comitato Etico: 173/2019

Elenco componenti del CE presenti alla discussione e votanti che hanno dichiarato assenza di conflitti di interessi di tipo diretto o indiretto:

Prof. Alessandro Mugelli, *Farmacologo*
 Dr.ssa Maria Bimbi, *Medico Medicina Generale*
 Dr. Antonio Ciccarone, *Fisico area radiologica*
 Dr.ssa Paola Cipriani, *Clinico neuropsichiatra infantile*
 Dr.ssa Maria Grazia Conti, *Pediatra di famiglia*
 Dr. Emanuele Crocetti, *Biostatistico*
 Dr.ssa Lorena Di Simone, *Farmacista del SSR*
 Prof. Giovanni Marelli, *Medico legale*
 Dr. Eugenio Paci, *Esperto clinico settore procedure diagnostiche e terapeutiche*
 Dr.ssa Silvia Paoli, *Rappr. area professioni sanitarie interessate alla sperimentazione*
 Prof.ssa Monica Toraldo di Francia, *Esperto di bioetica*

Elenco componenti del CE presenti non votanti:
 Sussistenza numero legale (n. 11 su 20)

Si ricorda che è obbligo notificare al Comitato Etico: data di arruolamento del primo paziente e stato di avanzamento dello studio, con cadenza semestrale e/o annuale, corredato da una relazione scritta; fine del periodo di arruolamento; data di conclusione dello studio a livello locale ed a livello globale e risultati dello studio, entro un anno dalla conclusione della stessa. Inoltre si ricorda di riferire immediatamente al Comitato: deviazioni dal protocollo, modifiche al protocollo.

Comitato Etico Pediatrico
 SEZIONE AUTONOMA DEL COMITATO ETICO REGIONALE
 PER LA SPERIMENTAZIONE CLINICA
 c/o AZIENDA OSPEDALIERO UNIVERSITARIA Meyer
 Viale Pieraccini, 24 - 50139 FIRENZE

Il Presidente

ANEXO 2

CLAIM: Checklist for Artificial Intelligence in Medical Imaging

Section / Topic	No	Item	
TITLE / ABSTRACT			
	1	Identification as a study of Artificial Intelligence methodology, specifying the category of technology used (e.g., deep learning)	SÍ
	2	Structured summary of study design, methods, results, and conclusions	SÍ
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the Artificial Intelligence approach	SÍ
	4	Study objectives and hypotheses	SÍ
METHODS			
<i>Study Design</i>	5	Prospective or retrospective study	SÍ
	6	Study goal, such as model creation, exploratory study, feasibility study, non-inferiority trial	SÍ
<i>Data</i>	7	Data sources	SÍ
	8	Eligibility criteria: how, where, and when potentially eligible participants or studies were identified (e.g., symptoms, results from previous tests, inclusion in registry, patient-care setting, location, dates)	SÍ
	9	Data pre-processing steps	SÍ
	10	Selection of data subsets, if applicable	SÍ
	11	Definitions of data elements, with references to Common Data Elements	SÍ
	12	De-identification methods	SÍ
	13	How missing data were handled	
<i>Ground Truth</i>	14	Definition of ground truth reference standard, in sufficient detail to allow replication	SÍ
	15	Rationale for choosing the reference standard (if alternatives exist)	SÍ

	16	Source of ground-truth annotations; qualifications and preparation of annotators	SÍ
	17	Annotation tools	SÍ
	18	Measurement of inter- and intrarater variability; methods to mitigate variability and/or resolve discrepancies	SÍ
<i>Data Partitions</i>	19	Intended sample size and how it was determined	SÍ
	20	How data were assigned to partitions; specify proportions	SÍ
	21	Level at which partitions are disjoint (e.g., image, study, patient, institution)	SÍ
<i>Model</i>	22	Detailed description of model, including inputs, outputs, all intermediate layers and connections	SÍ
	23	Software libraries, frameworks, and packages	SÍ
	24	Initialization of model parameters (e.g., randomization, transfer learning)	SÍ
<i>Training</i>	25	Details of training approach, including data augmentation, hyperparameters, number of models trained	SÍ
	26	Method of selecting the final model	SÍ
	27	Ensembling techniques, if applicable	
<i>Evaluation</i>	28	Metrics of model performance	SÍ
	29	Statistical measures of significance and uncertainty (e.g., confidence intervals)	SÍ
	30	Robustness or sensitivity analysis	SÍ
	31	Methods for explainability or interpretability (e.g., saliency maps), and how they were validated	SÍ
	32	Validation or testing on external data	SÍ
RESULTS			
<i>Data</i>	33	Flow of participants or cases, using a diagram to indicate inclusion and exclusion	SÍ
	34	Demographic and clinical characteristics of cases in each partition	SÍ
<i>Model performance</i>	35	Performance metrics for optimal model(s) on all data partitions	SÍ

	36	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	SÍ
	37	Failure analysis of incorrectly classified cases	SÍ
DISCUSSION			
	38	Study limitations, including potential bias, statistical uncertainty, and generalizability	SÍ
	39	Implications for practice, including the intended use and/or clinical role	SÍ
OTHER INFORMATION			
	40	Registration number and name of registry	SÍ
	41	Where the full study protocol can be accessed	SÍ
	42	Sources of funding and other support; role of funders	SÍ

Mongan J, Moy L, Kahn CE Jr. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): a guide for authors and reviewers. *Radiol Artif Intell* 2020; 2(2):e200029. <https://doi.org/10.1148/ryai.2020200029>

MAIC-10: Must Artificial Intelligence Criteria-10

Table 1 MAIC-10 checklist to assess quality of AI-based medical imaging research studies

Checklist item	Article section	Description	Reported
1. Clinical need	Introduction	The study is clearly put into context by describing the target clinical problem and any previous approaches in the literature	<input checked="" type="checkbox"/>
2. Study design	Materials and methods	The type of study (observational/interventional, single/multicentre) and inclusion/exclusion criteria are explicitly described, and a sample size estimate is given	<input checked="" type="checkbox"/>
3. Safety and privacy	Materials and methods	ELSI (ethical, legal, social implications), specifically including ethics committee approval and data de-identification issues, are discussed	<input checked="" type="checkbox"/>
4. Data curation	Materials and methods	Data extraction, cleaning, and transformation methods, including image pre-processing steps, are clearly described	<input checked="" type="checkbox"/>
5. Data annotation	Materials and methods	The ground truth reference is defined and the annotation process, including measures of inter/intra-observer variability, is described	<input checked="" type="checkbox"/>
6. Data partitioning	Materials and methods	Methods and criteria for data set splitting into train-tune-test-validation sets are indicated	<input checked="" type="checkbox"/>
7. AI model	Materials and methods, results	The AI model building methodology is sufficiently detailed by including used technologies (software and hardware), training-tuning-testing methods, performance metrics, and resulting AI model architecture	<input checked="" type="checkbox"/>
8. Robustness	Results, discussion	The generalizability of the AI model in real-world conditions is explicitly discussed	<input checked="" type="checkbox"/>
9. Explainability	Discussion	The interpretability of the model (including the use of uncertainty or confidence metrics) is explicitly discussed	<input checked="" type="checkbox"/>
10. Transparency	Discussion	Any possibility of access to original data sets and source code is clearly stated. Financing and conflicts of interest are detailed	<input checked="" type="checkbox"/>

Score:10

ANEXO 3

Article

Comparative Multicentric Evaluation of Inter-Observer Variability in Manual and Automatic Segmentation of Neuroblastic Tumors in Magnetic Resonance Images

Diana Veiga-Canuto ^{1,2,*}, Leonor Cerdà-Alberich ¹, Cinta Sangüesa Nebot ², Blanca Martínez de las Heras ³, Ulrike Pötschger ⁴, Michela Gabelloni ⁵, José Miguel Carot Sierra ⁶, Sabine Taschner-Mandl ⁴, Vanessa Düster ⁴, Adela Cañete ³, Ruth Ladenstein ⁴, Emanuele Neri ⁵ and Luis Martí-Bonmatí ^{1,2}



Citation: Veiga-Canuto, D.; Cerdà-Alberich, L.; Sangüesa Nebot, C.; Martínez de las Heras, B.; Pötschger, U.; Gabelloni, M.; Carot Sierra, J.M.; Taschner-Mandl, S.; Düster, V.; Cañete, A.; et al. Comparative Multicentric Evaluation of Inter-Observer Variability in Manual and Automatic Segmentation of Neuroblastic Tumors in Magnetic Resonance Images. *Cancers* **2022**, *14*, 3648. <https://doi.org/10.3390/cancers14153648>

Academic Editors: Hamid Khayyam, Ali Madani, Rahele Kafieh and Ali Hekmatnia

Received: 25 May 2022

Accepted: 26 July 2022

Published: 27 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

- ¹ Grupo de Investigación Biomédica en Imagen, Instituto de Investigación Sanitaria La Fe, Avenida Fernando Abril Martorell, 106 Torre A 7planta, 46026 Valencia, Spain; leonor_cerda@iislafe.es (L.C.-A.); marti_lui@gva.es (L.M.-B.)
 - ² Área Clínica de Imagen Médica, Hospital Universitario y Politécnico La Fe, Avenida Fernando Abril Martorell, 106 Torre A 7planta, 46026 Valencia, Spain; sanguesa_cin@gva.es
 - ³ Unidad de Oncohematología Pediátrica, Hospital Universitario y Politécnico La Fe, Avenida Fernando Abril Martorell, 106 Torre A 7planta, 46026 Valencia, Spain; blanca_martinez@iislafe.es (B.M.d.l.H.); canyete_ade@gva.es (A.C.)
 - ⁴ St. Anna Children's Cancer Research Institute, Zimmermannplatz 10, 1090 Vienna, Austria; ulrike.poetschger@ccri.at (U.P.); sabine.taschner@ccri.at (S.T.-M.); vanessa.duester@ccri.at (V.D.); ruth.ladenstein@ccri.at (R.L.)
 - ⁵ Academic Radiology, Department of Translational Research, University of Pisa, Via Roma, 67, 56126 Pisa, Italy; michela.gabelloni@med.unipi.it (M.G.); emanueleneri1@gmail.com (E.N.)
 - ⁶ Departamento de Estadística e Investigación Operativa Aplicadas y Calidad, Universitat Politècnica de València, Camí de Vera s/n, 46022 Valencia, Spain; jcarot@eio.upv.es
- * Correspondence: veiga_dia@gva.es

Simple Summary: Tumor segmentation is a key step in oncologic imaging processing and is a time-consuming process usually performed manually by radiologists. To facilitate it, there is growing interest in applying deep-learning segmentation algorithms. Thus, we explore the variability between two observers performing manual segmentation and use the state-of-the-art deep learning architecture nnU-Net to develop a model to detect and segment neuroblastic tumors on MR images. We were able to show that the variability between nnU-Net and manual segmentation is similar to the inter-observer variability in manual segmentation. Furthermore, we compared the time needed to manually segment the tumors from scratch with the time required for the automatic model to segment the same cases, with posterior human validation with manual adjustment when needed.

Abstract: Tumor segmentation is one of the key steps in imaging processing. The goals of this study were to assess the inter-observer variability in manual segmentation of neuroblastic tumors and to analyze whether the state-of-the-art deep learning architecture nnU-Net can provide a robust solution to detect and segment tumors on MR images. A retrospective multicenter study of 132 patients with neuroblastic tumors was performed. Dice Similarity Coefficient (DSC) and Area Under the Receiver Operating Characteristic Curve (AUC ROC) were used to compare segmentation sets. Two more metrics were elaborated to understand the direction of the errors: the modified version of False Positive (FPRm) and False Negative (FNR) rates. Two radiologists manually segmented 46 tumors and a comparative study was performed. nnU-Net was trained-tuned with 106 cases divided into five balanced folds to perform cross-validation. The five resulting models were used as an ensemble solution to measure training (n = 106) and validation (n = 26) performance, independently. The time needed by the model to automatically segment 20 cases was compared to the time required for manual segmentation. The median DSC for manual segmentation sets was 0.969 (±0.032 IQR). The median DSC for the automatic tool was 0.965 (±0.018 IQR). The automatic segmentation model achieved a better performance regarding the FPRm. MR images segmentation variability is similar between radiologists and nnU-Net. Time leverage when using the automatic model with posterior visual validation and manual adjustment corresponds to 92.8%.

Article

Independent Validation of a Deep Learning nnU-Net Tool for Neuroblastoma Detection and Segmentation in MR Images

Diana Veiga-Canuto ^{1,2,*} , Leonor Cerdà-Alberich ¹ , Ana Jiménez-Pastor ³ , José Miguel Carot Sierra ⁴ , Armando Gomis-Maya ¹ , Cinta Sangüesa-Nebot ², Matías Fernández-Patón ¹ , Blanca Martínez de las Heras ⁵, Sabine Taschner-Mandl ⁶ , Vanessa Düster ⁶, Ulrike Pötschger ⁶, Thorsten Simon ⁷ , Emanuele Neri ⁸ , Ángel Alberich-Bayarri ³, Adela Cañete ⁵, Barbara Hero ⁷, Ruth Ladenstein ⁶ and Luis Martí-Bonmati ^{1,2,*} 

¹ Grupo de Investigación Biomédica en Imagen, Instituto de Investigación Sanitaria La Fe, Avenida Fernando Abril Martorell, 106 Torre A 7planta, 46026 Valencia, Spain

² Área Clínica de Imagen Médica, Hospital Universitario y Politécnico La Fe, Avenida Fernando Abril Martorell, 106 Torre A 7planta, 46026 Valencia, Spain

³ Quantitative Imaging Biomarkers in Medicine, QUIBIM SL, 46026 Valencia, Spain

⁴ Departamento de Estadística e Investigación Operativa Aplicadas y Calidad, Universitat Politècnica de València, Camí de Vera s/n, 46022 Valencia, Spain

⁵ Unidad de Oncohematología Pediátrica, Hospital Universitario y Politécnico La Fe, Avenida Fernando Abril Martorell, 106 Torre A 7planta, 46026 Valencia, Spain

⁶ St. Anna Children's Cancer Research Institute, Zimmermannplatz 10, 1090 Vienna, Austria

⁷ Department of Pediatric Oncology and Hematology, University Children's Hospital of Cologne, Medical Faculty, University of Cologne, 50937 Cologne, Germany

⁸ Academic Radiology, Department of Translational Research, University of Pisa, Via Roma, 67, 56126 Pisa, Italy

* Correspondence: veiga_dia@gva.es (D.V.-C.); marti_lui@gva.es (L.M.-B.)



Citation: Veiga-Canuto, D.; Cerdà-Alberich, L.; Jiménez-Pastor, A.; Carot Sierra, J.M.; Gomis-Maya, A.; Sangüesa-Nebot, C.; Fernández-Patón, M.; Martínez de las Heras, B.; Taschner-Mandl, S.; Düster, V.; et al. Independent Validation of a Deep Learning nnU-Net Tool for Neuroblastoma Detection and Segmentation in MR Images. *Cancers* **2023**, *15*, 1622. <https://doi.org/10.3390/cancers15051622>

Academic Editor: Luis Montuenga

Received: 20 January 2023

Revised: 22 February 2023

Accepted: 5 March 2023

Published: 6 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Simple Summary: Tumor segmentation is a key step in oncologic imaging processing. We have recently developed a model to detect and segment neuroblastic tumors on MR images based on deep learning architecture nnU-Net. In this work, we performed an independent validation of the automatic segmentation tool with a large heterogeneous dataset. We reviewed the automatic segmentations and manually edited them when necessary. We were able to show that the automatic network was able to locate and segment the primary tumor on the T2 weighted images in the majority of cases, with an extremely high agreement between the automatic tool and the manually edited masks. The time needed for manual adjustment was very low.

Abstract: Objectives. To externally validate and assess the accuracy of a previously trained fully automatic nnU-Net CNN algorithm to identify and segment primary neuroblastoma tumors in MR images in a large children cohort. Methods. An international multicenter, multivendor imaging repository of patients with neuroblastic tumors was used to validate the performance of a trained Machine Learning (ML) tool to identify and delineate primary neuroblastoma tumors. The dataset was heterogeneous and completely independent from the one used to train and tune the model, consisting of 300 children with neuroblastic tumors having 535 MR T2-weighted sequences (486 sequences at diagnosis and 49 after finalization of the first phase of chemotherapy). The automatic segmentation algorithm was based on a nnU-Net architecture developed within the PRIMAGE project. For comparison, the segmentation masks were manually edited by an expert radiologist, and the time for the manual editing was recorded. Different overlaps and spatial metrics were calculated to compare both masks. Results. The median Dice Similarity Coefficient (DSC) was high 0.997; 0.944–1.000 (median; Q1–Q3). In 18 MR sequences (6%), the net was not able neither to identify nor segment the tumor. No differences were found regarding the MR magnetic field, type of T2 sequence, or tumor location. No significant differences in the performance of the net were found in patients with an MR performed after chemotherapy. The time for visual inspection of the generated masks was 7.9 ± 7.5 (mean \pm Standard Deviation (SD)) seconds. Those cases where manual editing was needed (136 masks) required 124 ± 120 s. Conclusions. The automatic CNN was able to locate and segment the primary tumor on the T2-weighted images in 94% of cases. There was an extremely high agreement between the automatic tool and the manually edited masks. This is the first study to validate an automatic segmentation model for neuroblastic



Imaging biomarkers and radiomics in pediatric oncology: a view from the PRIMAGE (PRedictive In silico Multiscale Analytics to support cancer personalized diaGnosis and prognosis, Empowered by imaging biomarkers) project

Diana Veiga-Canuto^{1,2} · Leonor Cerdá Alberich¹ · Matías Fernández-Patón¹ · Ana Jiménez Pastor³ · Jose Lozano-Montoya³ · Ana Miguel Blanco¹ · Blanca Martínez de las Heras⁴ · Cinta Sangüesa Nebot² · PRIMAGE Project consortium · Luis Martí-Bonmati^{1,2}

Received: 1 June 2023 / Revised: 1 September 2023 / Accepted: 3 September 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

This review paper presents the practical development of imaging biomarkers in the scope of the PRIMAGE (PRedictive In silico Multiscale Analytics to support cancer personalized diaGnosis and prognosis, Empowered by imaging biomarkers) project, as a noninvasive and reliable way to improve the diagnosis and prognosis in pediatric oncology. The PRIMAGE project is a European multi-center research initiative that focuses on developing medical imaging-derived artificial intelligence (AI) solutions designed to enhance overall management and decision-making for two types of pediatric cancer: neuroblastoma and diffuse intrinsic pontine glioma. To allow this, the PRIMAGE project has created an open-cloud platform that combines imaging, clinical, and molecular data together with AI models developed from this data, creating a comprehensive decision support environment for clinicians managing patients with these two cancers. In order to achieve this, a standardized data processing and analysis workflow was implemented to generate robust and reliable predictions for different clinical endpoints. Magnetic resonance (MR) image harmonization and registration was performed as part of the workflow. Subsequently, an automated tool for the detection and segmentation of tumors was trained and internally validated. The Dice similarity coefficient obtained for the independent validation dataset was 0.997, indicating compatibility with the manual segmentation variability. Following this, radiomics and deep features were extracted and correlated with clinical endpoints. Finally, reproducible and relevant imaging quantitative features were integrated with clinical and molecular data to enrich both the predictive models and a set of visual analytics tools, making the PRIMAGE platform a complete clinical decision aid system. In order to ensure the advancement of research in this field and to foster engagement with the wider research community, the PRIMAGE data repository and platform are currently being integrated into the European Federation for Cancer Images (EUCAIM), which is the largest European cancer imaging research infrastructure created to date.

✉ Diana Veiga-Canuto
dianaveigac@gmail.com

¹ Grupo de Investigación Biomédica en Imagen, Instituto de Investigación Sanitaria La Fe, Avenida Fernando Abril Martorell, 106 Torre A planta 7, 46026 Valencia, Spain

² Área Clínica de Imagen Médica, Área Clínica de Imagen Médica, Hospital Universitari i Politècnic La Fe, Avinguda Fernando Abril Martorell, 106 Torre E planta 0, 46026 Valencia, Spain

³ Quantitative Imaging Biomarkers in Medicine, QUIBIM SL, Valencia, Spain

⁴ Pediatric Oncology Department, Hospital Universitario y Politécnico La Fe, Avenida Fernando Abril Martorell, 106 Torre G planta 2, 46026 Valencia, Spain

Published online: 25 September 2023

Springer



MR Denoising Increases Radiomic Biomarker Precision and Reproducibility in Oncologic Imaging

Matías Fernández Patón¹ · Leonor Cerdá Alberich¹ · Cinta Sangüesa Nebot² · Blanca Martínez de las Heras³ · Diana Veiga Canuto² · Adela Cañete Nieto³ · Luis Martí-Bonmatí^{1,2}

Received: 5 November 2020 / Revised: 24 July 2021 / Accepted: 17 August 2021 / Published online: 10 September 2021
© Society for Imaging Informatics in Medicine 2021

Abstract

Several noise sources, such as the Johnson–Nyquist noise, affect MR images disturbing the visualization of structures and affecting the subsequent extraction of radiomic data. We evaluate the performance of 5 denoising filters (anisotropic diffusion filter (ADF), curvature flow filter (CFF), Gaussian filter (GF), non-local means filter (NLMF), and unbiased non-local means (UNLMF)), with 33 different settings, in T2-weighted MR images of phantoms ($N = 112$) and neuroblastoma patients ($N = 25$). Filters were discarded until the most optimal solutions were obtained according to 3 image quality metrics: peak signal-to-noise ratio (PSNR), edge-strength similarity-based image quality metric (ESSIM), and noise (standard deviation of the signal intensity of a region in the background area). The selected filters were ADFs and UNLMs. From them, 107 radiomics features preservation at 4 progressively added noise levels were studied. The ADF with a conductance of 1 and 2 iterations standardized the radiomic features, improving reproducibility and quality metrics.

Keywords Denoising · Image processing · Radiomics · Oncologic imaging biomarkers

Introduction

Magnetic resonance (MR)-reconstructed images from different scanners, series, and patients have different noise levels that reduce the quality of the images, affect the signal-to-noise ratio (SNR), and decrease reproducibility of imaging biomarkers. The noise reduces the visibility in low-contrast structures and blurs the edges, having an impact on both the qualitative radiological reporting and the quantitative measurements extracted from either feature extraction or dynamic signal fitting [1].

Reducing noise and increasing SNR and final image quality are common goals in precision imaging. At the acquisition level, MR parameters are adjusted to improve the SNR. Unfortunately, this usually results in increased acquisition time and in sensitivity to motion artifacts. New MR developments reconstructing the k -space with artificial intelligence methods have been proposed as an efficient way to acquire faster and almost noise-free images. In daily practice, MR images need to improve their quality by removing noise applying computer vision techniques known as denoising [2]. Denoising filters face the challenges of maximizing noise removal, smoothing homogenous regions, and preserving morphological detail [3].

The Johnson–Nyquist noise is one of the common source of noise, being induced by thermal fluctuations and modelled as a Rician distribution in the magnitude images [4, 5]. This noise can be approximated as Gaussian spreading in areas with high SNR. Noise is responsible of the low reproducibility of imaging biomarkers in the phenotyping, treatment prediction, and patient's prognostication in oncology, as image quality variability affects the related radiomics features and dynamic parameters.

We hypothesize that the use of MR denoising filters will increase the precision and reproducibility of extracted

✉ Matías Fernández Patón
matias_fernandez@iislafe.es

¹ Grupo de Investigación Biomédica en Imagen, Instituto de Investigación Sanitaria La Fe, Avenida Fernando Abril Martorell, 106 Torre A 7planta, 46026 Valencia, Spain

² Área Clínica de Imagen Médica, Hospital Universitario Y Politécnico La Fe, Avenida Fernando Abril Martorell, 106, 46026 Valencia, Spain

³ Unidad de Oncohematología Pediátrica, Hospital Universitario Y Politécnico La Fe, Avenida Fernando Abril Martorell, 106, 46026 Valencia, Spain

ORIGINAL ARTICLE

Open Access

A practical solution to estimate the sample size required for clinical prediction models generated from observational research on data



Carlos Baeza-Delgado¹, Leonor Cerdá Alberich¹, José Miguel Carot-Sierra², Diana Veiga-Canuto³, Blanca Martínez de las Heras⁴, Ben Raza^{1,4} and Luis Martí-Bonmatí^{1*} 

Abstract

Background: Estimating the required sample size is crucial when developing and validating clinical prediction models. However, there is no consensus about how to determine the sample size in such a setting. Here, the goal was to compare available methods to define a practical solution to sample size estimation for clinical predictive models, as applied to Horizon 2020 PRIMAGE as a case study.

Methods: Three different methods (Riley's; "rule of thumb" with 10 and 5 events per predictor) were employed to calculate the sample size required to develop predictive models to analyse the variation in sample size as a function of different parameters. Subsequently, the sample size for model validation was also estimated.

Results: To develop reliable predictive models, 1397 neuroblastoma patients are required, 1060 high-risk neuroblastoma patients and 1345 diffuse intrinsic pontine glioma (DIPG) patients. This sample size can be lowered by reducing the number of variables included in the model, by including direct measures of the outcome to be predicted and/or by increasing the follow-up period. For model validation, the estimated sample size resulted to be 326 patients for neuroblastoma, 246 for high-risk neuroblastoma, and 592 for DIPG.

Conclusions: Given the variability of the different sample sizes obtained, we recommend using methods based on epidemiological data and the nature of the results, as the results are tailored to the specific clinical problem. In addition, sample size can be reduced by lowering the number of parameter predictors, by including direct measures of the outcome of interest.

Keywords: Sample size calculation, Clinical predictive models, PRIMAGE, Paediatric oncology, Radiology

Key points

- Estimating the appropriate sample size in clinical prediction model development is mandatory to guarantee the robustness of the results.

- The selected method is designed to be applied to epidemiological data and based on the nature of outcomes.
- Strategies based on the selection and reduction of predictor variables are proposed to reduce sample size.
- The expected recruitment in PRIMAGE project fits the estimated sample size.

* Correspondence: marti_lui@gva.es

¹Biomedical Imaging Research Group (GBI230-PREBI) at La Fe Health Research Institute and the Imaging La Fe node of the Distributed Network for Biomedical Imaging (ReDIB) Unique Scientific and Technical Infrastructures (ICTS), Valencia, Spain
Full list of author information is available at the end of the article



© The Author(s) under exclusive licence to European Society of Radiology. 2022 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Article

A Confidence Habitats Methodology in MR Quantitative Diffusion for the Classification of Neuroblastic Tumors

Leonor Cerdá Alberich ^{1,*}, Cinta Sangüesa Nebot ², Angel Alberich-Bayarri ³, José Miguel Carot Sierra ⁴, Blanca Martínez de las Heras ⁵, Diana Veiga Canuto ², Adela Cañete ⁵ and Luis Martí-Bonmatí ^{1,2}

¹ Grupo de Investigación Biomédica en Imagen, Instituto de Investigación Sanitaria La Fe, Avenida Fernando Abril Martorell, 106 Torre A 7planta, 46026 Valencia, Spain; marti_lui@gva.es

² Área Clínica de Imagen Médica, Hospital Universitario y Politécnico La Fe, Avenida Fernando Abril Martorell, 106 Torre A 7planta, 46026 Valencia, Spain; sanguesa_cin@gva.es (C.S.N.); veiga_dia@gva.es (D.V.C.)

³ Quantitative Imaging Biomarkers in Medicine, QUIBIM SL. Edificio Europa, Av. d'Aragó, 30, Planta 12, 46021 Valencia, Spain; angel@quibim.com

⁴ Departamento de Estadística e Investigación Operativa Aplicadas y Calidad, Universitat Politècnica de València, Camí de Vera s/n, 46022 Valencia, Spain; jcarot@eio.upv.es

⁵ Unidad de Oncohematología Pediátrica, Hospital Universitario y Politécnico La Fe, Avenida Fernando Abril Martorell, 106 Torre A 7planta, 46026 Valencia, Spain; martinez_bladel@gva.es (B.M.d.l.H.); canyete_ade@gva.es (A.C.)

* Correspondence: leonor_cerda@iislafe.es; Tel.: +34-615224988

Received: 16 November 2020; Accepted: 18 December 2020; Published: 21 December 2020



Simple Summary: There is growing interest in applying quantitative diffusion techniques to magnetic resonance imaging for cancer diagnosis and treatment. These measurements are used as a surrogate marker of tumor cellularity and aggressiveness, although there may be factors that introduce some bias to these approaches. Thus, we explored a novel methodology based on confidence habitats and voxel uncertainty to improve the power of the apparent diffusion coefficient to discriminate between benign and malignant neuroblastic tumor profiles in children. We were able to show this offered an improved sensitivity and negative predictive value relative to standard voxel-based methodologies.

Abstract: Background/Aim: In recent years, the apparent diffusion coefficient (ADC) has been used in many oncology applications as a surrogate marker of tumor cellularity and aggressiveness, although several factors may introduce bias when calculating this coefficient. The goal of this study was to develop a novel methodology (Fit-Cluster-Fit) based on confidence habitats that could be applied to quantitative diffusion-weighted magnetic resonance images (DWIs) to enhance the power of ADC values to discriminate between benign and malignant neuroblastic tumor profiles in children. Methods: Histogram analysis and clustering-based algorithms were applied to DWIs from 33 patients to perform tumor voxel discrimination into two classes. Voxel uncertainties were quantified and incorporated to obtain a more reproducible and meaningful estimate of ADC values within a tumor habitat. Computational experiments were performed by smearing the ADC values in order to obtain confidence maps that help identify and remove noise from low-quality voxels within high-signal clustered regions. The proposed Fit-Cluster-Fit methodology was compared with two other methods: conventional voxel-based and a cluster-based strategy. Results: The cluster-based and Fit-Cluster-Fit models successfully differentiated benign and malignant neuroblastic tumor profiles when using values from the lower ADC habitat. In particular, the best sensitivity (91%) and specificity (89%) of all the combinations and methods explored was achieved by removing uncertainties at a 70% confidence threshold, improving standard voxel-based sensitivity and negative predictive values by 4% and

ANEXO 4

La presente Tesis Doctoral contó con la financiación del Instituto de Investigación Sanitaria La Fe (València, España), y del proyecto PRIMAGE (PRedictive In-silico Multiscale Analytics to support cancer personalized diaGnosis and prognosis, Empowered by imaging biomarkers), financiado por la Comisión Europea, *Grant Agreement 826494, Horizon 2020 SC1-DTH-2018-2020*.



Instituto de Investigación
Sanitaria La Fe

Dirección Científico, Máximo Vento Torres

D. Máximo Vento Torres, en calidad de Director Científico del Instituto de Investigación Sanitaria la Fe,

CERTIFICA:

Que **D./D^a. VEIGA CANUTO, DIANA**, con DNI 53761358P, figura como investigador/a en el proyecto de investigación gestionado por este Instituto de Investigación:

Ref: 2018/0228, P.I. Exp. 826494

Título del proyecto: PRIMAGE. PREDICTIVE IN-SILICO MULTISCALE ANALYTICS TO SUPPORT CANCER PERSONALIZED DIAGNOSIS AND PROGNOSIS, EMPOWERED BY IMAGING BIOMARKERS.

Entidad financiadora: COMISION EUROPEA

Investigador principal: MARTÍ BONMATÍ, LUIS

Fecha Inicio: 01/12/2018 - Fecha Fin: 30/11/2022

Cantidad concedida: 1.297.500,00 €

Y para que así conste, a petición del interesado, se expide el presente certificado en Valencia a 16 de abril de 2021.

Firmat per Máximo Vento Torres el
20/04/2021 11:58:12



Máximo Vento Torres
Director Científico

1

CSV:6L6GKPTH.1HX4YMSU:RXKQIA3Y URL de validació: <https://www.tramita.gva.es/csv-front/index.faces?cadena=6L6GKPTH.1HX4YMSU:RXKQIA3Y>



Contrato comunicado al Servicio Público de Empleo Estatal
FECHA: 08/10/2021 NUMERO: E4620210764842
Nr.de autorización administrativa para utilizar CONTRAT@: 1020032541

CONTRATO DE TRABAJO TEMPORAL

DATOS DE LA EMPRESA

Form fields for company data: CIF/NIF/NIE G97067557, D./Dña. GALDON TORNERO, SONIA, NIF/NIE 52708684K, EN CONCEPTO (1) DIRECTORA GENERAL, etc.

DATOS DE LA CUENTA DE COTIZACIÓN

Form fields for social security data: REGIMEN 0111, COD. PROV. 46, NUMERO 1184476, DIG. CONTR. 13, ACTIVIDAD ECONÓMICA Investigación y desarrollo, 72

DATOS DEL CENTRO DE TRABAJO

Form fields for center of work: PAIS ESPAÑA, MUNICIPIO VALENCIA, 724, C. POSTAL 46250

DATOS DEL/DE LA TRABAJADOR/A

Form fields for worker data: D./Dña. VEIGA CANUTO, DIANA, NIF/NIE (2) 53761358P, FECHA DE NACIMIENTO 30/04/1992, NACIONALIDAD ESPAÑA, 724, MUNICIPIO DEL DOMICILIO TORRENT, 46244, PAIS DOMICILIO ESPAÑA, 724

con la asistencia legal, en su caso, de D./Dña. en calidad de (2)

DECLARAN

Que reúnen los requisitos exigidos para la celebración del presente contrato y, en su consecuencia, acuerdan formalizarlo con arreglo a las siguientes:

CLÁUSULAS

PRIMERA: El/la trabajador/a prestará sus servicios como (3) INVESTIGADORA, incluido en el grupo profesional de GRUPO II INYES NO DOCTORA para la realización de las funciones (4) de acuerdo con el sistema de clasificación profesional vigente en la empresa.

En el centro de trabajo ubicado en (calle, nº y localidad) FERNANDO ABRIL MARTORELL 106 - VALENCIA 46026

TRABAJO A DISTANCIA (5).

SEGUNDA: La jornada de trabajo será:(6)

A tiempo completo: la jornada de trabajo será de 40 horas semanales, prestadas de LUNES a VIERNES, con los descansos establecidos legal o convencionalmente(7).

A tiempo parcial: la jornada de trabajo ordinaria será de ... horas al día, a la semana, al mes, al año(6), siendo esta jornada inferior a la de un trabajador a tiempo completo comparable (8)

La distribución del tiempo de trabajo será de (9) con- forme a lo previsto en el convenio Colectivo FUNDACION PARA LA INVESTIGACION DEL HOSPITAL UNIV Y POLITECNICO LA FE

En el caso de la jornada a tiempo parcial, existe pacto sobre la realización de horas complementarias SI NO Se establece un período de prueba de (10) 6 MESES hasta 03/10/2023 Cuando el convenio colectivo permita una duración mayor a la establecida legalmente, señálelo con una X:

CUARTA: El/la trabajador/a percibirá una retribución total de 2416,67 euros brutos(8) MENSUALES que se distribuyen en los siguientes conceptos salariales (11) POR TODOS LOS CONCEPTOS SALARIALES



CLAÚSULAS ESPECÍFICAS DE INVESTIGADORES

TIEMPO COMPLETO

CÓDIGO DE CONTRATO

4 0 1

PARA LA REALIZACIÓN DE UN PROYECTO ESPECÍFICO DE INVESTIGACIÓN CIENTÍFICA Y TÉCNICA

4 0 4

PERSONAL INVESTIGADOR PREDOCTORAL EN FORMACIÓN (R.D. 103/2019) (1)

4 2 0

DE ACCESO AL SISTEMA ESPAÑOL DE CIENCIA, TECNOLOGÍA E INNOVACIÓN. (1)

TIEMPO PARCIAL

CÓDIGO DE CONTRATO

5 0 1

PARA LA REALIZACIÓN DE UN PROYECTO ESPECÍFICO DE INVESTIGACIÓN CIENTÍFICA Y TÉCNICA

5 2 0

DE ACCESO AL SISTEMA ESPAÑOL DE CIENCIA, TECNOLOGÍA E INNOVACIÓN. (1) Y (2).

Que el/la empleador/a es (3) :

- Organismo Público de investigación de la Administración General del Estado.
- Organismo de Investigación de otra Administración Pública.
- Universidad Pública, perceptora de fondos cuyo destino incluya la contratación de personal investigador o para el desarrollo de los programas propios I+D+I.
- Universidades privadas y Universidades de la Iglesia Católica, cuando perciban fondos cuyo destino incluya la contratación de personal investigador.
- Entidades privadas sin ánimo de lucro que realicen actividades I+D tecnológico en los términos de la D.A. 1ª de la Ley 14/2011.
- Consorcios públicos y fundaciones del sector público en los términos de la D.A. 1ª de la Ley 14/2011.
- Otros organismos de investigación de la A.G. cuando realicen actividad investigadora y sean beneficiarios de ayudas y subvención es que incluyan personal investigador.
- Organismo de la A.G. del Estado de los contemplados en la D.A. 14ª de la Ley 14/2011 de 1 de junio.
- Otros

Indíquese la opción elegida :

- A Que el/la trabajador/a para la realización de un proyecto específico de investigación científica y técnica
- Personal investigador
 - Personal científico o técnico
- B Que el/la trabajador/a para ser personal investigador predoctoral en formación esta en posesión de :
Título de Licenciado, Arquitecto, Graduado Universitario de al menos 300 créditos o máster universitario o equivalente y hayan sido admitidos a un programa de doctorado (5).
- Que el/la trabajador/a que accede al sistema español de Ciencia, ecnología e Innovación:
- C Está en posesión del título de Doctor o equivalente que le capacitan para la practica profesional de este contrato (4).....

Que no ha estado contratado/a bajo esta modalidad en este u otro Organismo por tiempo superior a cinco años.

(1) Predoctoral, y de acceso al Sistema Español de Ciencia, Tecnología e Innovación no se les aplica la D.T. 8ª del E. de los Trabajadores.
 (2) Se aplicará lo establecido en el art. 11.1 del Estatuto de los Trabajadores, aprobado por Real Decreto Legislativo 2/2015 de 23 de octubre, (BOE de 24 de octubre).(Contrato en prácticas)
 (3) Indicar la entidad contratante.
 (4) Indicar la disciplina que correspond
 (5) El/la trabajador/a deberá entregar al empresario fotocopia compulsada del título, certificación de su solicitud o certificación acreditativa de la terminación de los estudios.
 (6) Deberá acompañar el escrito de admisión al programa de doctorado expedido por la unidad responsable de dicho programa o por la escuela de doctorado.