

Resumen

Análisis de características Radiómicas en imágenes de F-18-FDG PET y Resonancia Magnética Multiparamétrica como predictores de la respuesta al tratamiento de Quimioterapia Neoadyuvante en Cáncer de mama.

El propósito de este estudio fue evaluar si las características radiómicas derivadas de imágenes de F-18 FDG PET y los mapas de ADC obtenidos por resonancia magnética (RM) son capaces de predecir la respuesta patológica completa (pCR) a la quimioterapia neoadyuvante (QNA) en pacientes con cáncer de mama localmente avanzado. 30 pacientes sin tratamiento previo sometidas a PET con F18-FDG y RM fueron incluidas en este estudio. Los especímenes de la biopsia pretratamiento fueron usados para extraer las características inmunohistoquímicas del tumor. Los resultados histopatológicos del tumor resecado post QNA fueron usados para clasificar entre pCR y No-pCR. 1702 características radiómicas y 4 por biopsia fueron extraídas. La reproducibilidad ante la variabilidad intra- e inter-observador fue evaluada usando el coeficiente de correlación intraclass ICC>0.9. La reducción de la dimensionalidad fue realizada por selección de mínima redundancia y máxima relevancia. Tres métodos: Boruta, Wilcoxon y un modelo *Random Forest* fueron implementados para seleccionar las características más importantes. Un modelo de aprendizaje supervisado por *Random Forest* con validación cruzada por *Leave-One-Out* fue implementado. El mayor rendimiento fue obtenido por la característica de PET “wavelet-HHL_glcm_lmc1” (AUC=0.78; IC95%: 0.70-0.87; sensibilidad: 50%; especificidad: 91%) que también fue la de mayor robustez entre los modelos de selección. El método de muestreo sintético adaptativo (ADASYN) fue implementado para balancear las clases y el modelo combinado entre la expresión de receptores de estrógenos (RE)

obtenida por biopsia y la característica de PET “wavelet-HHL_glcm_Imc1” obtuvo el mejor rendimiento (AUC=0.95; sensibilidad=90%; especificidad=86%). El análisis radiómico combinado tiene el potencial para predecir la pCR a la QNA en pacientes con Ca de mama localmente avanzado y mejorar la estratificación preterapéutica. En este trabajo se ha desarrollado un flujo de trabajo radiómico robusto y reproducible para su implementación.

Palabras clave: Tomografía por emisión de positrones (PET), Resonancia Magnética (RM), radiomics, cáncer de mama, predicción de la respuesta patológica completa (pCR) basada en radiomics.

Abstract

18F FDG PET and Multiparametric Magnetic Resonance Imaging Based Radiomics for Prediction of Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer

Background: The aim of this study was to assess whether ¹⁸F FDG and multiparametric MRI-based radiomics analysis is able to predict pathological complete response (pCR) to neoadjuvant chemotherapy (NAC) in breast cancer patients.

Methods: A total of 30 female patients with locally advanced breast cancer proven by biopsy were included in this prospective study. All PET and MRI datasets were imported to dedicated software 3DSlicer (v. 4.11) for lesion annotation using a semi-automated method. Pretreatment biopsy specimens were used to determine tumour histology, tumor and nuclear grades, and immunohistochemical status. Histopathological results from surgical tumour specimens were used as the reference standard to distinguish between pCR and non-pCR. 1702 radiomics features were extracted using pyradiomics and 4 features from biopsy. The dimensionality reduction was achieved by minimum redundancy-maximum relevance. Three methods: Boruta, Wilcoxon and machine learning Random Forest were used to select the most important features. A supervised machine learning model with Random Forest and cross validated by Leave-One-Out was implemented. **Results:** The best performance was obtained by the PET feature “wavelet-HHL_glcM_Imc1” (AUC=0.78; IC95%: 0.70-0.87; sensibility=50%, specificity=91%) which was also the most robust among the selection models. The data were subsequently balanced using the method adaptive synthetic sampling (ADASYN) and better values in all metrics were obtained

(AUC=0.95). The adaptive synthetic sampling (ADASYN) method to balance the classes was implemented and the best performance was obtained by the combined model between the expression of estrogen receptors (ER) and the PET feature “wavelet-HHL_glcm_Imc1” (AUC=0.95; sensibility=90%; specificity=86%). **Conclusion:** multiparametric radiomic features demonstrated ability as imaging biomarkers to predict the pCR to NAC in locally advanced breast cancer patients and hence potentially enhance pretherapeutic patient stratification.

Keywords: Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI), Radiomics, Breast Cancer, radiomics-based prediction of pathologic complete response.

Contenido

	Pág.
Dedicatoria	V
Declaración de obra original	VII
Agradecimientos	VIII
Resumen	IX
Contenido	XIII
Lista de figuras	XVI
Lista de tablas	XX
Lista de Abreviaturas y Símbolos	XXI
Introducción	1
1. Fundamentación Teórica	5
1.1 Bases físicas de la Tomografía por Emisión de Positrones (PET)	5
1.1.1 Bases físicas y principio de funcionamiento.....	5
1.1.2 Correcciones cuantitativas	10
1.1.3 Reconstrucción de la imagen en PET.....	12
1.1.4 Cuantificación de las imágenes	13
1.1.5 Valor de captación estandarizado (SUV).....	15
1.1.6 Otras métricas volumétricas	16
1.1.7 F-18 FDG	17
1.1.8 Conclusiones	17
1.2 Imágenes por Resonancia Magnética (MRI).....	18
1.2.1 Bases físicas de la MRI.....	18
1.2.2 Formación de la imagen.....	24
1.2.3 Secuencias de pulsos.....	28
1.2.4 Imágenes pesadas por difusión (DWI).....	29
1.3 Radiomics	31
1.3.1 Flujo de trabajo en Radiomics.....	31
1.3.2 Adquisición de la imagen	32

1.3.3	Segmentación del ROI/VOI.....	32
1.3.4	Extracción de características radiómicas	32
1.3.5	Selección de características.....	35
1.3.6	Construcción y validación del modelo	35
2.	Materiales y Métodos.....	37
2.1	Conjunto de datos.....	37
2.1.1	Diseño de estudio.....	37
2.1.2	Criterios de inclusión y exclusión de pacientes.....	37
2.1.3	Adquisición de las imágenes.....	38
2.1.4	Preprocesamiento de las imágenes.....	40
2.1.5	Marcadores biológicos	41
2.1.6	Variable principal.....	41
2.1.7	Base de datos	41
2.2	Segmentación de las regiones de interés	41
2.2.1	Segmentación de las imágenes de PET	41
2.2.2	Segmentación de las imágenes de Resonancia Magnética	45
2.3	Extracción de características	47
2.4	Métodos de selección de características.....	48
2.4.1	Análisis del conjunto de datos.....	48
2.4.2	Evaluación del valor predictivo individual de las características biológicas por biopsia	49
2.4.3	Análisis de las características metabólicas de PET y evaluación de su valor predictivo individual.....	49
2.4.4	Reducción de características de acuerdo con la reproducibilidad intra- e inter-observador	49
2.4.5	Reducción de características por redundancia y máxima relevancia.....	50
2.4.6	Selección de las características más importantes	50
2.5	Construcción del modelo.....	52
2.6	Balanceo de datos usando ADASYN.....	53
2.7	Resumen del flujo de trabajo en radiomics	53
3.	Resultados	55
3.1	Características clínicas y demográficas	55
3.2	Análisis de características biológicas y metabólicas.....	56
3.2.1	Evaluación del valor predictivo individual de los marcadores biológicos..	56
3.2.2	Reproducibilidad de las métricas metabólicas de PET frente al posicionamiento del paciente.....	57
3.2.3	Evaluación del valor predictivo individual de las características metabólicas	59
3.3	Preprocesamiento y extracción de características radiómicas	61
3.4	Robustez del conjunto de características.....	61
3.4.1	Reducción de características según la reproducibilidad intra- e inter-observador.....	61
3.4.2	Reducción de características según la redundancia y relevancia	62
3.4.3	Selección de características por relevancia	63

3.5	Rendimiento de las predicciones de la pCR.....	65
3.6	Rendimiento de modelos predictivos usando ADASYN.....	69
4.	Discusión	71
5.	Conclusiones	79
5.1	Conclusiones.....	79
5.2	Perspectivas futuras	79
	Bibliografía.....	81

Lista de figuras

	Pág.
Figura 1-1: Aniquilación del par positrón (e ⁺)-electrón(e ⁻). Producto de la aniquilación se emiten dos fotones (γ_1 , γ_2) de 511 keV en sentido opuesto (A), los cuales son detectados en múltiples anillos de detectores (B). <i>Figura del autor</i>	6
Figura 1-2: Sistema de detección en coincidencia. Cada evento de detección simple produce un pulso eléctrico que es amplificado y analizado. Para que se considere una coincidencia ambos eventos deben estar dentro de la ventana energética, la ventana temporal y el rango angular permitido. <i>Figura tomada de [9]</i>	7
Figura 1-3: Esquema de la construcción de un sinograma a partir de las proyecciones de los ángulos de muestreo. (Izq.) Representación de dos proyecciones (líneas naranjas y azules). (Der.) Formación del sinograma $S\phi, r, z$, resaltando las filas de la matriz correspondientes a las proyecciones de la imagen de la izquierda y la posición que ocupa la LOR individual roja. El nivel de gris representa la cantidad total de eventos en una misma línea de respuesta. <i>Figura tomada de [9] y modificada por el autor</i>	8
Figura 1-4: Tipos de eventos de coincidencia. El punto negro corresponde a la aniquilación y la línea punteada a la LOR asignada. De izquierda a derecha se presentan: coincidencia verdadera (a), coincidencia dispersada (b), coincidencia aleatoria (c) y evento de detección múltiple (d). <i>Figura tomada de [8]</i>	9
Figura 1-5: Esquema del diseño de un PET/CT. El CT está posicionado enfrente del escáner PET y los centros están alineados. La camilla común favorece la corrección de las imágenes. <i>Figura tomada de [11]</i>	10

- Figura 1-6: Función de escalado para convertir las HU a los valores de atenuación lineal a 511 keV. Los coeficientes del tejido (μ_{mix}) dependen de la mezcla que lo componen: agua-aire o agua (μ_w)-hueso (μ_{cb}). *Figura tomada de [10]*..... 11
- Figura 1-7: Imagen PET/CT. a. Imagen de máxima proyección de intensidad (MIP) de un PET/CT con F-18 FDG en paciente con lesiones en la mama izquierda. b, c, d Imagen de corte axial a nivel de tórax de CT, de PET y fusión de PET/CT. 14
- Figura 1-8: Representación de los fenómenos que causan el efecto de volumen parcial. a. Resolución espacial finita. *Tomada de [14]*. b. Efecto de fracción de tejido en una lesión heterogénea. *Tomada de [15]*..... 15
- Figura 1-9: Generación del campo magnético principal B_0 en un sistema de MRI típico. Campo magnético horizontal producido en magneto de núcleo de aire con arreglo de bobinas paralelas sobre un cilindro de alrededor de 1 m de diámetro y 2 a 3 m de largo. DSV en el centro del cilindro. *Figura del autor*. 20
- Figura 1-10: Representación tridimensional de la distribución de espines en el cuerpo y su interacción con los campos magnéticos. A. En ausencia del campo la distribución es isotrópica. B. En presencia de B_0 aumenta la densidad de espines en dirección del campo ocasionando una magnetización neta M_0 . C. Cuando un pulso de RF de campo magnético homogéneo es aplicado, toda la distribución de espines rota. Representación en el sistema de referencia rotante con f_0 . *Tomada de [21]*. 21
- Figura 1-11: Comportamiento de la magnetización antes y después de la aplicación del pulso de radiofrecuencia. *Con permiso de J. N. Oshinski [24]*. Visto en el sistema de referencia fijo del laboratorio. B. Visto en un sistema de referencia rotante a la frecuencia de Larmor. C. Señal de RF recibida en la bobina receptora mientras M vuelve al equilibrio..... 23
- Figura 1-12: Diagrama de flujo del proceso de formación de la imagen en RM en 2D. A. Ejemplo de codificación de frecuencia y fase (matriz de adquisición 256×256) en un corte axial de cerebro. Los relojes a la izquierda representan el cambio de fase en cada fila. B. Señal codificada medida en la bobina receptora. Los cuadros de colores representan el muestreo discreto de la señal. C. Llenado Cartesiano

del espacio-k, de izquierda a derecha de acuerdo con el muestreo en (B), una fila por cada repetición con diferente fuerza de GPE (desde +128 a -128, 256 pasos). D. Imagen del espacio k lleno. E. Imagen final obtenida aplicando el valor absoluto de la FT inversa sobre el espacio-k. <i>Figura del autor</i>	27
Figura 1-13: Diagrama de tiempo de pulsos para la secuencia Spin-Echo. <i>Tomada de [18]</i>	29
Figura 1-14: Ejemplo de cálculo de las características radiómicas texturales. En la 1era fila, la matriz de la imagen adquirida en escala de grises. En la 2da fila, la matriz con los valores discretizados de la intensidad. En la 3era fila, se presentan tres diferentes algoritmos de cálculo de grupo de características GLCM, GLRLM y GLSZM formadas a partir de la matriz discretizada. En amarillo y verde se representa las forma de llenado de cada matriz. <i>Tomada de [32]</i>	34
Figura 2-1: Ejemplo de foco tumoral captante sin correlación morfológica clara entre el PET y el CT debido a la alta densidad del parénquima mamario. A la izquierda la imagen fusionada de PET/CT y a la derecha la imagen de CT.	42
Figura 2-2: Adquisición del fantoma NEMA de calidad de la imagen para la determinación del factor β del método de segmentación implementado. a. Imagen del fantoma NEMA de torso con inserto de 6 esferas rellenas de 37, 28, 22, 17, 13 y 10 mm de diámetro. b. Imagen del corte axial del fantoma con la segmentación de las esferas usando el β determinado.	44
Figura 2-3: Imagen de volumen tumoral con centro necrótico y el impacto sobre la definición del umbral de segmentación en PET. a. De izquierda a derecha: imagen de PET, imagen de CT e imagen fusionada de PET/CT mostrando el contraste de una lesión con centro necrótico en el parénquima mamario. b y c. El valor de SUV del centro necrótico es menor que el valor umbral definido para la segmentación. La inclusión del área en el ROI impacta en algunas métricas.	45
Figura 2-4: Imágenes de las secuencia de resonancia magnética usadas como referencia para la segmentación. a. Imagen DCE con t=2 min post contraste. b.	

Imagen T2W. c. Imagen de sustracción. d. Imagen DWI con b1100 y e. Imagen del mapa de ADC con b1100.....	46
Figura 2-5: Imágenes de resonancia magnética con lesión en parénquima mamario derecho con clip de biopsia que genera artefacto metálico en torno a él. a. Imagen de DWI con b1100 y b. Imagen de DCE con t=2 min post inyección de contraste.....	47
Figura 2-6: Flujo de trabajo del análisis radiómico implementado.....	54
Figura 3-1: Resultados del análisis del SUVmax (Izquierda) y SUVmean (Derecha) respecto del subtipo del cáncer de mama.	56
Figura 3-2: Resultado del análisis del nivel de receptor de estrógeno respecto de la respuesta al tratamiento.	57
Figura 3-3: Comparación entre el valor de SUVmáx medido en posición supino y prono con iguales parámetros de reconstrucción (STD: estándar).....	58
Figura 3-4: Comparación entre el valor del volumen medido en posición supino y posición prono con iguales parámetros de reconstrucción (STD: estándar). ..	58
Figura 3-5: Valor predictivo individual del SUVmax y SUVmean respecto de la respuesta al tratamiento.	59
Figura 3-6: Valor predictivo individual del MTV y el TLG respecto de la respuesta al tratamiento.....	60
Figura 3-7: Valor predictivo del Δ SUVmax respecto de la respuesta al tratamiento. .	60
Figura 3-8: Mapa de calor de la correlación del conjunto de datos reproducible ante la variabilidad en la segmentación. En orden de ubicación de izquierda a derecha y arriba-abajo, la primera característica corresponde a RE, seguida de todas las características de PET y por último las características del mapa de ADC.	63
Figura 3-9. Curva ROC para el modelo clasificador Random Forest entrenado con la característica de mayor rendimiento, P_wavelet-HHL_glcm_Imc1.	67
Figura 3-10: Matriz de confusión para el modelo clasificador entrenado con la característica de mayor rendimiento, P_wavelet-HHL_glcm_Imc1.	68
Figura 3-11: Curva ROC de 10 modelos clasificadores entrenados con datos aumentados por ADASYN usando características individuales.....	69

Lista de tablas

	Pág.
Tabla 2-1: Parámetros de adquisición para el protocolo de imágenes de resonancia magnética de las mamas usado en el conjunto de datos.	40
Tabla 3-1: Comparación demográfica entre grupos con respecto a la pCR.....	55
Tabla 3-2: Número de características de alta, moderada y baja robustez en cada grupo en PET y Resonancia Magnética.	62
Tabla 3-3. Características seleccionadas por el test Wilcoxon.....	64
Tabla 3-4. Características seleccionadas por el modelo de aprendizaje automático. .	65
Tabla 3-5: Rendimiento de la predicción para los modelos entrenados con cada una de las características individuales extraídas por el método de Wilcoxon.	66
Tabla 3-6: Rendimiento de la predicción para los modelos entrenados con cada una de las características individuales extraídas por el modelo de aprendizaje automático.....	67
Tabla 3-7. Rendimiento de los 3 mejores modelos predictivos entrenados usando parejas de características.	68
Tabla 3-8. Rendimiento de los 3 mejores modelos predictivos entrenados usando datos aumentados por ADASYN.	70
Tabla 3-9. Rendimiento de los 3 mejores modelos predictivos combinados usando datos aumentados por ADASYN.	70

Lista de Abreviaturas y Símbolos

Abreviatura	Término
^1H	Hidrógeno 1
^{18}F	Flúor 18
AAPM	Asociación Americana de Físicos Médicos
ADC	Coeficiente de difusión aparente
AUC	Área bajo la curva ROC
Ca	Cáncer
CDI	Carcinoma ductal invasivo
CLI	Carcinoma lobulillar infiltrante
CT	Tomografía Computada
DCE	Dinámica con realce de contraste
DWI	Imagen pesada por Difusión (Diffusion-weighted imaging)
DICOM	Digital Imaging and Communications in Medicine
EANM	Asociación Europea de Medicina Nuclear
FDG	2-Fluoro-2-desoxi-D-glucosa (fluorodesoxyglucosa)
FID	Decaimiento libre de inducción (Free Induction Decay)
FOV	Campo de visión de la imagen (Field Of View)
FWHM	Ancho completo a media altura (Fullwidth-at-Half-Maximum)
GLDM	Matriz de dependencia de la escala de grises
GLCM	Matriz de co-ocurrencia de escala de grises
GLRLM	Matriz de secuencia de datos de escala de grises
GLSZM	Matriz de tamaño de zona de escala de grises
GRE	Gradiente-Echo
IBSI	Iniciativa de estandarización de biomarcadores de imágenes
IC	Intervalo de confianza
IR	Recuperación de la inversión (Inversion Recovery)
LoG	Filtro LaPlaciano del Gaussiano
LOR	Línea de respuesta (Line Of Response)
MLEM	Maximum-Likelihood Expectation-Maximization
MRI	Imagenología por Resonancia Magnética
MTV	Volumen metabólico tumoral (Metabolic Tumor Volume)

Abreviatura	Término
NEMA	Asociación Nacional de Fabricantes Eléctricos
NGTDM	Matriz de diferencia de tono de grises en la vengidad
NSA	Número de señales promediadas (Number of signals averaged)
OMS	Organización Mundial de la Salud
OSEM	Ordered-Subsets Expectation-Maximization
pCR	Respuesta patológica completa
PET	Tomografía por Emisión de Positrones
PET/CT	Tomografía por Emisión de Positrones integrada con Tomografía Computada helicoidal multidetector
PSF	Función de dispersion puntual (Point Spread Function)
QIBA	Quantitative Imaging Biomarkers Alliance
QNA	Quimioterapia neoadyuvante
RCB	Carga tumoral residual
RF	Radiofrecuencia
<i>RF</i>	<i>Random Forest</i>
ROC	Características de funcionamiento del receptor
ROI	Región de interés (Region Of Interest)
RM	Resonancia Magnética
RMN	Resonancia Magnética Nuclear
RSNA	Sociedad Norte Americana de Radiología
SE	Spin-Echo
SLR	Supervivencia libre de recurrencia
SNR	Relación señal-ruido
SPAIR	(SPectral Attenuated Inversion Recovery)
SUV	Valor de captación estándar (Standardized Uptake Value)
TE	Tiempo de Eco
TLG	Glicólisis total de la lesión (Total Lesión Glycolysis)
TR	Tiempo de Repetición
VPN	Valor predictivo negativo
VPP	Valor predictivo positivo
VOI	Volumen de interés

Símbolos con letras latinas

Símbolo	Término	Unidad SI	Definición
b	Factor de gradiente en DWI	s/mm ²	
\vec{B}_0	Campo magnético principal	T	Ec. (1-2)

Símbolo	Término	Unidad SI	Definición
\vec{B}_1	Campo magnético oscilante	T	
f_0	Frecuencia de Larmor	Hz	Ec. (1-3)
G	Gradiente de Campo Magnético	mT/m	
G_{FE}	Gradiente de codificación de frecuencias	mT/m	
G_{PE}	Gradiente de codificación de fase	mT/m	
G_{SS}	Gradiente de selección de corte	mT/m	
h	Constante de Planck	J·s	6.626×10^{-34}
i	Intensidad de corriente eléctrica	A	
I	Espín nuclear		
k_B	Constante de Boltzmann	J/K	1.381×10^{-23}
\vec{M}	Vector de magnetización		
M_0	Magnetización neta en equilibrio		Ec. (1-5)
T1	Tiempo de relajación longitudinal	s	
T2	Tiempo de relajación transversal	s	

Símbolos con letras griegas

Símbolo	Término	Unidad SI	Definición
α	Ángulo	rad	
γ	Constante giromagnética	Hz/Tesla	μ/I
θ	Ángulo	°	
$\vec{\mu}$	Momento magnético	J/T	
μ_0	Permeabilidad del vacío	T·m/A	$4\pi \times 10^{-7}$
μ_x	Coefficiente de atenuación lineal	cm ⁻¹	
ϕ	Ángulo acimutal	rad	
Φ	Fase acumulada de los espines	rad	
ρ	Coefficiente de correlación de Spearman		

Introducción

El cáncer de mama es el cáncer de mayor incidencia y mortalidad entre las mujeres en todo el mundo. En 2020, representó el 24.5% de todos los nuevos casos entre todos los tipos de cáncer y el 15.5% de las muertes por cáncer en mujeres de todas las edades. En Latinoamérica, según estimaciones de la Agencia Internacional de Investigación sobre Cáncer (IARC, por sus siglas en inglés) y el Observatorio Global de Cáncer de la Organización Mundial de la Salud (OMS), Argentina es el país con más nuevos casos reportados por año por número de habitantes, con más de 22.000 casos, que representan el 16.8% de todos los tumores malignos en ambos sexos y el 32.1% en mujeres de todas las edades. Mientras que, en Colombia, la incidencia es del 13.7% entre todos los tumores malignos en ambos sexos y del 25.7% en mujeres [1].

La quimioterapia neoadyuvante (QNA) es normalmente la primera línea de tratamiento del cáncer de mama localmente avanzado, ya que permite reducir el volumen tumoral y generalmente mejora la supervivencia libre de recurrencia (SLR) [2], sin embargo, solo entre el 10-50% de los pacientes sometidos a la QNA alcanzarán la respuesta patológica completa. Modelos predictivos de la respuesta a la QNA han sido establecidos, sin embargo, carecen de suficiente exactitud debido a que el volumen tumoral es heterogéneo espacial y temporalmente (cambia en el tiempo), por lo tanto, como no es factible realizar una biopsia a cada vóxel del tumor, tampoco es posible caracterizarlo por completo por esta vía [3], [4]. Radiomics es una técnica con el potencial de complementar los modelos clínicos actuales para predecir la respuesta al tratamiento, a partir de la extracción de características multidimensionales de las imágenes radiológicas para la estimación de la heterogeneidad intratumoral [5]. Recientes investigaciones presentan resultados alentadores de modelos predictivos

con radiomics en distintos tipos de cáncer; sin embargo, en algunos de ellos la reproducibilidad está limitada por factores técnicos y metodológicos que han evitado su amplio uso en el ámbito clínico [6].

El objetivo principal de este trabajo es analizar la capacidad de las características radiómicas, extraídas de imágenes de tomografía por emisión de positrones (PET) con el radiofármaco ^{18}F -FDG y de imágenes del mapa de coeficientes de difusión aparente (ADC) obtenidas por resonancia magnética multiparamétrica, para predecir la respuesta patológica completa (pCR) al tratamiento de quimioterapia neoadyuvante en pacientes de cáncer de mama. La predicción de la pCR es de especial utilidad en un enfoque de medicina personalizada, ya que permitiría clasificar mejor qué grupo de pacientes tiene una alta probabilidad de respuesta al tratamiento y cuáles deberían ser sometidas a otra alternativa terapéutica, evitando los efectos colaterales de la quimioterapia y la posible progresión de la enfermedad. Una limitante del estudio es la cantidad de imágenes disponibles para el análisis.

Este documento está estructurado en cinco capítulos. En el capítulo uno, se presenta una introducción general a los principios físicos y técnicos del PET y la resonancia magnética. Asimismo, se presentan los conceptos básicos en torno a la radiómica, el flujo de trabajo para su implementación y sus limitaciones. En el capítulo dos, se presenta el marco metodológico seguido para el desarrollo del problema, basado en dos objetivos específicos: (i) la definición e implementación de un flujo de análisis radiómico robusto y reproducible, desde el preprocesamiento de las imágenes hasta la selección de las características radiómicas; y (ii) el establecimiento del modelo de aprendizaje supervisado para la clasificación de pacientes de acuerdo con la respuesta al tratamiento, usando las características con más alto valor predictivo y reproducibilidad. Además, en este capítulo se incorporan al análisis algunos marcadores metabólicos extraídos de la imagen de PET y marcadores biológicos de la anatomía patológica, con valor predictivo individual según la literatura científica, con el fin de aumentar la capacidad predictiva global del modelo clasificador.

En el capítulo tres, se presentan los resultados de la implementación del flujo de trabajo radiómico y el modelo de clasificación. Mientras que, en el capítulo cuatro se presenta la discusión en torno al análisis de los resultados. Se evalúa el proceso de reducción de la dimensionalidad para la selección de características y la capacidad del clasificador para predecir la respuesta al tratamiento de quimioterapia neoadyuvante.

Por su parte, el capítulo cinco, está dedicado a la conclusión general, preguntas que aún están pendientes y las perspectivas de trabajo futuro. Finalmente, se incluyen un conjunto de apéndices de diversa índole.

1. Fundamentación Teórica

En este capítulo se presenta una introducción a los aspectos físicos y los principios de funcionamiento de las técnicas radiológicas: Tomografía por Emisión de Positrones (PET, del inglés *Positron Emission Tomography*) e Imagenología por Resonancia Magnética (MRI, del inglés *Magnetic Resonance Imaging*). Del mismo modo se presentan los conceptos básicos de la técnica radiómica que darán cuenta de las ventajas y limitaciones de su implementación en el ámbito clínico.

1.1 Bases físicas de la Tomografía por Emisión de Positrones (PET)

1.1.1 Bases físicas y principio de funcionamiento

La tomografía por emisión de positrones más conocida como PET, por sus siglas en inglés, es una modalidad de imagen diagnóstica de medicina nuclear que permite obtener imágenes tridimensionales de la distribución y acumulación en los tejidos de una sustancia radiactiva administrada al paciente, llamada radiofármaco, compuesta por un isótopo emisor de positrones.

Los isótopos deficientes de neutrones pueden decaer emitiendo un positrón (e^+) -la antipartícula del electrón, con igual masa, pero carga opuesta- y un neutrino. Los positrones son emitidos con una pequeña cantidad de energía cinética la cual pierden rápidamente (en unos pocos milímetros) por colisiones con el medio hasta casi frenarse. Cuando la energía es lo suficientemente pequeña se produce la aniquilación con un electrón libre del tejido que resulta en la emisión de dos fotones, cada uno de 511 keV (equivalente a la masa en reposo del positrón y el electrón, manteniendo el

principio de conservación de la energía) y que viajan en sentido opuesto, casi exactamente a 180° , a menos que exista momento lineal residual antes de la aniquilación (manteniendo el principio de conservación del momento lineal). Tales emisiones ocurren sin una dirección preferencial, es decir, isotrópicamente [7].

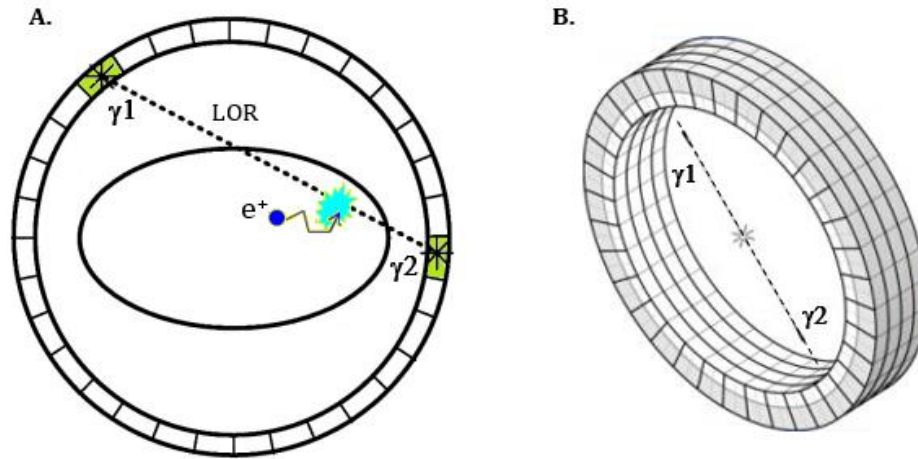


Figura 1-1: Aniquilación del par positrón (e^+)-electrón(e^-). Producto de la aniquilación se emiten dos fotones (γ_1 , γ_2) de 511 keV en sentido opuesto **A.** (*Figura del autor*), los cuales son detectados en múltiples anillos de detectores **B.** (*Figura tomada de y modificada por el autor.*)

Debido a lo anterior, el principio de funcionamiento del PET se basa en la detección en coincidencia de los pares de fotones emitidos por la aniquilación del par positrón-electrón usando una configuración cilíndrica de varios anillos de detectores que permiten determinar la línea en la que se produjo la aniquilación, como se muestra en la Figura 1-1. La línea que une los detectores donde se recibe la señal en coincidencia se denomina “línea de respuesta” (LOR, del inglés *Line Of Response*). Sin embargo, múltiples eventos de aniquilación ocurren simultáneamente, por lo tanto, es necesario reconocer cuáles de los fotones detectados pueden ser considerados coincidentes. Para ello, el equipo PET cuenta con un procesador de coincidencias que recibe la señal de cada bloque detector de forma individual y evalúa:

- (1) Que la amplitud del pulso eléctrico generado se encuentre dentro de un rango energético configurado en el equipo, centrado en 511 keV, denominado “ventana energética”.

- (2) Que otro evento de detección haya ocurrido antes o después de un intervalo de tiempo máximo (τ), igual al cociente de la división del diámetro del anillo entre la velocidad de la luz sumado a la resolución temporal del cristal centellador del detector, al cual se denomina “ventana de coincidencia”, de amplitud 2τ .
- (3) Que la línea de respuesta formada entre los detectores se encuentre dentro de un rango angular determinado por la electrónica, que solo permite a un detector individual entrar en coincidencia con un número finito de bloques opuestos. Esta característica define el campo de visión transversal del equipo (FOV, del inglés *Field Of View*) [8].

Si las tres condiciones se cumplen la coincidencia se considera válida, por lo tanto, el proceso de determinación de coincidencias, esquematizado en la Figura 1-2, es una forma de colimación electrónica.

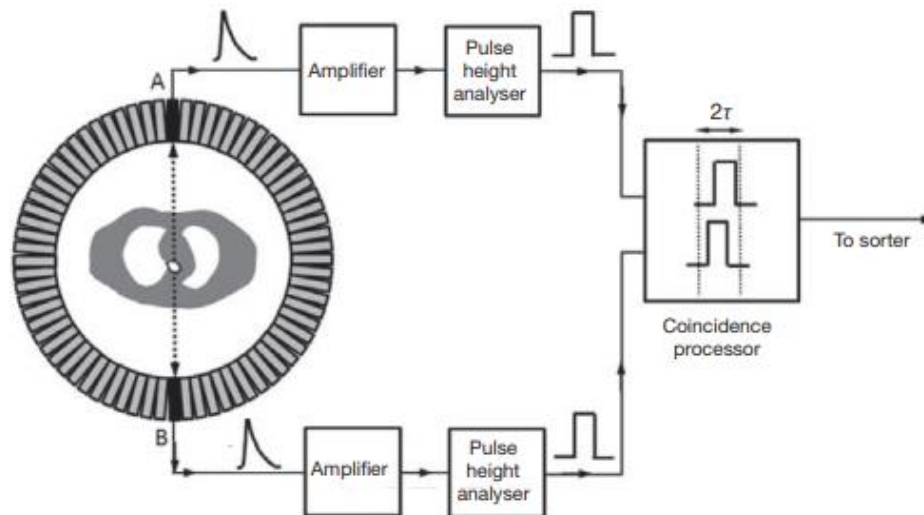


Figura 1-2: Sistema de detección en coincidencia. Cada evento de detección simple produce un pulso eléctrico que es amplificado y analizado. Para que se considere una coincidencia ambos eventos deben estar dentro de la ventana energética, la ventana temporal y el rango angular permitido. *Figura tomada de [9].*

Las LOR consideradas válidas se almacenan y ordenan en matrices bidimensionales, $S(\phi, r)$, conocidas como *sinogramas*, debido a que una fuente puntual traza un patrón de onda sinusoidal en la matriz, y que funcionan como un histograma en 2D. Cada LOR se identifica a partir del ángulo acimutal (ϕ) y la distancia radial (r) medida desde el

origen de coordenadas, ubicado en el centro del anillo, hasta la periferia. Por lo tanto, cada celda de la matriz representa la cantidad de eventos de coincidencia válidos en una línea de respuesta (Figura 1-3). Una tercera componente (z) define la posición longitudinal del sinograma generado para cada anillo de detectores. En los sistemas PET actuales que trabajan en modo 3D, una componente adicional (θ) define el ángulo de inclinación de los planos oblicuos entre diferentes anillos.

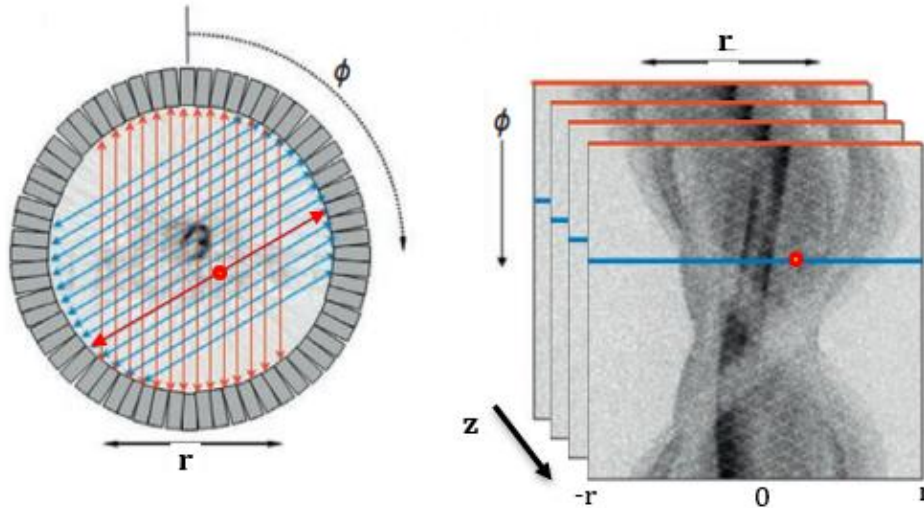


Figura 1-3: Esquema de la construcción de un sinograma a partir de las proyecciones de los ángulos de muestreo. (Izq.) Representación de dos proyecciones (líneas naranjas y azules). (Der.) Formación del sinograma $S(\phi, r, z)$, resaltando las filas de la matriz correspondientes a las proyecciones de la imagen de la izquierda y la posición que ocupa la LOR individual roja. El nivel de gris representa la cantidad total de eventos en una misma línea de respuesta. *Figura tomada de [9] y modificada por el autor.*

No obstante, algunos de los eventos validados como coincidencias y guardados en el sinograma son indeseados. Por ejemplo, aquellos eventos de coincidencia en los cuales al menos uno de los fotones sufre dispersión Compton, con desvío de su trayectoria y cambio en su energía antes de alcanzar el detector. En este caso si la energía del fotón dispersado se encuentra dentro de la ventana energética, será aceptado, pero la línea de respuesta asignada no será correcta. Este tipo de evento se denomina “coincidencia dispersada” (*scattered coincidence*, en inglés) (Figura 1-4b). Así mismo, aquellos eventos en los que la coincidencia se produce entre dos únicos fotones que provienen de eventos de aniquilación diferentes, pero que se detectan dentro de la ventana

temporal, mientras que los otros dos fotones relacionados no se detectan porque sufren dispersión y salen del FOV o sufren absorción por efecto fotoeléctrico en el cuerpo. En este caso la LOR asignada no corresponde a un evento de aniquilación real. Este tipo de evento se denomina “coincidencia aleatoria” (*random coincidence*, del inglés) (Figura 1-4c). Dado que el decaimiento radiactivo de dos núcleos son eventos independientes, la probabilidad de ocurrencia de una coincidencia aleatoria es el producto de las probabilidades de los eventos de aniquilación individuales. Por lo tanto, la tasa de ocurrencia de coincidencias aleatorias aumenta con el cuadrado de la tasa de desintegración de núcleos, que corresponde a la cantidad de material radiactivo. Los casos anteriormente descritos ocasionan deterioro en la calidad de la imagen final y errores cuantitativos. Las coincidencias que no corresponden a los casos anteriores se identifican como “coincidencias verdaderas” (*True coincidence*, del inglés) (Figura 1-4a).

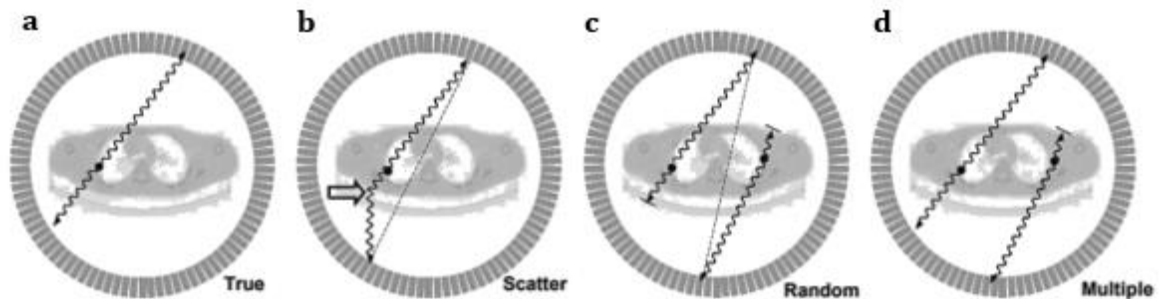


Figura 1-4: Tipos de eventos de coincidencia. El punto negro corresponde a la aniquilación y la línea punteada a la LOR asignada. De izquierda a derecha se presentan: coincidencia verdadera (a), coincidencia dispersada (b), coincidencia aleatoria (c) y evento de detección múltiple (d). *Figura tomada de [8].*

Por otra parte, existen eventos en los cuales un único fotón (evento simple) o más de dos fotones (evento múltiple, Figura 1-4d) son detectados dentro de la ventana temporal, en estos casos no es posible establecer una coincidencia clara y, por lo tanto, los eventos son rechazados. La subestimación de coincidencias verdaderas se da debido a la atenuación de los fotones por efecto fotoeléctrico.

1.1.2 Correcciones cuantitativas

Si bien, no es el propósito de este trabajo estudiar con detalle las correcciones cuantitativas de los datos, se hará una breve descripción de las mismas con el fin de brindar una visión general completa del funcionamiento del PET:

- **Normalización**

Esta corrección de software se aplica con el fin de homogeneizar la sensibilidad de las diferentes LORs, la cual se ve afectada por las variaciones en la eficiencia de los cristales centelladores individuales que componen los anillos de detectores y la distorsión geométrica ocasionada por el ángulo de impacto de los fotones en los detectores. Para aplicar la normalización se genera un sinograma por cada plano con factores de corrección para cada LOR, determinados experimentalmente.

- **Corrección de atenuación**

La corrección de atenuación es la corrección más importante en la generación de imágenes PET, ya que más del 60% de los fotones emitidos interactúan con el tejido [10]. Con el advenimiento de los equipos híbridos PET/CT, la corrección de atenuación se realiza a partir de las imágenes de tomografía computada (CT, del inglés *Computed Tomography*). En un PET/CT un escáner de CT se combina con el escáner de PET en un mismo *gantry* con una camilla común (Figura 1-5). La imagen de CT permite establecer un mapa de coeficientes de atenuación lineal (μ_x) de los tejidos frente a un haz de rayos X, originalmente polienergético, pero medido a una energía efectiva aproximada de 70 keV y expresado en unidades Hounsfield (nombradas así en honor a su inventor).

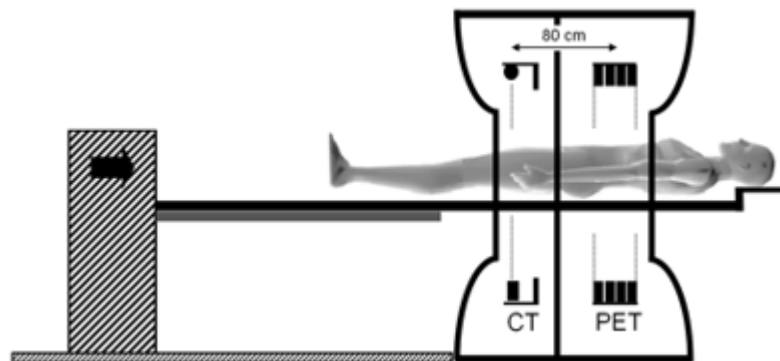


Figura 1-5: Esquema del diseño de un PET/CT. El CT está posicionado enfrente del escáner PET y los centros están alineados. La camilla común favorece la corrección de las imágenes. *Figura tomada de [11].*

Las unidades Hounsfield (HU) son unidades de una escala arbitraria que estandariza los coeficientes de atenuación de los tejidos con respecto al coeficiente de atenuación lineal del agua (μ_w), siendo el aire -1000 HU y el agua 0 HU. Dado que el valor de los coeficientes de atenuación lineal depende de la energía de la radiación, una función de escalado bilineal (Figura 1-6) es usada para convertir los coeficientes de la imagen de CT (~ 70 keV) en coeficientes a la energía de PET (511 keV). Esta función se basa en considerar a todos los tejidos con $\mu_x < \mu_w$ como una mezcla de aire y agua a diferentes concentraciones (k), y a aquellos con $\mu_x > \mu_w$ como una mezcla de agua y hueso, con el punto de quiebre en 60 HU [10].

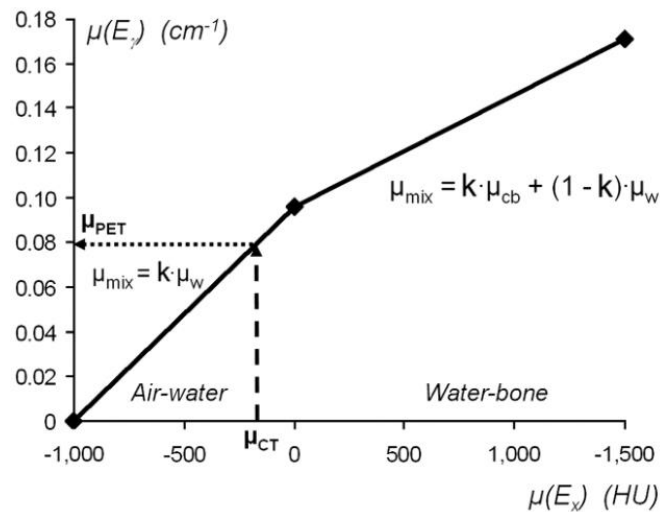


Figura 1-6: Función de escalado para convertir las HU a los valores de atenuación lineal a 511 keV. Los coeficientes del tejido (μ_{mix}) dependen de la mezcla que lo componen: agua-aire o agua (μ_w)-hueso (μ_{cb}). *Figura tomada de [10].*

La imagen de CT adquirida es muestreada al tamaño de vóxel de la imagen de PET y escalada con la función, para luego ser proyectada en un sinograma. Los factores de corrección de atenuación para cada LOR corresponden a la integral de los coeficientes de atenuación de los vóxeles a lo largo de la línea, sin importar la posición exacta en la que se produjo la aniquilación.

▪ Corrección de coincidencias dispersadas

Existe gran variedad de métodos propuestos para implementar la corrección de las coincidencias dispersadas, uno de los más exactos calcula directamente el promedio

de coincidencias dispersadas usando la fórmula de Klein-Nishina para modelar la probabilidad de que los fotones experimenten dispersión Compton, se desvíen en cierto ángulo y sean detectados por un detector particular, a partir del mapa de atenuación de la imagen de CT y una estimación de la distribución del radiofármaco en los tejidos [12]. Como resultado se genera un sinograma de coincidencias dispersadas.

- **Corrección de coincidencias aleatorias**

Los eventos aleatorios contribuyen a la pérdida de contraste y la disminución de la exactitud cuantitativa. Una forma de estimar la tasa de coincidencias aleatorias para cada LOR es multiplicando la tasa de eventos simples (N_1 y N_2) en los dos detectores opuestos y la ventana temporal de coincidencias (2τ). Como resultado se generan sinogramas de la estimación de coincidencias.

- **Corrección de tiempo muerto**

El tiempo muerto se refiere al tiempo que tarda el sistema en registrar (cristal centellador) y procesar el impacto de un fotón (electrónica), durante el cual un nuevo evento no puede ser procesado. La probabilidad de que esto ocurra aumenta a altas tasas de conteo y causa la disminución de la sensibilidad por la pérdida de fotones incidentes no contados. La corrección del tiempo muerto se puede implementar a través de un factor determinado experimentalmente midiendo una fuente de actividad conocida en decaimiento.

1.1.3 Reconstrucción de la imagen en PET

La calidad final de la imagen tomográfica reconstruida dependerá de la cantidad de proyecciones, la resolución de cada proyección y la cantidad de eventos detectados. En la actualidad los métodos iterativos son los de mayor aplicación para la reconstrucción de las imágenes de PET, cuya ventaja sobre los métodos analíticos radica en la mayor robustez frente al ruido y, por lo tanto, una mejor relación señal a ruido (SNR, del inglés *Signal-Noise Ratio*) en regiones de baja captación del radiofármaco (p. ej. En el contorno del cuerpo) [8]. Los más comunes son los métodos de maximización de la verosimilitud (EM, del inglés *Expectation-Maximization*):

- **Maximum Likelihood Expectation-Maximization (MLEM)**

El método MLEM maximiza una función de la verosimilitud de las proyecciones a partir de los datos medidos basada en la distribución de Poisson, haciendo que las proyecciones computadas sean tan similares como sea posible a las proyecciones medidas [13]. El flujo de trabajo de su implementación consiste en: (i) proponer una imagen inicial como solución, donde todos los voxels tienen un valor constante distinto de cero, (ii) simular las proyecciones en el espacio de los sinogramas, (iii) aplicar las correcciones cuantitativas de los datos (normalización, atenuación, coincidencias aleatorias, etc.), (iv) comparar los datos del sinograma corregido con los datos medidos, (v) ajustar la solución mediante una retroproyección del error y posterior multiplicación con la estimación anterior de la imagen y (vi) evaluar la similitud de la nueva solución con las proyecciones medidas. Si la similitud no es suficiente el ciclo vuelve a iniciar tomando como entrada (i) la solución ajustada en (v), realizando tantas iteraciones como sea necesario hasta que la solución converja.

- **Ordered-Subsets Expectation-Maximization (OSEM)**

El método OSEM es la versión acelerada del método MLEM. Para su implementación el conjunto de todas las proyecciones es dividido en subconjuntos y a cada uno de ellos consecutivamente se aplica el algoritmo descrito para el MLEM.

1.1.4 Cuantificación de las imágenes

En las imágenes de PET la intensidad del vóxel será proporcional a la cantidad de coincidencias detectadas corregidas por los factores descritos anteriormente (normalización, tiempo muerto, atenuación, aleatorios y dispersión). Además, es posible determinar un factor de conversión a concentración de actividad a partir de la adquisición de un fantoma cilíndrico relleno con una solución de actividad conocida, en un proceso llamado calibración cruzada con el activímetro. Dicho factor se guarda en la base de datos del equipo y es usado automáticamente durante la reconstrucción de las imágenes. Entonces, la imagen de PET es una representación tridimensional de la concentración de actividad del radiofármaco captada en el cuerpo del paciente,

donde el valor de cada vóxel está dado en kBq/mL. Un ejemplo de las imágenes de PET/CT se muestra en la Figura 1-7.

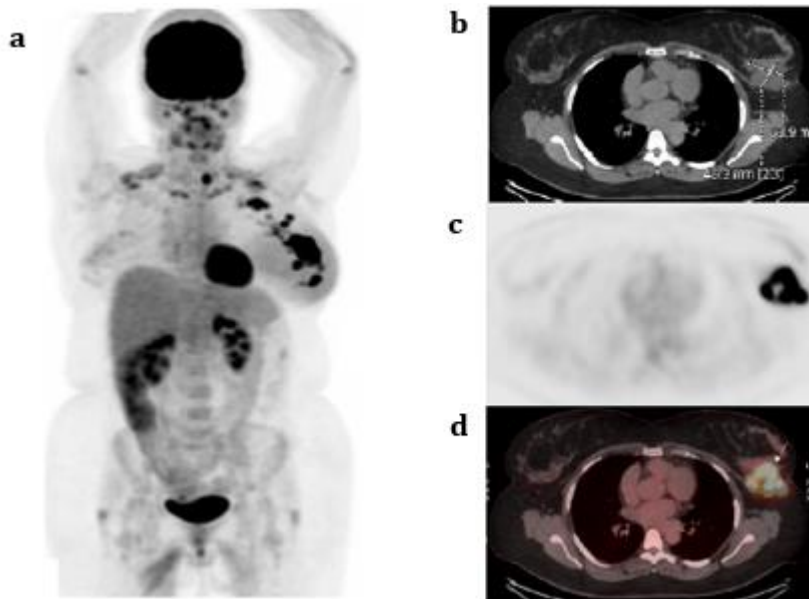


Figura 1-7: Imagen PET/CT. a. Imagen de máxima proyección de intensidad (MIP) de un PET/CT con F-18 FDG en paciente con lesiones en la mama izquierda. b, c, d Imagen de corte axial a nivel de tórax de CT, de PET y fusión de PET/CT.

Existen varios factores que afectan la cuantificación como el denominado *efecto de volumen parcial* (PVE, del inglés *Partial Volume Effect*). Este efecto es ocasionado por dos fenómenos: (1) la *resolución espacial finita* que resulta en bordes difuminados que se dispersan hacia afuera de la estructura y contaminan otras regiones (Figura 1-8a) y (2) el *efecto de fracción de tejido* que resulta en estimación errónea debido a que en un vóxel en el cual coexisten varios tejidos o hay captación heterogénea la intensidad se determina como la ponderación de los diferentes tejidos, ocasionando contaminación hacia afuera o hacia adentro de la región (Figura 1-8b). El PVE reduce la captación medida para objetos con dos o tres veces el tamaño de la resolución espacial del equipo (p. ej. Para objetos de 15 mm puede ser del 50%) [14]. Una posible solución es modelar la resolución espacial para cada vóxel usando la función de dispersión puntual (PSF, del inglés *Point Spread Function*) y convolucionar la respuesta con la estimación de la imagen, como en la aplicación de un filtro, dentro del algoritmo de reconstrucción. Sin embargo, esta técnica puede amplificar el nivel de ruido.

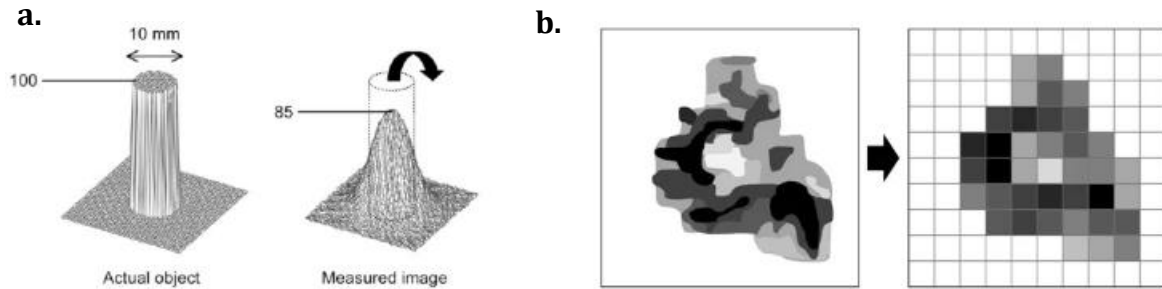


Figura 1-8: Representación de los fenómenos que causan el efecto de volumen parcial. **a.** Resolución espacial finita. *Tomada de [14].* **b.** Efecto de fracción de tejido en una lesión heterogénea. *Tomada de [15].*

1.1.5 Valor de captación estandarizado (SUV)

El valor de captación estandarizado (SUV, del inglés *Standardized Uptake Value*) es una medida semicuantitativa de la concentración de actividad del trazador en una región de interés (ROI, del inglés *Region Of Interest*) normalizada por la actividad total administrada al paciente corregida por el decaimiento (F_{dec}) entre la hora de la medición de la actividad inyectada y el inicio de la imagen PET y dividida por la masa del paciente:

$$SUV[\text{g/mL}] = \frac{\text{Concentración de Actividad} [\text{kBq/mL}]}{\text{Actividad inyectada} [\text{kBq}] \times F_{dec} / \text{Masa del paciente} [\text{g}]} \quad (1-1)$$

Por lo tanto, el valor de SUV medido de una lesión oncológica se mantiene invariable frente a cambios en la actividad administrada al paciente al repetir el estudio permitiendo así realizar comparaciones intrapaciente, por ejemplo, para seguimiento o evaluación de respuesta al tratamiento. Del mismo modo, la normalización por la masa del paciente permite realizar comparaciones del valor de SUV de una patología similar entre pacientes de diferente peso.

En la rutina clínica el SUV es usado para determinar malignidad y evaluar la respuesta al tratamiento. Puede calcularse de diferentes formas, como: el valor del píxel con la máxima actividad en el ROI (SUV_{\max}), el valor promedio de todos los píxeles en el ROI (SUV_{mean}) o el valor del máximo promedio en un volumen de 1 cm^3 dentro del volumen de interés (SUV_{peak}), el cual no necesariamente está centrado en el píxel de mayor

intensidad, entre otros. Sin embargo, existen factores biológicos y técnicos que pueden afectar el sesgo y la varianza de la medición.

▪ **Factores que afectan la medición del SUV en estudios con F-18 FDG:**

El SUV se ve afectado por: (i) *factores biológicos*, como: (1) el porcentaje de grasa del paciente, particularmente en pacientes obesos, ya que el tejido adiposo tiene baja captación del F-18 FDG y la estimación de SUV asume una distribución uniforme; (2) el nivel de glucosa en sangre, ya que compite con la captación de FDG y reduce el valor de SUV [16]; (3) los movimientos involuntarios del paciente, como la respiración, ya que en lesiones pulmonares el CT usado para la corrección de atenuación no corregistra exactamente con la imagen de PET que dura varios minutos en respiración libre y (4) lesiones benignas captantes de glucosa, como los procesos inflamatorios, que compiten por el FDG y causan falsos positivos. Por otra parte, (ii) *factores técnicos*, como: (1) la calibración cruzada con el activímetro, (2) la sincronización de relojes entre el activímetro y el equipo PET, ya que son usados para definir el tiempo en la corrección por decaimiento, (3) el algoritmo de reconstrucción y los parámetros (cantidad de iteraciones, subsets, post-filtros, etc.) empleados, (4) la aplicación de correcciones cuantitativas, (5) el efecto de volumen parcial, (6) la fórmula usada para definir la dosis a administrar, (7) el método de delineación del ROI, (8) la presencia de material de contraste usado para CT, que introduce errores en el cálculo del mapa de atenuación y (9) la variabilidad entre equipos PET, ya que la eficiencia de los componentes es diferente [17]. Por lo tanto, para que sean comparables los valores de SUV se debe protocolizar la técnica para que sea reproducible.

1.1.6 Otras métricas volumétricas

Existen otros parámetros medidos en las imágenes de PET para caracterizar cuantitativamente el metabolismo de una lesión, entre ellos:

▪ **Volumen metabólico tumoral (MTV)**

El MTV (del inglés, *Metabolic Tumor Volume*) es una medida del volumen tumoral con alto metabolismo y existen varios métodos para su determinación, por ejemplo, usando un porcentaje del SUV_{max} como umbral.

- **Nivel de glicólisis total de la lesión (TLG)**

El TLG (del inglés, *Total Lesión Glycolysis*) se define como el producto del SUV_{mean} del tumor y el MTV.

1.1.7 F-18 FDG

La 2-[^{18}F]-fluoro-2-deoxy-D-glucosa (F-18 FDG) es el radiofármaco más comúnmente usado en la rutina clínica de PET. Este análogo de la glucosa se incorpora a las células, pero no se metaboliza, sino que se acumula, por consiguiente, su biodistribución es un reflejo de la captación de glucosa en las células. El ^{18}F -FDG consta de un trazador (2-deoxy-D-glucosa, DG) ligado por medio de síntesis química al isótopo radiactivo emisor de positrones, F-18.

Debido a que las células tumorales son más activas metabólicamente por su alta tasa de reproducción, tienen una demanda más alta de energía producida a partir de la glucosa, por lo tanto, captan niveles más altos de FDG que los tejidos normales. Sin embargo, la captación de FDG no es específica del cáncer y puede verse incrementada en otros procesos como: inflamación, hiperactividad muscular o nerviosa y reparación de tejidos, por ejemplo, después de realizar actividad física. Por este motivo, una preparación rigurosa debe seguirse antes del estudio para evitar falsos positivos.

1.1.8 Conclusiones

La tomografía por emisión de positrones produce imágenes funcionales *in vivo* intrínsecamente cuantitativas, de las cuales es posible medir marcadores metabólicos que permiten caracterizar las lesiones tumorales, como el SUV, entre otros, usados especialmente en el seguimiento y evaluación de la respuesta al tratamiento en oncología. Tales marcadores metabólicos podrían aportar potencial predictivo por evaluar.

1.2 Imágenes por Resonancia Magnética (MRI)

En esta sección se presenta una breve introducción a los principios de funcionamiento de las imágenes por Resonancia Magnética (RM). Parte de esta revisión se realiza siguiendo el libro “*The Essential Physics of Medical Imaging*” de *Bushberg et al.* [18] y el curso en línea de *Allen Elster* [19].

1.2.1 Bases físicas de la MRI

La imagenología por Resonancia Magnética (MRI, por sus siglas en inglés) es una técnica radiológica no invasiva con una amplia variedad de aplicaciones debido a que permite obtener imágenes bi- y tridimensionales anatómicas, fisiológicas e incluso metabólicas del paciente con una alta resolución espacial (1.0 mm, que puede mejorar entre mayor sea el Campo Magnético) y mayor contraste entre tejidos blandos que la tomografía computada, sin el uso de radiación ionizante. Su funcionamiento se basa en la propiedad magnética que exhiben los núcleos del átomo de forma individual para selectivamente absorber energía y liberarla posteriormente a su entorno circundante al ser estimulados en presencia de un fuerte campo magnético externo, fenómeno denominado *Resonancia Magnética Nuclear* (RMN) [18]. Los nucleones, protones y neutrones, poseen una propiedad intrínseca llamada “momento angular de espín” (I) o simplemente “*espín*”, la cual está cuantizada en unidades semienteras y es igual a $I = 1/2$ para ambos. Sin embargo, los nucleones del mismo tipo tienden a aparearse entre sí, de acuerdo con el modelo nuclear de capas, eliminando la manifestación del espín total del núcleo ($I = 0$) cuando se encuentran en cantidades pares [20]. El espín genera un dipolo magnético en cada partícula, como si se tratara de un diminuto imán, cuya tendencia para interactuar con un campo magnético externo se representa por la magnitud vectorial “*Momento Magnético*” ($\vec{\mu}$). La magnitud de $\vec{\mu}$ a nivel nuclear también se ve afectada por el apareamiento de los nucleones que conforman el núcleo, por lo tanto, solo aquellos núcleos con espín diferente de cero pueden experimentar resonancia magnética. Sin embargo, la intensidad del momento magnético de un solo átomo no es suficiente para ser detectable y se necesita de una gran cantidad de ellos ($\sim 10^{15}$) para obtener una señal útil en MRI.

En resumen, se requieren tres elementos principales para producir Resonancia Magnética Nuclear: (1) Un fuerte campo magnético externo, que en adelante se denominará campo magnético principal (\vec{B}_0), (2) Núcleos atómicos con espín diferente de cero y abundancia suficiente para ser detectables y (3) Un campo magnético oscilante a la frecuencia de resonancia, perpendicular a B_0 , que excite los núcleos.

El campo principal B_0 en MRI, típicamente de 1.5 o 3.0 Teslas en la rutina clínica, generalmente es generado por una corriente eléctrica (i) estacionaria a través de una bobina mantenida a temperaturas de superconducción. La dirección de B_0 estará a lo largo del eje central de la bobina como describe la ley de la mano derecha (llamado por convención eje z). Mientras que, la intensidad del campo en un punto del eje estará determinada por la resolución de la ley de Biot-Savart, como en la Ecuación (1-2):

$$\vec{B}_{0z} = \frac{\mu_0 i N}{2L} \left[\frac{z + L/2}{\sqrt{R^2 + (z + L/2)^2}} - \frac{z - L/2}{\sqrt{R^2 + (z - L/2)^2}} \right] \vec{k} \quad (1-2)$$

Siendo N el número de espiras, L la longitud de la bobina, z la distancia del punto al centro de la bobina, R el radio de la bobina y μ_0 la permeabilidad del vacío.

Sin embargo, una única bobina no es suficiente para lograr la alta uniformidad de campo que se requiere para obtener imágenes de una región. En cambio, un arreglo de 6 a 10 bobinas paralelas, montadas sobre un cilindro y separadas por espacios, en las que fluye la corriente en el mismo sentido, permite obtener una región esférica de campo uniforme en el centro (isocentro del equipo) con una rápida disminución del campo marginal fuera de los extremos del cilindro (ver Figura 1-9). La homogeneidad de B_0 se cuantifica en partes por millón (ppm) sobre el diámetro del volumen esférico (DSV, del inglés *diameter of spherical volume*).

En MRI clínica es usado principalmente el núcleo de hidrógeno (^1H) debido a que posee el momento magnético más alto y, por tanto, la mayor sensibilidad a la RMN. Además,

tiene la mayor abundancia en los tejidos, predominantemente en las moléculas de agua y grasa. Características que resultan en un incremento de la relación señal-ruido (SNR).

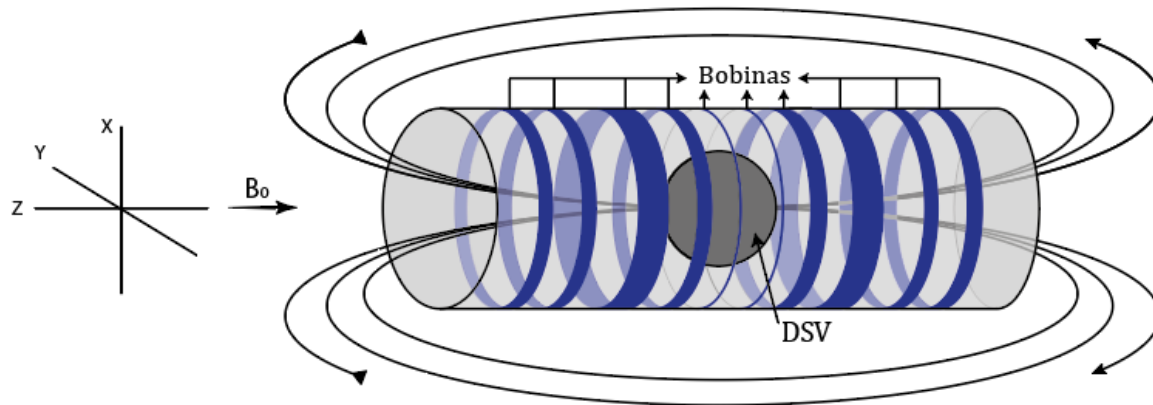


Figura 1-9: Generación del campo magnético principal B_0 en un sistema de MRI típico. Campo magnético horizontal producido en magneto de núcleo de aire con arreglo de bobinas paralelas sobre un cilindro de alrededor de 1 m de diámetro y 2 a 3 m de largo. DSV en el centro del cilindro. *Figura del autor.*

El núcleo de ^1H está formado por un único protón, por esta razón, en adelante, como es común en RMN se hará referencia a ellos como espines o protones indistintamente. Cuando el paciente ingresa en el resonador, los dipolos magnéticos de los protones que originalmente apuntaban en direcciones aleatorias con una distribución uniforme en ausencia del campo (ver Figura 1-10A), ahora, experimentan un torque que provoca la precesión del eje de los espines en torno a la dirección de \vec{B}_0 a una frecuencia común, llamada la *frecuencia de Larmor* (f_0) (ver ecuación (1-3)).

$$f_0 = \gamma \cdot B_0 \quad (1-3)$$

Donde γ es la constante giromagnética específica del núcleo (igual a 42.58 MHz/Tesla para el protón). Al mismo tiempo, los dipolos experimentan una leve tendencia a apuntar en dirección de \vec{B}_0 . Sin embargo, a la temperatura corporal, la energía asociada a la orientación individual de los espines es mucho menor que la energía asociada a las colisiones térmicas, por consiguiente, los espines individuales a menudo chocan entre sí y no alcanzarán una dirección estable, es decir, muchos seguirán apuntando aleatoriamente. Como resultado se obtiene una distribución esférica pero ligeramente

sesgada hacia \vec{B}_0 , la cual produce una magnetización macroscópica (\vec{M}_z), como se muestra en la Figura 1-10B. No obstante, la magnetización no es instantánea, sino que crece exponencialmente en el tiempo (t) a medida que los espines se orientan a su estado de baja energía. La energía del sistema disminuye por transferencia del movimiento térmico a los átomos cercanos hasta alcanzar el valor de equilibrio M_0 , un flujo de energía desde los espines hacia su ambiente, tal como lo describe la ecuación de Bloch (1-4). En ese momento la distribución de espines se mantiene rotando en el eje z sin precesar [21].

$$M_z(t) = M_0(1 - e^{-t/T1}) \quad (1-4)$$

Donde $T1$ es el tiempo requerido para que M_z crezca al 63% de su valor final, llamado tiempo de relajación longitudinal o relajación “espín-red” (en inglés, *spin-lattice*). El grado de alineación neta es proporcional a la fuerza del campo externo y al número de espines (N_s) como se muestra en la ecuación (1-5) :

$$M_0 \approx \gamma^2 h^2 N_s B_0 / 4k_B T \quad (1-5)$$

Donde h es la constante de Planck, k_B la constante de Boltzmann y T la temperatura absoluta.

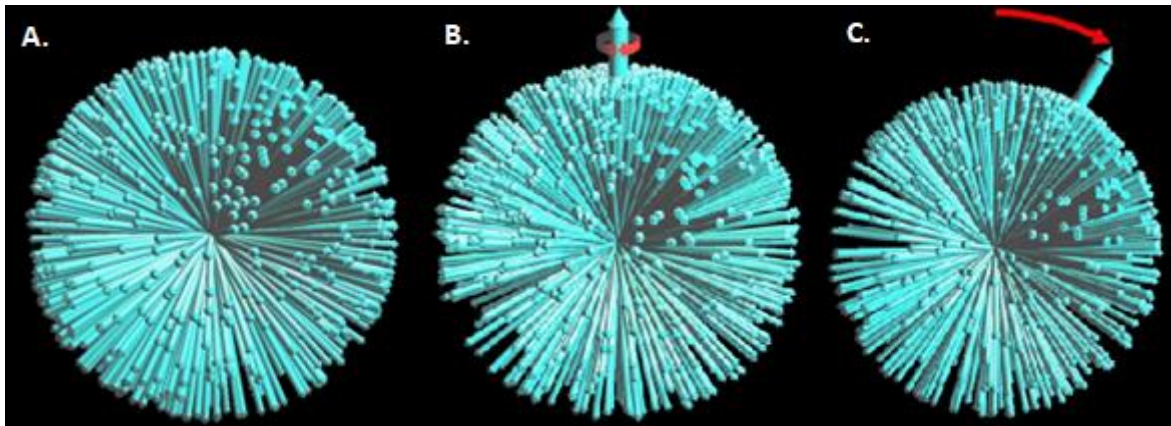


Figura 1-10: Representación tridimensional de la distribución de espines en el cuerpo y su interacción con los campos magnéticos. **A.** En ausencia del campo la distribución es isotrópica. **B.** En presencia de B_0 aumenta la densidad de espines en dirección del campo ocasionando una magnetización neta \vec{M}_0 . **C.** Cuando un pulso de RF de campo magnético homogéneo es aplicado, toda la distribución de espines rota. Representación en el sistema de referencia rotante con f_0 . Tomada de [21].

Cabe aclarar que aquí no se asume el enfoque más extendido entre los libros y cursos de resonancia magnética, que plantean la separación de los espines entre dos únicos autoestados (paralelo o antiparalelo) a lo largo de la dirección del campo magnético externo. Porque como se explica en [19], [21] y [22] se considera una interpretación errónea de los principios de la mecánica cuántica, ya que la función de onda no implica que el protón se encuentre exclusivamente en uno de sus dos autoestados ni que solo pueda cambiar del uno al otro. Por el contrario, de acuerdo con la interpretación de Copenhagen, los espines pueden estar en una combinación lineal de sus autoestados, y solo cuando se someten al proceso de medición la función de onda colapsa en alguno de los dos. Además, este enfoque es inadecuado, ya que en MRI no se mide la magnetización de los protones uno a uno como en el experimento de Stern-Gerlach sino la magnetización del conjunto. Por lo tanto, el vector \vec{M} opera bajo los principios de la física clásica como demostraron Feynman *et al.* [23] y puede tomar cualquier orientación con respecto al campo entre paralelo y antiparalelo pasando por todos los valores intermedios.

Ahora bien, para producir la señal de RM se requiere perturbar la alineación de \vec{M} . Esto se logra aplicando un segundo campo magnético (\vec{B}_1) perpendicular a \vec{B}_0 y oscilante a una frecuencia igual o muy cercana a la frecuencia de Larmor, en otros términos, un pulso de radiofrecuencia (RF). En la actualidad, el campo \vec{B}_1 es principalmente generado usando una bobina tipo jaula de pájaro, integrada o externa, llamada así por su forma, consistente de una estructura cilíndrica con dos espiras conductoras en los extremos conectadas por un número par de barras también conductoras. Una corriente alterna sinusoidal se aplica secuencialmente en cada barra con un cambio de fase sucesivo, similar al funcionamiento de un motor trifásico; de esta forma se produce un campo magnético rotante en el interior del cilindro, a una frecuencia cercana a la frecuencia de Larmor, altamente homogéneo y circularmente polarizado. Entonces, cuando el pulso de RF se enciende la energía es transmitida al conjunto de espines ocasionando la rotación de toda la distribución sin alterar la orientación

relativa de los espines individuales (ver Figura 1-10C), considerada una “*coherencia de fase*”, por esta razón, no cambia la magnitud de \vec{M} sino solo su dirección.

Una vez desalineado, el vector \vec{M} experimenta un torque debido a la interacción con \vec{B}_0 que ocasiona su precesión alrededor del eje z, pero, a su vez, experimenta un segundo torque debido a la interacción con \vec{B}_1 que ocasiona su rotación hacia el plano xy. Sin embargo, la magnitud de B_1 es mucho menor que la de B_0 y, por lo tanto, la frecuencia de rotación también es mucho menor que f_0 (del orden de decimas a unidades de ms para un giro de 90°). Como resultado \vec{M} describirá una trayectoria helicoidal como la mostrada en la Figura 1-11A. Para simplificar el análisis un cambio de sistema de referencia a un sistema rotante a la frecuencia de Larmor sobre el eje z es realizado, de esta forma, el efecto de precesión provocado por B_0 queda latente y ahora es fácil observar que por el efecto de \vec{B}_1 el vector \vec{M} describe un ángulo de inclinación (α) con respecto al eje z, que depende de la magnitud y la duración del pulso. Así mismo \vec{M} se puede descomponer en una magnetización longitudinal (M_z) y una transversal (M_{xy}), como se muestra en la Figura 1-11B.

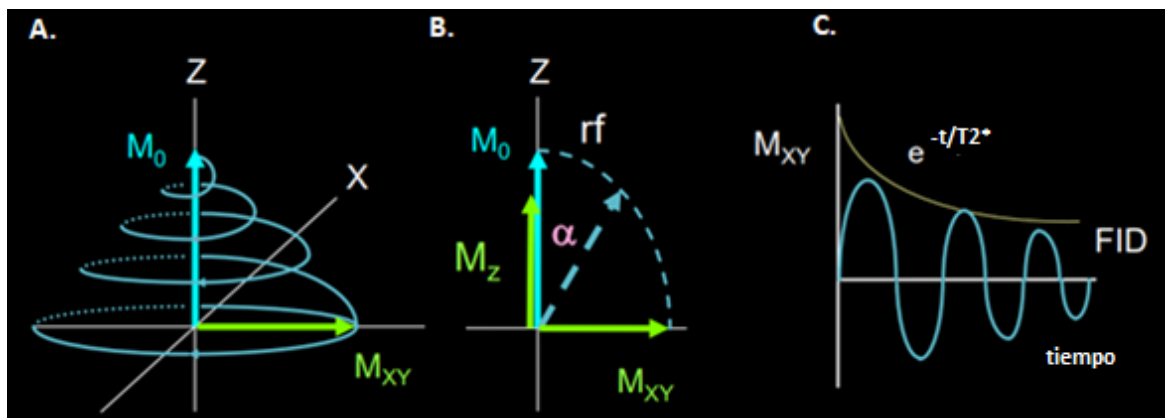


Figura 1-11: Comportamiento de la magnetización antes y después de la aplicación del pulso de radiofrecuencia. *Con permiso de J. N. Oshinski [24].* Visto en el sistema de referencia fijo del laboratorio. **B.** Visto en un sistema de referencia rotante a la frecuencia de Larmor. **C.** Señal de RF recibida en la bobina receptora mientras \vec{M} vuelve al equilibrio.

Cuando el pulso de RF se apaga, la magnetización retorna al equilibrio describiendo la misma trayectoria helicoidal de la Figura 1-11A. La magnetización longitudinal se

recupera por el proceso ya mencionado de relajación espín-red, como aparece en la ecuación (1-4) más el término $M_0 \cos \alpha \cdot e^{-t/T_1}$, para pulsos diferentes de 90° . Mientras que la magnetización transversal decae exponencialmente, debido a varios mecanismos de interacción entre espines, como describe la ecuación de Bloch (1-6):

$$M_{xy}(t) = M_0 e^{-t/T_2} \quad (1-6)$$

Donde T_2 es el tiempo requerido para que M_{xy} disminuya en un 63% de M_0 , llamado tiempo de relajación transversal o relajación “espín-espín”. Tanto T_1 como T_2 dependen de la estructura atómica del tejido. Además, a medida que M_{xy} rota la variación del flujo magnético induce un voltaje en la bobina receptora, como describe la ley de Faraday-Lenz, cuya señal es proporcional al número de protones excitados en la muestra y tiene la forma de una función sinusoidal a la frecuencia de Larmor, amortiguada por una envolvente exponencial negativa con una constante de tiempo T_2^* (ver Figura 1-11C). Donde T_2^* es la tasa de desfase transversal debido a T_2 más los efectos de la inhomogeneidad del campo magnético B_0 localmente. Esta señal recibida es llamada “decaimiento libre de inducción” (FID, del inglés *Free Induction Decay*).

1.2.2 Formación de la imagen

Para obtener imágenes de RM es indispensable determinar la localización espacial de donde proviene la señal. Tres conjuntos de bobinas enfrentadas, denominados bobinas gradientes, están integrados en el resonador con este fin, uno para cada eje (x, y, z). Cuando un conjunto se enciende, la corriente pasa en sentido opuesto a través de un par de bobinas produciendo un campo magnético en dirección z y de magnitud cero en el isocentro que incrementa linealmente hacia el exterior en la dirección de aplicación (incluyendo direcciones oblicuas), pero con polaridad contraria a cada lado. La superposición de este nuevo campo a B_0 genera un gradiente uniforme de campo magnético dependiente de la posición en la dirección de aplicación que, a su vez, provoca una variación lineal de la frecuencia de resonancia de los protones en función de la posición, como se expresa en la ecuación (1-7) para el eje z:

$$f(z) = \gamma(B_0 + z \cdot G_z) = f_0 + \gamma z \cdot G_z \quad (1-7)$$

Donde G_z es la fuerza del gradiente en el eje z , medida en mT/m y, z es la posición, medida desde $-z$ a $+z$, con $z = 0$ en el isocentro.

Entonces, para adquirir imágenes de RM en dos dimensiones (2D), primero se enciende simultáneamente un gradiente para modificar la frecuencia de precesión según la posición y un pulso de RF sintonizado en un rango de frecuencias (ancho de banda, Δf) para excitar de forma selectiva únicamente una sección del volumen a estudiar, a la que se denomina *corte*, cuyo espesor dependerá del ancho de banda del pulso y de la fuerza del gradiente. Es decir que la resonancia magnética ocurrirá solo en dicho corte. Por esta razón, este primer gradiente es llamado “gradiente de selección de corte” (G_{SS}). Cuando el G_{SS} se apaga la magnetización vuelve a precesar a f_0 , sin embargo, el cambio de frecuencias habrá ocasionado un desfase que podría dispersar la magnetización transversa y disminuir la señal, por tal motivo, un gradiente de polaridad inversa con la misma fuerza que G_{SS} , pero con la mitad de la duración es aplicado inmediatamente después para corregirlo.

Ahora, el problema ha sido reducido a la localización de la señal en el plano del corte. Para lograrlo, un segundo gradiente, denominado “*gradiente de codificación de fase*” (G_{PE}), es encendido temporalmente a lo largo de uno de los ejes del plano, modificando la frecuencia de precesión según la posición, lo que a su vez, provoca el desfase de la magnetización por filas con respecto al estado original, proporcionalmente a la fuerza del G_{PE} y su duración (t_G); como si se tratara de varias filas de relojes analógicos cuyas manecillas empiezan a girar desde el mismo punto, pero con diferente velocidad en cada fila. Al observar en un instante determinado, antes de completar un ciclo, las manecillas de una misma fila estarán en la misma posición, pero, en una posición diferente con respecto a las otras filas, es decir, habrá un desfase entre filas. Cuando el G_{PE} se apaga, la magnetización vuelve a precesar a la misma frecuencia f_0 en todo el corte, sin embargo, la fase acumulada (Φ) persiste en función de la posición. Después, un tercer gradiente, ortogonal a G_{PE} , se mantiene encendido durante la adquisición de

la señal para modificar la frecuencia de precesión de la magnetización según la posición en dirección perpendicular a la codificación de fase, razón por la cual es denominado “*gradiente de lectura*” o “*gradiente de codificación de frecuencia*” (G_{FE}). Por ejemplo, si por simplicidad se asume el corte en el plano xy , G_{PE} en el eje y y G_{FE} en el eje x , en cada celda volumétrica de la matriz formada, la magnetización precesará con diferente combinación de frecuencia y fase como en la Figura 1-12A.

La señal de resonancia recibida en la bobina, originalmente en MHz, es llevada a banda base –proceso equivalente a cambiar al sistema de referencia rotante a f_0 –, quedando en el orden de los kHz, para luego ser digitalizada y convertida a números complejos, con el fin de tener información de la fase o sentido de la rotación de la magnetización transversal en el plano. Esta señal digital compleja ($S(t)$) será la sumatoria de la magnetización en cada una de las celdas en que es dividido el corte por la codificación, proporcional a la densidad de protones, con una combinación única de frecuencia y fase (ver Figura 1-12B) como se expresa en la ecuación (1-8):

$$S(t) = \iint M_0(x, y) e^{-t/T_2^*} \cdot e^{j(\gamma G_{FE} \cdot x \cdot t + \gamma G_{PE} \cdot y \cdot t_G)} dx dy \quad (1-8)$$

Donde j es la unidad imaginaria. Realizando un cambio de variables:

$$S(k_x(t), k_y(t)) \propto \iint M_{\perp}(x, y) \cdot e^{-2\pi j(k_x \cdot x + k_y \cdot y)} dx dy \quad (1-9)$$

Donde $k_x = \gamma \int G_{FE} \cdot dt / 2\pi$ y $k_y = \gamma G_{PE} t_G / 2\pi$ son variables en el espacio de Fourier que representan la frecuencia espacial y M_{\perp} es la magnetización transversal en cada celda. La ecuación (1-9) corresponde a una Transformada de Fourier (FT) de la magnetización, lo cual significa que la señal se recibe directamente en el dominio frecuencial, sin embargo, aunque las componentes de frecuencia de la señal pueden ser decodificadas inequívocamente por la FT no es posible determinar la contribución individual de cada celda a la fase final con una sola señal, por consiguiente, es necesario repetir el proceso anterior cambiando la fuerza del gradiente G_{PE} , tantas veces como número de píxeles en la dirección de codificación de fase. Cada señal es muestreada en diferentes intervalos de tiempo durante la aplicación de G_{FE} y los datos ubicados en

una matriz compleja llamada “*espacio-k*”, la cual es simétrica respecto al origen y cuyo tamaño y espaciamiento estarán determinados por el tamaño de píxel y el tamaño del campo de visión de la imagen (FOV, del inglés *Field Of View*) configurados en el resonador. En el método más básico el espacio *k* se llena completo, de izquierda a derecha, una fila por cada señal medida en cada cambio del gradiente de fase (ver Figura 1-12C).

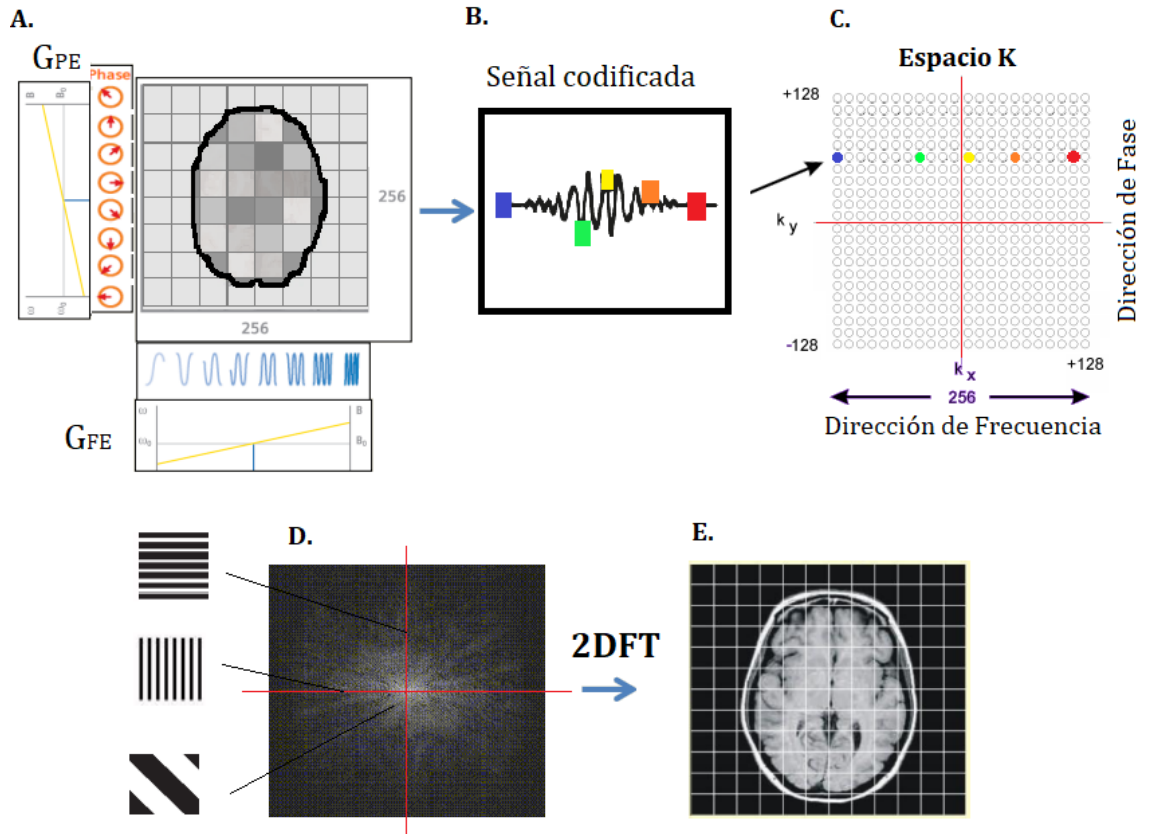


Figura 1-12: Diagrama de flujo del proceso de formación de la imagen en RM en 2D. **A.** Ejemplo de codificación de frecuencia y fase (matriz de adquisición 256×256) en un corte axial de cerebro. Los relojes a la izquierda representan el cambio de fase en cada fila. **B.** Señal codificada medida en la bobina receptora. Los cuadros de colores representan el muestreo discreto de la señal. **C.** Llenado Cartesiano del espacio-*k*, de izquierda a derecha de acuerdo con el muestreo en (B), una fila por cada repetición con diferente fuerza de G_{PE} (desde +128 a -128, 256 pasos). **D.** Imagen del espacio *k* lleno. **E.** Imagen final obtenida aplicando el valor absoluto de la FT inversa sobre el espacio-*k*. *Figura del autor.*

Una vez lleno el espacio-*k*, cada punto (k_x, k_y) corresponde a un patrón de frecuencias, donde: la distancia desde el origen (*k*) es el número de onda, el ángulo medido desde

$+k_y$ indica la orientación y la intensidad del punto el grado de aporte a la imagen final (ver Figura 1-12D). La reconstrucción de la imagen es realizada aplicando la transformada de Fourier inversa en 2D que produce dos imágenes, una de la componente real y otra de la componente imaginaria, a partir de las cuales se obtiene una imagen final de magnitud (ver Figura 1-12E) y una de fase. Las frecuencias espaciales son convertidas en valores de píxel. Cada píxel de la imagen mapea cada punto en el espacio k . Finalmente, el proceso completo es repetido para los demás cortes hasta completar el volumen estudiado.

1.2.3 Secuencias de pulsos

En esta sección se presentan las secuencias básicas de RM usando el diagrama de tiempo de pulsos, donde los gradientes se simbolizan idealmente como un rectángulo, cuya altura representa la fuerza del gradiente y el ancho representa el tiempo de aplicación. El tiempo entre el primer pulso de RF y la medición de la señal, también llamada “eco”, se denomina tiempo de Eco (TE). Mientras que el tiempo entre el primer pulso y la repetición de todo el paquete con el cambio de la codificación de fase se denomina tiempo de repetición (TR).

- **Secuencia Spin-Echo (SE)**

En esta secuencia primero un gradiente de selección de corte G_{SS} , por simplicidad en el eje z , y un pulso de RF de 90° con un ancho de banda fijo se aplican simultáneamente tumbando la magnetización M_0 sobre el plano xy , seguido de medio gradiente G_{SS} inverso para devolver la coherencia de fase. Luego, el gradiente G_{PE} es encendido con la mayor amplitud y se apaga dejando un desfase permanente. En ese momento, un nuevo pulso de RF de 180° es activado para rotar la magnetización 180° a la parte negativa del plano xy , donde los espines que se habían desfasado por efecto de $T2^*$ se volverán a poner en fase, produciendo un eco por la coherencia de fase. Justo ahí se encenderá el gradiente G_{FE} , con un lóbulo inverso previo para precorregir el desfase que genera su activación. Como resultado la señal recibida tendrá múltiples componentes de frecuencia con las fases generadas por G_{PE} . El paquete se repite n veces cambiando únicamente la fuerza del gradiente de fase. Al llenar todo el espacio

k, se repetirá todo el proceso para el siguiente corte, cambiando el ancho de banda de los pulsos de RF (Figura 1-13).

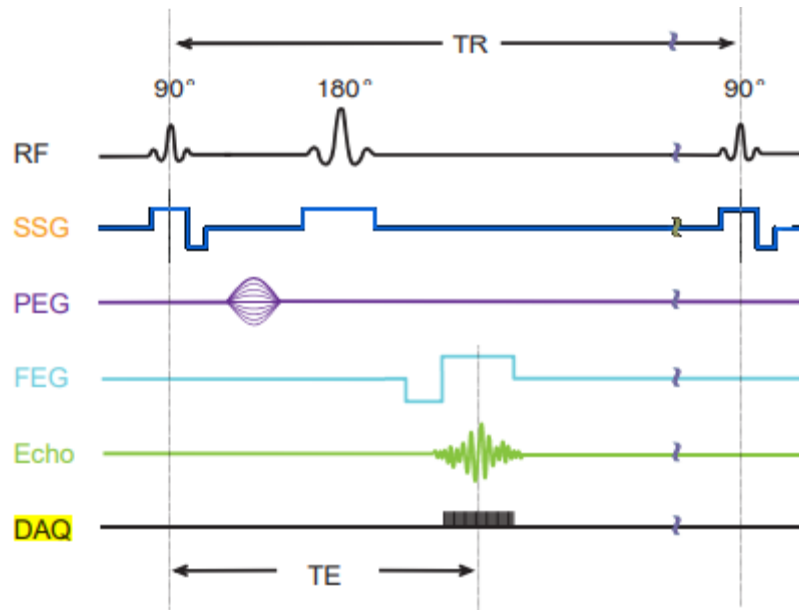


Figura 1-13: Diagrama de tiempo de pulsos para la secuencia Spin-Echo. Tomada de [18].

1.2.4 Imágenes pesadas por difusión (DWI)

Las imágenes pesadas por difusión (DWI, del inglés *Diffusion-Weighted Imaging*) reflejan el movimiento aleatorio de las moléculas de agua debido a la agitación térmica, conocido como movimiento Browniano, en los tejidos. La movilidad y, por lo tanto, la difusión es más restringida en tejidos con alta celularidad, ya que el espacio extracelular es reducido, como en los tejidos tumorales [19].

El flujo de las moléculas sobre una superficie en un tiempo determinado se conoce como coeficiente de difusión (D), medido en unidades de mm^2/s . Para generar la secuencia por difusión un par de gradientes simétricos son aplicados, uno a cada lado del pulso de 180° de una secuencia spin-echo de pulso individual (*single-shot*) pesada en T2. De esta forma, la fase acumulada por los espines de los átomos estacionarios durante el primer gradiente es reversada durante el segundo, por lo tanto, su señal no resulta afectada, mientras que, los espines del agua en movimiento si acumulan un cambio de fase debido al cambio de posición y, por lo tanto, su señal será más baja.

La fuerza y duración de los gradientes usados para las imágenes DW se refleja en el valor b [s/mm²]. Entonces, la señal (S) producida luego de los gradientes está descrita por la ecuación (1-10):

$$S = S_0 e^{-bD} \quad (1-10)$$

Donde, S_0 es la señal con los gradientes apagados, b_0 .

▪ **Coefficiente de difusión aparente (ADC)**

El ADC (del inglés *Apparent Diffusion Coefficient*) es la tasa de difusión medida u observada en un voxel.

La producción de las imágenes DW se da de acuerdo con la siguiente secuencia:

1. Se adquiere la secuencia de pulso descrita anteriormente con los gradientes dispersores apagados (b_0) o a muy bajo valor (b_{50}), obteniendo una imagen pesada en T2 que sirve como línea base.
2. Se adquiere la secuencia con los gradientes encendidos a lo largo de diferentes direcciones y en combinaciones de dirección e intensidad con el fin de recoger información sobre la difusión espacial.
3. Se calcula la imagen pesada por difusión como la media geométrica de las señales de difusión recogidas en las tres dimensiones.
4. Se calcula el mapa de ADC usando los datos de las imágenes DW y la imagen b_0 .

El mapa de ADC se crea dividiendo la señal de la imagen DW por la señal de cada punto correspondiente en la imagen b_0 para remover el pesaje en T2 y luego aplicando el logaritmo. En la imagen del mapa de ADC, los valores bajos que corresponden a zonas con restricción de la difusión (p. ej. Abscesos, isquemia, lesiones tumorales, etc.) aparecen oscuros, mientras que, valores grandes de ADC que corresponde a zonas libres de difusión aparecen brillantes, contrario a la imagen DW.

1.3 Radiomics

La radiómica (*radiomics*) consiste en la extracción de gran cantidad de información cuantitativa de las imágenes radiológicas digitales (CT, MRI, PET, etc.) a través de la aplicación de algoritmos de caracterización de la distribución espacial de la intensidad de los píxeles que conforman la imagen y su interrelación [25]. El objetivo del análisis radiómico es aportar información adicional de valor, imperceptible para el ojo humano durante la lectura radiológica, con el fin de mejorar el diagnóstico y la conducta de tratamiento de los pacientes, hacia un enfoque de medicina personalizada. En oncología, particularmente, radiomics tiene gran potencial debido a que es capaz de caracterizar la heterogeneidad del volumen tumoral completo, lo cual no es posible con la biopsia. Las características cuantitativas de la imagen son biomarcadores con capacidad para detectar y caracterizar una patología, e incluso predecir la respuesta al tratamiento, como lo presentan decenas de artículos publicados en la literatura científica en la última década [26], [27]. Cada característica o atributo extraído corresponde, desde el punto de vista estadístico, a una dimensión dentro de un conjunto de datos, por lo tanto, radiomics convierte una imagen digital en un conjunto de datos de alta dimensionalidad [25].

1.3.1 Flujo de trabajo en Radiomics

El flujo de trabajo para el análisis radiómico generalmente consiste en cinco pasos principales: la adquisición de la imagen, la segmentación de la región/volumen de interés (ROI/VOI), la extracción de características, la selección de características y la construcción y validación de un modelo. Cada uno de estos pasos presenta retos y limitaciones que pueden afectar su actuación [28]. Por esta razón es necesario estandarizar la implementación de un flujo radiómico robusto para garantizar la consistencia del modelo predictivo. Una iniciativa para el desarrollo, estandarización y optimización de protocolos para la implementación de imágenes cuantitativas es la Alianza de Biomarcadores de Imágenes Cuantitativas (QIBA, del inglés *Quantitative Imaging Biomarkers Alliance*) impulsada por la Sociedad Norteamericana de Radiología (RSNA, por sus siglas en inglés) desde el año 2007 [29].

1.3.2 Adquisición de la imagen

La adquisición de la imagen determina la calidad de los datos que entran al proceso, muchos factores impactan este paso: modelo de equipo, parámetros de adquisición, parámetros de reconstrucción de la imagen, preparación del paciente, etc. Entonces, un reto para la investigación multicéntrica es la homogeneización de parámetros para reducir el sesgo por factores procedimentales y la posterior armonización de las imágenes para reducir el sesgo por factores técnicos propios de los equipos [30]. En general, se recomienda establecer un protocolo de adquisición de imágenes que garanticen la calidad requerida y realizar controles de calidad periódicos a los equipos para monitorear su reproducibilidad.

1.3.3 Segmentación del ROI/VOI

Este paso es crítico debido a que en él se define la extensión sobre la cual se extraerán las características radiómicas. Diferentes técnicas manuales, semiautomáticas y automáticas han sido propuestas de acuerdo con la modalidad de imagen y región anatómica. La delineación manual es relativamente sencilla, pero requiere experticia en la tarea y consume mucho tiempo en grandes volúmenes de datos. Además, es más susceptible a la variabilidad inter- e intra-observador. La robustez frente a cambios en la segmentación siempre debe ser evaluada.

1.3.4 Extracción de características radiómicas

Este paso incluye el procesamiento de las imágenes (normalización, filtrado, agrupamiento por escala de grises, etc.) y la aplicación de los algoritmos matemáticos simples o con filtros para la extracción de las características radiómicas de un ROI/VOI. Dado que existen varias formas y fórmulas para el cálculo, se recomienda seguir un estándar como el publicado por Zwanenburg, A. *et al.* "Iniciativa de estandarización de biomarcadores de imagen (IBSI, del inglés *Image Biomarker Standardisation Initiative*) [31], con el fin de garantizar la reproducibilidad.

Para calcular algunas características se requiere primero discretizar la escala de grises de la imagen. Existen dos métodos según la definición del rango mínimo discreto (*bin*):

- **Ancho de bin fijo:**

$$X_{d,k} = \left\lfloor \frac{x_g - \min(x_g)}{w_b} \right\rfloor + 1 \quad (1-11)$$

- **Número de bins fijo:**

$$X_{d,k} = \begin{cases} \left\lfloor N_b \frac{x_g - \min(x_g)}{\max(x_g) - \min(x_g)} \right\rfloor + 1 & X_{d,k} < X_{g,k,max} \\ N_b & X_{d,k} = X_{g,k,max} \end{cases} \quad (1-12)$$

Donde x_g y $X_{d,k}$ son el valor de la intensidad de la escala de grises antes y después de la discretización, respectivamente. N_b es el número de bin y w_b es el ancho del bin.

Las características radiómicas principales pueden agruparse en cuatro categorías:

- **Características de forma:**

Están basadas en la forma geométrica de la ROI, entre ellas están, por ejemplo: el área superficial, el volumen, la compactibilidad.

- **Características de intensidad:**

También denominadas características estadísticas de primer orden. Están basadas en un histograma de la intensidad del píxel en la ROI sin tomar en cuenta como se relaciona con los píxeles vecinos. Entre ellas están, por ejemplo: el máximo, el mínimo, la curtosis.

- **Características de textura:**

Las características basadas en la textura, también denominadas “características estadísticas de segundo orden”, se calculan de la relación entre píxeles/vóxeles vecinos. Estas son de especial relevancia para determinar la heterogeneidad intratumoral. En la Figura 1-14 se presenta un ejemplo de tres grupos de características de este tipo. La matriz de co-ocurrencia de niveles de grises (GLCM, del inglés *Gray-Level Co-occurrence Matrix*) cuantifica el número de veces que se repite cierta combinación de intensidad de píxeles/vóxeles. La matriz de secuencia de niveles

de grises (GLRLM, del inglés *Gray-Level run-length*) cuantifica los píxeles/vóxeles consecutivos con la misma intensidad a lo largo de direcciones fijas. La matriz de tamaño de zona de niveles de grises (GLSZM, del inglés *Gray-Level Size Zone Matrix*) cuantifica los píxeles/vóxeles conectados en todas las direcciones.

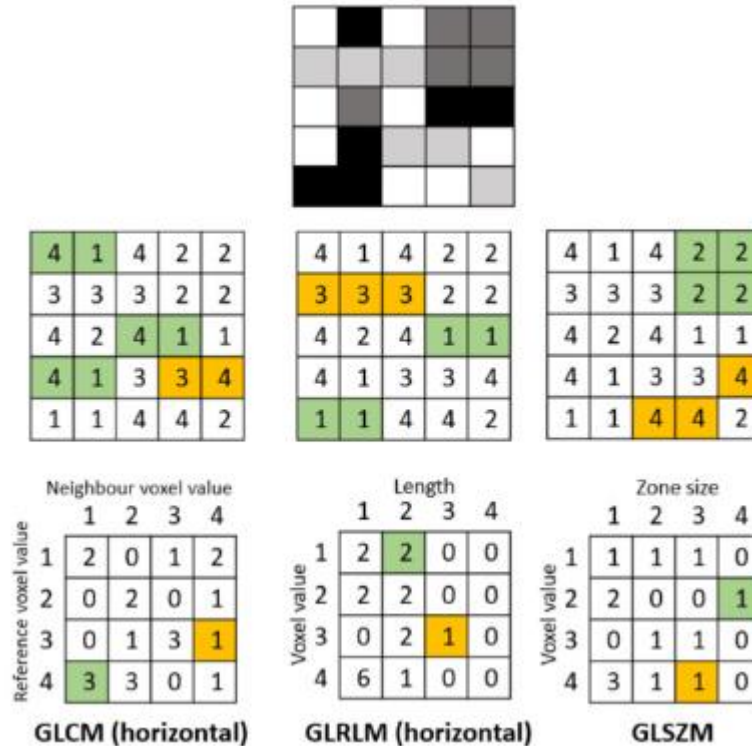


Figura 1-14: Ejemplo de cálculo de las características radiómicas texturales. En la 1era fila, la matriz de la imagen adquirida en escala de grises. En la 2da fila, la matriz con los valores discretizados de la intensidad. En la 3era fila, se presentan tres diferentes algoritmos de cálculo de grupo de características GLCM, GLRLM y GLSZM formadas a partir de la matriz discretizada. En amarillo y verde se representa las forma de llenado de cada matriz. Tomada de [32].

- **Características basadas en filtros:**

Para la extracción de estas características diferentes filtros son usados para preprocesar la imagen, entre ellos están los filtros: *Wavelet*, *Laplaciano del Gaussiano* (LoG), y algunos filtros simples como cuadrado, raíz cuadrada, logarítmico y exponencial. El LoG captura información textural a diferentes niveles, con una eficiencia excepcional en áreas de límites difusos. Mientras que, el Wavelet revela características sutiles de las características originales.

1.3.5 Selección de características

Debido a la gran cantidad de características extraídas por radiomics, hasta más de 1000 según la cantidad de algoritmos aplicados, es necesario implementar diferentes técnicas con el fin de seleccionar únicamente las características más robustas y reproducibles, eliminando las características redundantes que no aportan información nueva al modelo. Una limitante para la generalización de los modelos radiómicos es que las cohortes o grupos de datos suelen ser pequeños comparados con el número de características evaluadas, lo cual conduce a sobreajuste.

1.3.6 Construcción y validación del modelo

La construcción del modelo dependerá del objetivo que se persigue y los datos disponibles, puede ser un modelo de clasificación predictivo o discriminatorio. Si los datos son etiquetados, es decir, tienen asignada la variable de salida, es posible implementar un modelo de aprendizaje automático (*machine learning*) del tipo supervisado, por ejemplo: una base de datos de imágenes de animales en las cuales se ha etiquetado si la imagen corresponde a un perro o un gato; o como en nuestro caso, una base de datos clínicas etiquetada con el resultado de la respuesta al tratamiento. En el caso contrario, datos no etiquetados, se trata de un modelo no supervisado.

Finalmente, antes de que el modelo pueda ser aplicado se requiere una validación para evaluar su estabilidad y reproducibilidad, así como testear si es generalizable a nuevos datos o se ha sobreajustado a los datos de entrenamiento. La validación se realiza comúnmente dividiendo el conjunto de datos en dos grupos, conjunto de entrenamiento y conjunto de validación, de tal forma que el modelo se entrena con el primero y luego se evalúa en el segundo. Una proporción recomendada para la partición es 80-20, con datos balanceados. Otra alternativa es validar el modelo en un conjunto externo de datos nuevos, pero, obtener acceso a un gran volumen de datos no es sencillo en la práctica clínica. Por otra parte, si el número de pacientes en el conjunto de datos es muy pequeño, una alternativa es usar el método de “dejar uno afuera” (LOOCV, del inglés *Leave-One-Out Cross-Validation*), el cual entrena el modelo con

todas las filas de datos disponibles menos uno ($n - 1$), el cual se usa para la validación y se repite el proceso n veces hasta pasar por todos los datos.

2. Materiales y Métodos

2.1 Conjunto de datos

2.1.1 Diseño de estudio

El conjunto de imágenes y registros médicos usado en este estudio fue obtenido del protocolo prospectivo “APORTE DE MAMMI-PET, PET/TC y RMN EN EL MANEJO DEL CÁNCER DE MAMA LOCALMENTE AVANZADO”, un proyecto de colaboración multicéntrica para la investigación entre la Fundación Centro Diagnóstico Nuclear (FCDN) y el Instituto de Oncología Ángel H. Roffo, de la ciudad de Buenos Aires, Argentina, en el cual 50 pacientes femeninas con diagnóstico de cáncer de mama localmente avanzado, indicadas con tratamiento de quimioterapia neoadyuvante, fueron sometidas a la adquisición de imágenes radiológicas de 18F-FDG-PET/CT y Resonancia Magnética Multiparamétrica de mama previo al inicio y posterior al fin del tratamiento, entre enero de 2018 y diciembre de 2019. Todas las imágenes fueron adquiridas en las instalaciones de FCDN. Los resultados de los exámenes histopatológicos de los especímenes resecados quirúrgicamente fueron usados como referencia para etiquetar los datos en dos grupos: respuesta patológica completa (pCR) y no respuesta completa (no-pCR), utilizando los criterios propuestos por el centro MD Anderson Cancer Center [33]. El estudio fue conducido de acuerdo con la declaración de Helsinki y aprobado por el comité de ética del Instituto Roffo, el consentimiento informado firmado para el uso de las imágenes fue obtenido de todas las participantes.

2.1.2 Criterios de inclusión y exclusión de pacientes

Los criterios de inclusión para la selección de las 50 pacientes fueron los siguientes: (i) paciente mayor de 18 años con diagnóstico de cáncer de mama localmente avanzado

probado por biopsia sin metástasis distantes, (ii) indicación de tratamiento con quimioterapia neoadyuvante, (iii) no gestante ni lactante y, (iv) que el tamaño de la lesión primaria fuera superior a 1.5 cm para reducir el impacto del efecto de volumen parcial en la cuantificación. De ellas, 20 pacientes fueron excluidas debido a uno o varios de los siguientes criterios de exclusión: (i) la paciente fue sometida a la biopsia o alguna fase del proceso en una institución externa y la información no está disponible (ii) la paciente no terminó el tratamiento de QNA (2 pacientes fallecieron y 6 tuvieron que ser intervenidas antes de finalizar el tratamiento por progresión de la enfermedad), (iii) la paciente fue sometida a cirugía en una institución externa y la información no está disponible, o la pCR no fue evaluada, (iv) no se sometió a imágenes pretratamiento de resonancia magnética o PET, y (v) la calidad de las imágenes es pobre y no permite cuantificar correctamente. Finalmente, 30 pacientes fueron incluidas en este estudio.

2.1.3 Adquisición de las imágenes

Las imágenes PET fueron adquiridas en un tomógrafo PET/CT modelo Discovery 710, del fabricante General Electric (GE Healthcare, USA). Las pacientes recibieron instrucciones de preparación para el estudio de acuerdo con la segunda edición de las “guías para imágenes de tumores usando ^{18}F -FDG PET/CT” de la Asociación Europea de Medicina Nuclear (EANM, por sus siglas en inglés) [34]: ayuno de al menos 6 horas antes de la administración del radiofármaco, no realizar ejercicio al menos 24 horas antes, seguir una dieta estricta el día del estudio (p. ej. Comidas ricas en proteínas y vegetales y bajas en carbohidratos, no consumir azúcar, no ingerir bebidas con cafeína), entre otras medidas. El ^{18}F -FDG, cuya actividad fue estimada por relación cuadrática con el peso del paciente [35], fue administrado vía intravenosa en la vena cubital contralateral al tumor principal. El nivel de glucemia fue validado dentro de un rango de 80 mg/dL y 160 mg/dL antes de la inyección.

Las pacientes fueron sometidas al escaneo de PET en respiración libre (es decir, sin apnea ni gatillado), 60 minutos después de la inyección, en posición supino con los

brazos arriba de la cabeza, barriendo desde el vertex hasta el tercio medio del muslo, con un tiempo de adquisición de 2 minutos/cama. Una imagen adicional de la región torácica fue adquirida en posición prono con los brazos arriba de la cabeza usando una almohadilla de espuma para el posicionamiento y confort de las pacientes, de tal forma que los senos quedaran pendulares o colgantes. Previo al escaneo de cada PET, una imagen de CT de baja dosis en respiración libre (espesor de corte de 2.5 mm, 140 kVp) fue realizada para la corrección de atenuación y la fusión. Todas las imágenes PET fueron reconstruidas usando el algoritmo de reconstrucción iterativa 3D tipo OSEM, VUE Point (GE Healthcare), con correcciones por tiempo de vuelo ("FX") y modelado de la resolución espacial PSF ("SharpIR"), denominado con las siglas VPFXS. Todas las correcciones cuantitativas fueron activadas para la reconstrucción (atenuación, tiempo muerto, normalización, aleatorios, dispersión, etc.). Los parámetros de reconstrucción usados fueron 2 iteraciones y 24 subgrupos, con matrices de 192×192 voxels y un espesor de corte de 3.27 mm; seguido de un filtro de suavizado Gaussiano de 8.9 mm de ancho completo a media altura (FWHM, del inglés *Fullwidth-at-Half-Maximum*). Tales parámetros de reconstrucción fueron acreditados por el programa internacional EANM Research Ltd. (EARL) para FDG-PET/CT [36].

Por su parte, todas las imágenes de resonancia magnética fueron adquiridas en un resonador modelo Ingenia 3.0 Tesla (Philips Medical Systems, Best, The Netherlands). Las pacientes fueron instruidas para no aplicarse crema ni perfumes sobre el área de los senos, axilas ni brazos antes del estudio para evitar posibles artefactos en la imagen. El escaneo de resonancia fue realizado en posición prono con la bobina corporal como transmisor y una bobina dedicada del tipo arreglo de fase (*phased-array*) de 7 canales (dStream Breast 7ch, Philips Healthcare), cubriendo ambas mamas, como receptor. El protocolo de adquisición fue un protocolo estándar usado en la práctica clínica para explorar lesiones mamarias, incluyendo imágenes: axial T2W, axial T2 SPAIR, axial DWI ($b = 0, 400, 1100$ s/mm²), T1W 3D y una secuencia 3D dinámica pesada en T1 con saturación grasa, con realce de contraste (DCE, del inglés *Dynamic Contrast-Enhanced*), compuesta por una imagen pre-contraste y 6 imágenes

($t = 1, 2, 3, 4, 5$ y 6 minutos) después de la inyección intravenosa en bolo de gadobutrol (*Gadovist*; Bayer HealthCare, Germany) a 0.1 mmol/kg de masa corporal, administrado por una bomba de inyección a una tasa de 2.0 mL/s seguido por 20 mL de solución salina. Los parámetros de adquisición son detallados en la Tabla 2-1. El resonador fue sometido a controles de calidad mensuales (resolución espacial, perfil de corte, uniformidad de campo y linealidad espacial) siguiendo los protocolos del fabricante y del NEMA (siglas en inglés de *National Electrical Manufacturers Association*) [37], [38] durante la vigencia del proyecto.

Tabla 2-1: Parámetros de adquisición para el protocolo de imágenes de resonancia magnética de las mamas usado en el conjunto de datos.

Características	Secuencias				
	T2W	T2 SPAIR	DWI	T1W	DCE
Plano	Axial	Axial	Axial	Axial	Axial
Tipo adquisición	2D	2D	2D	3D	3D
Secuencia	SE	SE	SE-EPI	GRE (SP)	GRE (SP)
Corte (mm)	3	3	3	2	2
TR (ms)	3300	3500	11000	4.5	4.3
TE (ms)	120	80	86.8	2.2	2.1
ETL	30	20	55	43	38
Saturación grasa	NO	SPAIR	SPIR	NO	SPAIR
Ancho de banda (Hz/pixel)	291	291	2017	433	433
Ángulo de giro (°)	90	90	90	10	10
NSA	1	1	2	1	1

Abreviaciones: SP Spoiled, ETL Longitud del tren de ecos (Echo train length).

2.1.4 Preprocesamiento de las imágenes

Las imágenes de PET fueron convertidas automáticamente a valores de SUV normalizado con respecto a la masa corporal total directamente en la estación de trabajo (*Workstation*, GE Healthcare). Del mismo modo, los mapas de ADC para cada uno de los valores b fueron derivados de la imagen de difusión automáticamente en la consola principal del resonador magnético (Philips Healthcare).

2.1.5 Marcadores biológicos

Los datos de los marcadores biológicos: receptor de estrógeno (RE), receptor de progesterona (RP), HER2 y el índice Ki67 (marcador de proliferación celular) obtenidos de los resultados de la inmunohistoquímica al tejido biopsiado pretratamiento fueron suministrados por las mastólogas del Instituto Roffo. Para su evaluación los valores de RE y RP ≥ 1 , HER2 ≥ 1 y Ki67 $\geq 15\%$ fueron considerados positivos.

2.1.6 Variable principal

De la anatomía patológica del espécimen resecado en la cirugía post quimioterapia neoadyuvante la carga tumoral residual (RCB, del inglés *Residual Cancer Burden*) fue obtenida, en cuatro niveles, de 0 a 3, según la presencia de células cancerosas. Este valor fue usado para categorizar y etiquetar a los sujetos del estudio en dos grupos: respuesta patológica completa (pCR, cuando RCB=0) y no respuesta patológica completa (No-pCR, cuando RCB > 0). Entonces, la pCR fue definida como la ausencia de cáncer invasivo en el examen histopatológico del espécimen quirúrgico de la mama.

2.1.7 Base de datos

Una base de datos relacionando la información de los marcadores biológicos extraídos de la biopsia de cada paciente fue creada y etiquetada con los resultados de la respuesta patológica. Luego de ello los datos fueron desidentificados indexando la fila con un código consecutivo. Del mismo modo, todas las imágenes fueron anonimizadas y renombradas con el código consecutivo respectivo.

2.2 Segmentación de las regiones de interés

2.2.1 Segmentación de las imágenes de PET

La segmentación automática de las imágenes en PET es un problema abierto y vigente, debido a que por la resolución espacial y el efecto de volumen parcial que difumina los bordes es difícil establecer cuál es el contorno del tumor. En el caso de las lesiones mamarias, por las características de los tejidos que coexisten, muchas veces no existe

una correlación morfológica de los bordes del tumor en las imágenes de CT, lo cual hace más confusa definición del contorno, como en el caso de la Figura 2-1.

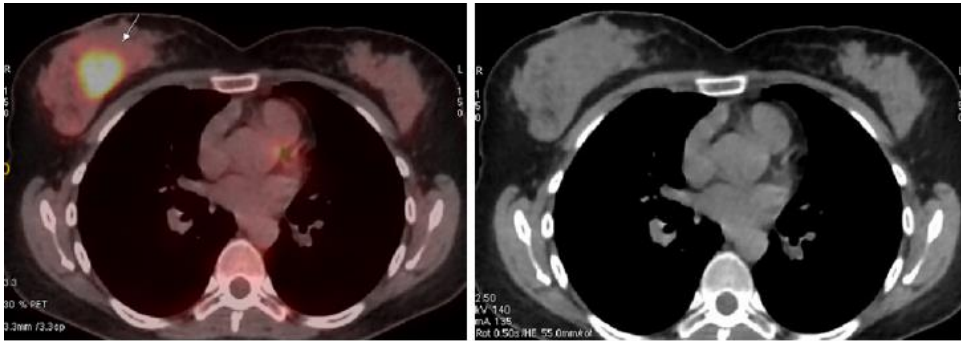


Figura 2-1: Ejemplo de foco tumoral captante sin correlación morfológica clara entre el PET y el CT debido a la alta densidad del parénquima mamario. A la izquierda la imagen fusionada de PET/CT y a la derecha la imagen de CT.

Muchos métodos han sido propuestos en los últimos años, con algoritmos de segmentación automáticos y semiautomáticos, una revisión del estado del arte se presenta en el reporte Task Group No. 211 de la AAPM [39]. Entre los métodos semiautomáticos más comúnmente utilizados se encuentra la segmentación basada en umbral fijo, por ejemplo, $SUV_{umbral}=2.5$ g/mL, dado que es el valor de referencia publicado en la literatura científica para diferenciar entre lesiones benignas y malignas en varias patologías, sin embargo, dicho criterio debe ser usado con especial cuidado ya que, como se explicó anteriormente, existen múltiples factores que afectan el valor medido de SUV. Otro método de umbral fijo consiste en usar arbitrariamente un porcentaje del SUVmax, por ejemplo, el 40% del SUVmax es un valor de uso extendido [39]. Sin embargo, los métodos de umbral fijo son susceptibles en gran medida a la relación señal a ruido (SNR) de la imagen, la relación de la intensidad tumor-fondo y el tamaño del tumor. Una alternativa son los métodos basados en umbrales adaptativos, los cuales minimizan estas dependencias al contemplarlas dentro del cálculo del umbral para cada lesión, con la limitante de que los parámetros para su implementación requieren ser calibrados para cada equipo PET, lo cual puede significar dificultades para la reproducibilidad [40].

En este estudio se ha propuesto implementar el método publicado por Nestle *et al.* [41], un método semiautomático basado en un algoritmo de umbral adaptativo que consiste en definir un valor de intensidad en el intervalo entre el fondo y el objeto según el nivel de captación del radiofármaco, todos los vóxeles con intensidad mayor o igual al umbral se incluirán como parte del ROI/VOI, como se describe en la ecuación (2-1).

$$T_{NESTLE} = \beta \cdot I_{70} + I_{fondo} \quad (2-1)$$

Donde T_{NESTLE} corresponde al umbral estimado, β es el parámetro de calibración específico para el equipo PET usado, que depende de las características técnicas del equipo, así como los parámetros de adquisición y reconstrucción de las imágenes; I_{70} es la intensidad promedio en una región definida por todos los vóxeles con intensidad mayor al 70% de la intensidad máxima; e I_{fondo} es el promedio de la intensidad en una capa de 2 vóxeles de espesor ubicada a una distancia de 6 vóxeles de la región definida por I_{70} . Todos los vóxeles con captación SUV >2.5 g/mL son excluidos del fondo.

Un coeficiente $\beta = 0.295$ fue encontrado óptimo basado en la adquisición de un fantoma NEMA de calidad de imagen que incluye 6 esferas de volúmenes 26.522, 11.494, 5.575, 2.572, 1.150 y 0.523 mL de mayor a menor (ver Figura 2-2a), rellenas con una solución de F-18, con una razón de actividad esfera a fondo de 10:1. Las imágenes del fantoma fueron adquiridas en el mismo equipo PET/CT Discovery 710 usado para las imágenes de los pacientes que hicieron parte del estudio y reconstruidas con los mismos parámetros acreditados en el programa EARL. El porcentaje de error obtenido en la medición de los volúmenes de las esferas sobre las imágenes, respecto de la dimensión real usando $\beta = 0.295$ fue de -0.1, 0.9, 1.8, 2.0, -5.8 y -107.6 % de mayor a menor respectivamente (Figura 2-2b). Producto de este resultado se reafirma el criterio de exclusión de lesiones menores a 15 mm de diámetro.

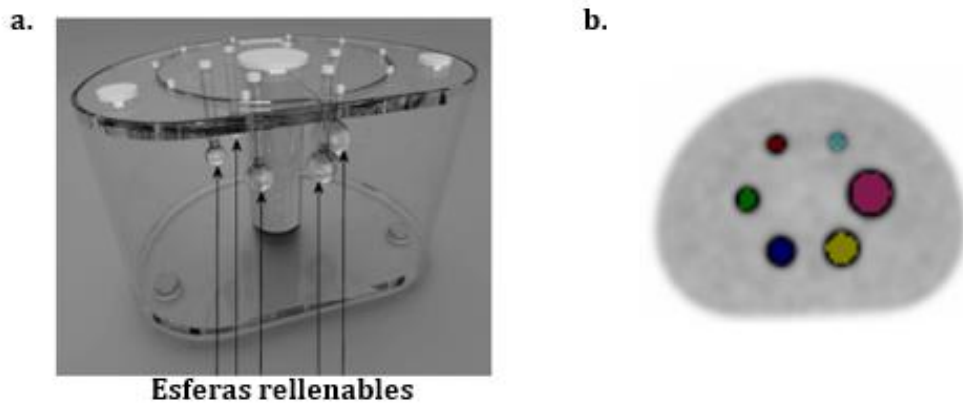


Figura 2-2: Adquisición del fantoma NEMA de calidad de la imagen para la determinación del factor β del método de segmentación implementado. **a.** Imagen del fantoma NEMA de torso con inserto de 6 esferas rellenas de 37, 28, 22, 17, 13 y 10 mm de diámetro. **b.** Imagen del corte axial del fantoma con la segmentación de las esferas usando el β determinado.

Para la implementación del método Nestle se usó el software libre LIFEx [42], en el cual primero se realiza un ROI que abarque todo el tumor y se introduce el valor de β como parámetro de entrada. Automáticamente el software calcula el valor umbral a partir del I_{70} y el I_{fondo} y ajusta el ROI. El umbral calculado fue implementado en el software 3D Slicer (versión 4.11.2021) [43] para obtener el volumen de segmentación. Los resultados fueron validados por una médico radiólogo, especialista en imagenología e intervencionismo mamario con 12 años de experiencia, con experiencia en lectura de PET/CT de mamas. Un caso de revisión fueron las lesiones con centro necrótico, ya que al segmentar por umbral la región necrótica queda por fuera del ROI (ver Figura 2-3). Por consenso (incluyendo el concepto de una mastóloga del Instituto Roffo, con más de 10 años de experiencia) se decidió incluir las áreas necróticas claramente definidas dentro de los límites del tumor debido a que representan parte de la heterogeneidad tumoral.

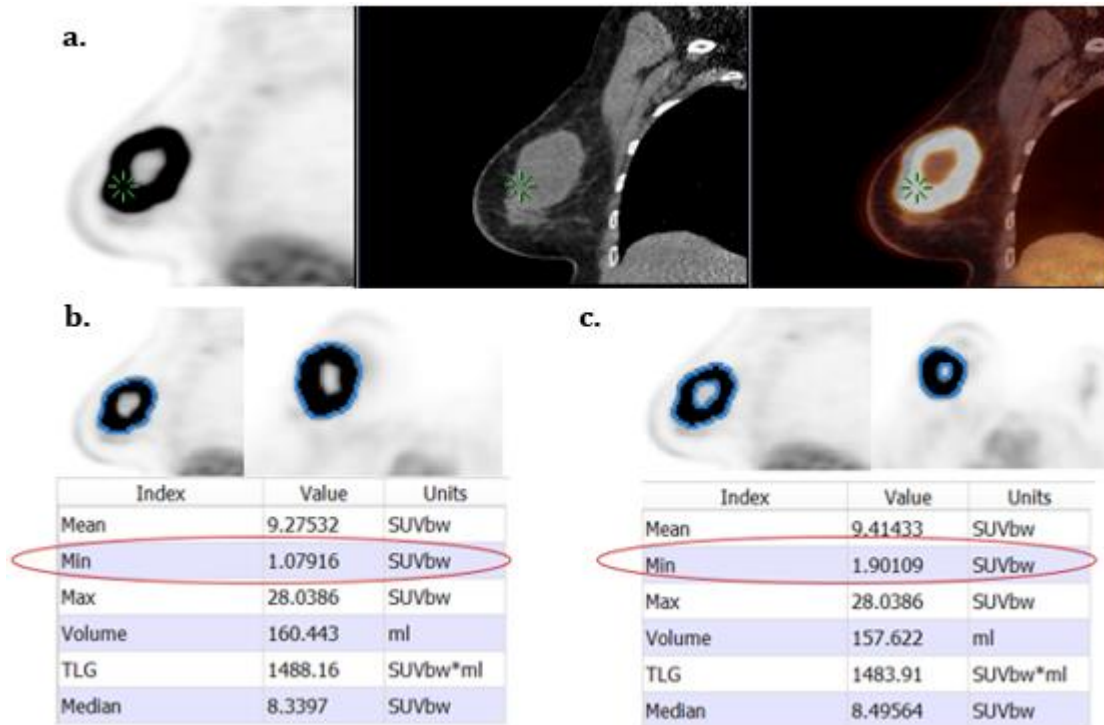


Figura 2-3: Imagen de volumen tumoral con centro necrótico y el impacto sobre la definición del umbral de segmentación en PET. **a.** De izquierda a derecha: imagen de PET, imagen de CT e imagen fusionada de PET/CT mostrando el contraste de una lesión con centro necrótico en el parénquima mamario. **b** y **c.** El valor de SUV del centro necrótico es menor que el valor umbral definido para la segmentación. La inclusión del área en el ROI impacta en algunas métricas.

2.2.2 Segmentación de las imágenes de Resonancia Magnética

Para la segmentación de las imágenes de resonancia magnética se probó inicialmente un método automático usando la extensión “NvidiaAIAssistedAnnotation” en el software 3D Slicer (versión 4.11), aplicado sobre la imagen con realce de contraste de 2 minutos post inyección. Sin embargo, debido a la ya mencionada variabilidad del contorno de los tumores mamarios (suaves, irregulares, espiculados, sin bordes definidos) muchas lesiones fueron incorrectamente delineadas. Por lo tanto, se decidió realizar segmentación manual de todas las imágenes de RM. Una médico radiólogo especialista en imagenología e intervencionismo mamario, con 12 años de experiencia, revisó y dibujó manualmente las ROI a lo largo de los márgenes visibles del tumor sobre las imágenes pesadas por difusión con b1100, para definir mejor los márgenes

del tumor y así aprovechar su alto contraste, usando el software 3D Slicer (versión 4.11). Posteriormente, las ROI fueron superpuestas sobre los mapas de ADC construidos con b1100 para la extracción de las características radiómicas. Adicionalmente, las imágenes: (i) pesada en T2 (T2W), (ii) dinámica DCE a t=2 minutos post contraste, (iii) Sustracción de la imagen en (ii) con la imagen antes de la inyección del contraste, fueron usadas como referencia para la confirmación de la localización del tumor en todos los casos (Figura 2-4).

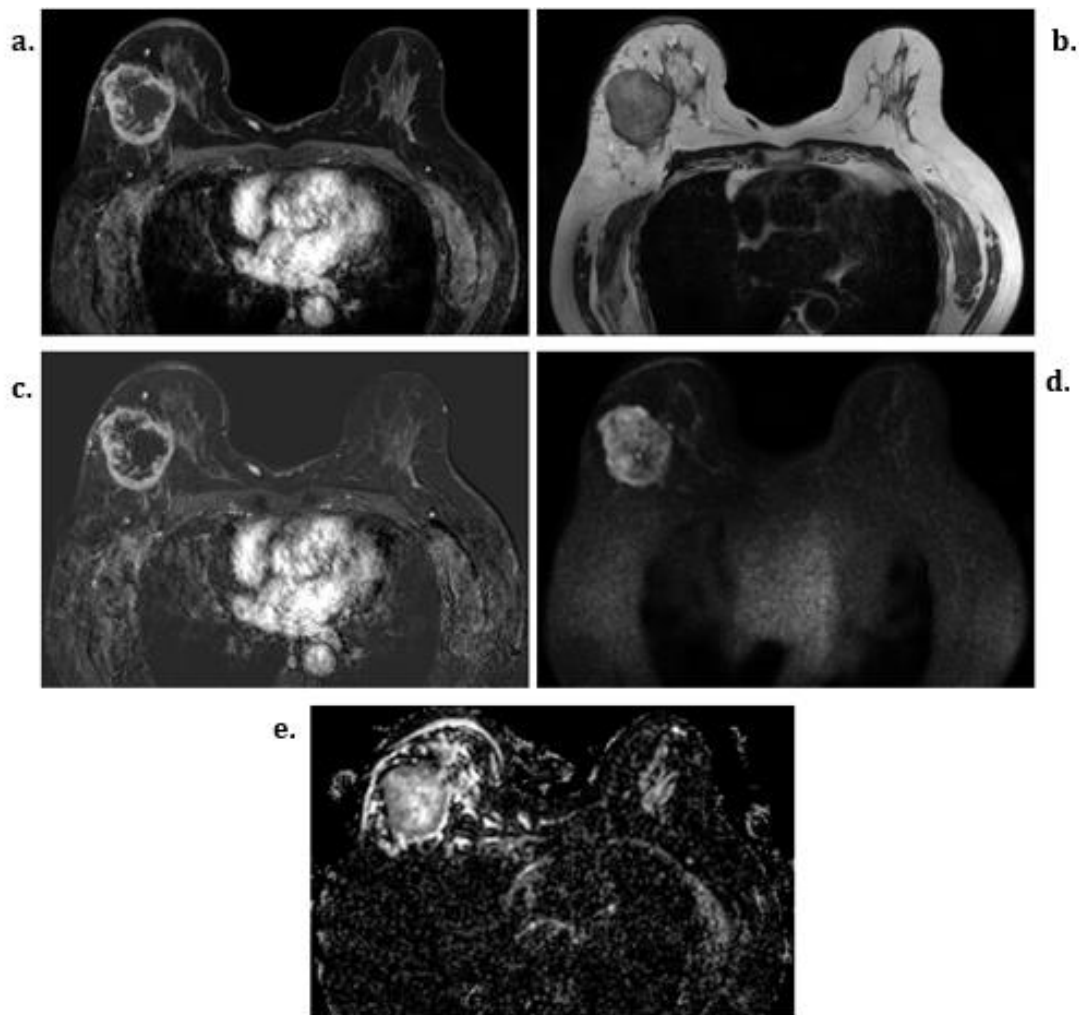


Figura 2-4: Imágenes de las secuencia de resonancia magnética usadas como referencia para la segmentación. **a.** Imagen DCE con t=2 min post contraste. **b.** Imagen T2W. **c.** Imagen de sustracción. **d.** Imagen DWI con b1100 y **e.** Imagen del mapa de ADC con b1100.

Durante la segmentación se evidenció que en las imágenes de 13 pacientes estaba presente el clip mamario usado para marcar el área de la biopsia, este dispositivo genera un artefacto metálico en las imágenes de resonancia magnética (ver Figura 2-5). Los artefactos fueron excluidos de la segmentación manteniendo una distancia de un píxel, incluso en las imágenes donde no había artefactos visibles, pero se conocía la ubicación del marcador.

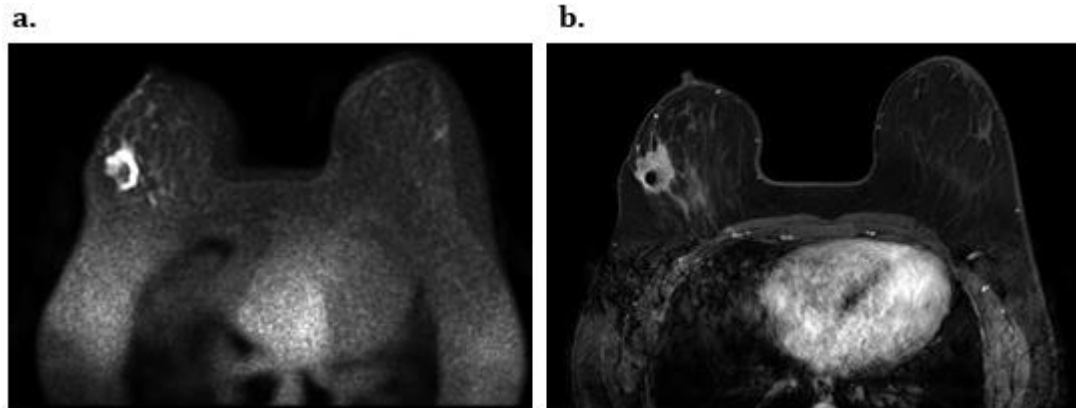


Figura 2-5: Imágenes de resonancia magnética con lesión en parénquima mamario derecho con clip de biopsia que genera artefacto metálico en torno a él. **a.** Imagen de DWI con b1100 y **b.** Imagen de DCE con t=2 min post inyección de contraste.

2.3 Extracción de características

Las características radiómicas fueron extraídas usando el paquete de código abierto *pyradiomics* [44], escrito en el lenguaje de programación *Python* y ejecutado en el software 3D Slicer (versión 4.11) por medio de la extensión “*Radiomics*”. La definición matemática de las características en *pyradiomics* [45] está acorde con la IBSI. Los parámetros configurados en 3D Slicer fueron: la inclusión de todas las características originales (107 en total), la cuantificación de las características texturales a partir de una transformación tipo wavelet en 3D (744 en total) y el forzamiento a la construcción de matrices simétricas para la extracción de la característica textural GLCM. Ningún preprocesamiento fue aplicado a las imágenes debido a que las características de textura tienden a ser sensibles a las perturbaciones en la intensidad local [46]. Por la misma razón, tampoco fue aplicada normalización alguna, la cual,

además, se considera innecesaria en este caso por tratarse de imágenes calibradas en valor de SUV[g/mL] en PET y en [mm²/s] en los mapas de ADC.

Un ancho de bin igual a 0.25 fue definido para el cálculo en las imágenes PET en posición prono, de acuerdo con los valores recomendados en la literatura científica [47], [48], que normalmente oscilan entre 0.1 y 0.5. Mientras que, para las imágenes de los mapas de ADC, tomando en cuenta que el rango promedio del valor de ADC fue [0, 2.72×10^{-3} mm²/s], un ancho de bin de 0.03 fue definido de tal forma que se produjera un total de 90 bins, en concordancia con la recomendación de manejar entre 8 a 128 bins [49].

Los marcadores metabólicos de PET (Valor de captación estandarizado máximo, promedio y pico [SUV_{max}], [SUV_{mean}] y [SUV_{peak}], así como, el volumen metabólico tumoral [MTV] y el nivel de glicólisis de la lesión [TLG]) fueron también computados en 3D Slicer usando la extensión “PET-IndiC” [50] sobre las imágenes adquiridas en la posición prono y en la posición supino.

2.4 Métodos de selección de características

Debido a la gran cantidad de características que son extraídas con la técnica radiómica es necesario reducir la dimensionalidad de los datos, de tal forma que el modelo use como entrada las características más relevantes y con la mínima redundancia entre sí. De esta forma se busca garantizar que el modelo sea generalizable, consistente y no se sobreajuste a los datos de entrenamiento. El proceso de selección de características fue implementado en *Python* 3.10.

2.4.1 Análisis del conjunto de datos

En primera instancia una prueba de normalidad multivariante de *Mardia* [51], basada en la distancia de Mahalanobis (una medida de la distancia entre un punto de datos y el centro del conjunto de datos multivariado) fue aplicada a la matriz formada por todas las características, con el fin de conocer que estadísticos son aplicables para su

análisis, cuyo resultado fue negativo. Además, una prueba de normalidad de D'Agostino y Pearson fue aplicada a cada característica de forma individual. Menos del 50% de las variables cumplieron el criterio de normalidad. Este dato fue tenido en cuenta en los análisis estadísticos realizados posteriormente.

2.4.2 Evaluación del valor predictivo individual de las características biológicas por biopsia

La asociación entre las características extraídas por biopsia y la respuesta patológica a la QNA fue evaluada a través de la gráfica de los marcadores biológicos respecto de los distintos tipos de respuesta patológica. El valor predictivo individual fue establecido a partir del análisis del área bajo la curva ROC.

2.4.3 Análisis de las características metabólicas de PET y evaluación de su valor predictivo individual

Los valores de SUVmax y de los volúmenes medidos para las lesiones segmentadas en las imágenes de PET en posición prono y en posición supino fueron comparados con el fin de establecer la variabilidad asociada al posicionamiento del paciente durante la adquisición. Adicionalmente, la capacidad individual de los marcadores metabólicos de PET para separar los grupos según la respuesta patológica fue evaluada.

2.4.4 Reducción de características de acuerdo con la reproducibilidad intra- e inter-observador

Una necesidad fundamental para lograr resultados confiables basados en marcadores radiómicos es que los valores sean estables ante diferentes condiciones y que las medidas obtenidas bajo las mismas condiciones sean consistentes. Algunas publicaciones han evaluado como se ve afectada la robustez por factores como: el escáner [52], los parámetros de adquisición de la imagen [53] y los observadores [54]. En este estudio la estabilidad intra- e inter-observador al momento de la segmentación de las imágenes fue evaluada a partir de la dilatación (aumento) y erosión (disminución) del tamaño de los ROI dibujados en un píxel, moviendo los bordes, con

el fin de simular la variabilidad [55]. Para las lesiones segmentadas en PET la dilatación y erosión fue realizada variando el valor β .

Entonces, el primer paso para reducir la dimensionalidad del conjunto de datos fue usar el Coeficiente de Correlación Intraclase (ICC, del inglés *Intraclass Correlation Coefficient*) (basado en una tasa de observadores $k=1$, de acuerdo absoluto, 2-factores modelo de efecto mixto) en los tres pares de combinaciones (original/dilatación, original/erosión, dilatación/erosión) para evaluar la estabilidad inter e intra-observador de las características. El ICC y sus intervalos de confianza al 95% fue calculado con el código desarrollado por Zoeller [56], usando el software MATLAB (versión R2021b) [57]. Una característica fue definida como altamente reproducible si el ICC(2,1) >0.9 [58], aquellas que no cumplieron con el criterio fueron excluidas.

2.4.5 Reducción de características por redundancia y máxima relevancia

En un segundo paso, la técnica de selección de características por correlación fue implementada. Para ello, una matriz de correlación fue construida usando el coeficiente de correlación de Spearman (ρ) como criterio de distancia. Cada par de características altamente correlacionadas ($\rho \geq 0.8$) fueron consideradas redundantes y, por lo tanto, una de ellas fue eliminada. Para tomar la decisión, el área bajo la curva ROC (AUC, del inglés *Area Under Curve*) respecto de la etiqueta de respuesta al tratamiento fue estimada para cada característica. Aquella con mayor distancia de la línea de referencia aleatoria (AUC=0.5) fue conservada por considerarse más relevante para el modelo.

2.4.6 Selección de las características más importantes

En un tercer paso, la reducción de características por relevancia fue implementada basada en tres enfoques diferentes: un método de envoltura, un método estadístico y un método de selección por modelo de aprendizaje automático. Los 3 fueron implementados usando la librería *Scikit-Learn* en *Python 3.10*

- **Enfoque 1: Algoritmo Boruta:**

El algoritmo Boruta, basado en un modelo de “bosques aleatorios” (*Random Forest*), fue implementado para el proceso de selección de características [59]. Este método genera características artificiales a partir de las características originales, las cuales son combinadas y evaluadas en un modelo clasificador. En él, las características originales con un nivel de importancia para la clasificación mayor que el de todas las características artificiales creadas asociadas fueron seleccionadas. Por defecto el algoritmo Boruta genera una lista de características de alta relevancia y otra de relevancia intermedia, en este estudio solo la primera lista fue usada para el análisis.

Random Forest se basa en el ensamblaje de múltiples modelos de árboles de decisión para obtener una predicción más robusta y precisa. La construcción de múltiples árboles de decisión y la agregación de sus predicciones tiende a reducir la varianza del modelo en comparación con un solo árbol de decisión. En la implementación del algoritmo de clasificación los parámetros configurados fueron: número de estimadores ($n_estimators=200$) y máxima profundidad del árbol ($max_depth=5$), estos fueron determinados por búsqueda de hiperparámetros a partir de una cuadrícula de valores la cual es evaluada hasta encontrar la mejor combinación posible, denominada “GridSearch”, los demás parámetros fueron establecidos por defecto. Durante la evaluación de importancia con el algoritmo de selección *BorutaPy*, un número de estimadores automático, el cual permite un mejor despliegue del modelo, con un máximo de 1000 iteraciones fue configurado.

- **Enfoque 2: Selección de características basado en el test Wilcoxon (WLCX):**

En este enfoque el test de Wilcoxon, prueba no paramétrica, fue usado para evaluar la importancia de las características en el conjunto de datos. El test de Wilcoxon no hace suposiciones sobre la distribución de los datos, por lo tanto, no requiere que sigan una distribución normal. Para su implementación, la diferencia entre las distribuciones de las características para las dos clases (pCR y No-pCR) fue evaluada según el test de Wilcoxon asignando una puntuación que representaba su nivel de importancia con

respecto a las etiquetas de clase. Este método es útil para seleccionar las características más relevantes para predecir la respuesta al tratamiento. Las 10 características más relevantes fueron seleccionadas.

▪ **Enfoque 3: Selección por modelo de aprendizaje automático:**

En este enfoque un modelo de aprendizaje automático tipo Random Forest fue entrenado para evaluar la importancia de cada característica en la predicción de la variable objetivo. Una búsqueda de hiperparámetros del tipo GridSearch fue implementada, con los siguientes resultados: mínimo de hojas por muestra (`min_samples_leaf=4`), mínimo de divisiones por muestra (`min_samples_split=2`) y número de estimadores (`n_estimators=200`). Las 10 características de mayor importancia para el modelo fueron extraídas para el análisis.

2.5 Construcción del modelo

Un modelo clasificador de aprendizaje automático supervisado fue implementado. El algoritmo de clasificación analiza los datos usados para entrenamiento e infiere una hipótesis de una función que caracteriza los datos, de tal forma que pueda predecir la clase o categoría de nuevos datos. En este estudio el algoritmo de “bosque aleatorio” (“*RandomForest*”) [60] fue propuesto usando la librería *Scikit-Learn* de *Python* 3.10.

Dos enfoques fueron usados también para la implementación del modelo, con una y dos variables, el rendimiento fue evaluado y comparado entre modelos entrenados. En ambos enfoques los siguientes pasos fueron ejecutados por el algoritmo:

1. Identificar el grupo de características que harán parte del modelo. En el caso de dos variables, primero se buscaron todas las combinaciones posibles de pares de características (grupos de 2 conformados a partir de las 10 variables).
2. Entrenar un modelo de clasificación para cada combinación de características utilizando *RandomForest* con validación cruzada del tipo *leave-one-out*.
3. Calcular las predicciones del modelo y las etiquetas reales para cada iteración de LOO.

4. Calcular las métricas de evaluación del modelo (precisión, sensibilidad (recall), F1-score, área bajo la curva ROC (AUC) y matriz de confusión) para cada característica.
5. La curva ROC para cada modelo fue graficada y una matriz de confusión fue construida con los resultados de las predicciones para estimar la especificidad, el valor predictivo positivo (VPP) y el valor predictivo negativo (VPN) del modelo.

2.6 Balanceo de datos usando ADASYN

Un experimento adicional con aumentación de datos fue llevado a cabo justo antes del entrenamiento del modelo predictivo. Dado que las clases en el conjunto de datos estaban desbalanceadas y los pacientes que no respondieron al tratamiento de QNA superaban en gran medida a aquellos que tuvieron respuesta patológica completa (22:8), la clase minoritaria fue aumentada en el conjunto de características más importantes hasta obtener un paciente menos que la clase mayoritaria usando el método de muestreo sintético adaptativo (ADASYN, del inglés *Adaptive Synthetic Sampling*) [61]. El método ADASYN genera muestras sintéticas para la clase minoritaria basada en la distribución de densidad local, lo que permite capturar mejor la distribución original del conjunto de datos. Los modelos predictivo del paso anterior fueron entrenados nuevamente con el fin de evaluar el efecto de la aumentación sobre el rendimiento.

2.7 Resumen del flujo de trabajo en radiomics

El flujo de trabajo implementado para el análisis radiómico se esquematiza en la Figura 2-6.

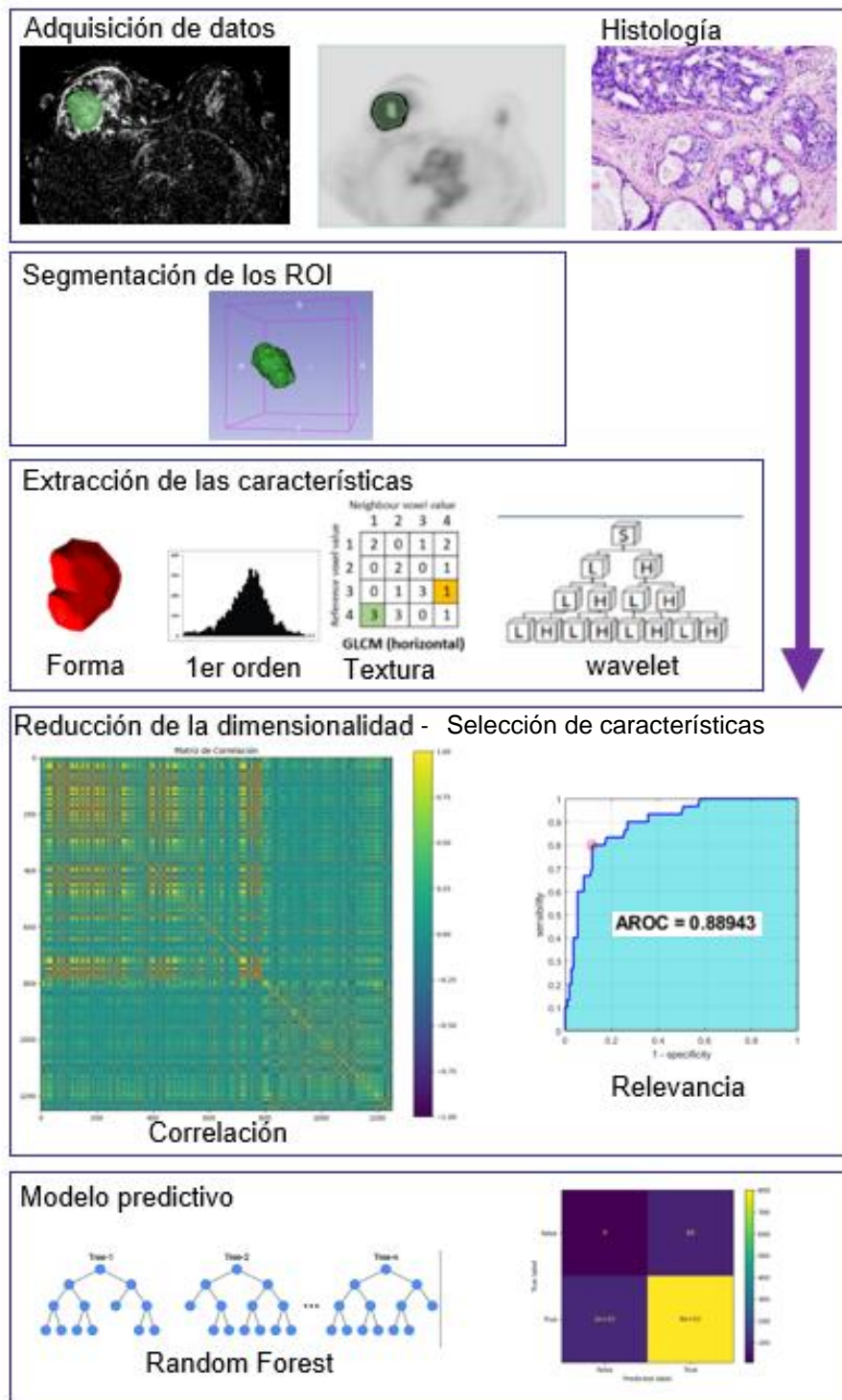


Figura 2-6: Flujo de trabajo del análisis radiómico implementado.

3. Resultados

3.1 Características clínicas y demográficas

De acuerdo con los criterios de inclusión y exclusión, 50 pacientes fueron enroladas inicialmente en el estudio, de las cuales 30 pacientes fueron incluidas en la muestra final. La mayoría de los tumores fueron del tipo histológico carcinoma ductal invasivo con el 70%, de los cuales el 23.8% alcanzó la pCR. Los datos demográficos se presentan en la Tabla 3-1.

Tabla 3-1: Comparación demográfica entre grupos con respecto a la pCR.

Características		Respuesta patológica (N=30)	
		pCR (n=8)	No pCR (n=22)
Edad (prom años ± desv.)		47.0 ± 10.4	46.4 ± 9.3
RE (%)	Positivo	3 (37.5)	17 (77.3)
	Negativo	5 (62.5)	5 (22.7)
RP (%)	Positivo	3 (37.5)	16 (72.7)
	Negativo	5 (62.5)	6 (27.3)
HER2 (%)	Positivo	5(62.5)	3(13.6)
	Negativo	3(37.5)	19(86.4)
Ki67 (%)	Positivo	7(87.5)	5 (22.7)
	Negativo	1(12.5)	17 (77.3)
Tipo	CDI	5(62.5)	16(72.7)
	CLI	3(37.5)	4(18.2)
	Otros		2(9.1)
	Luminal A	2(0.25)	4(18.2)
	Luminal B	0	10(45.5)
	L/H	3(37.5)	3(13.6)
	Triple Neg.	2(0.25)	4(18.2)
	HER puro	1(12.5)	1(4.5)

Ninguna lesión del subtipo luminal B alcanzó la respuesta completa al tratamiento. La combinación de mayor porcentaje representativo entre las lesiones que no alcanzaron la pCR fue RE y RP positivo, HER2 negativo y Ki67 negativo, que corresponde con la clasificación de los subtipos luminal A y luminal B.

Por otra parte, al relacionar los marcadores metabólicos de PET con los subtipos de cáncer de mama se evidenció que las lesiones triple negativo mostraron un SUV_{max} y SUV_{mean} significativamente más alto respecto de los luminales ($p < 0.001$ y $p < 0.002$), consistente con lo reportado por *Romeo et al.* en [62] (ver Figura 3-1). No hubo diferencias significativas entre las otras métricas metabólicas y los subtipos tumorales.

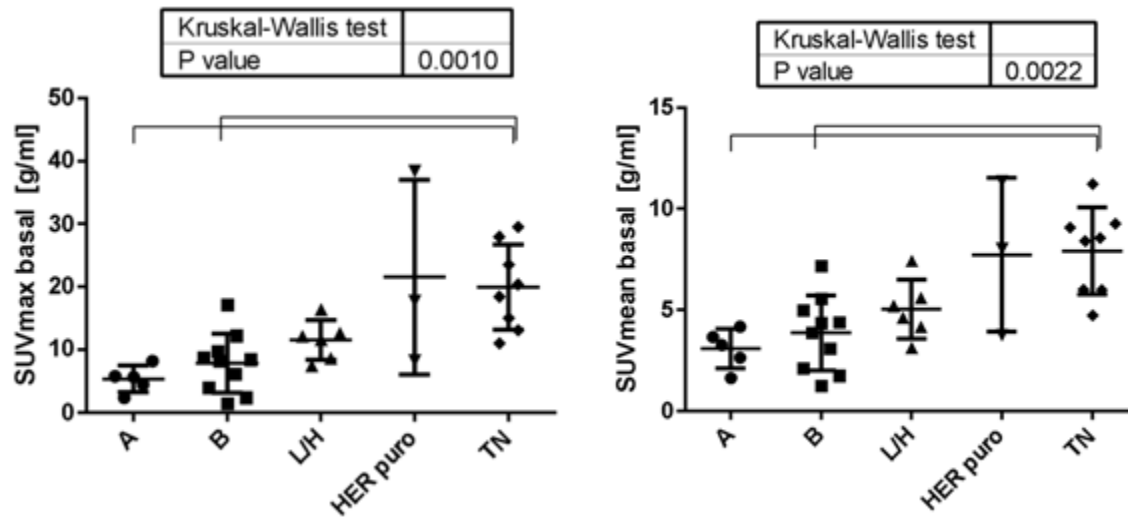


Figura 3-1: Resultados del análisis del SUV_{max} (Izquierda) y SUV_{mean} (Derecha) respecto del subtipo del cáncer de mama.

3.2 Análisis de características biológicas y metabólicas

3.2.1 Evaluación del valor predictivo individual de los marcadores biológicos

Una regresión logística multinomial fue realizada para evaluar el valor de los marcadores biológicos como predictores de la respuesta patológica. Un valor p de 0.05 fue considerado significativo. Un alto nivel de expresión de receptores de estrógeno fue significativamente asociado con la no respuesta patológica completa (ver Figura

3-2). Mientras que, no hubo diferencias significativas en la expresión de receptores de progesterona ni Ki67 entre los distintos tipos de respuesta patológica.

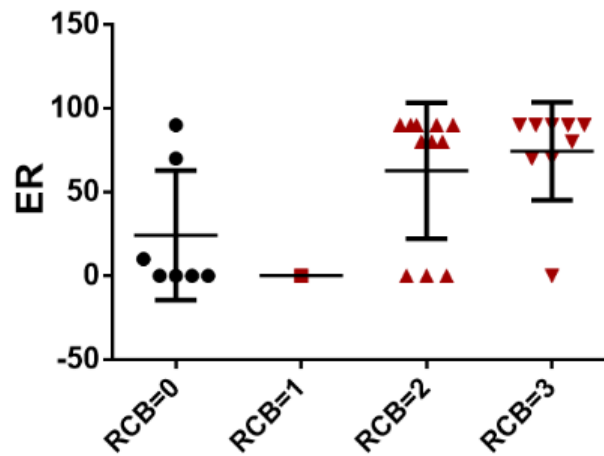


Figura 3-2: Resultado del análisis del nivel de receptor de estrógeno respecto de la respuesta al tratamiento.

3.2.2 Reproducibilidad de las métricas metabólicas de PET frente al posicionamiento del paciente

Los valores de SUVmax medidos en la posición prono versus la posición supino estuvieron altamente correlacionados (ver Figura 3-3). Una relación lineal directa existe entre ellos, un coeficiente $R^2=0.985$ y una pendiente $m = 0.93$ fueron obtenidos. Del mismo modo, los volúmenes medidos en cada posición fueron comparados (ver Figura 3-4) obteniendo un coeficiente $R^2=0.997$ y una pendiente $m = 1.01$. Por lo tanto, un alto grado de similitud entre la cuantificación de las imágenes en las dos posiciones fue observado.

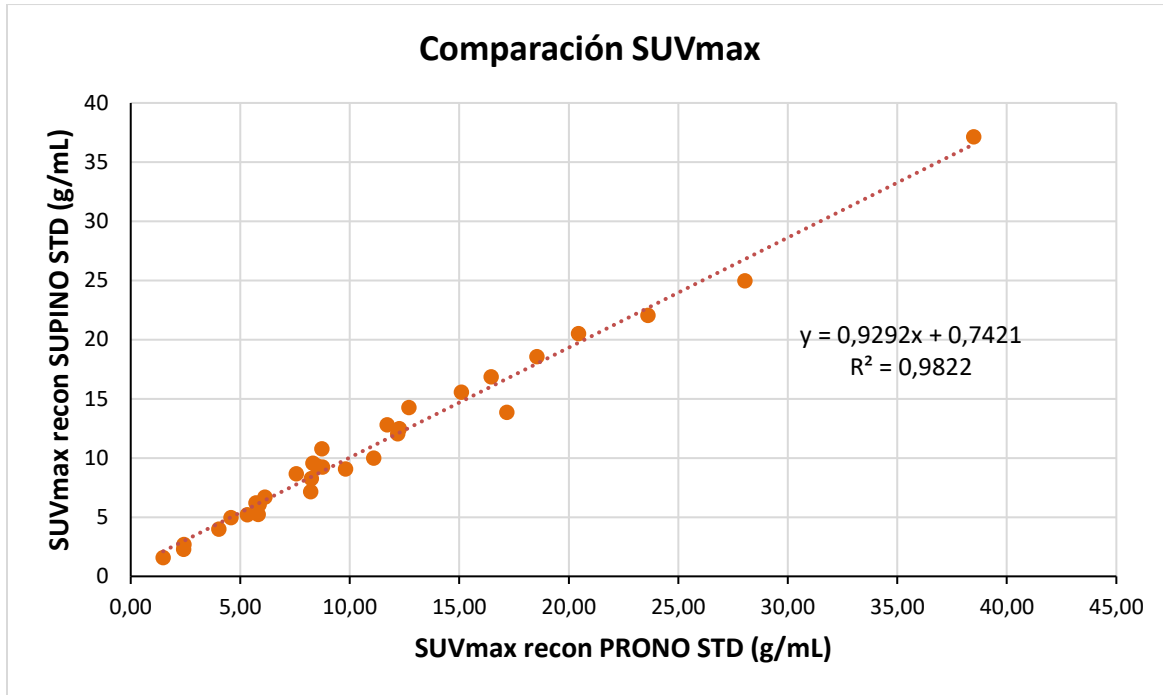


Figura 3-3: Comparación entre el valor de SUVmáx medido en posición supino y prono con iguales parámetros de reconstrucción (STD: estándar).

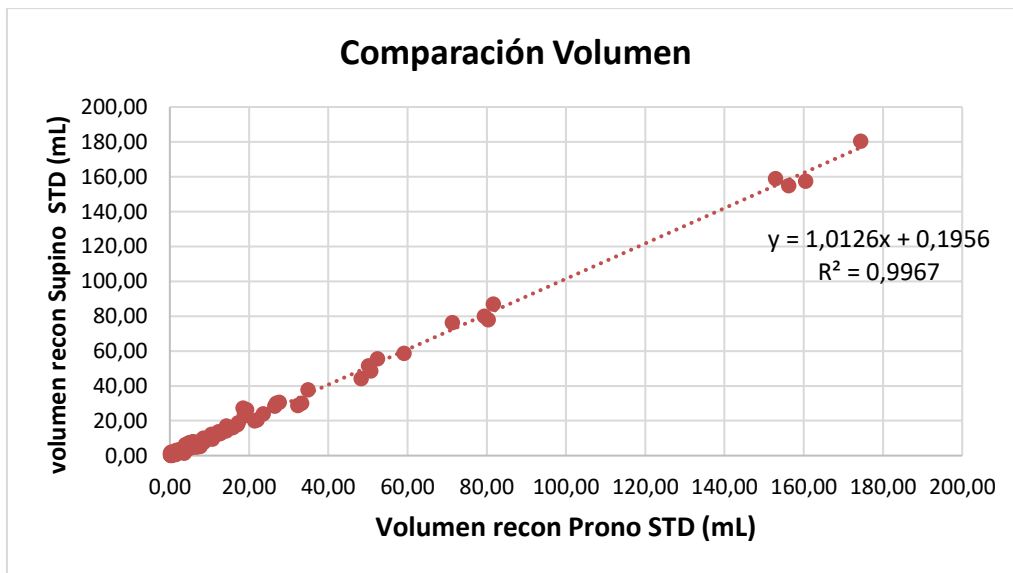


Figura 3-4: Comparación entre el valor del volumen medido en posición supino y posición prono con iguales parámetros de reconstrucción (STD: estándar).

3.2.3 Evaluación del valor predictivo individual de las características metabólicas

La capacidad de los biomarcadores extraídos de las imágenes de PET en posición prono para separar los grupos de respuesta fue evaluada. El estadístico de Mann-Whitney U fue determinado y un valor p de 0.05 fue considerado significativo. No se encontraron diferencias significativas en la medida de SUVmax ni SUVmean entre los grupos (Figura 3-5). De igual forma, el análisis de las métricas TLG y MTV tampoco mostró diferencias significativas entre los grupos de respuesta patológica (Figura 3-6).

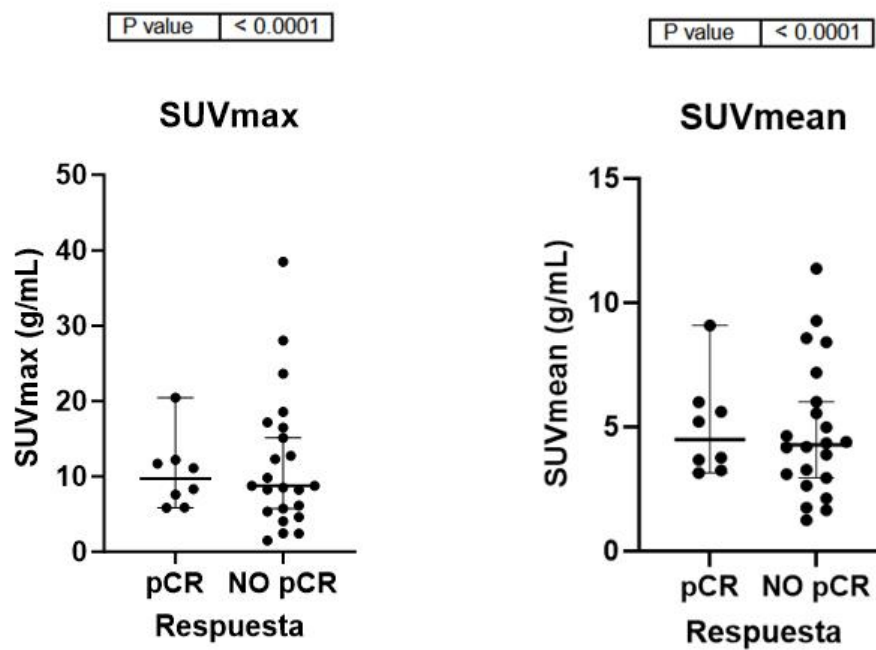


Figura 3-5: Valor predictivo individual del SUVmax y SUVmean respecto de la respuesta al tratamiento.

Por otra parte, la capacidad predictiva individual de la respuesta metabólica por PET fue evaluada, analizando el cambio relativo en la captación tumoral del F-18 FDG antes y después del tratamiento con el Δ SUVmax, calculado como en la ecuación (3-1):

$$\Delta SUVmax = 100 \times (SUVmax_{post} - SUVmax_{pre}) / SUVmax_{pre} \quad (3-1)$$

El Δ SUVmax fue significativamente más alto para el grupo que no responde al tratamiento (ver Figura 3-7).

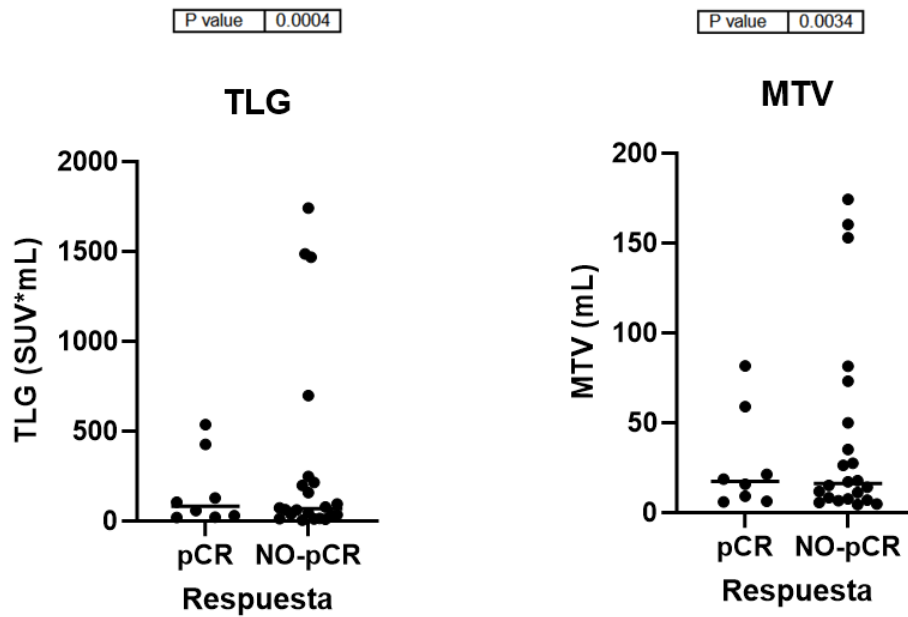


Figura 3-6: Valor predictivo individual del MTV y el TLG respecto de la respuesta al tratamiento.

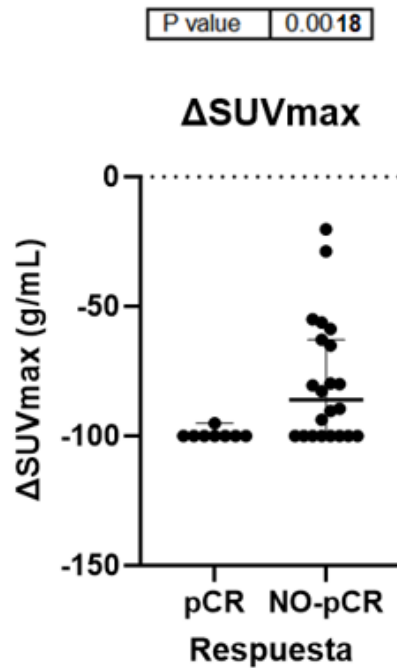


Figura 3-7: Valor predictivo del Δ SUVmax respecto de la respuesta al tratamiento.

3.3 Preprocesamiento y extracción de características radiómicas

Un total de 851 características radiómicas de las imágenes de PET e igual número de característica del mapa de ADC con b1100 fueron extraídas. Entre ellas se incluyen 107 características originales (sin filtro), compuestas por: 14 de forma (*shape*), 18 de estadística de primer orden y 75 texturales (24 GLCM, 16 GLSZM, 16 GLRLM, 14 GLDM y 5 NGTDM); y, 744 con filtro tipo wavelet, repartidas en 8 grupos de 93 características (LLL, LLH, LHL, LHH, HLL, HLH, HHL y HHH) donde L y H son señales de baja (L: Low) y alta (H: High) frecuencia, respectivamente. Por ejemplo, LLH es un volumen de imágenes transformadas usando los filtros pasa bajo en el eje X e Y, y un filtro pasa alto en Z.

Al conjunto de las características radiómicas se unió el valor de expresión de receptores de estrógeno (RE), que fue la única de las características obtenidas de la biopsia que mostró valor predictivo individual, para conformar un conjunto de características en una matriz con 30 observaciones, 1703 dimensiones y una etiqueta binaria de respuesta al tratamiento de la QNA.

3.4 Robustez del conjunto de características

3.4.1 Reducción de características según la reproducibilidad intra- e inter- observador

El efecto de la variabilidad intra- e inter-observador al momento de la segmentación fue evaluado a partir del cálculo del ICC, Una característica fue considerada altamente estable si el valor del ICC (2, 1) > 0.90. Los resultados de esta reducción se muestran en la Tabla 3-2. En total, 52.9% (450/851) de las características del mapa de ADC y 92.2% (785/851) de las características de PET fueron independientes de la variabilidad aplicada a la segmentación. 1235 características mostraron estabilidad.

Tabla 3-2: Número de características de alta, moderada y baja robustez en cada grupo en PET y Resonancia Magnética.

Grupo de características	PET			Mapa ADC		
	<i>Alta</i> (<i>ICC>0.9</i>)	<i>Media</i> (<i>ICC 0.5-0.9</i>)	<i>Baja</i> (<i>ICC<0.5</i>)	<i>Alta</i> (<i>ICC>0.9</i>)	<i>Media</i> (<i>ICC 0.5-0.9</i>)	<i>Baja</i> (<i>ICC<0.5</i>)
Forma	14	0	0	13	1	0
1er orden	18	0	0	7	11	0
GLCM	24	0	0	10	14	0
GLDM	13	1	0	7	6	1
GLSZM	14	2	0	7	8	1
NGTDM	5	0	0	1	4	0
GLRML	16	0	0	5	11	0
Wavelet	681	53	10	410	320	14

3.4.2 Reducción de características según la redundancia y relevancia

La correlación entre las características resultantes de la etapa anterior es mostrada en el mapa de calor de la Figura 3-8. La primera posición de la matriz (columna y fila) corresponde a la característica de expresión de receptores de estrógeno, la cual muestra poca relación con todas las demás, luego se encuentran las características extraídas de las imágenes de PET, las cuales muestran un alto grado de correlación entre sí (redundancia), seguidas de las características extraídas del mapa de ADC. Usando como umbral, el coeficiente de correlación de Spearman, $\rho \geq 0.8$, 125 características sobrevivieron a este análisis, 61.6% (77) fueron de PET, 37.6% (47) del mapa de ADC y 0.8% de la biopsia (1); 88.8% (111) del total fueron de tipo wavelet entre ambos grupos (70 y 41, respectivamente).

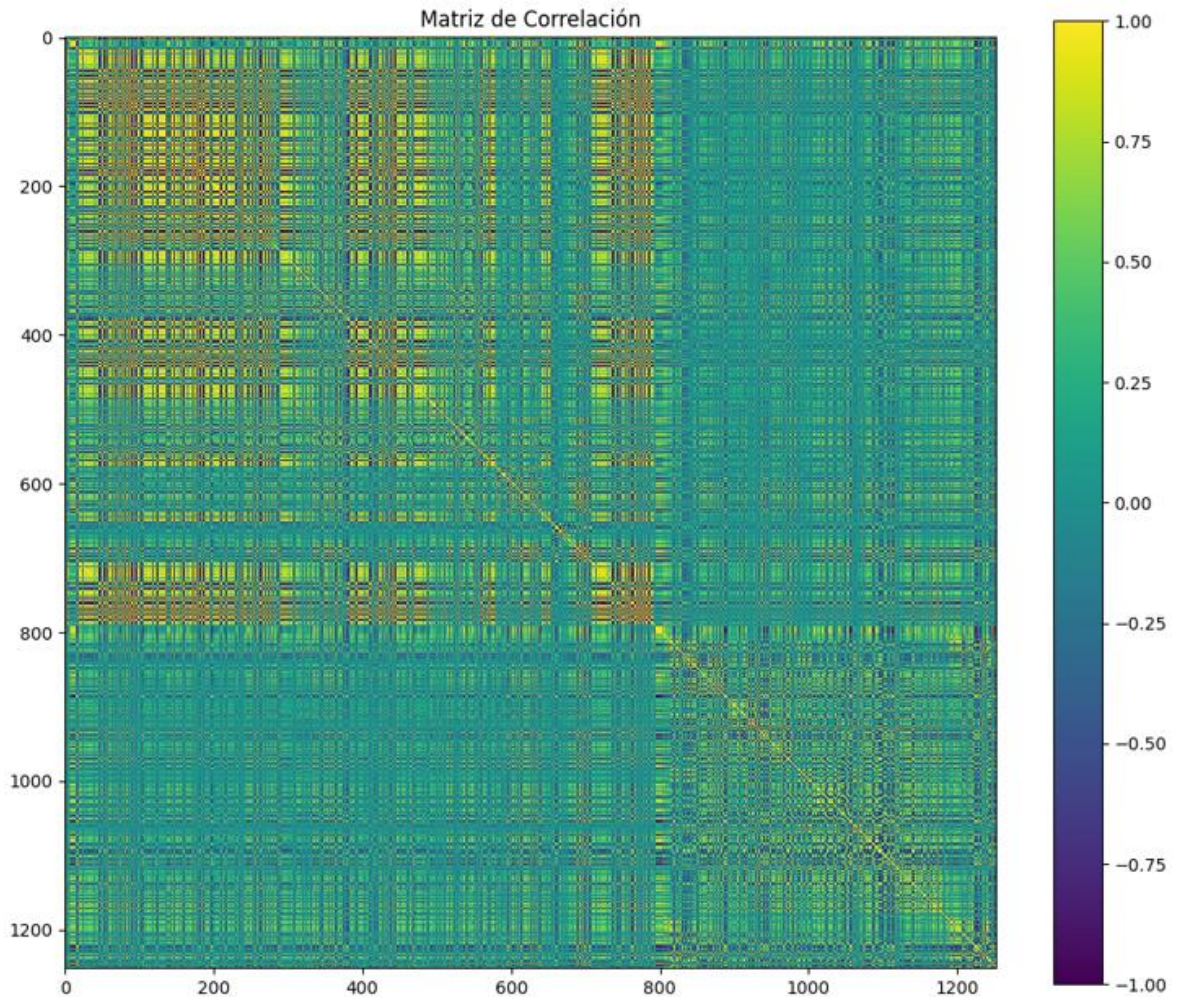


Figura 3-8: Mapa de calor de la correlación del conjunto de datos reproducible ante la variabilidad en la segmentación. En orden de ubicación de izquierda a derecha y arriba-abajo, la primera característica corresponde a RE, seguida de todas las características de PET y por último las características del mapa de ADC.

3.4.3 Selección de características por relevancia

El algoritmo Boruta nominó una única característica como de alta relevancia: 'wavelet-HHL_glcm_Imc1' extraída de las imágenes de PET. Todas las demás fueron rechazadas. Su tiempo de ejecución fue aproximadamente 10 segundos.

Usando el método del test de Wilcoxon, las 10 características más relevantes fueron seleccionadas, las cuales se listan en la Tabla 3-3, entre ellas está la nominada por la implementación de Boruta. En la tabla, a las características de PET se agregó el prefijo

("P_") para identificarlas y diferenciarlas de las características del mapa de ADC a las cuales se le agregó el prefijo ("M_"). Del total, 6 características fueron extraídas de las imágenes de PET, 3 de los mapas de ADC y 1 de los datos de la biopsia. Cabe anotar que 7 de las 10 características fueron del tipo wavelet. El tiempo de la ejecución computacional del método fue alrededor de 2 segundos.

Tabla 3-3. Características seleccionadas por el test Wilcoxon.

Categoría	Características seleccionadas	Acrónimo
Características PET	P_original_gldm_SmallDependence LowGrayLevelEmphasis	P_o_gldm_SDLGLE
	P_wavelet-LHH_gldm_ DependenceNonUniformityNormalized	P_w-LHH_gldm_DNUN
	P_wavelet-HLH_firstorder_Median	P_w-HLH_fo_Median
	P_wavelet- HHL_glcm_DifferenceEntropy	P_w-HHL_glcm_DE
	P_wavelet-HHL_glcm_Imc1	P_w-HHL_glcm_Imc1
	P_wavelet-HHH_firstorder_Mean	P_w-HHH_fo_Mean
Características mapa ADC	M_original_firstorder_10Percentile	M_o_fo_10P
	M_wavelet-LLH_firstorder_Mean	M_w-LLH_fo_Mean
	M_wavelet-LLL_firstorder_Minimum	M_w- LLL_fo_Minimum
Características clínicas	RE	

Por su parte, usando el modelo de aprendizaje automático tipo Random Forest con la configuración de hiperparámetros: min_samples_leaf=4, min_samples_split=2, n_estimators=200, las 10 características de mayor importancia para la clasificación fueron las presentadas en la Tabla 3-4. Del total, 4 características fueron extraídas de las imágenes de PET ("P_"), 5 de los mapas de ADC ("M_") y 1 de los datos de la biopsia. Cabe anotar que 7 de las 10 características fueron del tipo wavelet. El tiempo computacional del método fue de 1.5 minutos.

Con los 3 métodos, la característica 'P_wavelet-HHL_glcm_Imc1' fue seleccionada. Mientras que, Wilcoxon y el modelo de aprendizaje coincidieron en 6 de las 10 características extraídas (2 de PET, 3 del mapa de ADC y RE).

Tabla 3-4. Características seleccionadas por el modelo de aprendizaje automático.

Categoría	Características seleccionadas	Acrónimo
Características PET	P_wavelet-HHL_glcm_Imc1	P_w-HHL_glcm_Imc1
	P_wavelet-HHH_firstorder_Median	P_w-HHH_fo_Median
	P_original_firstorder_Kurtosis	P_o_fo_Kurtosis
	P_wavelet-HLH_firstorder_Median	P_w-HLH_fo_Median
Características mapa ADC	M_wavelet-LLL_firstorder_Minimum	M_w-LLL_fo_Minimum
	M_wavelet-LLH_firstorder_Mean	M_w-LLH_fo_Mean
	M_wavelet-LLL_glcm_ClusterShade	M_w-LLL_glcm_CS
	M_original_firstorder_10Percentile	M_o_fo_10P
	M_wavelet-LLL_glcm_MCC	M_w-LLL_glcm_MCC
Características clínicas	RE	

3.5 Rendimiento de las predicciones de la pCR

Las métricas del rendimiento resultantes de Leave-One-Out en los modelos predictivos usando cada una de las características de forma individual se despliegan en la Tabla 3-5 para las características extraídas por el método de Wilcoxon y en la Tabla 3-6 para las características extraídas por modelo de aprendizaje automático restantes, no comunes con las del método de Wilcoxon. El tiempo de ejecución del modelo iterando en las 10 características fue de 1.5 minutos aproximadamente. Los hiperparámetros configurados para el modelo fueron: min_samples_leaf=4, min_samples_split=2, n_estimators=200.

La característica con mejor rendimiento fue P_wavelet-HHL_glcm_Imc1, de AUC=0.78 con un intervalo de confianza (IC) al 95% de (0.70-0.87), la cual además fue extraída por los 3 métodos de selección implementados, por lo tanto, se considera robusta. La gráfica de la curva ROC de dicha característica se presenta en la Figura 3-9 y la matriz

de confusión correspondiente en la Figura 3-10. La sensibilidad y la especificidad fueron de 50% y 91% respectivamente, mientras que el VPP y el VPN fueron 67% y 83% respectivamente. Esto quiere decir que el modelo es muy bueno para predecir aquellas pacientes que no responderán al tratamiento con un pequeño margen de falsos positivos.

Por su parte, las características extraídas del mapa de ADC en general tuvieron un rendimiento pobre (Tabla 3-5 y Tabla 3-6), la mejor de ellas alcanzó apenas un “rendimiento justo” con un AUC =0.62.

Tabla 3-5: Rendimiento de la predicción para los modelos entrenados con cada una de las características individuales extraídas por el método de Wilcoxon.

Característica	AUC	Exactitud	Sensibilidad	Especificidad	VPP	VPN
RE	0.58	0.73	0.63	0.77	0.50	0.85
P_o_gldm_SDLGLE	0.60	0.57	0.00	0.77	0.00	0.68
P_w-LHH_gldm_DNUN	0.37	0.73	0.00	1.00	-	0.73
P_w-HLH_fo_Median	0.64	0.80	0.50	0.91	0.67	0.83
P_w-HHL_glcm_DE	0.39	0.67	0.00	0.91	0.00	0.71
P_w-HHL_glcm_Imc1	0.78	0.80	0.50	0.91	0.67	0.83
P_w-HHH_fo_Mean	0.64	0.57	0.00	0.77	0.00	0.68
M_o_fo_10P	0.62	0.77	0.38	0.91	0.60	0.80
M_w-LLH_fo_Mean	0.57	0.73	0.38	0.86	0.50	0.79
M_w-LLL_fo_Minimum	0.57	0.77	0.50	0.86	0.57	0.83

Tabla 3-6: Rendimiento de la predicción para los modelos entrenados con cada una de las características individuales extraídas por el modelo de aprendizaje automático.

Característica	AUC	Exactitud	Sensibilidad	Especificidad	VPP	VPN
P_w- HHH_fo_Median	0.67	0.77	0.63	0.82	0.56	0.86
P_o_fo_Kurtosis	0.50	0.70	0.38	0.82	0.43	0.78
M_w-LLL_glcm_CS	0.48	0.63	0.13	0.82	0.20	0.72
M_w- LLL_glcm_MCC	0.41	0.63	0.50	0.91	0.67	0.83

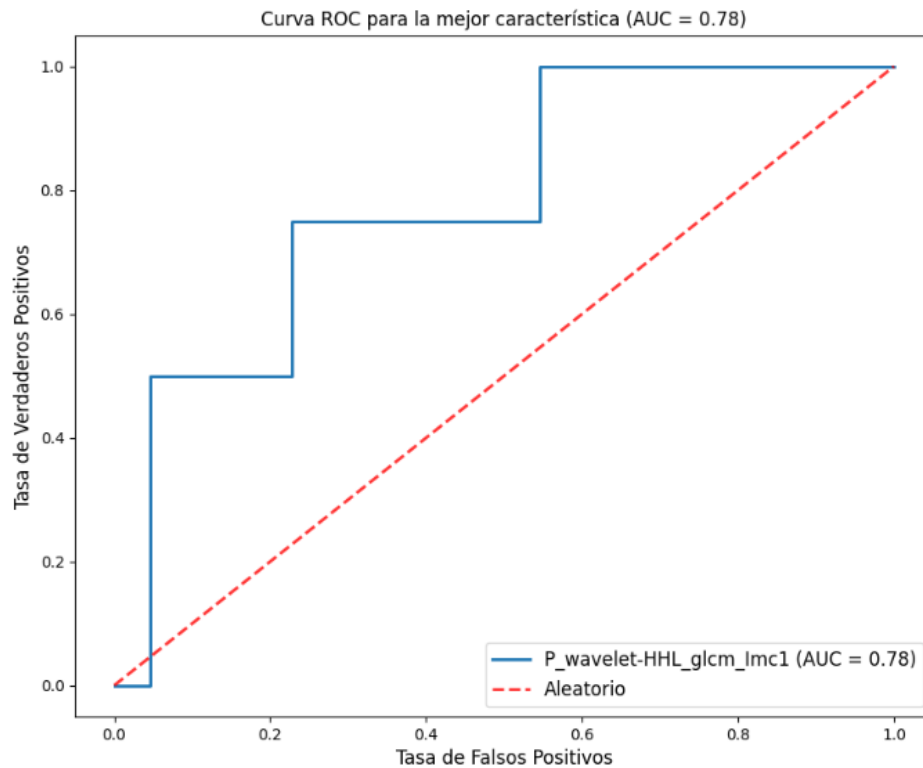


Figura 3-9. Curva ROC para el modelo clasificador Random Forest entrenado con la característica de mayor rendimiento, P_wavelet-HHL_glcm_Imc1.

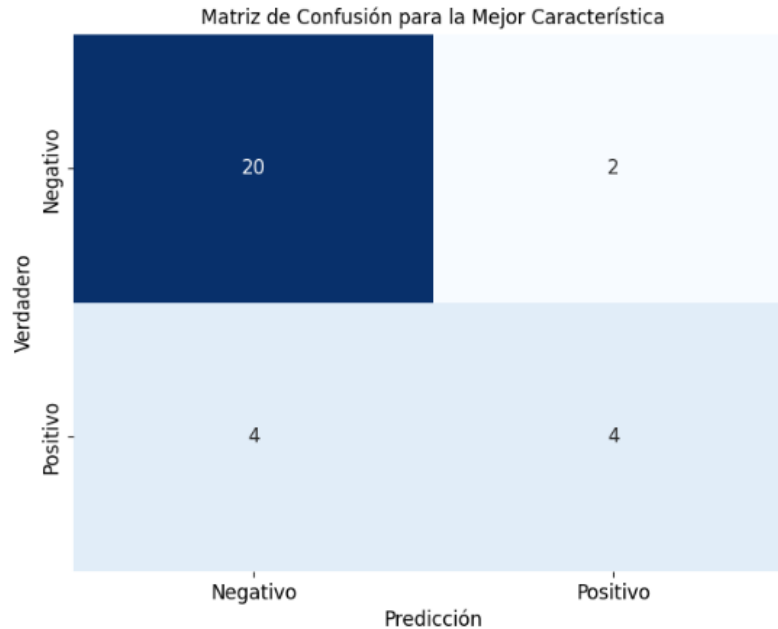


Figura 3-10: Matriz de confusión para el modelo clasificador entrenado con la característica de mayor rendimiento, P_wavelet-HHL_glcm_Imc1.

Por su parte, las métricas del rendimiento de los 3 mejores modelos predictivos entrenados con la combinación de dos características, seleccionadas de los dos conjuntos de datos por separado (Tabla 3-3 y Tabla 3-4), se despliegan en la Tabla 3-7. La combinación con mejor rendimiento fue “P_wavelet-HHL_glcm_Imc1, P_wavelet-HHH_firstorder_Mean” con un AUC=0.86 con IC95% (0.78-0.94).

Tabla 3-7. Rendimiento de los 3 mejores modelos predictivos entrenados usando parejas de características.

Característica	AUC	Exactitud	Sensibilidad	Especificidad	VPP	VPN
RE, P_w-HHL_glcm_Imc1	0.84	0.80	0.38	0.95	0.75	0.81
P_w-HHL_glcm_Imc1, P_w-HHH_fo_Mean	0.86	0.87	0.63	0.95	0.83	0.88
P_w-HHL_glcm_Imc1, M_o_fo_10P	0.82	0.83	0.38	1.00	1.00	0.81

3.6 Rendimiento de modelos predictivos usando ADASYN

Para el conjunto de datos con aumentación de datos sintéticos, el mejor resultado fue alcanzado por la característica clínica RE con un AUC=0.80 de IC95% [0.70-0.91] y sensibilidad 0.86, mientras que, para la característica wavelet-HHL_glcm_Imc1 el AUC bajó de 0.78 a 0.74, la sensibilidad subió a 67%, pero la especificidad bajó a 77%. Las gráficas de la curva ROC para los modelos entrenados con las características individuales se presentan en la Figura 3-11. Las métricas de los 3 modelos con mejor rendimiento se presentan en la Tabla 3-8.

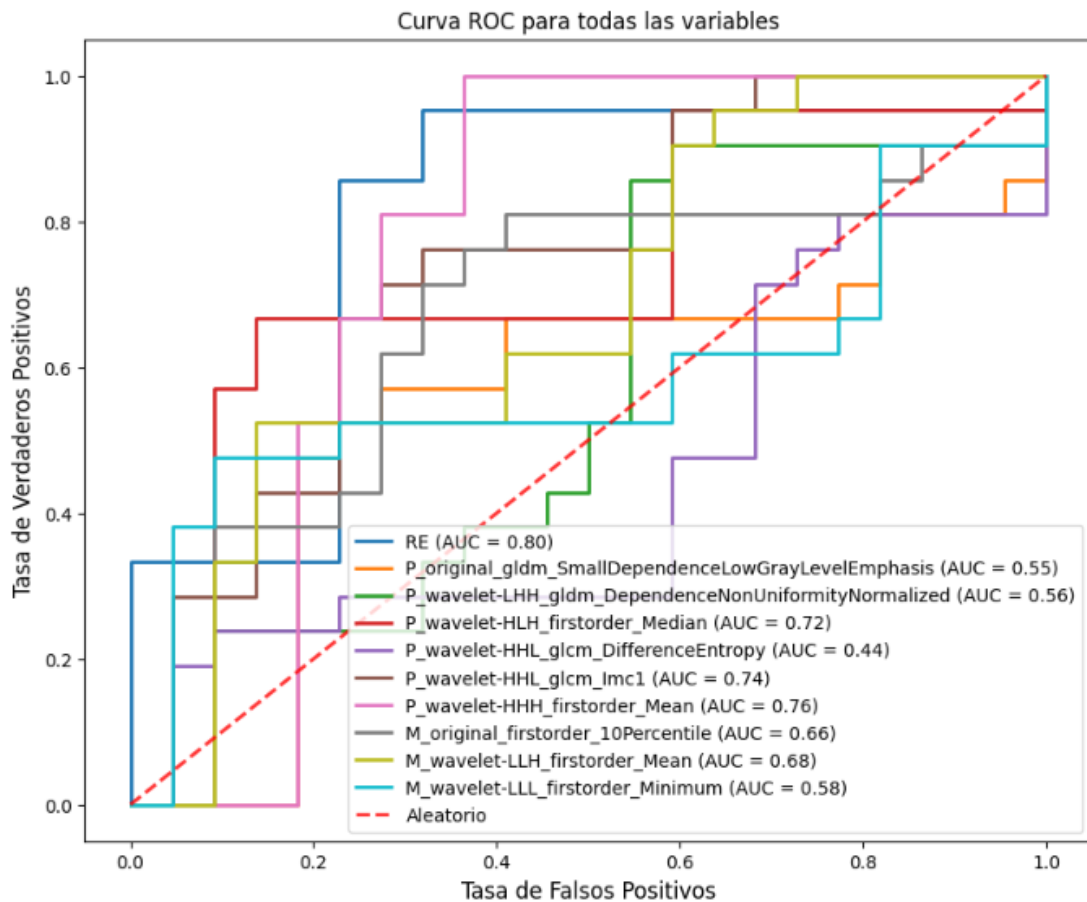


Figura 3-11: Curva ROC de 10 modelos clasificadores entrenados con datos aumentados por ADASYN usando características individuales.

Tabla 3-8. Rendimiento de los 3 mejores modelos predictivos entrenados usando datos aumentados por ADASYN.

Característica	AUC	Exactitud	Sensibilidad	Especificidad	VPP	VPN
RE	0.80	0.77	0.86	0.68	0.72	0.83
P_w-HHH_fo _Mean	0.76	0.74	0.86	0.63	0.69	0.82
P_w- HHL_glcm_Imc1	0.74	0.72	0.67	0.77	0.74	0.71

Las métricas de los 3 modelos con mejor rendimiento para los modelos combinados se presentan en la Tabla 3-9.

Tabla 3-9. Rendimiento de los 3 mejores modelos predictivos combinados usando datos aumentados por ADASYN.

Característica	AUC	Exactitud	Sensibilidad	Especificidad	VPP	VPN
P_w-HHL_ glcm_Imc1, RE	0.95	0.88	0.90	0.86	0.86	0.90
P_w-HHL_ glcm_Imc1, M_o_fo_10P	0.90	0.79	0.86	0.73	0.75	0.84
M_w- LLL_glcm_CS, RE	0.89	0.86	0.76	0.96	0.94	0.81

4. Discusión

En este estudio se evaluó la capacidad de características radiómicas extraídas de imágenes de PET y los mapas de ADC por resonancia magnética, adquiridas pretratamiento, como predictores de la respuesta al tratamiento de la QNA. Los principales hallazgos son: 1) Los modelos predictivos entrenados con una sola característica demostraron un buen rendimiento ($0.7 \leq \text{AUC} < 0.8$) para la clasificación entre pCR y No-pCR, mientras que los modelos combinados con dos características mostraron un rendimiento muy bueno ($0.8 \leq \text{AUC} < 0.9$). 2) El balance de clases usando un método de aumentación tiene el potencial de mejorar el rendimiento del modelo para la predicción de la clase minoritaria. 3) Se construyó un flujo de trabajo de análisis radiómico robusto para clasificar los pacientes con Ca de mama que resultarán en pCR a la quimioterapia neoadyuvante a partir de imágenes pretratamiento de PET y el mapa de ADC, más datos de inmunohistoquímica.

Con respecto al diseño del protocolo de adquisición de imágenes, la posición del paciente, prono o supino, durante la adquisición de las imágenes de PET resultó ser insignificante para la cuantificación de biomarcadores como el volumen de interés y el SUVmax. Iguales hallazgos fueron publicados por A. A. van Loevezijn *et al.* al estudiar el impacto de la posición para la estadificación de nódulos axilares del Ca de mama [63]. En nuestro análisis radiómico fueron usadas las imágenes en prono por considerar que con las mamas colgantes existía menor riesgo de superposición de estructuras en los planos usados para la segmentación y mayor comparabilidad con las imágenes de resonancia magnética. Por lo tanto, debido a la ventaja observada, se recomienda la adquisición de las imágenes de PET en posición prono para los análisis radiómicos que usan varias modalidades de imágenes. No obstante, nuevos desarrollos

en las antenas receptoras en RM permiten la adquisición del estudio mamario en posición supino con buena calidad de imagen, en tal caso la imagen de PET también debería ser adquirida en supino. Por otra parte, la presencia del clip metálico detectado en 13 de las imágenes de RM, es un aspecto para tener en cuenta puesto que introduce variabilidad indeseada y evitable en la cuantificación de las características radiómicas. El impacto del clip no pudo ser evaluado en este estudio por falta de datos comparativos, tampoco se encontraron publicaciones relacionadas con este tema en la literatura científica. Se recomienda programar la adquisición de las imágenes siempre antes de la biopsia.

Un análisis exploratorio de la capacidad de los biomarcadores de PET y las características de la inmunohistoquímica para separar los grupos de respuesta fue realizado. Ninguno de los biomarcadores mostró diferencias considerables entre grupos, hallazgos similares fueron publicados por Choi *et al.* [64]. El ΔSUV_{max} fue analizado y resultó ser significativamente más alto para el grupo que no responde al tratamiento, alineado con los resultados publicados por Choi *et al.* [64] y por Sengoz *et al.* [65]. Dicha característica no fue incluida en el conjunto de datos debido a que el objetivo del presente estudio era predecir la respuesta a partir de características pretratamiento. Sin embargo, puede ser objeto de análisis de otro estudio.

En el paso de la definición del ROI, si bien existen algoritmos de segmentación basados en *deep learning* capaces de realizar difíciles tareas de segmentación sin la necesidad de ajuste manual en regiones como el cerebro [66], en las lesiones mamarias la gran variabilidad entre pacientes en cuanto a los márgenes del tumor y la composición del tejido mamario (relación del porcentaje de grasa y tejido fibroglandular) dificultan la tarea de segmentación. Un algoritmo de segmentación para lesiones de Ca de mama en imágenes de RM, basado en redes neuronales convolucionales (Res-UNet) fue publicado por Yue *et al.* en 2022 [67] con resultados de ICC>0.8 para la medida del volumen tumoral entre el modelo y los radiólogos, sin embargo, en dicha publicación no se especifica el subtipo de las lesiones evaluadas, ni el impacto de la composición

de las mamas o de las diferentes geometrías de los bordes del tumor sobre el rendimiento del algoritmo. Por lo tanto, en este estudio la segmentación manual, considerada aún el *gold standard* para lesiones mamarias en RM, fue implementada por una médico radióloga experta. En general, el uso de métodos de segmentación automática en RM mamaria debería ser supervisada por un radiólogo experto. Por su parte, la segmentación en PET fue realizada con el método semiautomático por umbral adaptativo, Nestle. Para PET los métodos automáticos y semiautomáticos brindan mejor desempeño y son más robustos que la segmentación manual debido al efecto de volumen parcial y la difuminación de los bordes [39], por lo tanto, la segmentación manual debería ser evitada para radiomics en PET. Bashir *et al.* encontraron que los métodos semiautomáticos basados en umbral conducen a la mayor reproducibilidad cuando son usados por diferentes observadores [68]. En general, con relación a la segmentación un aspecto que aún permanece poco claro para el flujo de análisis radiómico es si el ROI debería incluir únicamente el componente sólido del tumor o cubrir adicionalmente regiones quísticas, hemorrágicas, necróticas o inflamadas intra o peritumorales. Por ejemplo, Braman *et al.* [69] demostraron que características radiómicas extraídas de las regiones peritumorales en imágenes de DCE-MRI contribuyen a la predicción de la pCR a la QNA. En el presente estudio se incluyeron las regiones necróticas intratumorales, mientras que las regiones peritumorales fueron excluidas.

En el paso de extracción de características radiómicas, un tamaño de bin fijo fue usado para la discretización, siguiendo la recomendación de la IBSI [31], debido a que la discretización absoluta (por tamaño de bin) ha demostrado proveer características más reproducibles que la discretización relativa (por número de bin) tanto en PET como en los mapas de ADC [54], [70]. La robustez interobservador medida a través del ICC>0.9 demostró que la mayoría de las características radiómicas de PET tuvieron alta estabilidad (92.2%), mientras que, apenas el 52% de las características del mapa de ADC. Otros estudios encontraron alrededor del 45% de las características del mapa de ADC con alta repetibilidad [71], [72]. De las categorías de extracción se evidenció

que las características extraídas después de aplicar el filtro wavelet proporcionaron más información a los resultados finales, representando el 70% de las variables usadas para el entrenamiento de los modelos predictivos con mejor rendimiento.

En el paso de reducción de la dimensionalidad y selección de características no existe un consenso sobre el análisis estadístico óptimo o la metodología de machine-learning para obtener los mejores resultados, esta decisión depende en gran medida de los datos y el objetivo del estudio [73]. Los métodos de selección de características se agrupan en tres categorías principales: métodos de filtrado, métodos embebidos y métodos envolventes, según cuando se realiza la selección. Tres métodos para la selección de las características más importantes fueron implementados: Boruta (envolvente), Wilcoxon (de filtrado) y un modelo *random forest* (embebido), sin embargo, ninguno de ellos toma en cuenta la multicolinealidad, por lo tanto, permiten el paso de información redundante. En general la mayoría de los métodos de selección por importancia no tienen en cuenta la redundancia. Por este motivo recomendamos realizar la eliminación de las características redundantes previamente en el flujo de trabajo. En este estudio la reducción de la dimensionalidad por mínima redundancia ($\rho \leq 0.8$) y máxima relevancia (mayor AUC individual) fue implementada previamente, disminuyendo el conjunto de variables de 1235 a solo 125. El método Boruta únicamente identificó como importante una característica, mientras que, de los otros dos se extrajeron las primeras 10 características en orden de importancia, coincidiendo en 6 de ellas. El método de Wilcoxon encontró la mayor cantidad de características que hicieron parte de los modelos con mejor rendimiento, en unos pocos segundos de ejecución.

Un modelo de clasificación *random forest* fue elegido debido a que es de fácil explicabilidad, rápido de entrenar y que el conjunto de datos es pequeño, en comparación con modelos más complejos como Xgboost que para el caso de estudio podría sobreaprender la clasificación de los datos de entrenamiento y no ser generalizable a nuevos datos. Mientras que con *random forest*, la construcción de

múltiples árboles de decisión y la agregación de sus predicciones tiende a reducir la varianza del modelo en comparación con un solo árbol de decisión y ayuda a mitigar el sobreajuste. Una división en grupos de entrenamiento y validación no fue considerada apropiada por la cantidad de datos, en cambio, se implementó la validación por leave-one-out. De los resultados obtenidos se sugiere como parte del flujo de trabajo una combinación del test de Wilcoxon para selección de características y un modelo de clasificación del tipo random forest, consistente con los resultados publicados por Parmar *et al.* [73] quienes encontraron en esta combinación el más alto rendimiento pronóstico y la mayor estabilidad contra la perturbación de los datos entre 14 métodos de selección y 12 modelos de clasificación evaluados.

El rendimiento de los experimentos fue evaluado usando 3 métricas: el área bajo la curva característica de funcionamiento del receptor (ROC), la sensibilidad y la especificidad. El AUC refleja la habilidad del modelo para mostrar una diferencia constante en los datos con diferentes niveles. El mejor rendimiento individual para el conjunto de datos originales se obtuvo con la característica del tipo textural post filtro wavelet, denominada “wavelet-HHL_glcm_Imc1” (AUC de 0.78; IC95%: 0.70-0.87; sensibilidad: 50%; especificidad: 91%). Por lo tanto, el modelo es muy bueno para predecir aquellas pacientes que no responderán al tratamiento con un pequeño margen de falsos positivos. Además, esta fue la característica más robusta, ya que fue seleccionada por los tres métodos implementados. El mejor rendimiento combinado se obtuvo de wavelet-HHL_glcm_Imc1 y wavelet-HHH_firstorder_Mean (AUC de 0.86; IC95%: 0.78-0.94; sensibilidad: 63%; especificidad: 95%) ambas características de PET post filtro wavelet, seguido por RE y wavelet-HHL_glcm_Imc1 (AUC de 0.84; IC95%: 0.77-0.91; sensibilidad: 38%; especificidad: 95%). Nuevamente aparece la característica “wavelet-HHL_glcm_Imc1” como de gran potencial predictivo, aunque la sensibilidad de los modelos que conforma sea baja, su especificidad es bastante alta, por lo tanto, podría ser de utilidad para predecir aquellos pacientes que no responderán al tratamiento. En general, los modelos con características derivadas de PET mostraron una más alta especificidad en predecir la pCR después de la QNA en

pacientes con Ca de mama, alineado con los hallazgos reportados por Guarín *et al.* [74] quienes encontraron que los reportes clínicos de las imágenes de PET tenían un 100% de especificidad para predecir la pCR en este mismo grupo de pacientes y consistente con otras publicaciones [75], [75], [76].

El mejor rendimiento considerando la aumentación de datos fue alcanzado por la característica RE (AUC de 0.80; IC95%: 0.70-0.91; sensibilidad: 86%; especificidad: 91%). Mientras que, el mejor rendimiento de modelos combinados fue de las características wavelet-HHL_glcm_Imc1 y RE (AUC de 0.95; IC95%: 0.85-1.00; sensibilidad: 90%; especificidad: 86%). El balanceo de clases en los datos mediante aumentación mejoró el rendimiento de los modelos entrenados.

Una dificultad en la implementación del análisis radiómico es que la estabilidad de las características radiómicas se puede ver afectada por diferentes parámetros durante el flujo de trabajo, como: la variabilidad en los protocolos de adquisición de las imágenes, la reconstrucción, el preprocesamiento, la variabilidad intra e inter-observador al momento de la segmentación, la discretización de las imágenes e incluso las fórmulas matemáticas usadas para la extracción de las características con software *in-house*. En general, los estudios radiómicos reportados en la literatura científica son todavía muy inhomogéneos y, en muchos, la información publicada es incompleta con relación al flujo de trabajo seguido, lo que dificulta la reproducibilidad requerida para su introducción en la práctica clínica, como lo describen análisis previos [77]. Sin embargo, una tendencia de estandarización ha comenzado como lo sugiere Urso *et al.* [26]. Por esta razón, con el objetivo de generar un flujo de trabajo reproducible con potencial aplicación en la rutina clínica este estudio cumplió con la mayoría de los puntos de control que debe tener un estudio de validación: (1) se implementó un estudio prospectivo con protocolos de imágenes estandarizados usando un único equipo PET y un único resonador para la adquisición de todas las imágenes, lo que elimina la varianza que introduce el uso de diferente tecnología, aunque en la actualidad este problema puede ser bien manejado usando la herramienta de

armonización de datos ComBat [78]; (2) los equipos fueron sometidos a las pruebas de control de calidad periódicas de acuerdo con las recomendaciones del fabricante; (3) se evaluó la reproducibilidad de las características radiómicas ante la variabilidad intra- e inter-observador en la delimitación de los ROIs usando el ICC y el reporte se realizó según lo recomendado por *Koo & Li* [79]; (4) el cálculo de las características radiómicas se realizó usando *pyradiomics*, software de libre acceso, cuyas fórmulas están estandarizadas en concordancia con la iniciativa IBSI; (5) se implementó un método de reducción de características por mínima redundancia y máxima relevancia, además de varios métodos de selección de características por importancia para reducir la dimensionalidad y evitar potencial sobreajuste. Además, (6) la adquisición de las imágenes, preprocesamiento y extracción de las características radiómicas fueron reportadas de acuerdo con las directrices del QIBA y la IBSI. Por lo tanto, los modelos y características descritos en este trabajo pueden ser fácilmente aplicados a futuros conjuntos de datos, lo que permitiría una validación independiente de la precisión de los mismos.

Varios análisis han determinado el poder predictivo combinado de las características de PET y MRI [76], [80]. Aunque la secuencia DCE de RM es la más sensible para la detección del Ca de mama y ha sido evaluada en análisis radiómicos previos [69], la evaluación de los mapas de ADC es de importancia en la actualidad, ya que permitiría usar imágenes libres del uso de contraste [81].

A nuestro leal saber y entender, este estudio es uno de los primeros en utilizar simultáneamente características radiómicas de F18-FDG PET y mapas de ADC más datos de inmunohistoquímica para la predicción de la pCR a la QNA. La configuración establecida para analizar cada característica individualmente y en combinación permitió obtener más información sobre la eficacia predictiva de los modelos basados en datos multiparamétricos, ya que, si bien todas las características extraídas de los mapas de ADC mostraron un potencial predictivo bastante pobre, el análisis combinado con características de PET proporcionó modelos con valores valiosos de

AUC. Del mismo modo, el modelo combinado con una característica radiómica de PET y el RE (inmunohistoquímica) fue el de mejor rendimiento. Esto respalda la hipótesis de que el uso de características multiparamétricas puede proporcionar una mirada más completa a la caracterización del Ca de mama y una herramienta no invasiva para la predicción de la pCR en pacientes con Ca de mama localmente avanzado.

Aunque los resultados son prometedores, existen algunas limitaciones en este estudio: principalmente debido al limitado tamaño de la población y la distribución desbalanceada de las clases con una baja tasa de respondedores a la QNA (73.3% de no respondedores), para ello, se aplicó el método de validación cruzada "Leave One Out" y se evaluó el balance por aumentación de datos sintéticos usando ADASYN. Por otra parte, si bien la implementación del método de segmentación *Nestle* es conveniente en las imágenes de PET, requiere la disponibilidad de un fantoma NEMA de calidad de imagen, con el cual no cuentan todos los centros clínicos.

En resumen, este estudio es una investigación exploratoria inicial que demuestra el potencial valor de las características radiómicas extraídas de PET y el mapa de ADC más los resultados de inmunohistoquímica para predecir la pCR a la quimioterapia neoadyuvante en pacientes de Ca de mama usando un modelo machine learning. Grandes estudios prospectivos y ensayos multicéntricos con igual distribución de los subtipos moleculares son necesarios para evaluar la robustez de las características radiómicas.

5. Conclusiones

5.1 Conclusiones

En este trabajo se desarrolló e implementó un flujo de análisis radiómico para la predicción de la pCR. Se implementaron tres métodos de selección de características por importancia y un método de aumentación de datos. Una combinación de selección de características más importantes con el test de Wilcoxon y un modelo predictivo random forest es sugerida.

Los resultados de este estudio demostraron el potencial del análisis combinado de características radiómicas metabólicas (PET) y funcionales (RM) más características clínicas para conformar modelos predictivos para apoyar las decisiones clínicas, predecir de forma precisa y no invasiva la pCR a la QNA en pacientes con Ca de mama, así como, potencialmente mejorar la estratificación preterapéutica de los pacientes. Sin embargo, un mayor número de casos es requerido para validar estos hallazgos.

5.2 Perspectivas futuras

Este estudio incluyó 1703 características radiómicas, de las cuales solo 125 fueron encontradas estables. Esto indica que muchas características no son apropiadas para estudios clínicos adicionales, haciendo evidente el valor de la evaluación de la estabilidad. Por lo tanto, una investigación más profunda en esta área es necesaria.

Este trabajo presenta una investigación exhaustiva de las características radiómicas que puede servir como punto de partida para múltiples trabajos en torno a la predicción de la respuesta a la QNA en pacientes con cáncer de mama. Uno de ellos

podría ser la evaluación de la estabilidad de las características frente al cambio de la definición del ancho de bin al momento de la extracción. Una segunda propuesta podría ser el uso de radiomics en las imágenes de DCE en combinación con características radiómicas derivadas de PET y la biopsia. Una tercera investigación de interés podría ser la construcción de modelos predictivos con las características extraídas en este trabajo, usando otras herramientas de aprendizaje automático (machine learning) o profundo (Deep learning) y la comparación del rendimiento. Sin embargo, una limitante es la poca cantidad de datos disponibles, por lo tanto, se sugiere primero ampliar el conjunto de datos de muestra. Los datos usados en este estudio podrían estar disponibles previa solicitud.

La característica “wavelet-HHL_glcm_Imc1” extraída de PET proporcionó muy buen rendimiento de los modelos que conformaba y demostró robustez dentro del análisis, valdría la pena la evaluación de su poder predictivo en un conjunto de datos más grande.

Bibliografía

- [1] International Agency for Research on Cancer, “Global Cancer Observatory: Cancer Today”. Consultado: el 10 de noviembre de 2023. [En línea]. Disponible en: <http://gco.iarc.fr/today>
- [2] L. M. Spring *et al.*, “Pathologic Complete Response after Neoadjuvant Chemotherapy and Impact on Breast Cancer Recurrence and Survival: A Comprehensive Meta-analysis”, *Clinical Cancer Research*, vol. 26, núm. 12, pp. 2838–2848, jun. 2020, doi: 10.1158/1078-0432.CCR-19-3492.
- [3] A. Marusyk, V. Almendro, y K. Polyak, “Intra-tumour heterogeneity: a looking glass for cancer?”, *Nat Rev Cancer*, vol. 12, núm. 5, pp. 323–334, may 2012, doi: 10.1038/nrc3261.
- [4] R. Fisher, L. Pusztai, y C. Swanton, “Cancer heterogeneity: implications for targeted therapeutics”, *Br J Cancer*, vol. 108, núm. 3, pp. 479–485, feb. 2013, doi: 10.1038/bjc.2012.581.
- [5] R. J. Gillies, P. E. Kinahan, y H. Hricak, “Radiomics: Images Are More than Pictures, They Are Data”, *Radiology*, vol. 278, núm. 2, pp. 563–577, feb. 2016, doi: 10.1148/radiol.2015151169.
- [6] S. S. F. Yip y H. J. W. L. Aerts, “Applications and limitations of radiomics”, *Phys Med Biol*, vol. 61, núm. 13, pp. R150–R166, jul. 2016, doi: 10.1088/0031-9155/61/13/R150.
- [7] B. Bendriem y D. W. Townsend, *The Theory and Practice of 3D PET*. Dordrecht: Springer Netherlands, 1998. doi: 10.1007/978-94-017-3475-2.
- [8] D. L. Bailey, D. W. Townsend, P. E. Valk, y M. N. Maisey, *Positron Emission Tomography: Basic Sciences*. Springer London, 2005. doi: <https://doi.org/10.1007/b136169>.

- [9] D. L. Bailey, J. L. Humm, A. Todd-Pokropek, y A. van Aswegen, *Nuclear medicine physics: A handbook for teachers and students*. 2014. [En línea]. Disponible en: <https://api.semanticscholar.org/CorpusID:183658714>
- [10] D. W. Townsend, “Dual-Modality Imaging: Combining Anatomy and Function”, *Journal of Nuclear Medicine*, vol. 49, núm. 6, p. 938, jun. 2008, doi: 10.2967/jnumed.108.051276.
- [11] D. W. Townsend, T. Beyer, y T. M. Blodgett, “PET/CT scanners: A hardware approach to image fusion”, *Semin Nucl Med*, vol. 33, núm. 3, pp. 193–204, jul. 2003, doi: 10.1053/snuc.2003.127314.
- [12] John M Ollinger, “Model-based scatter correction for fully 3D PET”, *Phys Med Biol*, vol. 41, núm. 1, p. 153, 1996, doi: 10.1088/0031-9155/41/1/012.
- [13] L. A. Shepp y Y. Vardi, “Maximum Likelihood Reconstruction for Emission Tomography”, *IEEE Trans Med Imaging*, vol. 1, núm. 2, pp. 113–122, 1982, doi: 10.1109/TMI.1982.4307558.
- [14] M. Soret, S. L. Bacharach, y I. Buvat, “Partial-Volume Effect in PET Tumor Imaging”, *Journal of Nuclear Medicine*, vol. 48, núm. 6, p. 932, jun. 2007, doi: 10.2967/jnumed.106.035774.
- [15] J. M. M. Rogasch, F. Hofheinz, L. van Heek, C.-A. Voltin, R. Boellaard, y C. Kobe, “Influences on PET Quantification and Interpretation”, *Diagnostics*, vol. 12, núm. 2, p. 451, feb. 2022, doi: 10.3390/diagnostics12020451.
- [16] M. Eskinian *et al.*, “Effect of blood glucose level on standardized uptake value (SUV) in 18F- FDG PET-scan: a systematic review and meta-analysis of 20,807 individual SUV measurements”, *Eur J Nucl Med Mol Imaging*, vol. 46, núm. 1, pp. 224–237, ene. 2019, doi: 10.1007/s00259-018-4194-x.
- [17] M. C. Adams, T. G. Turkington, J. M. Wilson, y T. Z. Wong, “A Systematic Review of the Factors Affecting Accuracy of SUV Measurements”, *American Journal of Roentgenology*, vol. 195, núm. 2, pp. 310–320, ago. 2010, doi: 10.2214/AJR.10.4923.

-
- [18] J. T. Bushberg, J. A. Seibert, E. M. Leidholdt Jr., y J. M. Boone, *The Essential Physics of Medical Imaging*, 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins, 2012. doi: 10.1118/1.4811156.
- [19] A. D. Elster, “Questions and Answers in MRI”. Consultado: el 7 de agosto de 2022. [En línea]. Disponible en: <https://www.mriquestions.com>
- [20] A. B. Cohen, *Concepts of Nuclear Physics*. Tata McGraw-Hil, 1971.
- [21] L. G. Hanson, “Introduction to Magnetic Resonance Imaging Techniques”, Copenhagen, ago. 2009. Consultado: el 7 de agosto de 2022. [En línea]. Disponible en: <https://eprints.drcomr.dk/37/>
- [22] L. G. Hanson, “Is quantum mechanics necessary for understanding magnetic resonance?”, *Concepts in Magnetic Resonance Part A*, vol. 32A, núm. 5, pp. 329–340, 2008, doi: <https://doi.org/10.1002/cmra.20123>.
- [23] R. P. Feynman, Jr. Vernon Frank L., y R. W. Hellwarth, “Geometrical Representation of the Schrödinger Equation for Solving Maser Problems”, *J Appl Phys*, vol. 28, núm. 1, pp. 49–52, ene. 1957, doi: 10.1063/1.1722572.
- [24] J. N. Oshinski, “MR Physics I: Magnetization, resonant frequency, relaxation and the MR signal”, en *College on Medical Physics: Medical Imaging Physics and Technology. Principles of Optimisation, Safety and Education Development for Building Capacity in Developing Countries*, Trieste: ICTP, sep. 2022.
- [25] R. J. Gillies, P. E. Kinahan, y H. Hricak, “Radiomics: Images are more than pictures, they are data”, *Radiology*, vol. 278, núm. 2, pp. 563–577, feb. 2016, doi: 10.1148/radiol.2015151169.
- [26] L. Urso *et al.*, “PET-Derived Radiomics and Artificial Intelligence in Breast Cancer: A Systematic Review”, *Int J Mol Sci*, vol. 23, núm. 21, p. 13409, nov. 2022, doi: 10.3390/ijms232113409.
- [27] S.-H. Lee, H. Park, y E. S. Ko, “Radiomics in Breast Imaging from Techniques to Clinical Applications: A Review”, *Korean J Radiol*, vol. 21, núm. 7, p. 779, 2020, doi: 10.3348/kjr.2019.0855.

- [28] V. Kumar *et al.*, “Radiomics: the process and the challenges”, *Magn Reson Imaging*, vol. 30, núm. 9, pp. 1234–1248, nov. 2012, doi: 10.1016/j.mri.2012.06.010.
- [29] RSNA, “Quantitative Imaging Biomarkers Alliance”. Consultado: el 27 de enero de 2021. [En línea]. Disponible en: <https://www.rsna.org/research/quantitative-imaging-biomarkers-alliance>
- [30] E. J. Limkin *et al.*, “Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology”, *Annals of Oncology*, vol. 28, núm. 6, pp. 1191–1206, jun. 2017, doi: 10.1093/annonc/mdx034.
- [31] A. Zwanenburg, S. Leger, M. Vallières, y S. Löck, “Image biomarker standardisation initiative”, dic. 2016, doi: 10.1148/radiol.2020191145.
- [32] K. Cheng, A. Lin, J. Yuvaraj, S. J. Nicholls, y D. T. L. Wong, “Cardiac Computed Tomography Radiomics for the Non-Invasive Assessment of Coronary Inflammation”, *Cells*, vol. 10, núm. 4, p. 879, abr. 2021, doi: 10.3390/cells10040879.
- [33] The University of Texas MD Anderson Cancer Center, “Residual Cancer Burden Calculator”. Consultado: el 10 de noviembre de 2022. [En línea]. Disponible en: <https://www3.mdanderson.org/app/medcalc/index.cfm?pagename=jsconvert3>
- [34] R. Boellaard *et al.*, “FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0”, *Eur J Nucl Med Mol Imaging*, vol. 42, núm. 2, pp. 328–354, feb. 2015, doi: 10.1007/s00259-014-2961-x.
- [35] D. Koopman *et al.*, “Technical note: how to determine the FDG activity for tumour PET imaging that satisfies European guidelines”, *EJNMMI Phys*, vol. 3, núm. 1, p. 22, dic. 2016, doi: 10.1186/s40658-016-0158-z.
- [36] EANM, “18F Accreditation specifications”. Consultado: el 10 de noviembre de 2023. [En línea]. Disponible en: <https://earl.eanm.org/accreditation-specifications/>

- [37] National Electrical Manufacturers Association, “NEMA MS 3-2008 (R2014) Determination Of Image Uniformity In Diagnostic Magnetic Resonance Images”, 2014.
- [38] National Electrical Manufacturers Association, “NEMA MS 5, 2018 Edition, 2018 - Determination of Slice Thickness in Diagnostic Magnetic Resonance Imaging”, ene. 2018.
- [39] M. Hatt *et al.*, “Classification and evaluation strategies of auto-segmentation approaches for PET: Report of AAPM task group No. 211”, *Med Phys*, vol. 44, núm. 6, pp. e1–e42, 2017, doi: 10.1002/mp.12124.
- [40] M. Tamal, “Intensity threshold based solid tumour segmentation method for Positron Emission Tomography (PET) images: A review”, *Heliyon*, vol. 6, núm. 10, p. e05267, oct. 2020, doi: 10.1016/j.heliyon.2020.e05267.
- [41] U. Nestle *et al.*, “Comparison of Different Methods for Delineation of ^{18}F -FDG PET-Positive Tissue for Target Volume Definition in Radiotherapy of Patients with Non-Small Cell Lung Cancer”, *Journal of Nuclear Medicine*, vol. 46, núm. 8, p. 1342, ago. 2005, [En línea]. Disponible en: <http://jnm.snmjournals.org/content/46/8/1342.abstract>
- [42] C. Nioche *et al.*, “LIFEx: A Freeware for Radiomic Feature Calculation in Multimodality Imaging to Accelerate Advances in the Characterization of Tumor Heterogeneity”, *Cancer Res*, vol. 78, núm. 16, pp. 4786–4789, ago. 2018, doi: 10.1158/0008-5472.CAN-18-0125.
- [43] A. Fedorov *et al.*, “3D Slicer as an image computing platform for the Quantitative Imaging Network”, *Magn Reson Imaging*, 2012, doi: 10.1016/j.mri.2012.05.001.
- [44] J. J. M. van Griethuysen *et al.*, “Computational Radiomics System to Decode the Radiographic Phenotype”, *Cancer Res*, vol. 77, núm. 21, pp. e104–e107, nov. 2017, doi: 10.1158/0008-5472.CAN-17-0339.
- [45] Pyradiomics Community, “Radiomic Features”. Consultado: el 9 de octubre de 2022. [En línea]. Disponible en: <http://github.com/radiomics/pyradiomics> Revision 6a761c4e

- [46] A. Traverso, L. Wee, A. Dekker, y R. Gillies, “Repeatability and Reproducibility of Radiomic Features: A Systematic Review”, *International Journal of Radiation Oncology*Biography*Physics*, vol. 102, núm. 4, pp. 1143–1158, nov. 2018, doi: 10.1016/j.ijrobp.2018.05.053.
- [47] R. T. H. Leijenaar *et al.*, “The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis”, *Sci Rep*, vol. 5, núm. 1, p. 11075, ago. 2015, doi: 10.1038/srep11075.
- [48] E. Pfaehler *et al.*, “Repeatability of ^{18}F -FDG PET radiomic features: A phantom study to explore sensitivity to image reconstruction settings, noise, and delineation method”, *Med Phys*, vol. 46, núm. 2, pp. 665–678, feb. 2019, doi: 10.1002/mp.13322.
- [49] P. Brynolfsson *et al.*, “Haralick texture features from apparent diffusion coefficient (ADC) MRI images depend on imaging and pre-processing parameters”, *Sci Rep*, vol. 7, núm. 1, p. 4041, jun. 2017, doi: 10.1038/s41598-017-04151-4.
- [50] E. Ulrich, “PET-IndiC Extension”. Iowa. Consultado: el 27 de agosto de 2021. [En línea]. Disponible en: <https://www.slicer.org/wiki/Documentation/Nightly/Extensions/PET-IndiC>
- [51] K. V. Mardia, J. T. Kent, y J. M. Bibby, “Multivariate Analysis”, *Academic Press*, 1979.
- [52] J. Peerlings *et al.*, “Stability of radiomics features in apparent diffusion coefficient maps from a multi-centre test-retest trial”, *Sci Rep*, vol. 9, núm. 1, p. 4800, mar. 2019, doi: 10.1038/s41598-019-41344-5.
- [53] G. Collewet, M. Strzelecki, y F. Mariette, “Influence of MRI acquisition protocols and image intensity normalization methods on texture classification”, *Magn Reson Imaging*, vol. 22, núm. 1, pp. 81–91, ene. 2004, doi: 10.1016/j.mri.2003.09.001.
- [54] L. Duron *et al.*, “Gray-level discretization impacts reproducible MRI radiomics texture features”, *PLoS One*, vol. 14, núm. 3, p. e0213459, mar. 2019, doi: 10.1371/journal.pone.0213459.

- [55] R. Cattell, S. Chen, y C. Huang, “Robustness of radiomic features in magnetic resonance imaging: review and a phantom study”, *Vis Comput Ind Biomed Art*, vol. 2, núm. 1, p. 19, dic. 2019, doi: 10.1186/s42492-019-0025-6.
- [56] T. Zoeller, “Intraclass correlation coefficient with confidence intervals”. Consultado: el 29 de octubre de 2023. [En línea]. Disponible en: <https://la.mathworks.com/matlabcentral/fileexchange/26885-intraclass-correlation-coefficient-with-confidence-intervals>
- [57] The MathWorks INC., “MATLAB”. Natick, MA, USA, 2021. Consultado: el 9 de octubre de 2022. [En línea]. Disponible en: <https://www.mathworks.com/products/matlab.html>
- [58] C. Spick *et al.*, “Diffusion-weighted MRI of breast lesions: a prospective clinical investigation of the quantitative imaging biomarker characteristics of reproducibility, repeatability, and diagnostic accuracy”, *NMR Biomed*, vol. 29, núm. 10, pp. 1445–1453, oct. 2016, doi: 10.1002/nbm.3596.
- [59] M. B. Kursya y W. R. Rudnicki, “Feature Selection with the **Boruta** Package”, *J Stat Softw*, vol. 36, núm. 11, 2010, doi: 10.18637/jss.v036.i11.
- [60] V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan, y B. P. Feuston, “Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling”, *J Chem Inf Comput Sci*, vol. 43, núm. 6, pp. 1947–1958, nov. 2003, doi: 10.1021/ci034160g.
- [61] H. He, Y. Bai, E. A. Garcia, y S. Li, “ADASYN: Adaptive synthetic sampling approach for imbalanced learning”, en *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 2008, pp. 1322–1328. doi: 10.1109/IJCNN.2008.4633969.
- [62] V. Romeo *et al.*, “A Simultaneous Multiparametric 18F-FDG PET/MRI Radiomics Model for the Diagnosis of Triple Negative Breast Cancer”, *Cancers (Basel)*, vol. 14, núm. 16, p. 3944, ago. 2022, doi: 10.3390/cancers14163944.
- [63] A. A. van Loevezijn *et al.*, “[18F]FDG-PET/CT in prone compared to supine position for optimal axillary staging and treatment in clinically node-positive

- breast cancer patients with neoadjuvant systemic therapy”, *EJNMMI Res*, vol. 11, núm. 1, p. 78, dic. 2021, doi: 10.1186/s13550-021-00824-4.
- [64] J. H. Choi *et al.*, “Early prediction of neoadjuvant chemotherapy response for advanced breast cancer using PET/MRI image deep learning”, *Sci Rep*, vol. 10, núm. 1, p. 21149, dic. 2020, doi: 10.1038/s41598-020-77875-5.
- [65] T. Sengoz *et al.*, “Role of F-18 FDG PET/CT in Predicting Response to Neoadjuvant Chemotherapy in Invasive Ductal Breast Cancer”, *Eur J Breast Health*, vol. 19, núm. 2, pp. 159–165, abr. 2023, doi: 10.4274/ejbh.galenos.2023.2023-1-3.
- [66] R. Ranjbarzadeh, A. Bagherian Kasgari, S. Jafarzadeh Ghouschi, S. Anari, M. Naseri, y M. Bendeche, “Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images”, *Sci Rep*, vol. 11, núm. 1, p. 10930, may 2021, doi: 10.1038/s41598-021-90428-8.
- [67] W. Yue *et al.*, “Deep learning-based automatic segmentation for size and volumetric measurement of breast cancer on magnetic resonance imaging”, *Front Oncol*, vol. 12, ago. 2022, doi: 10.3389/fonc.2022.984626.
- [68] U. Bashir *et al.*, “The effects of segmentation algorithms on the measurement of 18F-FDG PET texture parameters in non-small cell lung cancer”, *EJNMMI Res*, vol. 7, núm. 1, p. 60, 2017, doi: 10.1186/s13550-017-0310-3.
- [69] N. M. Braman *et al.*, “Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI”, *Breast Cancer Research*, vol. 19, núm. 1, p. 57, dic. 2017, doi: 10.1186/s13058-017-0846-1.
- [70] R. T. H. Leijenaar *et al.*, “The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis”, *Sci Rep*, vol. 5, núm. 1, p. 11075, ago. 2015, doi: 10.1038/srep11075.
- [71] A. Traverso *et al.*, “Sensitivity of radiomic features to inter-observer variability and image pre-processing in Apparent Diffusion Coefficient (ADC) maps of cervix cancer patients”, *Radiotherapy and Oncology*, vol. 143, pp. 88–94, feb. 2020, doi: 10.1016/j.radonc.2019.08.008.

- [72] 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, núm. 1, p. 14163, ago. 2020, doi: 10.1038/s41598-020-70940-z.
- [73] C. Parmar, P. Grossmann, J. Bussink, P. Lambin, y H. J. W. L. Aerts, “Machine Learning methods for Quantitative Radiomic Biomarkers”, *Sci Rep*, vol. 5, núm. 1, p. 13087, ago. 2015, doi: 10.1038/srep13087.
- [74] M. Guarin *et al.*, “18F-FDG-PET/CT and MRI in the Assessment of Neoadjuvant Chemotherapy Treatment Response in Breast Cancer. Correlation with Pathological Response”, San Antonio: San Antonio Breast Cancer Symposium, 2020.
- [75] H. Li *et al.*, “MRI and PET/CT for evaluation of the pathological response to neoadjuvant chemotherapy in breast cancer: A systematic review and meta-analysis”, *The Breast*, vol. 40, pp. 106–115, 2018, doi: <https://doi.org/10.1016/j.breast.2018.04.018>.
- [76] H. Yoon, Y. Kim, J. Chung, y B. S. Kim, “Predicting neo-adjuvant chemotherapy response and progression-free survival of locally advanced breast cancer using textural features of intratumoral heterogeneity on F-18 <sc>FDG PET</sc> / <sc>CT</sc> and diffusion-weighted <sc>MR</sc> imaging”, *Breast J*, vol. 25, núm. 3, pp. 373–380, may 2019, doi: 10.1111/tbj.13032.
- [77] M. Sollini *et al.*, “PET/CT radiomics in breast cancer: Mind the step”, *Methods*, vol. 188, pp. 122–132, abr. 2021, doi: 10.1016/j.ymeth.2020.01.007.
- [78] H. Horng *et al.*, “Generalized ComBat harmonization methods for radiomic features with multi-modal distributions and multiple batch effects”, *Sci Rep*, vol. 12, núm. 1, p. 4493, mar. 2022, doi: 10.1038/s41598-022-08412-9.
- [79] T. K. Koo y M. Y. Li, “A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research”, *J Chiropr Med*, vol. 15, núm. 2, pp. 155–163, jun. 2016, doi: 10.1016/j.jcm.2016.02.012.
- [80] S. ying Huang *et al.*, “Exploration of PET and MRI radiomic features for decoding breast cancer phenotypes and prognosis”, *NPJ Breast Cancer*, vol. 4, núm. 1, 2018, doi: 10.1038/s41523-018-0078-2.

-
- [81] D. Leithner *et al.*, “Radiomic Signatures Derived from Diffusion-Weighted Imaging for the Assessment of Breast Cancer Receptor Status and Molecular Subtypes”, *Mol Imaging Biol*, vol. 22, núm. 2, pp. 453–461, abr. 2020, doi: 10.1007/s11307-019-01383-w.