

anatómicas de interés.

5. **Postprocesado morfológico:** se corrigen huecos y se eliminan pequeños objetos aislados mediante operaciones de cierre y apertura morfológica, garantizando la continuidad anatómica.

En este estudio, se han extraído de forma sistemática las máscaras correspondientes a la sustancia gris y sustancia blanca para todos los casos de la base de datos. La herramienta SPM12, en la que se basa CTseg, cuenta con una amplia validación en la literatura, que avala su precisión y fiabilidad tanto en poblaciones sanas como en pacientes con patología neurológica (Adduru et al., 2020; Dos Santos et al., 2024; Yelanchezian et al., 2025).

En el contexto de este estudio, la presencia de hemorragia no interfiere de forma sustancial en la segmentación de la materia gris y blanca. Esto se debe a que las áreas con alta atenuación, producidas por la sangre, son clasificadas en su mayoría como LCR dentro del esquema de SPM12, minimizando así el impacto sobre la radiómica extraída de las estructuras parenquimatosas.

La Figura 3.7 muestra un ejemplo representativo de segmentación: el azul corresponde a materia gris, el gris oscuro a materia blanca, el naranja a LCR, el gris claro al cráneo y el morado al tejido blando extracraneal. Las etiquetas verdes y rojas indican las regiones de HSA y sus variantes, evidenciando cómo el sangrado queda claramente diferenciado de las estructuras cerebrales segmentadas.

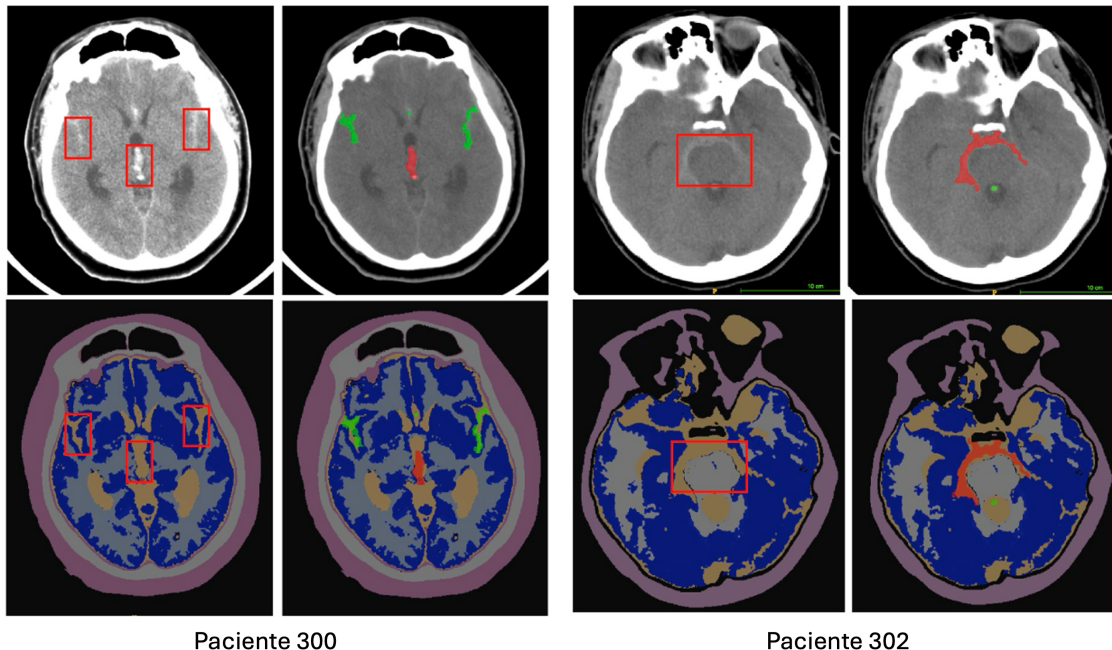


Figura 3.7: Segmentación del tejido cerebral utilizando CTseg en dos pacientes ejemplo. La columna derecha de cada paciente muestra la imagen original con la segmentación del cerebro. En la columna izquierda también se representa la segmentación de la hemorragia (Urbanos et al., 2025).

3.2.3 Preprocesado de datos

El preprocesado de datos se plantea de forma diferenciada según las características de cada modalidad y los objetivos de cada flujo analítico. En el ámbito de la imagen, se incluyen transformaciones específicas para la calibración espectral en HSI de GBM, el entrenamiento del algoritmo de segmentación automática en HSAa y el análisis radiómico en HSAa. En cuanto a los datos clínicos, el preprocesado se orienta al manejo de valores faltantes y a la codificación de variables.

Preprocesado de imagen

Imagen Hiperespectral en glioblastoma

Las imágenes hiperespectrales se preprocesan con el objetivo de homogeneizar las firmas espectrales de las imágenes etiquetadas. La cadena de preprocesamiento consta de cuatro pasos principales:

1. **Creación de cubos hiperespectrales:** Las imágenes sin procesar obtenidas con la cámara *snapshot* se convierten en cubos hiperespectrales. El área activa de la imagen cruda tiene una resolución de 2045×1085 píxeles y está compuesta por 25 longitudes de onda dispuestas en un mosaico repetitivo de 5×5 píxeles. Por lo tanto, el cubo hiperespectral resultante presenta una resolución espacial de 419×217 píxeles y una resolución espectral de 25 bandas.
2. **Calibración:** Este paso es necesario para mantener la reproducibilidad de los datos, independientemente de las diferentes condiciones de iluminación en cada cirugía. Para calibrar las imágenes, se capturan imágenes de referencia en blanco y en negro:
 - *Referencia blanca:* obtenida mediante un objetivo de reflectancia difusa lambertiana (SphereOptics GmbH, 2024), que refleja el 95 % de la luz incidente en todo el rango espectral cubierto por la cámara. La captura se realiza colocando el objetivo a la misma distancia y orientación que la superficie cerebral in vivo, bajo las mismas condiciones de iluminación.
 - *Referencia negra:* obtenida cubriendo la lente de la cámara.

La imagen calibrada banda a banda (I_c) se obtiene mediante:

$$I_c = \frac{I - D}{W - D} \quad (3.1)$$

donde I es una banda del cubo hiperespectral capturado, D la banda correspondiente de la referencia negra y W la misma banda de la referencia blanca.

3. **Corrección espectral:** La elevada sensibilidad de la cámara instantánea hace que las bandas espectrales vecinas se solapen parcialmente, provocando que la señal capturada en una banda incluya información procedente de bandas adyacentes (diafonía espectral). Esto genera armónicos secundarios que no se pueden eliminar con filtros ópticos. Este efecto se mitiga multiplicando la señal por una matriz de corrección espectral (SCM), proporcionada por el fabricante:

$$I_{sc} = I_c \times SCM \quad (3.2)$$

donde SCM es una matriz de 25×25 cuyos coeficientes corrigen la respuesta espectral de cada banda virtual.

4. **Normalización de datos:** Debido a la curvatura natural del cerebro, los píxeles no se encuentran a la misma altura, lo que provoca variaciones en la intensidad de luz recogida por la cámara. Para evitar que estas variaciones influyan en la clasificación de tejidos, se aplica una normalización que preserve la forma de la firma espectral, independientemente de su amplitud.

Se calcula el coeficiente de normalización de cada píxel como el valor RMS de su firma espectral:

$$coef[r, c] = \sqrt{\frac{1}{B} \sum_{b=1}^B (I_{sc}[r, c, b])^2} \quad (3.3)$$

donde B es el número de bandas, r la fila y c la columna.

La normalización del cubo corregido espectralmente se obtiene mediante:

$$I_{norm}[r, c, b] = \frac{I_{sc}[r, c, b]}{coef[r, c]} \quad (3.4)$$

TC en Hemorragia Subaracnoidea Aneurismática – Segmentación automática

Como primer paso, todas las imágenes en formato DICOM se convierten a formato NIfTI, con el fin de facilitar su manejo y garantizar la compatibilidad con las herramientas de procesamiento empleadas. En el caso de las máscaras de segmentación, todas las hemorragias (independientemente de su tipo o localización) se unifican bajo una única etiqueta (valor 1), ya que el objetivo de esta fase es entrenar el modelo para segmentar la clase global de hemorragia, sin establecer distinciones entre sus distintos subtipos.

En la fase de segmentación automática de la HSAa se emplea el marco **nnU-Net**, descrito en la Sección 2.7.2. Esta arquitectura de segmentación totalmente automatizada integra un conjunto de rutinas de preprocesamiento que se ajustan dinámicamente a las características del conjunto de datos de entrada. Dichas rutinas forman parte del flujo interno de **nnU-Net** y se ejecutan de manera sistemática antes del entrenamiento y la inferencia, con el propósito de homogeneizar la información espacial y de intensidades entre pacientes, favoreciendo así un aprendizaje más estable y generalizable del modelo.

El preprocesamiento automático de **nnU-Net** incluye los siguientes pasos principales:

- **Normalización de intensidades:** para datos de TC, **nnU-Net** aplica una normalización basada en la distribución de intensidades dentro de la máscara de tejido, utilizando valores de media y DE calculados a partir del conjunto de entrenamiento. Las intensidades se reescalan para reducir la variabilidad debida a diferencias en calibración y protocolos de adquisición, manteniendo la información relevante para el contraste tisular.
- **Resampling a espaciamiento isotrópico:** todas las imágenes se reinterpolan a un *voxel spacing* isotrópico definido automáticamente en función de la resolución mediana del conjunto de entrenamiento. Este paso garantiza que las dimensiones físicas de los

vóxeles sean homogéneas, evitando que el modelo aprenda sesgos relacionados con la anisotropía del espaciado original.

- **Ajuste de tamaño y recorte automático:** las imágenes se recortan a la región que contiene al paciente para eliminar zonas sin información relevante (aire, mesa de exploración) y se ajustan a un tamaño de entrada común mediante operaciones de *padding* o recorte centrado, según corresponda.
- **Generación y almacenamiento de metadatos:** durante el preprocesamiento, nnU-Net guarda la información de transformación aplicada (espaciado original, valores de normalización, coordenadas de recorte) en archivos auxiliares, lo que permite revertir las transformaciones o transferir las predicciones a las coordenadas originales del paciente.
- **Estandarización de máscaras de referencia:** las segmentaciones manuales (*ground truth*) se reinterpolan y ajustan al mismo espacio y tamaño que las imágenes preprocesadas, asegurando la correspondencia vóxel a vóxel entre la imagen de entrada y la etiqueta.

La principal ventaja de este enfoque radica en que nnU-Net adapta automáticamente los parámetros de preprocesamiento a las propiedades estadísticas del conjunto de datos, evitando la necesidad de ajustes manuales y garantizando reproducibilidad. Asimismo, esta estandarización facilita la combinación de datos heterogéneos, como en este estudio, donde se integraron imágenes de la cohorte local y de bases de datos públicas con diferentes resoluciones, protocolos y características de adquisición.

TC en Hemorragia Subaracnoidea Aneurismática – Modelado radiómico

En el flujo de trabajo radiómico, las imágenes de TC se someten a un preprocesamiento específico orientado a garantizar la comparabilidad y reproducibilidad de las características extraídas entre pacientes, en consonancia con las directrices de la guía *TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis)* (Collins et al., 2015b). Dicha guía establece criterios destinados a asegurar la claridad, reproducibilidad y transparencia metodológica en el desarrollo y evaluación de modelos predictivos. El cumplimiento de estas recomendaciones se detalla en el Anexo 2 (Sección B.2), donde se especifican los ítems aplicados en el presente estudio.

Los pasos de preprocesamiento implementados son los siguientes:

- **Resampling a espaciamiento isotrópico:** Todas las imágenes se interpolan a un tamaño de vóxel uniforme de $1 \times 1 \times 1 \text{ mm}^3$, utilizando interpolación bilineal para las imágenes y *nearest neighbor* para las máscaras de segmentación. La interpolación bilineal calcula la intensidad en una nueva coordenada (x, y, z) como combinación ponderada de los valores de los vóxeles vecinos:

$$I'(x, y, z) = \sum_{i=0}^1 \sum_{j=0}^1 \sum_{k=0}^1 w_{ijk} I(x_i, y_j, z_k),$$

donde w_{ijk} son los coeficientes de interpolación que dependen de la distancia relativa a los vóxeles originales. En el caso de las máscaras, se emplea interpolación de vecino más

cercano:

$$M'(x, y, z) = M\left(\arg \min_{(x_i, y_j, z_k)} \|(x, y, z) - (x_i, y_j, z_k)\|\right),$$

preservando así etiquetas discretas sin generar valores intermedios inválidos (Thévenaz et al., 2002).

- **Discretización de intensidades:** Las intensidades en UH se discretizan empleando un ancho de bin fijo de 25 UH. El valor discretizado I_d se obtiene como:

$$I_d = \left\lfloor \frac{I}{\Delta} \right\rfloor, \quad \Delta = 25 \text{ UH},$$

donde I es la intensidad original y Δ el tamaño de bin. Este procedimiento reduce la sensibilidad a fluctuaciones menores y facilita la comparabilidad entre estudios.

- **Normalización Z:** Antes de la extracción de características, se aplica una normalización tipo Z (*z-score*) a todas las imágenes, transformando cada intensidad I según:

$$I_z = \frac{I - \mu}{\sigma},$$

donde μ es la media de intensidades y σ la DE de la imagen. Esta estandarización garantiza distribuciones comparables entre pacientes y reduce el impacto de variaciones técnicas o biológicas ajenas a la patología.

Preprocesado de datos clínicos

La base de datos clínica inicial incluye un total de 49 variables recogidas de forma sistemática en el momento del ingreso hospitalario. Dada la heterogeneidad en la disponibilidad de estos registros a lo largo del periodo de estudio (2007–2024) y las variaciones en la práctica clínica, se ha aplicado un proceso de depuración orientado a minimizar el impacto de los datos faltantes y garantizar la consistencia del conjunto final.

Filtrado y selección de variables.

Primero, para la selección de variables se aplica un umbral de exclusión basado en la proporción de valores ausentes por variable. La literatura recomienda evitar análisis multivariados cuando faltan más del 10–20 % de los datos, dado que proporciones superiores pueden introducir sesgos y comprometer la validez de las conclusiones (Madley-Dowd et al., 2019; Yeatts y Martin, 2015). En consecuencia, en este estudio se adopta un umbral intermedio del 15 % como criterio de descarte. Además, a partir de la *fecha de nacimiento* y la *fecha de HSA*, se calculó de forma sistemática la edad del paciente al evento, como medida homogénea y clínicamente relevante de un factor pronóstico bien establecido.

Tratamiento de valores ausentes.

Para abordar la presencia de datos faltantes en las variables clínicas, se ha utilizado el método de imputación múltiple por ecuaciones encadenadas (*Multiple Imputation by Chained Equations*, MICE) (van Buuren y Groothuis-Oudshoorn, 2011), técnica ampliamente validada en investigación clínica y epidemiológica. A diferencia de la imputación simple, MICE genera

varias versiones completas del conjunto de datos mediante un proceso iterativo en el que cada variable con valores ausentes se modela a partir del resto. En este estudio se realizan cinco imputaciones independientes, empleando modelos de regresión lineal para variables continuas y regresión logística para categóricas. Las estimaciones finales se combinan conforme a las reglas de Rubin (Rubin, 2018), que integran la variabilidad dentro y entre imputaciones para obtener una estimación global del parámetro y su varianza total.

Evaluación de la comparabilidad entre particiones de datos.

Finalmente, se verifican la comparabilidad entre los conjuntos de entrenamiento/validación y test tras la partición de datos. Para las variables categóricas se aplica la prueba de chi-cuadrado (Agresti et al., 1996), estimando el tamaño del efecto mediante el coeficiente Phi en tablas 2×2 y la V de Cramér en tablas con más categorías (Cramér, 1999). Para las variables continuas, se emplean pruebas t para muestras independientes (Student, 1908), complementadas con la d de Cohen como medida de tamaño del efecto (Cohen, 2013). Este análisis permite identificar posibles desequilibrios entre particiones y garantizar la comparabilidad clínica de las cohortes antes del modelado.

3.3 Extracción de características

La Figura 3.8 ilustra el tercer bloque metodológico, centrado en la obtención de descriptores cuantitativos mediante la extracción de firmas espectrales en HSI y características radiómicas en TC, garantizando la reproducibilidad conforme a la guía IBSI.

3.3 Extracción de características

- **3.3.1 Firmas espectrales (HSI):** obtención de curvas espectrales por tejido.
- **3.3.2 Características radiómicas (TC):** extracción de radiómica de primer, segundo y tercer orden.

Figura 3.8: Bloque 3.3. Extracción de características: obtención de descriptores espectrales (HSI) y radiómicos (TC).

3.3.1 Firmas espectrales (HSI)

Como se explica en la Sección 2.6.2, la firma espectral de un píxel se define como el vector formado por sus valores de reflectancia en todas las longitudes de onda capturadas por el cubo hiperespectral. Esta información constituye la unidad básica de análisis y clasificación en HSI, ya que describe de forma única la interacción de la luz con el tejido en un punto determinado.

En este estudio, la extracción de firmas espectrales se realiza a partir de las máscaras de anotación generadas con la herramienta descrita en la Sección 3.2.2. Dichas máscaras, alineadas y superpuestas sobre el cubo hiperespectral preprocesado, permiten identificar los píxeles correspondientes a cada categoría tisular definida por el especialista.

Procedimiento.

1. Se carga el cubo hiperespectral normalizado I_{norm} junto con su máscara de anotación correspondiente.
2. Para cada píxel (r, c) incluido en una máscara de clase, se extrae la secuencia ordenada de intensidades $I_{norm}[r, c, b]$, donde $b \in \{1, \dots, B\}$ corresponde a las B bandas espectrales.
3. Cada vector espectral se almacena junto con su etiqueta de clase y los metadatos del paciente, garantizando la trazabilidad completa de su procedencia.

3.3.2 Características radiómicas (TC)

La extracción de características radiómicas se ha llevado a cabo utilizando la librería `PyRadiomics v3.0,1` (Van Griethuysen et al., 2017), una herramienta ampliamente utilizada para la cuantificación de biomarcadores radiómicos. Previamente a esta fase, todas las imágenes y máscaras de segmentación se han sometido al proceso de preprocesamiento descrito en la Sección 3.2.3, con el fin de garantizar la homogeneidad espacial y la reproducibilidad de las características radiómicas.

Tipos de características extraídas

Se han extraído tres categorías principales de características radiómicas, que en conjunto proporcionan una descripción multidimensional de la morfología, la intensidad y la heterogeneidad interna de las regiones segmentadas:

- **Características de forma (*shape*):** cuantifican propiedades geométricas tridimensionales de las regiones de interés, tales como volumen, área superficial, compacidad, esfericidad, elongación y planicidad. Estas métricas son independientes de las intensidades y reflejan directamente la morfología de la estructura segmentada.
- **Estadísticas de primer orden (*first-order*):** describen la distribución de intensidades de vóxel sin considerar su relación espacial. Incluyen métricas como media, mediana, DE, asimetría, curtosis, energía, entropía y percentiles.
- **Texturas de segundo orden (*textural*):** capturan relaciones espaciales entre intensidades, reflejando la heterogeneidad interna de la lesión o tejido. Incluyen características derivadas de:
 - *GLCM*: contraste, homogeneidad, correlación, energía, entropía, entre otros.
 - *GLRLM*: longitudes de corridas, uniformidad y variabilidad en la repetición de intensidades.
 - *GLSZM*: tamaño y homogeneidad de zonas de intensidad uniforme.
 - *NGTDM*: diferencia de tonos de gris respecto a su vecindad.
 - *GLDM*: grado de dependencia entre intensidades de vóxeles vecinos.

Características derivadas mediante filtrado (radiómica de tercer orden).

Con el objetivo de aumentar la sensibilidad del análisis a patrones específicos y mejorar la capacidad de generalización de los modelos, se han aplicado transformaciones de filtrado a las imágenes originales antes de la extracción de características, generando así radiómica de tercer orden:

- **Filtro cuadrado:** eleva al cuadrado el valor de cada vóxel y lo escala linealmente, conservando los valores negativos. Resalta intensidades altas, incrementando el peso de regiones hiperdensas.
- **Filtro exponencial:** aplica la función $e^{|x|}$ sobre cada vóxel y reescala posteriormente, conservando la simetría de la distribución. Útil para resaltar diferencias sutiles en intensidades medias.
- **Filtro logarítmico:** calcula $\log(\text{intensidad} + 1)$ y reescala, atenuando el rango dinámico de intensidades muy altas, lo que permite un mejor contraste en estructuras con alta heterogeneidad.
- **Filtro *Wavelet*:** emplea la transformada de Haar en tres dimensiones para descomponer la imagen en ocho sub-bandas: LLL, LLH, LHL, LHH, HLL, HLH, HHL y HHH (L: paso bajo, H: paso alto). Cada sub-banda enfatiza diferentes combinaciones de frecuencias espaciales y orientaciones, extrayendo patrones texturales en distintas escalas.
- **Filtro Laplaciano de Gauss (*LoG*):** resalta bordes y transiciones bruscas de intensidad, calculando el laplaciano de la imagen suavizada mediante un núcleo gaussiano. Se aplicaron tres valores de σ (0,5; 3; 5 mm), abarcando desde detalles finos (sigma bajo) hasta estructuras más gruesas (sigma alto).

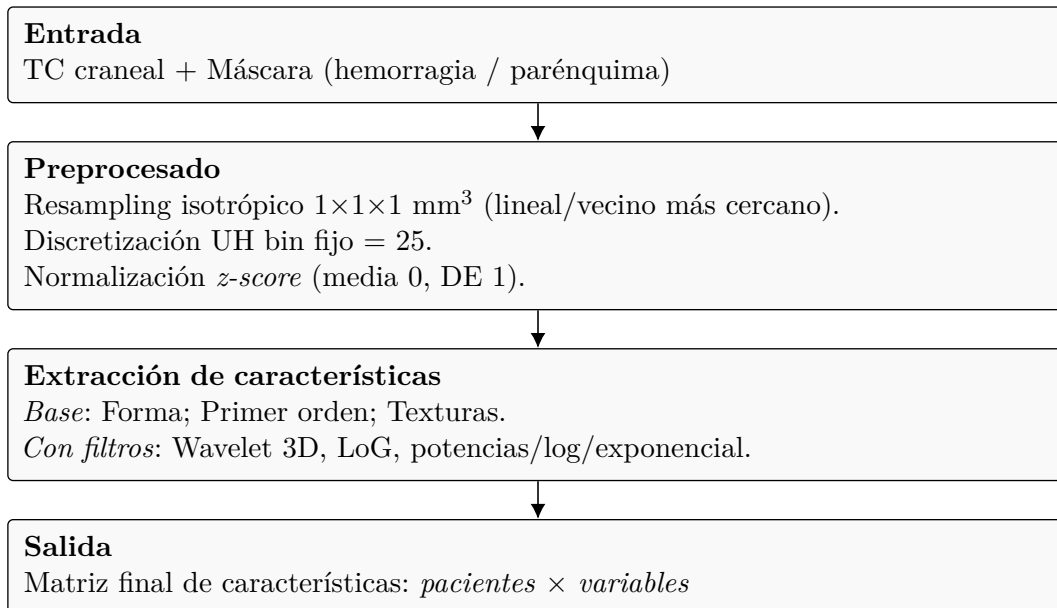


Figura 3.9: Flujo simplificado de radiómica: entrada, preprocesado, extracción y salida.

3.4 Modelado y evaluación

El cuarto bloque metodológico (Figura 3.10) resume el desarrollo, validación e interpretación de los modelos predictivos, incluyendo los experimentos de clasificación tisular en HSI-GBM, segmentación automática en TC-HSAa y predicción radiómica multimodal, junto con las métricas empleadas para su evaluación.

Para la HSI en GBM, se presenta el modelado orientado a la clasificación tisular, incluyendo el entrenamiento, la validación y la evaluación de distintos algoritmos. Finalmente, se especifican las métricas de evaluación adoptadas para cada tarea, con su definición y justificación según el contexto clínico y metodológico.

En el caso de la HSAa en TC, se detalla el diseño experimental de la segmentación automática de la hemorragia, las arquitecturas y configuraciones utilizadas, su evaluación cuantitativa y cualitativa, y la comparación con modelos externos, así como su comparación con escalas clínicas.

En el análisis radiómico, se describen los procedimientos de construcción de modelos predictivos integrando variables de imagen y clínicas, las estrategias de validación empleadas, y el análisis de interpretabilidad y comparación y relación con escalas convencionales.

Los detalles de implementación a nivel *hardware* y *software* utilizados para llevar a cabo estos experimentos se encuentran detallados en el Anexo D.

3.4 Modelado y evaluación

- **3.4.1 Clasificación tisular (HSI-GBM):** Diseño experimental: experimentos intra/interpaciente; algoritmos utilizados; estrategias de entrenamiento.
- **3.4.2 Segmentación (HSAa-TC):** diseño y escenarios experimentales; arquitectura y configuración de modelos: *nnU-Net*; evaluación cuantitativa, cualitativa y comparación con modelos externos. Evaluación del impacto clínico: relación con escala de Fisher modificada.
- **3.4.3 Modelos radiómicos (TC):** diseño experimental, flujo de desarrollo y validación de modelos, integración multimodal, interpretabilidad y comparación con escalas clínicas.
- **3.4.4 Métricas de evaluación:** Justificación de métricas utilizadas.

Figura 3.10: Bloque 3.4. Modelado y evaluación: desarrollo, validación e interpretación de modelos para clasificación, segmentación y predicción radiómica.

3.4.1 Clasificación tisular en HSI GBM

Esta subsección describe el modelado aplicado a la clasificación de tejidos en HSI intraoperatorias de GBM. Se presentan los algoritmos evaluados, el diseño experimental, las estrategias de entrenamiento y validación de los modelos desarrollados.

Diseño experimental

En este trabajo se emplea un esquema de validación cruzada estratificada de 5 *folds*, asegurando que cada división mantiene la proporción original de píxeles por clase mediante *muestreo aleatorio estratificado*. Esta técnica consiste en dividir el conjunto de datos en varias particiones o pliegues (*folds*). En cada iteración, el modelo se entrena con todos los *folds* menos uno y se evalúa en el *fold* restante, que en este contexto actúa como conjunto de **test interno**. El proceso se repite tantas veces como *folds* se hayan definido, de modo que cada partición es utilizada una vez como test interno. El rendimiento global del modelo se obtiene promediando los resultados de todas las iteraciones, lo que proporciona una estimación más robusta que una única división entrenamiento–test.

El objetivo principal de esta estrategia es evitar el sesgo por filtrado de información (*data leakage*) y obtener una estimación más realista de la capacidad de generalización de los modelos. Para ello, se aplica validación cruzada anidada (*nested cross-validation*), que combina dos (o más) niveles jerárquicos de validación, como se ilustra en la Figura 3.11:

- El **bucle externo** se destina a estimar el rendimiento del modelo en datos no vistos durante el entrenamiento.
- El **bucle interno** se utiliza exclusivamente para la selección del modelo, sin acceso a los datos reservados en el bucle externo.

En función del origen de los datos de entrenamiento y test, se plantearon dos configuraciones:

1. **Validación cruzada doble (5×5)**: esquema con dos bucles anidados. El bucle interno se utilizó para la selección del mejor modelo, mientras que el bucle externo se destinó a su evaluación.
2. **Validación cruzada triple (5×5×5)**: incorpora un bucle adicional previo que, para cada paciente, divide aleatoriamente los píxeles en 80 % entrenamiento y 20 % test definitivo. Así, el bucle más externo controla la separación paciente a paciente, mientras que los bucles internos realizan la selección y evaluación del modelo. Esta variante se realiza cuando entrenamiento y test proceden del mismo paciente, ya que una división aleatoria simple puede introducir sesgo.

Definición de experimentos

El análisis de HSI en GBM busca establecer métodos automáticos de clasificación tisular que permitan diferenciar entre tejido sano, tumoral, duramadre y componentes vasculares durante la cirugía. En este contexto, se han diseñado dos experimentos con enfoques complementarios representados en la Figura 3.12.

Experimento A — Clasificación intrapaciente Este experimento constituye una primera aproximación destinada a valorar el potencial discriminativo de los clasificadores en condiciones controladas, sin introducir variabilidad interpaciente de forma explícita. Este planteamiento permite estimar el rendimiento máximo alcanzable cuando los datos de entrenamiento y test son muy similares. En cada paciente, el conjunto de datos se divide en un 80 % para

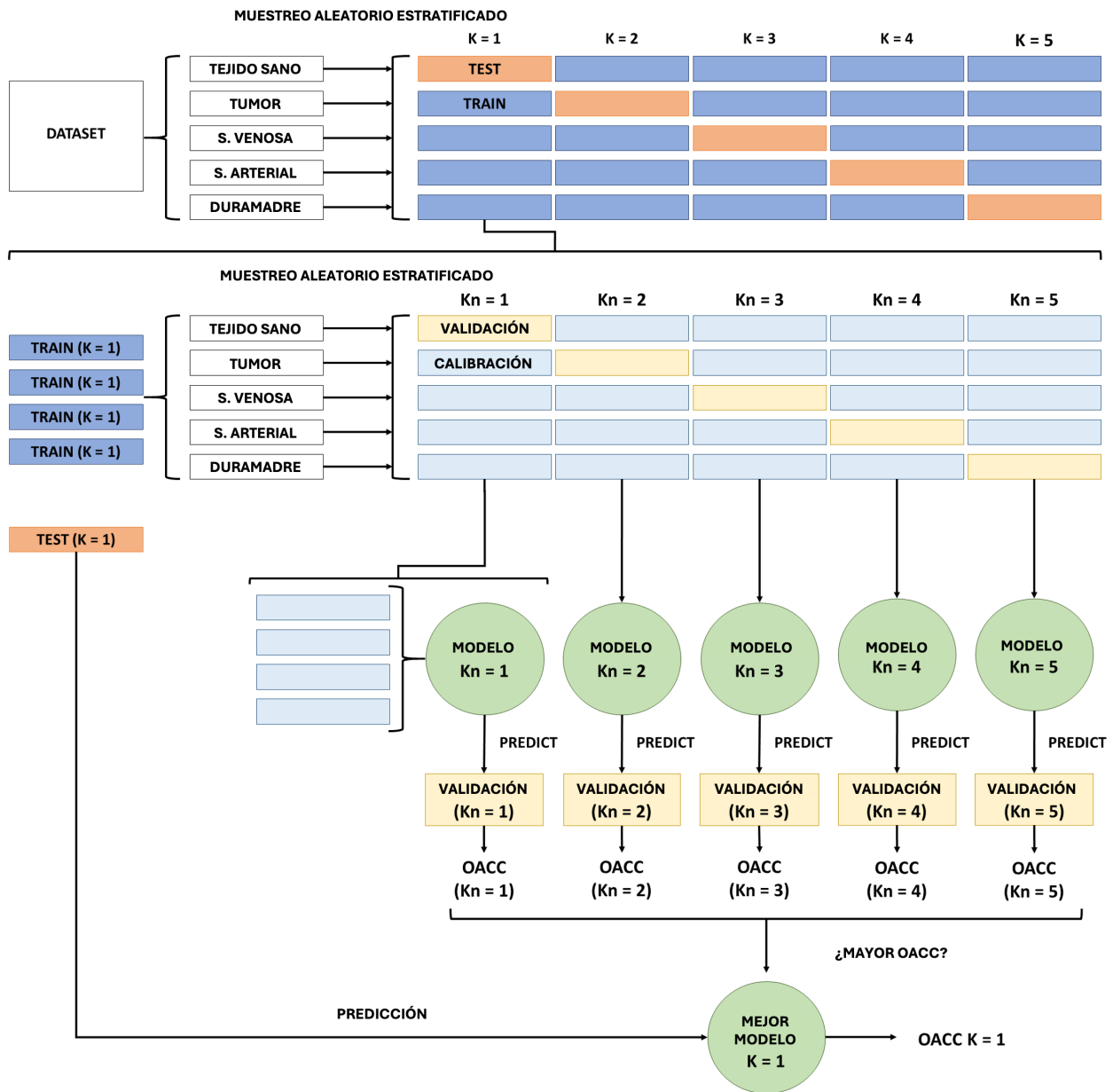


Figura 3.11: Esquema del procedimiento de validación cruzada anidada doble utilizado para el entrenamiento de los modelos. Por simplicidad, únicamente se ilustran los bucles externo e interno correspondientes a la combinación $K = 1$; el resto de combinaciones K_n siguen un procedimiento análogo (Urbanos et al., 2021).

entrenamiento y un 20 % para test, manteniendo la proporción de píxeles por clase mediante muestreo estratificado. Para seleccionar la partición de entrenamiento y evitar sesgos derivados de la elección aleatoria, se aplica una validación cruzada anidada triple ($5 \times 5 \times 5$ *folds*): el bucle más externo fija la separación entrenamiento–test, el bucle intermedio actúa como test interno y el bucle más interno como validación interna.

El modelo resultante de este proceso para cada algoritmo se utiliza posteriormente para clasificar la totalidad de las imágenes del paciente correspondiente. Este planteamiento proporciona una estimación del rendimiento máximo esperable en escenarios en los que las características del conjunto de test externo son muy próximas a las del conjunto de entrenamiento.

Experimento B — Clasificación entre pacientes Este experimento se concibe como un paso intermedio hacia un escenario más realista, en el que se incluye explícitamente la variabilidad entre pacientes. Aquí, cada modelo se entrena con datos de varios individuos y se evalúa en un paciente no visto previamente, lo que proporciona una medida de la capacidad de generalización en condiciones próximas a la práctica clínica. Se ha implementado un esquema *leave-one-patient-out*, en el que cada iteración deja fuera a un paciente como conjunto de test externo y entrena el modelo con las imágenes de los demás. El paciente excluido se mantiene completamente aislado durante el proceso, utilizándose únicamente para la evaluación final. La selección del modelo en cada iteración se realiza mediante validación cruzada anidada doble (5×5 *folds*) sobre el conjunto de entrenamiento. Este diseño proporciona una estimación más realista de la capacidad de generalización del modelo a nuevos pacientes, reproduciendo un escenario clínico en el que no se dispone de ejemplos previos del individuo evaluado.

Estrategias de entrenamiento

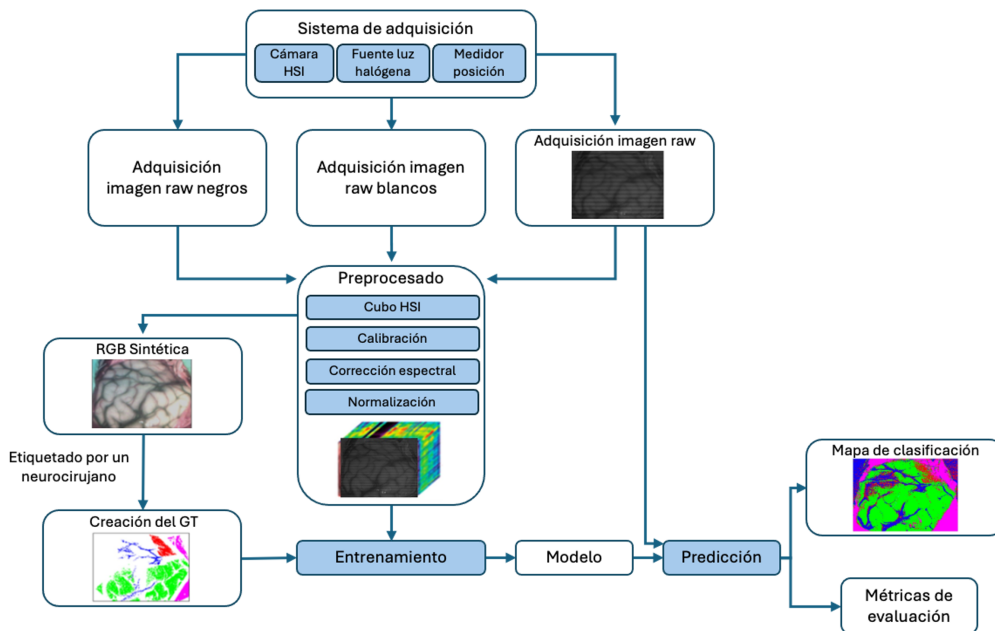
Se han seleccionado tres algoritmos supervisados ampliamente empleados en clasificación de imágenes médicas: SVM, RF y 3D-CNN (Fabelo et al., 2016; Shen et al., 2017).

Cada uno de ellos aporta distintas fortalezas en cuanto a capacidad de generalización, interpretabilidad y manejo de información multidimensional, por lo que su inclusión permite una evaluación comparativa equilibrada entre métodos basados en aprendizaje clásico y profundo.

Máquina de vectores de soporte (*Support Vector Machine*) La fundamentación teórica y los principios de funcionamiento de SVM se describen en la Sección 2.7.1. En este estudio se ha optado por un clasificador SVM lineal, dado que su simplicidad y robustez lo convierten en una referencia habitual en tareas de clasificación espectral en imagen biomédica. Trabajos previos en detección intraoperatoria de cáncer cerebral (Fabelo, Ortega, Lazcano et al., 2018) han mostrado su efectividad en contextos análogos. Los parámetros específicos empleados durante su entrenamiento se detallan en el Anexo A, Tabla A.3.

Bosques aleatorios (*Random Forest*) El algoritmo RF, descrito teóricamente en la Sección 2.7.1, se incluye por su reconocida capacidad para modelar relaciones no lineales y su

Experimento A



Experimento B

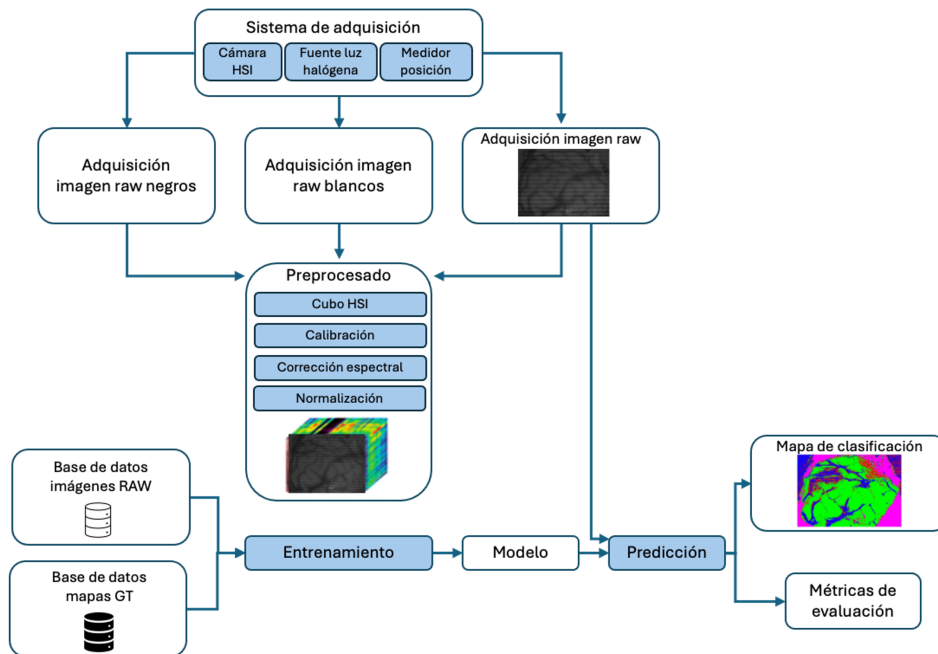


Figura 3.12: Esquema de los dos diseños experimentales. En el Experimento A (clasificación intrapaciente), el modelo se entrena y evalúa dentro de cada paciente, estimando el rendimiento en condiciones óptimas. En el Experimento B (clasificación entre pacientes), se aplica un enfoque *leave-one-patient-out* para valorar la capacidad de generalización a nuevos individuos (Urbanos et al., 2021).

resistencia al sobreajuste. Su uso ha demostrado un rendimiento competitivo en estudios previos de clasificación hiperespectral (Black et al., 2024; Leon et al., 2023), donde la combinación de múltiples árboles de decisión favorece la estabilidad del modelo frente a la variabilidad espectral de los tejidos. La configuración empleada, junto con los valores de los hiperparámetros, se presenta igualmente en el Anexo A, Tabla A.3.

Red neuronal convolucional tridimensional (3D-CNN) Finalmente, se ha implementado una arquitectura basada en redes convolucionales tridimensionales, diseñada para explotar la información espacial y espectral de forma conjunta. Este enfoque permite analizar simultáneamente la estructura local del tejido en el plano espacial y su comportamiento espectral a lo largo de las 25 longitudes de onda registradas. El modelo procesa parches tridimensionales centrados en el píxel de interés y extrae representaciones jerárquicas mediante bloques convolucionales 3D, seguidos de operaciones de normalización y activación no lineal. Posteriormente, las características extraídas se integran en capas densas con regularización por *dropout*, finalizando en una capa *softmax* para la clasificación multiclase (Figura 3.13). Los detalles completos de la arquitectura y los parámetros de entrenamiento se recogen en el Anexo A, Tabla A.3.

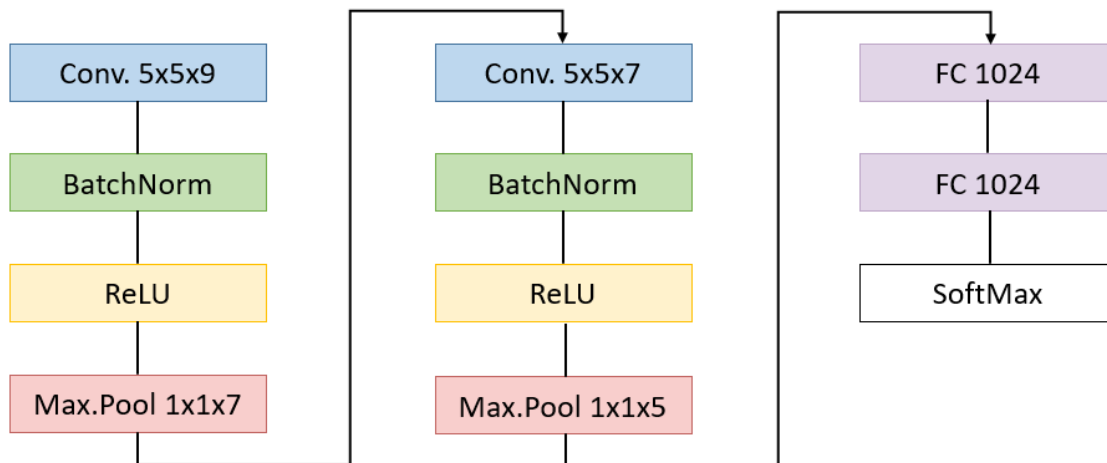


Figura 3.13: Esquema de la arquitectura 3D-CNN empleada para la clasificación hiperespectral de tejidos cerebrales.

Procedimiento de entrenamiento y evaluación con validación cruzada anidada

El flujo de entrenamiento y evaluación se estructura siguiendo el esquema de validación cruzada anidada descrito en la Sección 3.4.1, adaptando la estructura a las variantes doble o triple según el diseño del experimento.

Bucle externo — Evaluación en test

1. El conjunto de datos se divide en 5 *folds* externos.

2. En cada iteración, 4 *folds* se utilizan para entrenamiento y 1 se reserva como conjunto de **test**. El conjunto de test se mantiene completamente aislado durante todo el proceso, sin intervenir en la selección ni en el ajuste del modelo.

Bucle interno — Selección de hiperparámetros

1. Los 4 *folds* de entrenamiento del bucle externo se subdividen de nuevo en 5 *folds* internos, manteniendo la proporción de píxeles por clase.
2. En cada división, se entrena un modelo con los pliegues de entrenamiento y se evalúa sobre el pliegue retenido como **validación**.
3. Se comparan los resultados en validación y se selecciona la configuración de hiperparámetros con mejor rendimiento basado en la métrica precisión global (OACC, de sus singlas en inglés *Overall Accuracy*).

Entrenamiento final y evaluación

1. Con la configuración óptima seleccionada en el bucle interno, el modelo se reentrena utilizando los 4 *folds* externos completos.
2. Este modelo final se evalúa sobre el conjunto de **test** reservado en el bucle externo, generando una estimación de rendimiento libre de sesgo.

Validación cruzada triple (cuando aplica)

En los casos en los que los datos de entrenamiento y prueba proceden del mismo paciente, se introduce un bucle adicional previo a la validación cruzada. En este bucle más externo, el conjunto de datos de cada paciente se divide aleatoriamente en un 80 % para **entrenamiento** y un 20 % para **test**. De este modo, la validación cruzada anidada (5×5) se aplica únicamente sobre el 80 % de entrenamiento, quedando el 20 % reservado como conjunto de evaluación final. Este procedimiento añade un nivel extra de control sobre la partición inicial, reduciendo el riesgo de que la selección aleatoria de datos favorezca artificialmente el rendimiento del modelo.

La Figura 3.14 muestra el flujo completo de modelado en los dos experimentos realizados, desde la partición y validación de datos hasta el cálculo de las métricas de evaluación (descritas en la Sección 3.4.4).

3.4.2 Segmentación automática de Hemorragia Subaracnoidea Aneurismática

Esta subsección presenta el diseño experimental y la configuración de los modelos desarrollados para la segmentación automática de la HSAa en TC. Se describen la arquitectura utilizada así como las estrategias de validación y comparación con modelos externos. La evaluación incluye análisis cuantitativo y cualitativo de resultados y una exploración del impacto clínico mediante su relación con la escala de Fisher modificada.

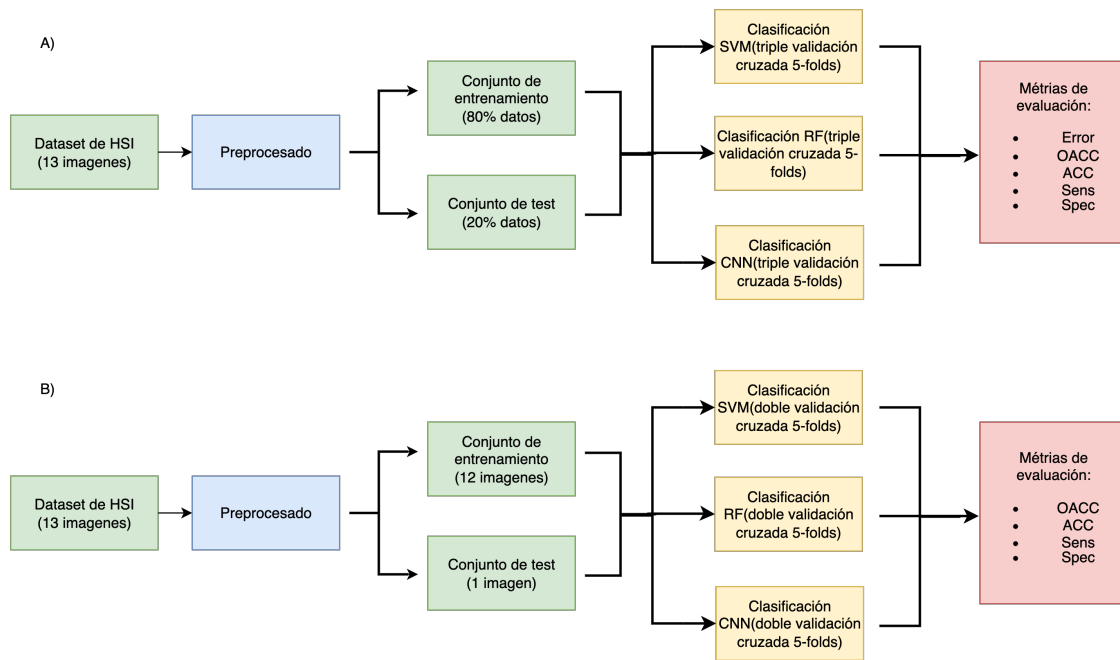


Figura 3.14: Flujos de procesamiento y validación cruzada aplicados a la clasificación tisular con imágenes hiperespectrales. (a) División clásica del conjunto de datos en entrenamiento (80 %) y test externo (20 %), seguida de validación cruzada triple ($5 \times 5 \times 5$) en los clasificadores SVM, RF y CNN. (b) Estrategia *leave-one-out*, utilizando 12 imágenes para entrenamiento y 1 para test externo, con validación cruzada doble (5×5) en los mismos clasificadores Urbanos et al., 2021.

Diseño experimental

La Figura 3.3 representa de forma esquemática el flujo metodológico seguido para el desarrollo y validación del modelo de segmentación automática de hemorragias cerebrales a partir de TC craneales sin contraste. El proceso se estructura en tres fases principales: entrada de datos, modelización y obtención del resultado final.

En primer lugar, las imágenes de TC craneal sin contraste constituyen la entrada del sistema. Estas exploraciones se seleccionan según los criterios de inclusión del estudio (véase Sección 3.1.2) para garantizar la calidad y homogeneidad necesarias antes de su incorporación al modelo.

El dataset completo se ha dividido inicialmente en dos subconjuntos: un 80 % destinado al entrenamiento y validación del modelo y un 20 % reservado como test, que se mantiene completamente aislado de cualquier fase de optimización o ajuste, con el objetivo de evaluar el rendimiento del modelo sobre datos no vistos.

Para el entrenamiento se ha aplicado una validación cruzada de cinco particiones (5-fold cross-validation), en la que, de forma iterativa, cuatro folds se emplean para el entrenamiento y uno para la validación. Este procedimiento se repite cinco veces, generando cinco modelos independientes (Modelo 1–Modelo 5), cada uno de los cuales produce una predicción sobre el conjunto de test. Finalmente, las segmentaciones obtenidas de los distintos modelos se combinan mediante un promediado de predicciones, con el fin de obtener una segmentación final más estable y menos sensible a la variabilidad de cada iteración.

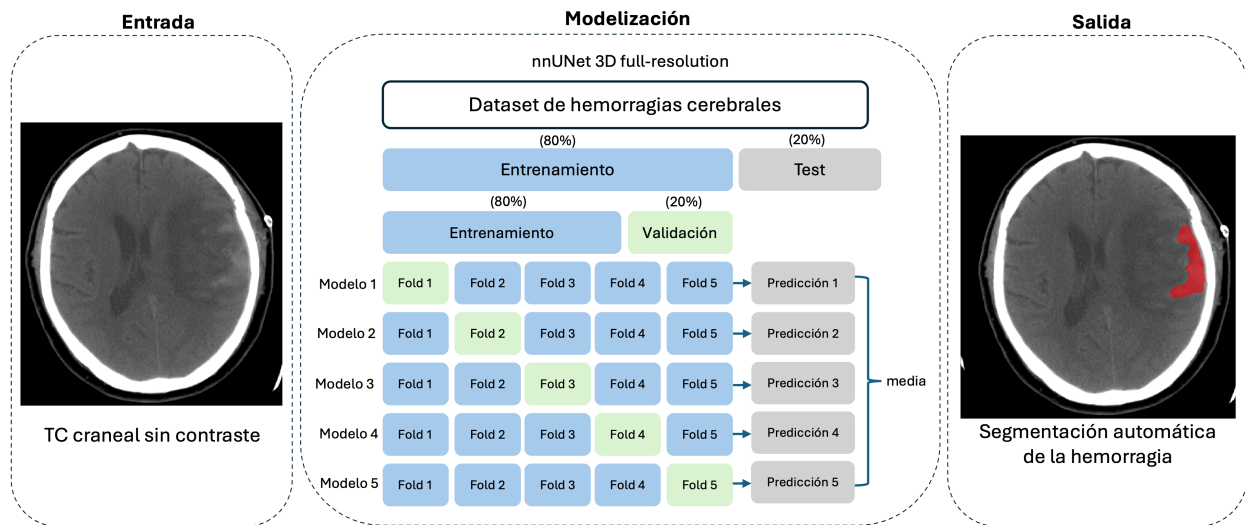


Figura 3.15: Esquema del flujo de validación cruzada de 5 *folds* para el entrenamiento y evaluación del modelo de segmentación automática, mostrando la TC craneal como entrada y la segmentación de hemorragia como salida.

Definición de escenarios experimentales.

Además, el diseño experimental evalúa cómo el tipo de hemorragias incluidas en el entrenamiento influye en el rendimiento del modelo de segmentación automática.

- **Experimento 1:** Se ha combinado el conjunto institucional del H12O y uno público. Se ha realizado una estratificación por centro, reservando el 20 % de los casos de cada centro para test. Este escenario busca cuantificar el rendimiento del modelo en datos institucionales y públicos, empleando únicamente hemorragias vinculadas a la rotura aneurismática (HSA, HIV y HIP).
- **Experimento 2:** Se amplía el conjunto de entrenamiento incorporando dos datasets públicos que incluyen hemorragias de origen traumático, con el objetivo de evaluar si la introducción de mayor heterogeneidad favorece la capacidad de generalización del modelo. El conjunto de *test* independiente se mantiene idéntico al del Exp. 1, garantizando la comparabilidad directa entre ambos escenarios.

Arquitectura y configuración de modelos

Como se ha comentado en anteriores secciones, la segmentación automática se ha implementado utilizando el marco *nnU-Net* (Isensee et al., 2021), en su variante `3d_fullres`. Esta configuración se basa en una arquitectura U-Net 3D modificada que incorpora:

- **Conexiones residuales (*residual connections*):** facilitan el flujo de gradientes en redes profundas permitiendo una convergencia más estable.
- **Normalización por lotes (*batch normalization*):** aplicada tras cada capa convolucional para estabilizar la distribución de activaciones y acelerar el entrenamiento.
- **Supervisión profunda (*deep supervision*):** permite calcular la función de pérdida en múltiples escalas de resolución, incentivando que las capas intermedias aprendan representaciones útiles.
- **Función de pérdida combinada:** se emplea una suma ponderada de *Soft Dice loss*, que optimiza la superposición espacial entre la predicción y la referencia, y *Cross-Entropy loss*, que favorece una clasificación precisa a nivel de vóxel. Esta combinación permite equilibrar la sensibilidad a las regiones minoritarias (típicas en lesiones pequeñas) con la capacidad de penalizar errores de clasificación globales, evitando el sesgo que puede aparecer al emplear cada función por separado:

$$\mathcal{L}_{\text{total}} = \lambda \mathcal{L}_{\text{Dice}} + (1 - \lambda) \mathcal{L}_{\text{CE}} \quad (3.5)$$

donde $\lambda \in [0, 1]$ controla el peso relativo de cada término, \mathcal{L}_{CE} es la entropía cruzada estándar y $\mathcal{L}_{\text{Dice}}$ se define como:

$$\mathcal{L}_{\text{CE}} = -\frac{1}{N} \sum_{i=1}^N [g_i \log(p_i) + (1 - g_i) \log(1 - p_i)] \quad (3.6)$$

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2 \sum_{i=1}^N p_i g_i + \epsilon}{\sum_{i=1}^N p_i + \sum_{i=1}^N g_i + \epsilon} \quad (3.7)$$

donde p_i y g_i representan, respectivamente, la probabilidad predicha y la etiqueta real para el vóxel i , N es el número total de vóxeles, y ϵ es un valor de suavizado para evitar divisiones por cero.

La arquitectura generada permite alcanzar un compromiso eficiente entre cobertura anatómica y coste computacional, garantizando la reproducibilidad de los experimentos y la comparabilidad entre configuraciones.

Los detalles específicos de los hiperparámetros empleados (incluyendo tamaños de entrada, espaciado final, política de aprendizaje, parámetros del optimizador y estrategias de aumento de datos) se presentan de forma completa en el Anexo B, Tabla B.1, que incluye las configuraciones utilizadas en los Experimentos 1 y 2. Ambas configuraciones mantuvieron una estructura común, diferenciándose principalmente en el tamaño de los volúmenes de entrada y el espaciado final, ajustados automáticamente por *nnU-Net* en función de las dimensiones promedio y la resolución nativa de los datos empleados en cada experimento.

Comparativa con modelos externos

Con el objetivo de contextualizar el rendimiento de los modelos propuestos frente a soluciones ya existentes, se ha llevado a cabo una evaluación comparativa con cinco algoritmos de segmentación automática disponibles de la literatura (véase la Sección 2.8.2):

1. **Blast-CT** (Monteiro et al., 2020): modelo multiclase basado en CNN 3D, entrenado para segmentar lesiones cerebrales traumáticas.
2. **SMCCH** (*Subarachnoid Hemorrhage Segmentation Challenge Model*) (García-García et al., 2024): modelo basado en una arquitectura *Vision Transformer* (ViT) con bloques *Swin-UNet* para la segmentación de HSAa.
3. **ISBI 2024 – Segmentación de hemorragias minúsculas** (Gong et al., 2024): modelo especializado en la detección y segmentación de volúmenes hemorrágicos inferiores a 1 ml, con alta sensibilidad para lesiones de difícil identificación.
4. **ICH 2023** (Rusche et al., 2023): algoritmo de segmentación de hemorragias intracraneales desarrollado utilizando *nnU-Net*, incluido como módulo en la herramienta *TotalSegmentator* (Wasserthal et al., 2023).
5. **Viola-UNet v2** (Q. Liu et al., 2023): arquitectura ganadora del desafío INSTANCE 2022, optimizada para segmentación de hemorragias intracraneales.

Para asegurar una comparabilidad justa y minimizar sesgos derivados del tratamiento previo de los datos, todas las inferencias se realizan aplicando el preprocesamiento recomendado o requerido por cada uno de los algoritmos.

Las salidas de cada modelo externo se evalúan siguiendo exactamente el mismo protocolo cuantitativo y cualitativo, lo que garantizó que las métricas obtenidas fueran directamente comparables entre todos los algoritmos. Este procedimiento no solo permite estimar con precisión el rendimiento relativo de los modelos desarrollados, sino también valorar su potencial de aplicación clínica frente a otras soluciones propuestas en literatura.

Impacto clínico : relación con escala de Fisher modificada

Con el objetivo de evaluar el valor pronóstico relativo de una medida cualitativa ampliamente utilizada en la práctica clínica (la *escala de Fisher modificada*) frente a una medida cuantitativa

derivada de imagen (el volumen hemorrágico), se ha analizado su capacidad para predecir el resultado funcional a los 6 meses (GOS dicotomizada en *buen resultado* [4–5] y *mal resultado* [1–3]).

El volumen hemorrágico se obtiene tanto a partir de segmentaciones manuales como automáticas, con el fin de determinar si la aproximación automatizada podía sustituir o complementar a la medición manual. La comparación se ha realizado mediante modelos de regresión logística binaria:

1. **Análisis univariable:** se evalúa, por separado, la escala de Fisher modificada, el volumen hemorrágico manual y el volumen hemorrágico automático como predictores independientes.
2. **Análisis multivariable:** se incluyen como covariables la edad, la escala WFNS al ingreso y el grado de Fisher modificada o el volumen hemorrágico, con el fin de determinar si la segmentación (manual o automática) aporta valor explicativo adicional tras el ajuste por factores clínicos relevantes.

Estos resultados contribuyen a determinar si la segmentación automática del volumen hemorrágico puede convertirse en una herramienta clínica complementaria a la escala de Fisher modificada en la predicción de desenlaces.

En los modelos multivariados, los coeficientes de la regresión logística se interpretan en forma de **odds ratios** (OR). El OR cuantifica cómo cambian las probabilidades relativas de presentar un mal resultado funcional por cada unidad de incremento en la variable, manteniendo constantes las demás covariables:

- **OR = 1:** la variable no modifica el riesgo; su presencia o aumento no se asocia con cambios en el desenlace.
- **OR > 1:** la variable se asocia a un incremento en las probabilidades relativas de mal resultado (factor de riesgo).
- **OR < 1:** la variable se asocia a una disminución en las probabilidades relativas de mal resultado (factor protector).

Esta forma de representar los resultados permite comparar de manera homogénea el impacto relativo de los distintos predictores incluidos en la regresión multivariable, y facilita la interpretación clínica de los modelos.

3.4.3 Modelos predictivos basados en radiómica

En este apartado se detalla la construcción de modelos predictivos utilizando características radiómicas extraídas de las imágenes de TC de HSAa, integradas con variables clínicas. El flujo de trabajo combina la extracción de características radiómicas de las regiones segmentadas, la integración con variables clínicas relevantes y la aplicación de algoritmos supervisados de aprendizaje automático bajo un esquema de validación robusto y reproducible.

El objetivo no solo es construir modelos de alto rendimiento, sino también garantizar su reproducibilidad, interpretabilidad y aplicabilidad clínica. Los resultados se comparan con

escalas clínicas establecidas, incluyendo la SEBES, para valorar la utilidad clínica del enfoque radiómico.

Diseño experimental

El diseño experimental se plantea con el objetivo de obtener estimaciones robustas del rendimiento del modelo, minimizando el riesgo de sobreajuste y garantizando su capacidad de generalización ante datos no vistos. Para ello, se emplea una estrategia de validación cruzada anidada (*nested cross-validation*), complementada con una validación externa independiente en el tiempo.

La validación cruzada se implementa mediante cinco *folds* estratificados, preservando la proporción original de clases en cada partición. En cada iteración, el modelo se entrena con el 80% de los casos y se evalúa sobre el 20% restante, que permanece completamente aislado durante la optimización interna. Además, la división entre entrenamiento y test se repite con tres semillas aleatorias diferentes, dando lugar a un total de quince entrenamientos independientes (5 *folds* × 3 semillas). La métrica final se obtiene como la media de los resultados de todas las repeticiones, proporcionando una estimación más estable del rendimiento global.

La metodología completa para el desarrollo y la evaluación de los modelos predictivos se ilustra de forma esquemática en la Figura 3.16.

En el **bucle externo (outer loop)**, los datos de cada partición se normalizan mediante un *Standard Scaler* ajustado exclusivamente sobre el conjunto de entrenamiento, evitando así fugas de información y preservando la independencia estadística respecto al conjunto de test. Posteriormente, en el **bucle interno (inner loop)** se lleva a cabo la selección de hiperparámetros y el ajuste del modelo, aplicando estrategias de búsqueda sistemática (*grid search*) basadas en la optimización del área bajo la curva ROC (AUC, *Area Under the Curve*).

Ajuste de parámetros del modelo en el bucle interno

Tratamiento de *outliers* El filtrado de observaciones atípicas se ha realizado mediante el algoritmo *Isolation Forest* (F. T. Liu et al., 2008), que aísla las instancias anómalas mediante árboles binarios aleatorios. Dado que las observaciones atípicas suelen requerir menos divisiones para su aislamiento, el algoritmo permite identificar casos extremos con una alta eficiencia computacional. En este estudio, el módulo de detección de *outliers* se ha integrado de forma opcional dentro del flujo de preprocesamiento, de modo que el proceso de entrenamiento se ejecutó tanto con como sin dicha etapa, con el objetivo de evaluar su impacto sobre la estabilidad y el rendimiento final de los modelos.

Equilibrio de clases Para mitigar el sesgo hacia la clase mayoritaria, se ha empleado el algoritmo *Adaptive Synthetic Sampling* (ADASYN) (He et al., 2008), una técnica de sobremuestreo adaptativo que genera instancias sintéticas de la clase minoritaria en función de la dificultad local de clasificación. El número de ejemplos sintéticos generados para cada instancia minoritaria se pondera según su complejidad local, de modo que los casos más difíciles de clasificar contribuyen en mayor medida a la expansión sintética del conjunto minoritario. La aplicación o no de ADASYN se evaluó experimentalmente dentro del proceso de validación

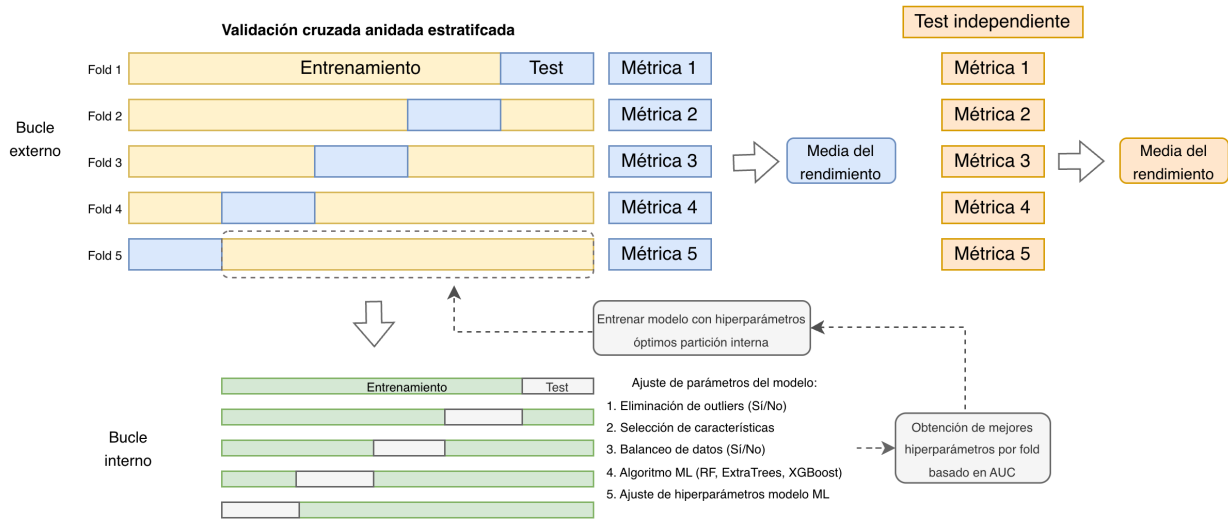


Figura 3.16: Esquema del procedimiento de validación cruzada anidada. El bucle externo (amarillo) divide los datos en cinco particiones, utilizando en cada iteración una como **test externo** (azul). En el bucle interno (verde) se optimizan los hiperparámetros y se entrena el modelo, seleccionando la configuración con mejor AUC medio. Los modelos finales se evalúan sobre sus particiones de test externas y, posteriormente, en un **test independiente** para estimar la capacidad de generalización.

interna para determinar su efecto sobre las métricas de discriminación y calibración.

Algoritmos evaluados Se han evaluado tres algoritmos supervisados ampliamente utilizados en clasificación de imágenes médicas y datos radiómicos: RF (Breiman, 2001), ExtraTrees (Geurts et al., 2006), y XGBoost (T. Chen y Guestrin, 2016). Su fundamentación teórica se describe en la Sección 2.7.1.

Optimización de hiperparámetros El ajuste de hiperparámetros se realiza mediante búsqueda exhaustiva (*grid search*), optimizando según el AUC obtenido. Este método consiste en evaluar de forma sistemática todas las combinaciones posibles de valores dentro de un espacio de parámetros predefinido, ejecutando un entrenamiento y validación cruzada para cada configuración. La descripción detallada de los espacios de búsqueda y configuraciones específicas de hiperparámetros se incluye en el Anexo B.5.

Finalmente, a partir de los hiperparámetros óptimos identificados en la validación interna (*bucle interno*), se procedió al entrenamiento del modelo correspondiente a cada partición externa (*bucle externo*). En esta fase, el conjunto de entrenamiento se ajusta exclusivamente con los parámetros seleccionados, evaluándose el rendimiento sobre el subconjunto de test externo asociado a cada *fold*.

Las métricas globales de desempeño se han obtenido como la media de los resultados en los cinco conjuntos de test del esquema de validación cruzada anidada, proporcionando así una

estimación robusta y libre de optimismo del rendimiento general del modelo.

Validación externa independiente

Más allá del esquema de validación cruzada anidada, se ha incorporado un conjunto de test externo adicional (test independiente), compuesto por estudios obtenidos en un periodo temporal posterior al del desarrollo del modelo. Este conjunto permanece completamente ajeno a las fases de entrenamiento y validación, y su inclusión permite simular una validación prospectiva.

El uso de esta cohorte posterior se fundamenta en un principio clave para la traslación clínica de modelos predictivos: evaluar la robustez temporal frente a variaciones del entorno, tales como la evolución tecnológica de los equipos de TC, la modificación de protocolos de adquisición o cambios en la población atendida.

Integración de información multimodal

La información empleada en los modelos procede de dos dominios complementarios: (i) características radiómicas cuantitativas extraídas de las imágenes de TC, y (ii) variables clínicas relevantes relacionadas con el estado basal y la evolución del paciente.

Con el fin de evaluar la contribución individual y combinada de ambos dominios, se plantean tres estrategias experimentales:

- **Modelos clínicos puros**, basados exclusivamente en variables tabulares derivadas de la HCE (véase Anexo B, Sección B.2).
- **Modelos radiómicos**, contruidos a partir de las características cuantitativas extraídas de las máscaras segmentadas. Estas incluyen tanto las regiones hemorrágicas como las segmentaciones tisulares (materia gris y blanca), abarcando atributos de forma, textura y primer orden.
- **Modelos combinados o multimodales**, en los que se integran simultáneamente las variables clínicas y las radiómicas correspondientes, buscando capturar la información complementaria de ambos niveles.

El análisis de estas configuraciones permite explorar el valor añadido de la radiómica frente a las escalas clínicas tradicionales, así como las sinergias potenciales derivadas de la combinación de ambas fuentes de información.

Interpretabilidad

Como se ha descrito en la Sección 2.6.1, los biomarcadores de imagen (entre los que se incluyen las características radiómicas) se obtienen mediante un proceso estructurado que incluye la adquisición, segmentación, preprocesado y cálculo de descriptores matemáticos. Una vez obtenidos, estas datos se utilizan como entrada para modelos de IA que buscan predecir desenlaces clínicos, diagnóstico o estratificar pacientes.

No obstante, la potencia predictiva de estos modelos no es suficiente para su adopción clínica si no va acompañada de explicabilidad. En medicina, es necesario entender qué variables están

impulsando la predicción y si esas variables son coherentes desde un punto de vista clínico o biológico. Es aquí donde toma importancia la interpretabilidad de los modelos de IA, y en particular el uso de métodos como SHAP (*SHapley Additive exPlanations*).

SHAP es una técnica basada en teoría de juegos cooperativos que asigna a cada característica un valor cuantitativo que refleja su contribución al resultado del modelo, tanto a nivel global (importancia media de cada variable en el conjunto de datos) como a nivel local (influencia en una predicción concreta). Este enfoque se fundamenta en los valores de *Shapley*, que determinan el aporte marginal de una característica considerando todas las posibles combinaciones con el resto del conjunto de variables (Lundberg y Lee, 2017).

Aplicado a la radiómica y a cualquier otro tipo de variables incluidas en el modelo, SHAP permite:

- Determinar qué características (por ejemplo, textura GLCM, morfología, primer orden o variables clínicas como edad, biomarcadores sanguíneos, etc.) son las más influyentes en la clasificación o predicción de un resultado clínico.
- Explorar la dirección del efecto: si una característica concreta aumenta o reduce la probabilidad de un desenlace (por ejemplo, mal pronóstico).
- Explicar predicciones individuales por caso.
- Evaluar la robustez y fiabilidad del modelo, identificando posibles sesgos o sobreajuste.

Comparación con escalas clínicas

Con el objetivo de contextualizar los resultados obtenidos y establecer una referencia frente a herramientas clínicas consolidadas, este estudio compara el rendimiento de los modelos radiómicos con diversas escalas pronósticas clásicamente empleadas en la HSAa. Entre ellas destacan la *SEBES* y *SEBES6c*, ampliamente utilizada para cuantificar el edema cerebral temprano, y la *ictWFNS* (Hofmann et al., 2022), una versión revisada de la escala WFNS que integra información clínica y radiológica.

La puntuación *ictWFNS*, calculada por un neurocirujano experto para todos los casos del conjunto de test independiente, constituye un índice compuesto que pondera la WFNS original junto con los hallazgos radiológicos derivados de tres escalas clásicas: SEBES, Le Roux e Hijdra. Esta escala se adopta como referencia de comparación al reflejar tanto el estado neurológico basal como la extensión del daño cerebral visible en la TC.

Para estimar su valor predictivo frente a los modelos desarrollados, se ajustan modelos de regresión logística univariados para cada desenlace clínico de interés, permitiendo comparar la capacidad pronóstica de la radiómica con respecto a escalas clínicas previamente validadas. Esta comparación se extiende a las puntuaciones *SEBES* y *SEBES6c*, evaluando su rendimiento discriminativo frente a los desenlaces de mortalidad, pronóstico funcional desfavorable, vasoespasmos e hidrocefalia, y explorando el posible valor añadido de integrar variables clínicas adicionales en modelos multivariados. En este contexto, se calculan los *odds ratios* (OR) correspondientes, con el fin de proporcionar una interpretación clínica de los modelos combinados.

Bajo la hipótesis de que la SEBES, basada en la apreciación visual de patrones en la TC, podría tener una manifestación cuantitativa en el espacio radiómico, se desarrolla un modelo denominado **SEBES radiómico**. Este modelo busca replicar la lógica subyacente de la escala clínica a partir de descriptores cuantitativos de textura.

Para explorar dicha correspondencia, se analiza la correlación de rangos de Spearman (Spearman, 1987) entre cada característica radiómica y la puntuación SEBES. Posteriormente, se aplica una regresión logística penalizada mediante el método LASSO (*Least Absolute Shrinkage and Selection Operator*) (Tibshirani, 1996), que introduce una penalización L_1 sobre los coeficientes y promueve la selección automática de las variables más relevantes. La estrategia de optimización se plantea de forma incremental: las características se ordenan según su relevancia predictiva y se incorporan progresivamente en bloques de cinco, evaluando el rendimiento del modelo en cada iteración. Esta aproximación permite identificar un subconjunto óptimo de variables radiómicas y, simultáneamente, controlar el sobreajuste asociado a la alta dimensionalidad del espacio de características.

El modelo resultante, *SEBES radiómico*, se emplea posteriormente como base para la construcción de los modelos predictivos específicos de cada desenlace clínico mediante regresión logística.

3.4.4 Métricas de evaluación

La selección de métricas en este trabajo se fundamenta en la necesidad de evaluar el rendimiento de los modelos desde una perspectiva integral, que contemple no solo la precisión técnica, sino también la relevancia clínica y la interpretabilidad de los resultados. De este modo, las métricas se estructuran en dos niveles complementarios: cuantitativo y cualitativo.

Métricas cuantitativas

Las métricas cuantitativas se emplearon para comparar objetivamente el desempeño entre modelos, tanto en tareas de segmentación como de clasificación y predicción clínica. Los resultados se resumen mediante la media o mediana acompañadas de su intervalo intercuartílico (IQR), y las diferencias se contrastan mediante intervalos de confianza del 95

Dado que el trabajo aborda problemas de naturaleza diversa (clasificación tisular, segmentación automática y predicción de desenlaces clínicos), las métricas se agrupan en tres familias principales, descritas con detalle en el Anexo C (Sección C):

- (i) métricas derivadas de la matriz de confusión, orientadas a la evaluación de tareas de clasificación y predicción;
- (ii) métricas específicas de segmentación, centradas en la concordancia espacial y volumétrica;
- (iii) métricas estadísticas, destinadas a la interpretación de modelos de regresión logística.

En primer lugar, las métricas basadas en la matriz de confusión se seleccionaron por su capacidad para describir el comportamiento discriminativo de los modelos en escenarios con alta heterogeneidad y desbalance de clases, frecuentes tanto en la clasificación de tejidos mediante HSI como en la predicción de desenlaces clínicos en HSAa. Indicadores como la

OACC, la precisión (ACC), la exactitud balanceada (BalAcc), la precisión, la sensibilidad (Sens), la especificidad (Spec), el *F1-score* y el área bajo la curva ROC (AUC) permiten evaluar simultáneamente la tasa de aciertos, la robustez ante desequilibrios y la capacidad discriminativa global del modelo. La elección de estas métricas se apoya en su amplia aceptación en la literatura sobre inteligencia artificial aplicada a imagen médica (Rajpurkar et al., 2022) y en su adecuación para entornos clínicos con alta variabilidad interpaciente. Además, el uso del AUC sigue las recomendaciones del estándar TRIPOD para la validación de modelos diagnósticos (Collins et al., 2015b).

En segundo lugar, las métricas específicas de segmentación se aplicaron para evaluar la coherencia geométrica y volumétrica de las máscaras generadas en los experimentos de segmentación automática de HSAa. El coeficiente de similitud de Dice (DSC) se adoptó como métrica principal por su sensibilidad al solapamiento entre regiones predichas y de referencia, mientras que medidas basadas en distancia, como la distancia de Hausdorff (HD) y la distancia media de superficie (ASD, de sus siglas en inglés *Average Surface Distance*), complementaron la valoración al reflejar discrepancias en los contornos. Estas métricas se combinaron con evaluaciones volumétricas, incluyendo el coeficiente de correlación intraclase (ICC) y los gráficos de Bland–Altman (Kalra, 2017), que aportan información sobre la concordancia cuantitativa y la reproducibilidad interobservador. La combinación de índices de solapamiento, distancia y volumen proporciona una valoración más completa y robusta, en consonancia con las directrices de reproducibilidad y validez establecidas por la guía IBSI (Zwanenburg et al., 2016).

Por último, para los modelos predictivos radiómicos se emplearon métricas estadísticas propias de la regresión logística, orientadas a evaluar la significación, la calibración y la capacidad explicativa de los modelos multivariados. Parámetros como los *odds ratios* (OR), el pseudo- R^2 , la log-verosimilitud y el AUC permiten estimar el peso relativo de cada predictor, la capacidad del modelo para discriminar entre grupos clínicos y su valor añadido frente a escalas clínicas convencionales (Van Calster et al., 2019). El uso de estas métricas favorece la interpretación clínica y posibilita una comparación transparente entre modelos basados en IA y herramientas pronósticas tradicionales.

Métricas cualitativas

Complementariamente, los modelos de segmentación automática de HSAa se someten a una evaluación cualitativa por expertos con el objetivo de valorar la precisión geométrica y la aceptabilidad clínica de las segmentaciones generadas. Esta revisión se diseña como un *estudio ciego independiente* para minimizar sesgos, seleccionando 15 casos aleatorios del conjunto H120 que no pertenecen a los conjuntos de entrenamiento, validación ni test, y que carecen de segmentaciones manuales previas.

Siete neurocirujanos con experiencia en patología neurovascular evalúan las segmentaciones obtenidas por los distintos modelos (descritos en la Sección 3.4.2). Todas las máscaras se anonimizan y se etiquetan con identificadores aleatorios para evitar el reconocimiento del modelo generador. Cada evaluador puntúa cinco criterios en una escala de 1 (muy deficiente) a 5 (excelente):

1. **Cobertura anatómica:** grado en que la segmentación abarca la totalidad de la región

hemorrágica real.

2. **Sobresgmentación:** inclusión de regiones no hemorrágicas.
3. **Subsegmentación:** omisión de áreas hemorrágicas presentes en la referencia.
4. **Consistencia 3D:** coherencia de la segmentación a lo largo de cortes contiguos.
5. **Calidad global:** impresión general sobre la utilidad clínica de la segmentación.

Las valoraciones se agregaron mediante el **método de Borda**, una técnica de agregación de preferencias ampliamente utilizada en contextos de votación y evaluación multicriterio (Young, 1974). En este método, cada modelo recibe un rango por criterio (1 = mejor, N = peor), y los rangos se suman entre todos los evaluadores y criterios. La suma total constituye la *puntuación Borda*, donde un valor menor indica un desempeño global superior. Este enfoque permite integrar, en una única medida sintética, valoraciones procedentes de múltiples dimensiones, equilibrando las posibles discrepancias entre criterios y evitando que un rendimiento extremo en un solo aspecto condicione el resultado final.

Capítulo 4

Resultados

Este capítulo reúne los resultados obtenidos en las diferentes fases metodológicas de este trabajo, organizados de forma progresiva: desde la caracterización inicial de las bases de datos hasta la validación de los modelos predictivos. En primer lugar, se describe la conformación de las cohortes analizadas, con el número de pacientes incluidos y excluidos, los resultados de la segmentación manual de referencia y un resumen de las principales características clínicas y demográficas de la población. A continuación, se presentan los hallazgos relacionados con la extracción de características, tanto en el análisis espectral aplicado a GBM mediante HSI como en la cuantificación radiómica de la HSAa a partir de TC. Finalmente, se muestran los resultados de los experimentos de modelado y evaluación, que abarcan la clasificación tisular en GBM, la segmentación automática de hemorragias, el desarrollo y validación de modelos predictivos radiómicos, su interpretación y la comparación con escalas clínicas convencionales. Con esta estructura se pretende ofrecer una visión clara y ordenada de los hallazgos, destacando cómo cada bloque metodológico contribuye a dar respuesta a las hipótesis planteadas en la investigación.

4.1 Caracterización de las bases de datos

Antes de abordar los experimentos de entrenamiento y validación de los modelos, se ha realizado una caracterización detallada de las imágenes disponibles en la base de datos, evaluando tanto las propiedades técnicas de adquisición como la distribución de cortes efectivamente segmentados. Esta caracterización no solo permite contextualizar la complejidad del problema, sino también valorar la heterogeneidad inherente a los datos, un aspecto clave para interpretar el rendimiento de los modelos desarrollados.

Primero, en la Sección 4.1.1 se detallan los resultados de la caracterización de la base de datos empleada en el problema de clasificación tisular de GBM mediante HSI. Posteriormente, la Sección 4.1.2 recoge la caracterización de la base de datos utilizada en los experimentos de segmentación de HSAa en TC. Finalmente, en la Sección 4.1.3 se exponen los resultados y las decisiones que condujeron a la configuración definitiva de la base de datos de HSAa en TC para los análisis radiómicos.

4.1.1 Clasificación tisular en glioblastoma (Imagen Hiperespectral)

La base de datos empleada en este estudio se ha construido a partir de las adquisiciones hiperespectrales descritas en la Sección 3.1.4, obtenidas *in vivo* durante procedimientos neuroquirúrgicos en pacientes con diagnóstico de GBM. Todas las capturas se realizan tras la craneotomía, en condiciones intraoperatorias controladas y con el lecho quirúrgico completamente expuesto, de modo que sea posible la visualización directa tanto de la superficie cortical como del tejido tumoral y estructuras cerebrales adyacentes.

El conjunto final incluye un total de **13** cubos hiperespectrales correspondientes a **12** pacientes. Esta muestra, aunque limitada en tamaño debido a la complejidad técnica y logística inherente a la adquisición intraoperatoria, resulta representativa para explorar la viabilidad de la clasificación tisular mediante HSI en un escenario clínico real.

En la mayoría de los casos se cuenta con un único cubo hiperespectral, una vez el tumor se encontraba visible en el campo quirúrgico, ya fuera inmediatamente después de la craneotomía o durante el proceso de resección. Sin embargo, en un paciente se han registrado dos adquisiciones en momentos diferentes: la primera, tras la craneotomía y la segunda, una vez realizada la resección, en el propio lecho quirúrgico, donde aún persistían restos tumorales macroscópicos. Todos los cubos han sido considerados válidos para el análisis, al cumplir los criterios de calidad en términos de calibración espectral, iluminación y ausencia de artefactos relevantes.

La Figura 4.1 presenta la imagen RGB sintética, junto con el mapa de *ground truth* correspondiente para todos los pacientes incluidos en este estudio. En los mapas de referencia, el tumor se representa en rojo, los vasos sanguíneos (arteriales y venosos) en azul oscuro, el tejido sano en verde, la duramadre en rosa y el hueso en azul claro; esta última clase no se consideró en los análisis posteriores. Todos los mapas han sido elaborados por el neurocirujano responsable de la intervención de cada paciente.

Tal y como se mencionó en la Sección 3.2.1, la segmentación manual se ha realizado sobre distintos tejidos cerebrales visibles intraoperatoriamente (tejido sano, tumor, sangre venosa, sangre arterial y duramadre), generando así las máscaras de referencia empleadas posteriormente en los procesos de entrenamiento y validación de los algoritmos de clasificación tisular. La Tabla 4.1 muestra la distribución de los píxeles etiquetados en cada clase tisular para todos los pacientes incluidos en el estudio. Se observa una marcada variabilidad interpaciente en el número total de píxeles anotados, oscilando entre poco más de mil (p. ej., ID47C1) y casi diez mil (ID56). En términos relativos, el tejido sano y el tumor constituyen las clases con mayor representación, aunque con proporciones muy heterogéneas: algunos casos presentan un claro predominio de tejido sano (ID29, 67,9%), mientras que en otros el tumor alcanza valores cercanos o superiores al 40% (ID50, ID34). Las clases vasculares (sangre venosa y arterial) muestran una presencia más limitada en la mayoría de los pacientes, si bien destacan ciertos casos con porcentajes elevados de sangre venosa (23,1% en ID50) o arterial (12,5% en ID47C2). La duramadre aparece representada de manera consistente en prácticamente todos los pacientes, con proporciones que oscilan entre el 7–60% del total de píxeles, alcanzando su máximo en ID56. En conjunto, se refleja un conjunto de datos con una importante variabilidad interindividual que constituye un escenario realista para la evaluación de modelos de clasificación tisular.

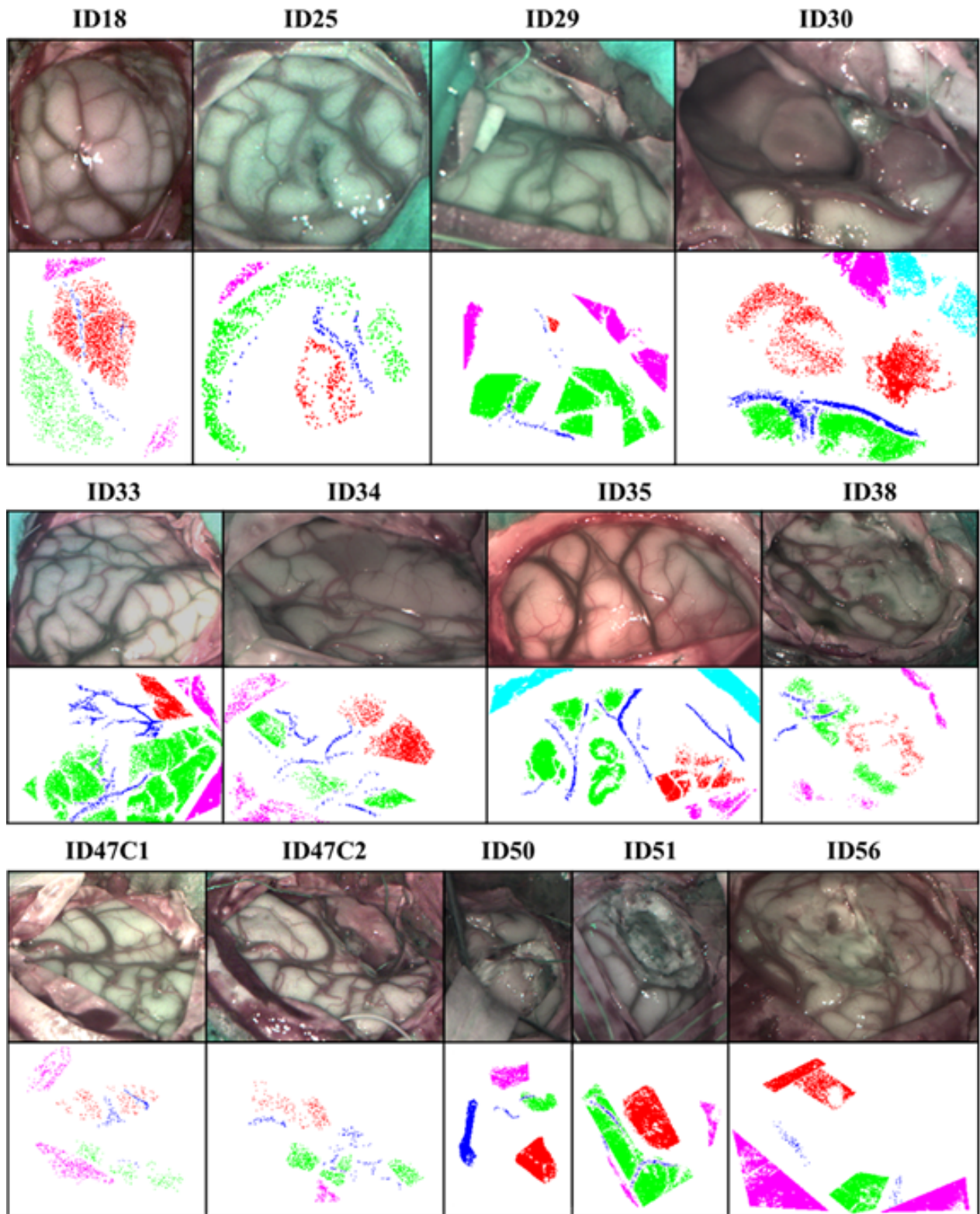


Figura 4.1: Imagen RGB sintética generada a partir de cubos hiperespectrales y mapas de *ground truth* con un color por clase para todos los pacientes incluidos.

Tabla 4.1: Conjunto de datos con el número total de píxeles etiquetados y su porcentaje por clase y paciente.

Paciente	Tejido sano	Tumor	Sangre venosa	Sangre arterial	Duramadre	Total
ID18	648 (24,02 %)	1587 (58,84 %)	79 (2,92 %)	14 (0,52 %)	369 (13,68 %)	2697
ID25	801 (66,42 %)	206 (17,08 %)	90 (7,46 %)	15 (1,24 %)	94 (7,79 %)	1206
ID29	3752 (67,92 %)	64 (1,16 %)	98 (1,77 %)	11 (0,20 %)	1599 (28,95 %)	5524
ID30	2587 (34,23 %)	2737 (36,21 %)	487 (6,44 %)	381 (5,04 %)	1366 (18,07 %)	7558
ID33	6671 (66,95 %)	973 (9,76 %)	842 (8,45 %)	35 (0,35 %)	1443 (14,48 %)	9964
ID34	1186 (31,32 %)	1464 (38,66 %)	181 (4,78 %)	176 (4,65 %)	780 (20,60 %)	3787
ID35	3864 (57,38 %)	1389 (20,63 %)	837 (12,43 %)	58 (0,86 %)	586 (8,70 %)	6734
ID38	1740 (51,85 %)	487 (14,51 %)	113 (3,37 %)	191 (5,69 %)	825 (24,58 %)	3356
ID47C1	174 (17,17 %)	160 (15,79 %)	58 (5,73 %)	54 (5,33 %)	567 (55,97 %)	1013
ID47C2	715 (60,19 %)	182 (15,32 %)	14 (1,18 %)	148 (12,46 %)	129 (10,86 %)	1188
ID50	410 (13,32 %)	1282 (41,64 %)	712 (23,12 %)	47 (1,53 %)	628 (20,40 %)	3079
ID51	3888 (53,09 %)	2731 (37,29 %)	95 (1,30 %)	85 (1,16 %)	525 (7,17 %)	7324
ID56	1920 (19,35 %)	1967 (19,83 %)	75 (0,76 %)	37 (0,37 %)	5922 (59,69 %)	9921

4.1.2 Segmentación automática de Hemorragia Subaracnoidea Aneurismática (TC)

Tamaño muestral

La cohorte principal utilizada en este trabajo corresponde a una base de datos local del H120, formada por **539 pacientes** con escáner TC entre enero de 2007 y junio de 2024. Dentro de esta cohorte local, **223 casos** cuentan con segmentaciones manuales completas realizadas por clínicos expertos, que se han utilizado como *ground truth* para entrenar y validar el algoritmo de segmentación automática de la hemorragia (véase Sección 3.2.2). En estas segmentaciones manuales, **223 estudios** incluyen segmentaciones de **HSA**, **218** incluyen segmentaciones de **HIV** y **1** caso incluye segmentación de **HIP**. Cada tipo de hemorragia recibe una etiqueta numérica única (1, 2 o 3) que indica su localización.

Para mejorar la capacidad de generalización del modelo de segmentación automática y ampliar la diversidad de casos, la cohorte local se ha complementado con bases de datos públicas de HIC, seleccionadas por su relevancia y calidad de anotaciones. Estas bases de datos aportan imágenes procedentes de distintos entornos clínicos, protocolos de adquisición y etiologías, y se han empleado para aumentar el conjunto de entrenamiento del algoritmo:

- **Segmentación de hemorragias de la *Radiological Society of North America*, (RSNA, HemSeg-200)** (Song et al., 2024): 222 TC anotadas del RSNA 2019 *Brain CT Hemorrhage Challenge*, que incluyen hemorragias por traumatismo, accidente cerebrovascular y rotura aneurismática. Incluye segmentaciones de HIP (114) y HIV (108), aunque las etiquetas de las hemorragias no están diferenciadas.
- **Conjunto de datos de segmentación de hemorragias cerebrales (BHSD)** (B. Wu et al., 2023): 192 TC de pacientes con TCE, con segmentaciones de HSA (109), HIP (127), HIV (104), HSD (70) y HED (22). Las etiquetas están diferenciadas por tipo de hemorragia.
- **Conjunto de datos del Hospital Universitario Al Hilla (Irak)** (M. Hssayeni et al., 2020): 82 TC obtenidas en el TCE, con variabilidad demográfica y técnica. Incluye

segmentaciones de HSA (7), HIP (16), HIV (5), HSD (4) y HED (21). Las etiquetas que distinguen las hemorragias no están diferenciadas.

La distribución de los casos y sus correspondientes segmentaciones, desglosada por fuente de datos y tipo de hemorragia, se presenta en la Tabla 4.2. Asimismo, la Figura 4.2 muestra ejemplos representativos de las segmentaciones manuales realizadas sobre los distintos conjuntos de datos empleados.

Tabla 4.2: Distribución de casos con segmentaciones manuales por fuente de datos y tipo de hemorragia.

Fuente de datos	Total anotados	HSA	HIV	HIP	HSD	HED
H12O (local)	223	223	218	1	–	–
HemSeg200	222	–	108	114	–	–
BHSD	192	109	104	127	70	22
Al Hilla	36	7	5	16	4	21

Como se ha comentado en la Sección 3.4.2, se han realizado dos experimentos, representados en la Figura 4.3.

- **Experimento 1:** Esta configuración combina el conjunto de datos institucional del H12O con el dataset público HemSeg200 (n total = 445). El conjunto de entrenamiento incluye 356 volúmenes (178 de H12O y 178 de HemSeg200), mientras que el conjunto de test se compone de 89 volúmenes (45 de H12O y 44 de HemSeg200).
- **Experimento 2:** En este escenario, se utilizan todos los conjuntos de datos disponibles para la fase de entrenamiento (n total = 722), manteniendo sin cambios el conjunto de test definido en el Experimento 1.

Información a nivel de imagen

Los parámetros de adquisición de las imágenes de TC pueden condicionar los resultados del modelo de segmentación automática, por lo que en este apartado se describe su distribución respecto a algunos parámetros de imagen que afectan a las segmentaciones de hemorragia. En la cohorte local (H12O), las imágenes se adquieren dentro del flujo de trabajo diagnóstico rutinario en el H12O mediante un escáner Philips Brilliance 6. Cada imagen consta con un tamaño matricial fijo de 512×512 con una resolución en X e Y isotrópica. Por otro lado, la Figura 4.4 resume de forma gráfica la variabilidad observada en el tamaño de píxel, el grosor de corte y el número de cortes segmentados por paciente. En el panel (a) se aprecia que el tamaño de píxel en el plano axial presenta una variabilidad moderada, con un predominio de valores en torno a 0,45–0,50 mm, aunque existen casos con resoluciones más finas (0,35–0,40 mm) y otros más groseros ($>0,55$ mm). En el panel (b), el grosor de corte muestra una fuerte concentración en 2 mm (55 % de los casos), seguido de 3 mm (24 %), mientras que otros valores son mucho menos frecuentes.

El número total de cortes por imagen (panel c) oscila entre aproximadamente 20 y 160, con un pico en torno a 100–110 cortes, lo que indica diferencias sustanciales en la extensión

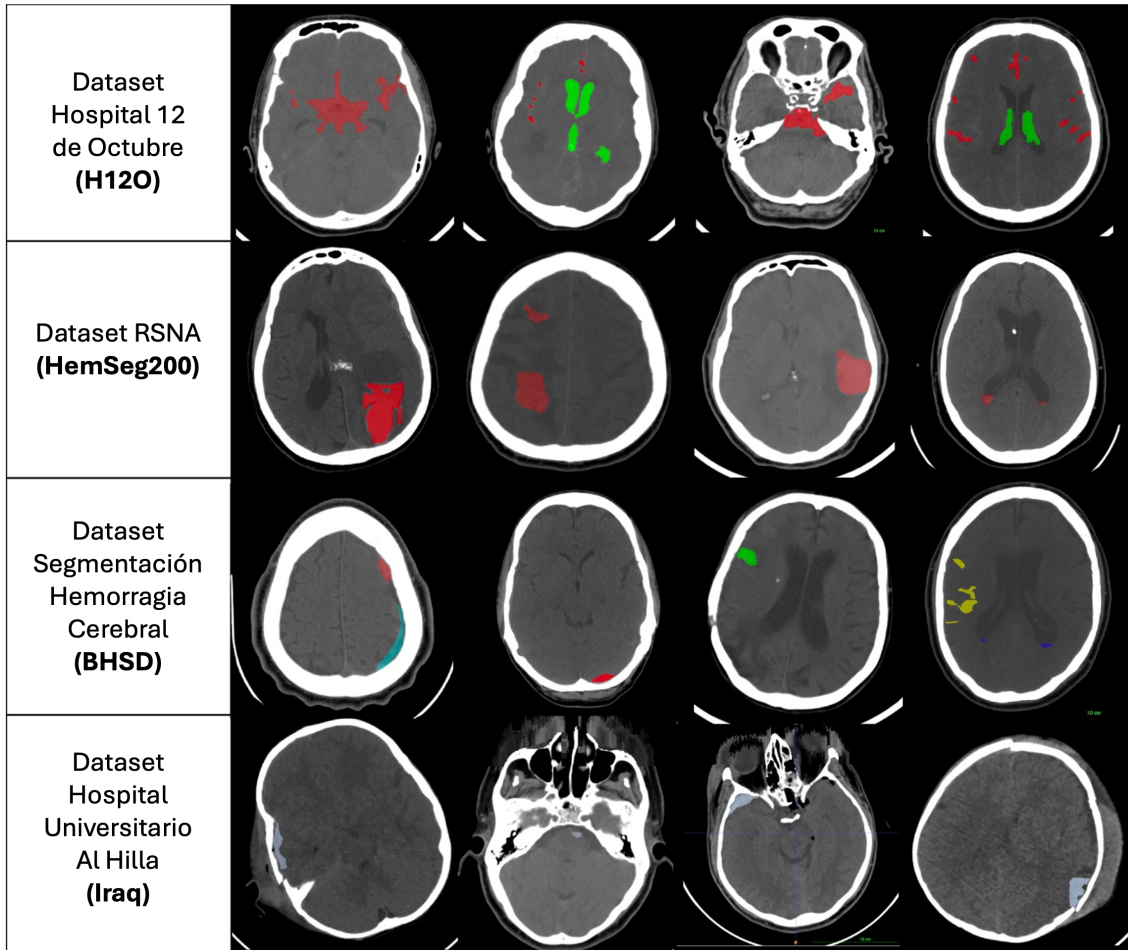
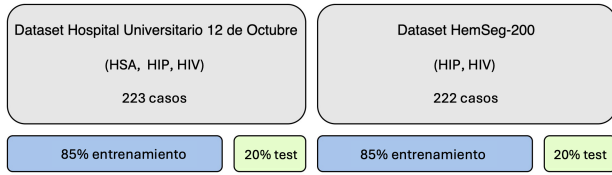


Figura 4.2: Ejemplos de segmentaciones manuales en cada uno de los datasets usados en el estudio. En la primera fila, segmentaciones del H12O (en rojo HSA, en verde IVH); en la segunda fila el dataset HemSeg200 (IPH y IVH); en la tercera fila el dataset BHSD (en cyan HED; en rojo HSD; en verde IPH; en amarillo HSA; en azul marino HIV); en la última fila el dataset de Al Hilla (HIV, IPH, HSD, HED y HSA).

A) Experimento 1



B) Experimento 2

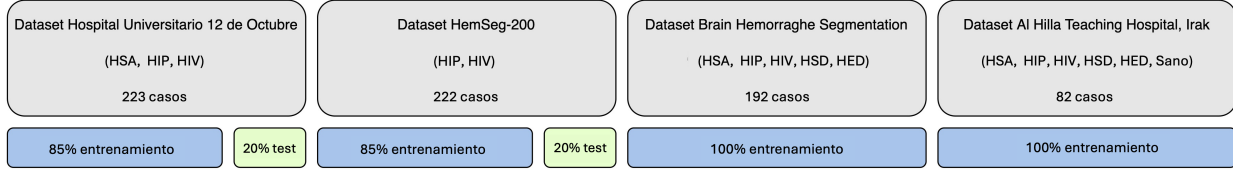


Figura 4.3: Esquema de división de los conjuntos de datos en los distintos experimentos, detallando las particiones aplicadas para las fases de entrenamiento, validación y evaluación del modelo.

volumétrica de las exploraciones. Finalmente, el panel (d) muestra el número de cortes que contienen segmentaciones manuales, con una distribución amplia que va desde menos de 10 hasta más de 80 cortes segmentados por caso.

Además de la base H12O, en este trabajo se han incorporado conjuntos de datos públicos descritos en la bibliografía, empleados para enriquecer el entrenamiento de los modelos.

- El conjunto HemSeg200 comprende un total de 222 volúmenes de TC, de los cuales 114 incluyen segmentaciones de HIP y 108 de HIV. Todas las exploraciones tienen un tamaño matricial fijo de 512×512 píxeles, con entre 24 y 56 cortes por volumen. La separación entre píxeles varía entre 0,39 y 0,63 mm, mientras que el grosor de corte oscila entre 3,0 y 6,5 mm. Este conjunto, junto con H12O, constituyó la base del Experimento 1.
- El BHSD contiene 192 volúmenes de TC cerebral, cada uno formado por entre 24 y 60 cortes, con tamaño matricial fijo de 512×512 píxeles. A diferencia de HemSeg200, este dataset incluye una mayor diversidad de tipos hemorrágicos anotados: 22 casos de HED, 127 de HIP, 104 de HIV, 109 de HSA y 70 de HSD. La resolución en el plano varía entre 0,37 y 0,68 mm, y el grosor de corte entre 0,5 y 6,5 mm. Esta cohorte se incorporó al conjunto de entrenamiento del Experimento 2.
- Finalmente, el conjunto del Hospital Universitario Al Hilla (Irak) está compuesto por 82 volúmenes de TC, cada uno con entre 31 y 58 cortes y tamaño matricial fijo de 512×512 píxeles. El grosor de corte es constante de 5 mm, mientras que el tamaño de píxel oscila entre 0,31 y 0,49 mm. De estas exploraciones, 36 incluyen segmentaciones, correspondientes a 21 casos de HED, 16 de HIP, 5 de HIV, 7 de HSA y 4 de HSD. Este conjunto se integra en el entrenamiento del Experimento 2, aportando ejemplos procedentes de un entorno clínico distinto y con características técnicas específicas.

Con el fin de facilitar la comparación entre las distintas cohortes empleadas, en la Tabla 4.3 se resumen sus principales características técnicas.

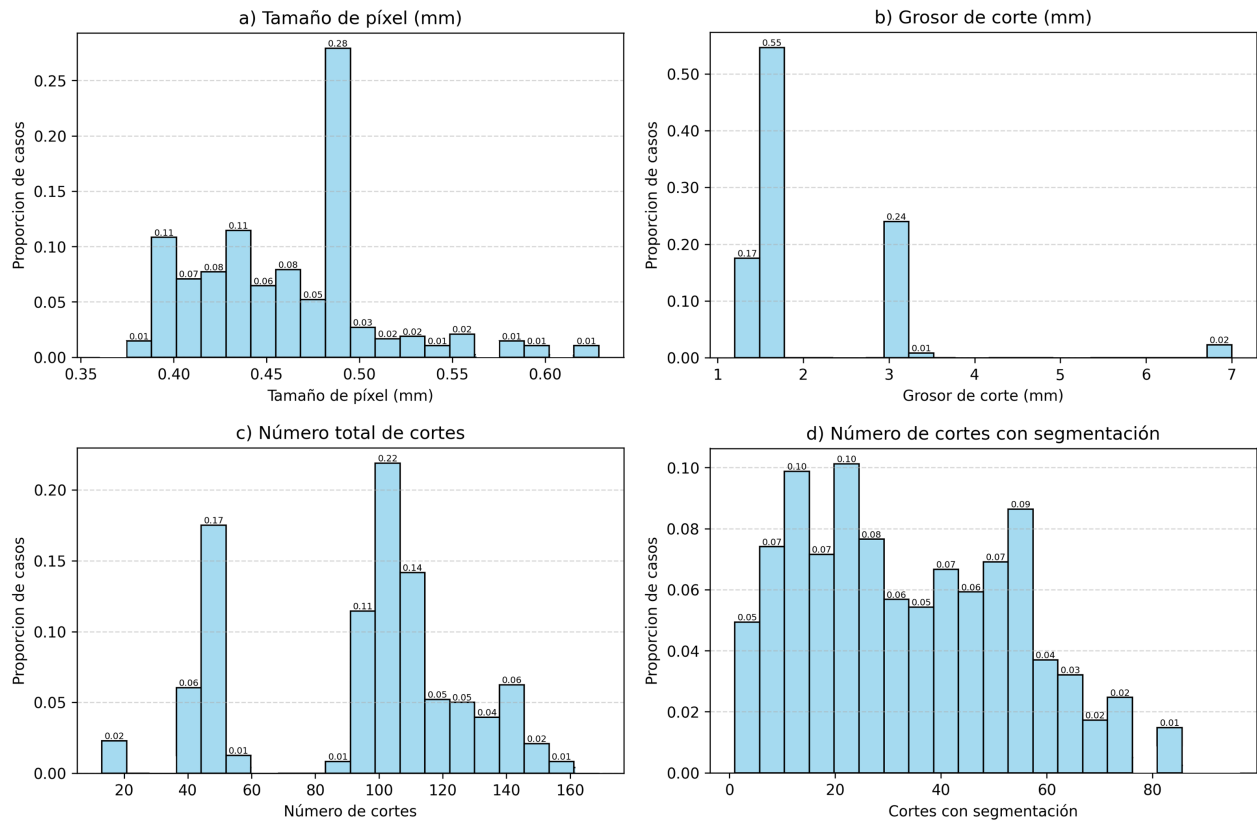


Figura 4.4: Histogramas de las características técnicas y de anotación de la base de datos de imagen del H120. (a) Distribución del tamaño de píxel (mm) (b) Distribución del grosor de corte (mm) (c) Distribución del número total de cortes por exploración (d) Distribución del número de cortes con segmentaciones manuales

Tabla 4.3: Características principales de los conjuntos de las bases de datos utilizadas para la segmentación del sangrado.

Dataset	Nº volúmenes	Cortes por TC	Tamaño píxel (mm)	Grosor (mm)
H120	223	13–169	0,36–0,73	1,2-7
HemSeg200	222	24–56	0,39–0,63	3,0–6,5
BHSD	192	24–60	0,37-0,68	0,5-6.5
Al Hilla (Irak)	82	31-58	0,31-0,49	5,0

La Figura 4.5 muestra la distribución de tipos de hemorragia en las diferentes cohortes empleadas. En el conjunto H12O predominan los casos de HSA (n=223) y HIV (n=218), con un único caso de HIP. En el Experimento 1 se observa la suma de los casos de HemSeg, resultando en HIV (n=326) y HIP (n=109) y manteniéndose el número de HSA estable (n=223). Finalmente, el Experimento 2 constituye la cohorte más amplia y heterogénea, incluyendo un mayor número de casos de HSA (n=339), HIV (n=435) y HIP (n=252), además de la inclusión de HED (n=43) y HSD (n=74), lo que aporta una representación más completa de los distintos subtipos hemorrágicos provenientes de distintas enfermedades.

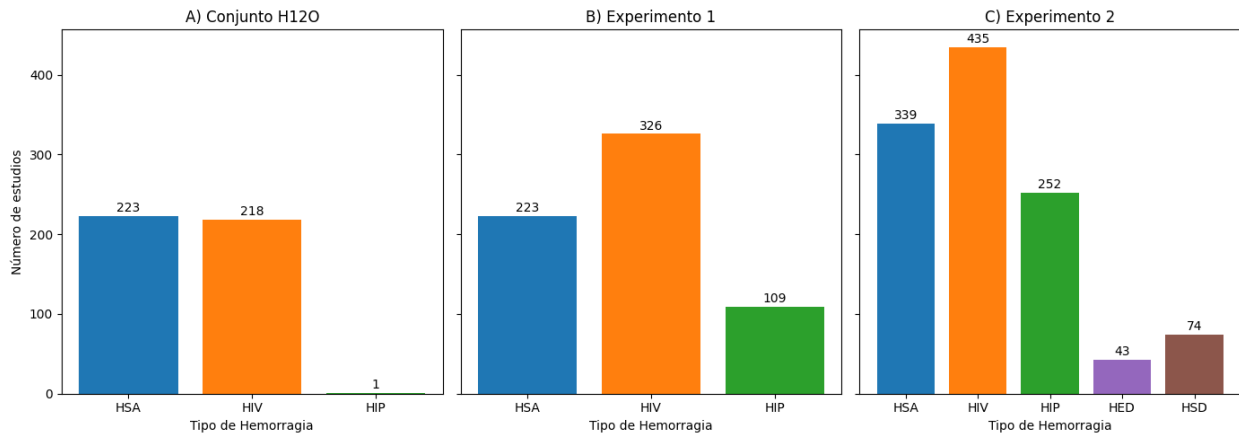


Figura 4.5: Distribución del número de estudios según el tipo de hemorragia en los tres conjuntos analizados. Se muestran los casos disponibles en el conjunto H12O (A), en el Experimento 1 (B) y en el Experimento 2 (C).

Información clínica

Variables clínicas y Desenlaces pronósticos

Además de las imágenes, en la cohorte local se han recopilado variables clínicas y radiológicas relevantes para el análisis pronóstico. La base de datos clínica inicial contiene un total de 49 variables recogidas de forma sistemática en el momento del ingreso hospitalario a lo largo del periodo de estudio (2007–2024). Estas variables, descritas en el Anexo B, en la Tabla B.2 incluyen datos demográficos (edad, sexo), antecedentes y factores de riesgo vascular, puntuaciones en escalas clínicas, escalas radiológicas (Fisher modificado) y parámetros analíticos iniciales (p. ej., glucosa, sodio), así como variables de seguimiento, como la escala GOS, la presencia de hidrocefalia o de vasoespasmio, que se emplearán como desenlaces pronósticos a predecir mediante modelos de aprendizaje automático.

La Tabla 4.4 resume la información clínica básica de 223 los pacientes del conjunto H12O usados para entrenar este modelo de segmentación automática. La edad media de los pacientes es de 54 ± 12 años, con una ligera predominancia de mujeres (63,3%) frente a hombres (36,7%).

En términos de pronóstico funcional a los 6 meses (escala de GOS), el 43,3% de los pacientes alcanza una recuperación favorable (GOS=5), mientras que el 30% ha fallecido (GOS=1). El

resto se distribuye entre valores intermedios: 14,7 % en GOS=3, 11,5 % en GOS=4 y apenas un 0,45 % en GOS=2.

La clasificación WFNS al ingreso muestra un predominio de grados bajos de severidad (35,0 % en WFNS=1 y 19,4 % en WFNS=2), aunque una proporción importante de pacientes presenta afectación grave (26,7 % en WFNS=5).

Por su parte, la escala de Fisher modificada refleja que más de la mitad de los casos corresponden a grado 3 (52,4 %), seguidos de un 33,1 % en grado 4. Los grados bajos son menos frecuentes (7,8 % en grado 2 y 6,7 % en grado 1). No se observan casos con grado 0, ya que todos los pacientes presentan hemorragia segmentada.

En conjunto, estos datos muestran una cohorte clínicamente heterogénea, tanto en severidad inicial como en evolución funcional, lo que proporciona un marco sólido para la evaluación comparativa de los modelos de segmentación y predicción.

Tabla 4.4: Distribución de las principales variables clínicas asociadas a los pacientes del conjunto H12O (n=223) para el desarrollo del algoritmo de segmentación de hemorragia.

Variable	Categoría / Descripción	Distribución	
Sexo	Mujer	63,3 %	
	Hombre	36,7 %	
GOS (6 meses)	1	30,0 %	
	2	0,45 %	
	3	14,7 %	
	4	11,5 %	
	5	43,3 %	
WFNS	1	35,0 %	
	2	19,4 %	
	3	4,1 %	
	4	14,7 %	
	5	26,7 %	
Escala de Fisher modificada	0	0 %	
	1	6,7 %	
	2	7,8 %	
	3	52,4 %	
	4	33,1 %	
	Edad	Media \pm DE	54 \pm 12 años

4.1.3 Modelos predictivos radiómicos en HSAa (TC)

Para el desarrollo de los modelos radiómicos se ha empleado la base de datos de HSAa completa del H12O, incluyendo tanto las imágenes como todos los datos clínicos recopilados. A partir de esta base de datos institucional de pacientes con diagnóstico confirmado de HSAa, se han definido dos subconjuntos con propósitos diferenciados:

- El conjunto para el desarrollo de modelos, integrado por 403 pacientes diagnosticados entre enero de 2007 y diciembre de 2023, se ha empleado para el entrenamiento, validación y test dentro de la metodología de validación cruzada anidada (véase Sección 3.4.3).
- El conjunto de test independiente, formado por una cohorte temporal distinta de 41 pacientes diagnosticados entre enero de 2023 y junio de 2024.

El periodo de solapamiento parcial entre ambos conjuntos permitió disponer de un volumen suficiente de datos para la fase de entrenamiento, mientras que el intervalo temporal más reciente del test independiente permitió simular una validación prospectiva, evaluando la estabilidad del modelo frente a posibles cambios tecnológicos o clínicos.

La Figura 4.6 resume el proceso de selección y estratificación de los pacientes del H12O, diferenciando la cohorte de desarrollo de modelos (panel A) del conjunto de test independiente (panel B). Se cuenta con un total de 498 pacientes con TC y datos clínicos. De ellos, se han excluido 95 por presentar HSA idiopática, es decir, casos en los que no se identificó una causa subyacente tras un estudio diagnóstico completo. Finalmente, se incluyen 403 pacientes (80,9 %) para el análisis radiómico y la predicción de desenlaces. Se han evaluado tres desenlaces principales: pronóstico funcional a los 6 meses (GOS), mortalidad y complicaciones asociadas (vasoespasmos e hidrocefalia). El número de pacientes incluidos en cada análisis, así como las exclusiones por ausencia de datos o fallecimiento precoz, se detallan en cada rama del diagrama. En total, 190 pacientes presentan un mal resultado funcional, 132 fallecen, 125 desarrollan vasoespasmos (96 clínico y 29 radiológico) y 33 hidrocefalia.

El panel B refleja el conjunto de test independiente, igualmente procedente del H12O, compuesto por 41 pacientes adicionales. En esta cohorte se aplican los mismos criterios de estratificación, obteniéndose 41 casos válidos para el análisis de resultado funcional y mortalidad, y 30 para vasoespasmos e hidrocefalia tras excluir a los pacientes que fallecen antes de una semana y no han sufrido el desenlace. En esta cohorte, 27 pacientes presentan un mal resultado funcional, 13 fallecen, 12 desarrollan vasoespasmos radiológico y 3 hidrocefalia. En conjunto, la Figura 4.6 sintetiza la organización de la base de datos y la disponibilidad final de pacientes para cada tarea de predicción.

Información a nivel de imagen

En relación con los protocolos de adquisición, la mayoría de las TC de la cohorte de entrenamiento se obtienen con un escáner Philips Brilliance 6 (n=380, Philips Healthcare, Ámsterdam, Países Bajos), mientras que un número menor procede de un GE Revolution EVO (n=23, GE Healthcare, Chicago, EE. UU.). Entre los parámetros de adquisición, el tamaño de píxel y el grosor de corte son especialmente relevantes, ya que determinan la resolución espacial de la imagen y pueden influir en la robustez y reproducibilidad de las características radiómicas.

En la cohorte de desarrollo de modelos, el tamaño de píxel en el plano axial muestra una distribución relativamente homogénea, con valores concentrados entre 0,39 y 0,50 mm y un predominio en torno a 0,47–0,49 mm. El grosor de corte es más variable, aunque con predominio de cortes finos de 1.5 mm (74 % de los casos) y, en menor medida, de 3 mm (23 %), mientras que otros valores son mucho más escasos (<3 %).

En el conjunto de test independiente, las exploraciones procedieron mayoritariamente de un escáner GE Revolution EVO (n=38, GE Healthcare, Chicago, EE. UU.), con un pequeño número adquirido en un Philips Brilliance 6 (n=3, Philips Healthcare, Ámsterdam, Países Bajos). En este caso, el tamaño de píxel también se mantiene en rangos similares a los de la cohorte de entrenamiento, con valores más concentrados alrededor de 0,48–0,50 mm, lo que indica una elevada homogeneidad. Sin embargo, el grosor de corte es más uniforme: el 100 %

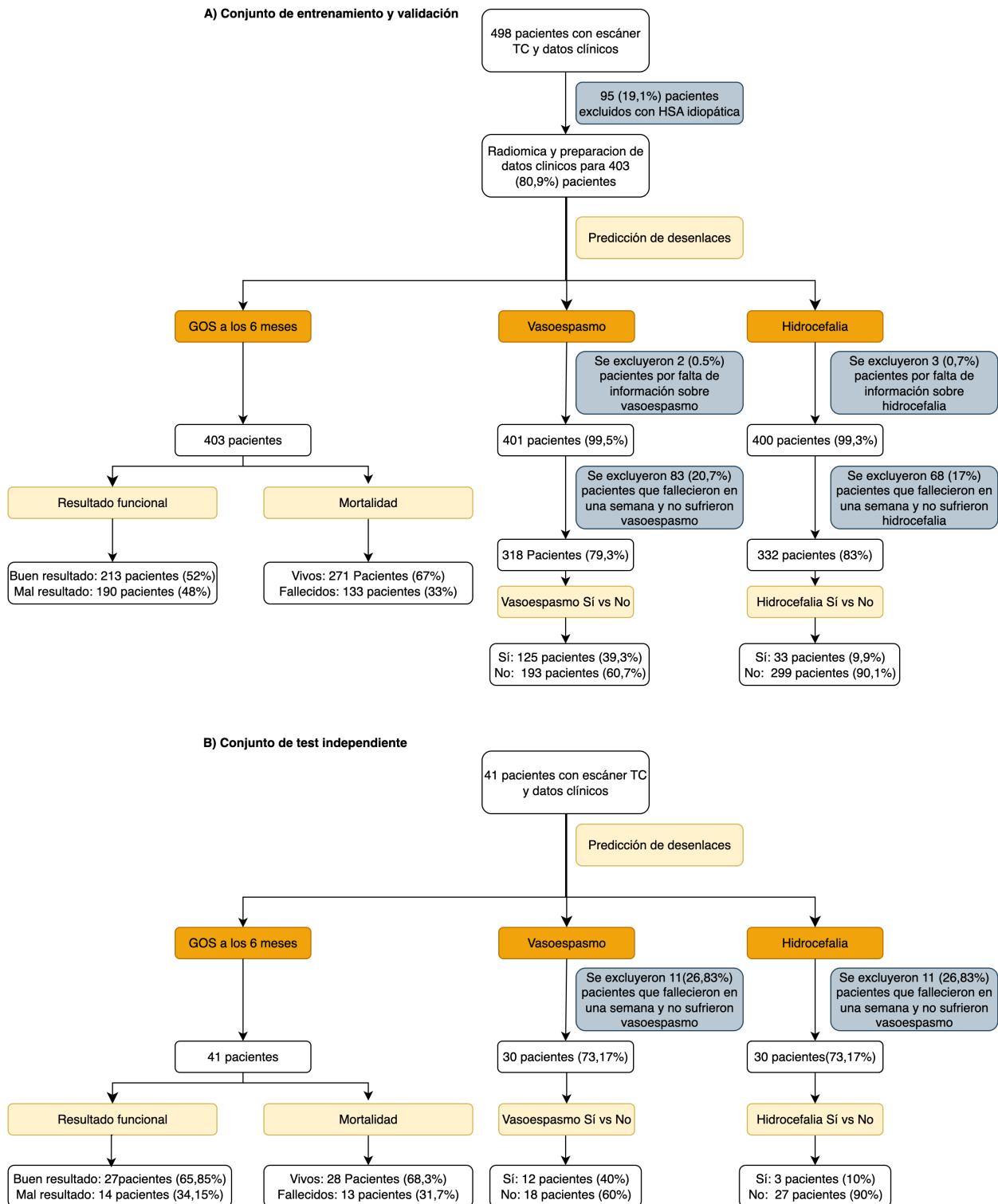
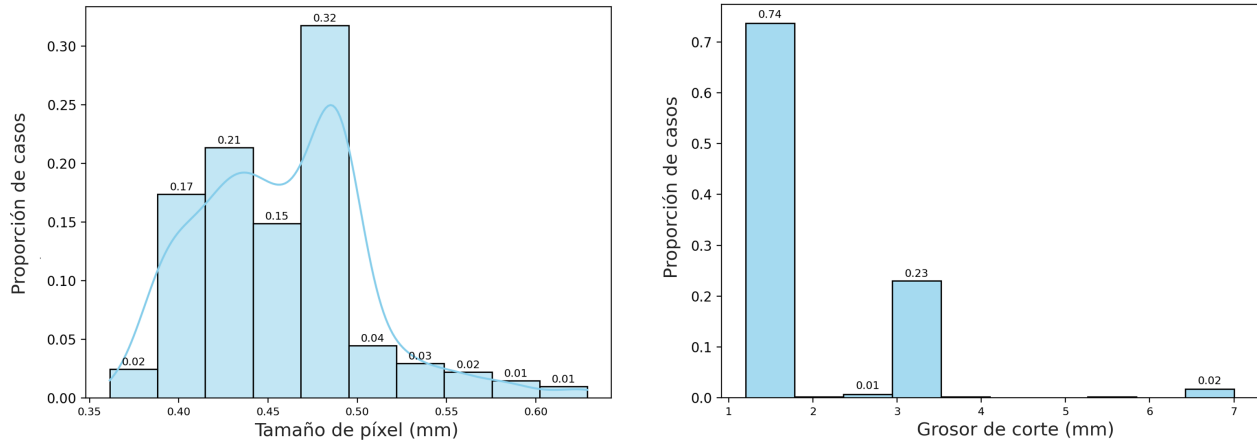


Figura 4.6: Diagramas de flujo de pacientes según los diferentes desenlaces (resultado funcional, mortalidad, vasoespasm e hidrocefalia). a) Conjunto de desarrollo del modelo (entrenamiento, validación y test) y b) Conjunto de test independiente

de los estudios presentan cortes de 1.5 mm.

La Figura 4.7 ilustra la distribución del tamaño de píxel y grosor de corte en el conjunto de desarrollo del modelo y por otro lado, en el test independiente.

a) Conjunto de entrenamiento y validación



b) Conjunto de test independiente

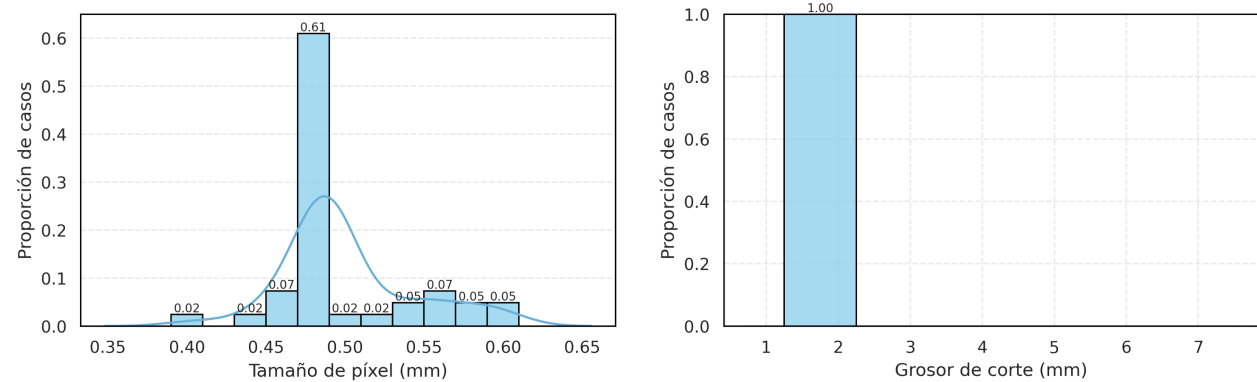


Figura 4.7: Distribución de los parámetros de adquisición en las cohortes utilizadas para los modelos predictivos de desenlaces clínicos. Histogramas del tamaño de píxel (izquierda) y grosor de corte (derecha) en a) Conjunto de desarrollo del modelo y b) Conjunto de test independiente.

Información clínica

Como se menciona anteriormente, la base de datos de información clínica cuenta con 49 variables clínicas. A partir de esa base de datos, se ha llevado a cabo una depuración sistemática de las variables recogidas en el H120. En primer lugar, se ha evaluado el porcentaje de valores nulos en cada variable clínica. Aquellas con más de un 15% de datos ausentes se excluyen del análisis. Como resultado, 20 variables clínicas superan este filtrado. En estas 20 variables se aplica la imputación basada en MICE únicamente sobre las variables clínicas que no corresponden a los desenlaces a predecir (hidrocefalia, vasoespasma y GOS). La Figura 4.8 muestra el porcentaje de valores nulos correspondientes a variables clínicas del conjunto original de datos. Se representan únicamente aquellas variables con significado clínico o analítico relevante, excluyendo campos técnicos o administrativos (por ejemplo, fechas o

parámetros de adquisición).

Finalmente, se ha decidido restringir el análisis exclusivamente a las variables disponibles en el momento del diagnóstico inicial, descartando aquellas que se derivan de decisiones terapéuticas o de la evolución intrahospitalaria. En consecuencia, variables como *Tratamiento*, *Resangrado* o *Ingreso en UCI*, si bien clínicamente relevantes para el pronóstico, no ha sido consideradas en el desarrollo de los modelos predictivos, puesto que no forman parte de la información que un clínico tendría en el momento de realizar la primera valoración diagnóstica.

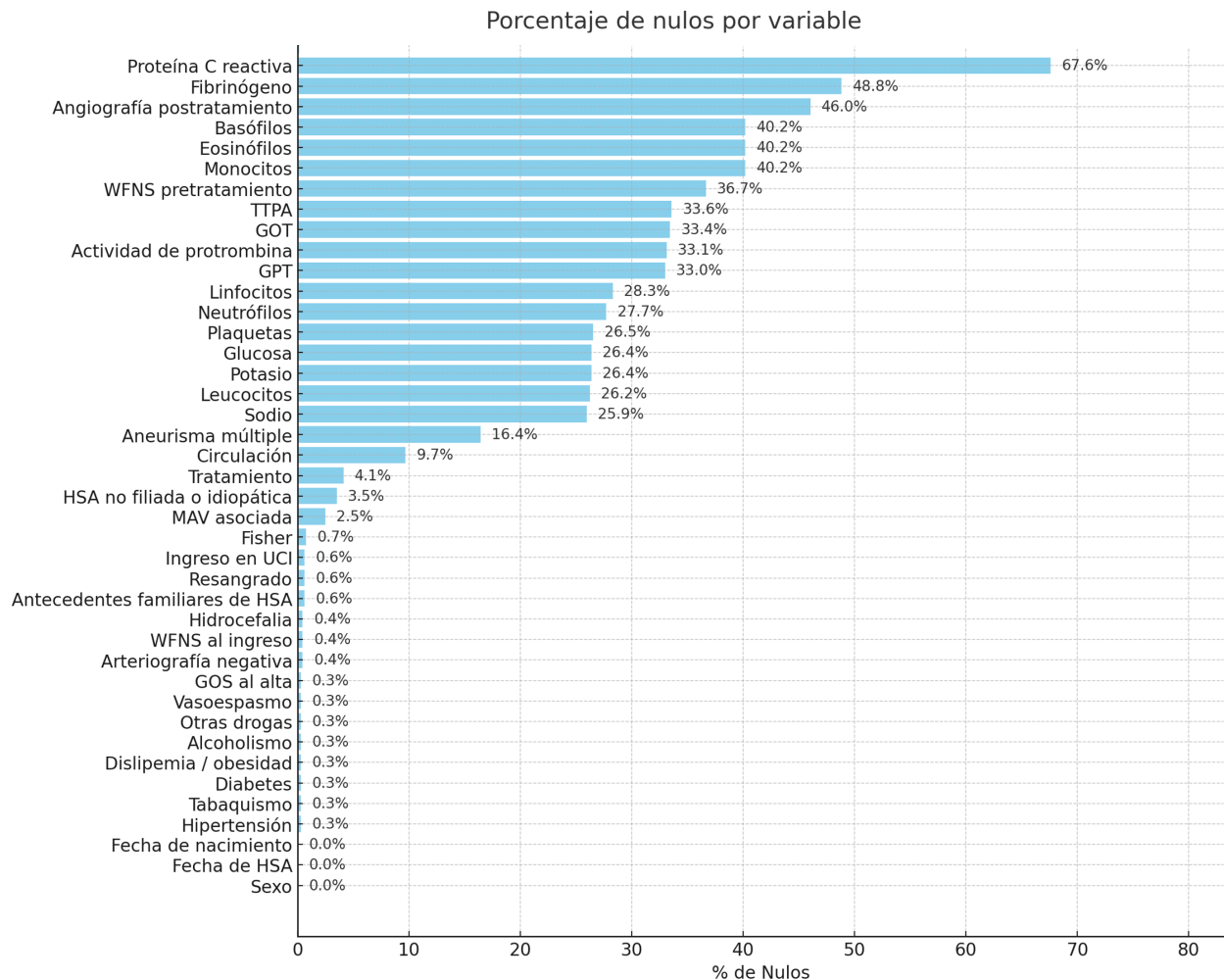


Figura 4.8: Porcentaje de valores nulos en la base de datos de variables clínicas relevantes del H12O.

Como resultado del proceso de filtrado, la Tabla 4.5 presenta las variables finalmente incluidas, junto con su naturaleza (numérica o categórica) y una breve descripción.

Para mayor información, en el Anexo B se presenta la Tabla B.2, donde se representan las variables clínicas categóricas y continuas incluidas en el estudio, con los valores que pueden tomar o descripción correspondiente.

Las variables de referencia, que recogen los valores esperados en población sana, se describen

en detalle en la Tabla B.3 del Anexo B, donde se especifican los rangos normales de las principales magnitudes analíticas.

Por su parte, en la Tabla B.4 del mismo anexo se incluyen los criterios diagnósticos y umbrales clínicos empleados para distintas variables incluidas en este estudio.

Tabla 4.5: Variables clínicas incluidas en el conjunto final tras aplicar el criterio de exclusión por datos faltantes.

Variable	Tipo	Descripción
Edad	Numérica	Edad del paciente al ingreso (años)
Género	Categórica	Sexo biológico (hombre/mujer)
Hipertensión arterial	Categórica	Antecedente de hipertensión arterial
Tabaquismo activo	Categórica	Consumo actual de tabaco
Diabetes	Categórica	Antecedente de diabetes mellitus
Dislipemia/obesidad	Categórica	Antecedente de dislipemia y/o obesidad
Alcoholismo	Categórica	Trastorno por consumo de alcohol documentado
Consumo de otras drogas	Categórica	Consumo activo de drogas recreativas
Antecedentes familiares de HSA	Categórica	Familiar de primer grado con antecedente de HSA
WFNS al ingreso	Categórica	Escala de la World Federation of Neurosurgical Societies al ingreso
Escala de Fisher modificada	Categórica	Grado de Fisher modificado en la TC inicial
AVM asociada	Categórica	Diagnóstico de malformación arteriovenosa asociada
Circulación	Categórica	Localización de la circulación aneurismática (anterior/posterior)
Glucemia al ingreso	Numérica	Glucosa sérica (mg/dL) al ingreso
Potasio al ingreso	Numérica	Concentración sérica de potasio (mmol/L)
Sodio al ingreso	Numérica	Concentración sérica de sodio (mmol/L)
Plaquetas al ingreso	Numérica	Recuento plaquetario ($\times 10^3/\mu\text{L}$)
Leucocitos al ingreso	Numérica	Recuento leucocitario ($\times 10^3/\mu\text{L}$)
Neutrófilos al ingreso	Numérica	Recuento absoluto de neutrófilos ($\times 10^3/\mu\text{L}$)
Linfocitos al ingreso	Numérica	Recuento absoluto de linfocitos ($\times 10^3/\mu\text{L}$)

En la Tabla 4.7 se presentan las características clínicas de las cohortes incluidas, diferenciando la distribución en el conjunto de desarrollo de modelos y en el de test independiente.

En el conjunto de desarrollo de modelos, los pacientes muestran una edad media de 52 ± 18 años, con predominio femenino (68,5%). Los valores de glucosa son moderadamente elevados (151 ± 56 mg/dL), mientras que el recuento plaquetario se sitúa dentro de la normalidad ($398 \pm 101 \times 10^9/L$). En cuanto a electrolitos, las concentraciones medias de potasio ($3,8 \pm 0,6$ mmol/L) y sodio ($139,6 \pm 4,9$ mmol/L) se mantienen en rangos fisiológicos habituales. Por su parte, los recuentos de leucocitos y linfocitos muestran medias elevadas y una alta dispersión: leucocitos $29,6 \pm 357,9 \times 10^9/L$, neutrófilos $18,1 \pm 160,9 \times 10^9/L$ y linfocitos $9,2 \pm 165,7 \times 10^9/L$. Los valores de referencia fisiológicos empleados para contextualizar estas variables se detallan en la Tabla B.3 en el Anexo B.

Tabla 4.7: Características clínicas y comparación entre cohortes.

Variable	Train/Val	Test	p	Tamaño efecto	Sig.
<i>Variables categóricas (%)</i>					
Sexo (Mujer)	68,5 %	58,5 %	0,05	0,15	No
Hipertensión (Sí)	41,7 %	51,2 %	0,03	0,22	Sí
Tabaquismo (Sí)	28,5 %	43,9 %	0,01	0,31	Sí
Diabetes (Sí)	18,6 %	14,6 %	0,33	0,10	No
Dislipemia/Obesidad (Sí)	17,9 %	48,8 %	$\leq 0,001$	0,64	Sí
Alcoholismo (Sí)	5,7 %	14,6 %	0,01	0,27	Sí
Otros fármacos (Sí)	3,2 %	4,9 %	0,57	0,08	No
Antecedentes familiares HSA	1,5 %	0,0 %	0,34	0,09	No
WFNS (≥ 4)	38,4 %	61,0 %	$\leq 0,001$	0,54	Sí
Fisher modificado (≥ 3)	79,4 %	90,2 %	0,03	0,21	Sí
MAV asociada (Sí)	1,4 %	0,0 %	0,43	0,08	No
Circulación posterior	9,0 %	0,0 %	0,04	0,24	Sí
<i>Variables numéricas (media \pm DE)</i>					
Edad (años)	52 \pm 18	59 \pm 14	0,03	-0,43	Sí
Glucosa (mg/dL)	151 \pm 56	186 \pm 63	0,02	-0,58	Sí
Potasio (mmol/L)	3,8 \pm 0,6	3,7 \pm 0,5	0,33	0,23	No
Sodio (mmol/L)	139,6 \pm 4,9	140,2 \pm 4,9	0,52	-0,14	No
Plaquetas ($10^9/L$)	398 \pm 101	238 \pm 53	$\leq 0,001$	5,69	Sí
Leucocitos ($10^9/L$)	29,6 \pm 357,9	13,2 \pm 4,5	0,47	0,06	No
Neutrófilos ($10^9/L$)	18,1 \pm 160,9	10,6 \pm 4,4	0,59	0,07	No
Linfocitos ($10^9/L$)	9,2 \pm 165,7	1,8 \pm 1,2	0,72	0,06	No

En cuanto a la severidad inicial, el 38,4% presenta WFNS ≥ 4 y el 79,4% un Fisher modificado ≥ 3 . Los factores de riesgo cardiovascular tienen una prevalencia moderada: hipertensión en el 41,7%, tabaquismo en el 28,5%, diabetes en el 18,6% y dislipemia/obesidad en el 17,9%. El consumo de alcohol es menos frecuente (5,7%), al igual que el uso de otras drogas (3,2%) y los antecedentes familiares de HSA (1,5%). La presencia de malformación arteriovenosa (MAV) asociada es poco común (1,4%). La mayoría de aneurismas se localizan en circulación anterior (90,9%), con un 9,0% en circulación posterior.

El análisis comparativo con la cohorte de test independiente evidencia diferencias significativas en múltiples variables clínicas (Tabla 4.7). Los pacientes del test independiente son de mayor edad (59 ± 14 años; $p = 0,03$) y con una distribución por sexos más equilibrada (58,5% mujeres). La glucosa es significativamente más alta (186 ± 63 mg/dL; $p = 0,02$) y el recuento plaquetario notablemente más bajo ($238 \pm 53 \times 10^9/L$; $p < 0,001$). Aunque el potasio ($3,7 \pm 0,5$ mmol/L) y sodio ($140,2 \pm 4,9$ mmol/L) no muestran diferencias relevantes, los recuentos leucocitarios y diferenciales son más bajos y homogéneos (leucocitos $13,2 \pm 4,5$, neutrófilos $10,6 \pm 4,4$ y linfocitos $1,8 \pm 1,2 \times 10^9/L$), sin alcanzar significación estadística.

En términos de severidad, la cohorte de test independiente presenta un perfil más desfavorable, con un 61,0% de pacientes con WFNS ≥ 4 ($p < 0,001$) y un 90,2% con Fisher modificado ≥ 3 ($p = 0,03$). También se observa una mayor prevalencia de factores de riesgo: hipertensión (51,2%; $p = 0,03$), tabaquismo (43,9%; $p = 0,01$), dislipemia/obesidad (48,8%; $p < 0,001$) y alcoholismo (14,6%; $p = 0,01$). La diabetes es similar entre cohortes (14,6%; NS), mientras que los antecedentes familiares de HSA, el consumo de otras drogas y la presencia de MAV asociada son anecdóticos o ausentes. Por último, ningún paciente del grupo test independiente presenta aneurismas en circulación posterior.

En conjunto, la cohorte de test independiente concentró pacientes de mayor edad, con mayor prevalencia de factores de riesgo vascular, cifras más altas de glucosa, plaquetopenia significativa y una mayor gravedad clínica al ingreso. Estos hallazgos sugieren un perfil clínico más adverso respecto a la cohorte de desarrollo de modelos.

4.2 Extracción de características

Esta sección se organiza en dos bloques principales, correspondientes a las modalidades de imagen empleadas en el estudio.

En primer lugar, en la Sección 4.2.1 se presentan las firmas espectrales obtenidas a partir de las segmentaciones manuales de las HSI en pacientes con GBM. Estas firmas reflejan la variación de la reflectancia en función de la longitud de onda y permiten explorar las diferencias ópticas entre distintos tipos de tejido.

En segundo lugar, en la Sección 4.2.2 se analizan las características radiómicas extraídas de escáneres TC de pacientes con HSAa. Dichas características se han obtenido tanto de las segmentaciones automáticas realizadas por un algoritmo previamente publicado y entrenado específicamente para HSAa (SMCCH, descrito en la Sección 3.4.2), como de las segmentaciones generadas por el mejor modelo desarrollado en este trabajo (Experimento 2), identificado como el más sólido en términos cuantitativos y cualitativos (véase Sección 4.3.2).

4.2.1 Firmas espectrales en glioblastoma (Imagen Hiperespectral)

A través de la segmentación manual, se aíslan las regiones correspondientes a tejido sano, tumor, sangre venosa, sangre arterial y duramadre, a partir de las cuales se calculan las curvas promedio y sus intervalos de variabilidad.

Las Figuras 4.9 y 4.10 muestran las firmas espectrales medias obtenidas tras la normalización de la reflectancia para cada clase tisular. Estas representan tanto la curva promedio como la variabilidad intra-clase. En la Figura 4.9, las curvas promedio de cada clase se presentan de forma independiente, acompañadas de un área sombreada que refleja la dispersión intra-clase. Esta representación permite valorar la consistencia dentro de cada grupo y las diferencias en la morfología espectral. Por otro lado, la Figura 4.10 integra en un mismo gráfico las curvas promedio de las cinco clases, lo que facilita la comparación directa entre tejidos y resalta sus contrastes espectrales.

Se observa que las curvas de tejido sano y tumor presentan patrones muy similares en todo el espectro. Por su parte, la sangre arterial y venosa muestran picos más marcados en determinadas longitudes de onda (como 891,89 nm). La duramadre, en cambio, exhibe una tendencia descendente más acusada en la región de mayores longitudes de onda, separándose del resto de tejidos.

Cabe destacar que la distribución de los píxeles anotados para cada clase tisular (Tabla 4.1) se refleja en las firmas espectrales medias, con curvas más estables en las clases mayoritarias y una mayor dispersión relativa en las minoritarias. El tejido sano y el tumor concentran la mayoría de los píxeles segmentados, mientras que las clases vasculares y la duramadre están menos representadas.

4.2.2 Características radiómicas en Hemorragia Subaracnoidea Aneurismática (TC)

Como se ha comentado en la Sección 3.4.3, se han explorado varias configuraciones para integrar diferentes fuentes de información, evaluando su impacto individual y combinado en la capacidad predictiva:

1. **Solo variables clínicas:** modelos desarrollados exclusivamente con las 20 variables clínicas seleccionadas tras el preprocesamiento y depuración inicial.
2. **Solo características radiómicas:**
 - Segmentación de hemorragias: 1379 características radiómicas extraídas de las máscaras de hemorragia.
 - Segmentación de materia gris y blanca: 2758 características radiómicas obtenidas de las máscaras tisulares.
3. **Combinación de variables clínicas y radiómicas:**
 - Segmentación de hemorragias: 1399 características resultantes de la combinación de variables clínicas y radiómicas de hemorragia.

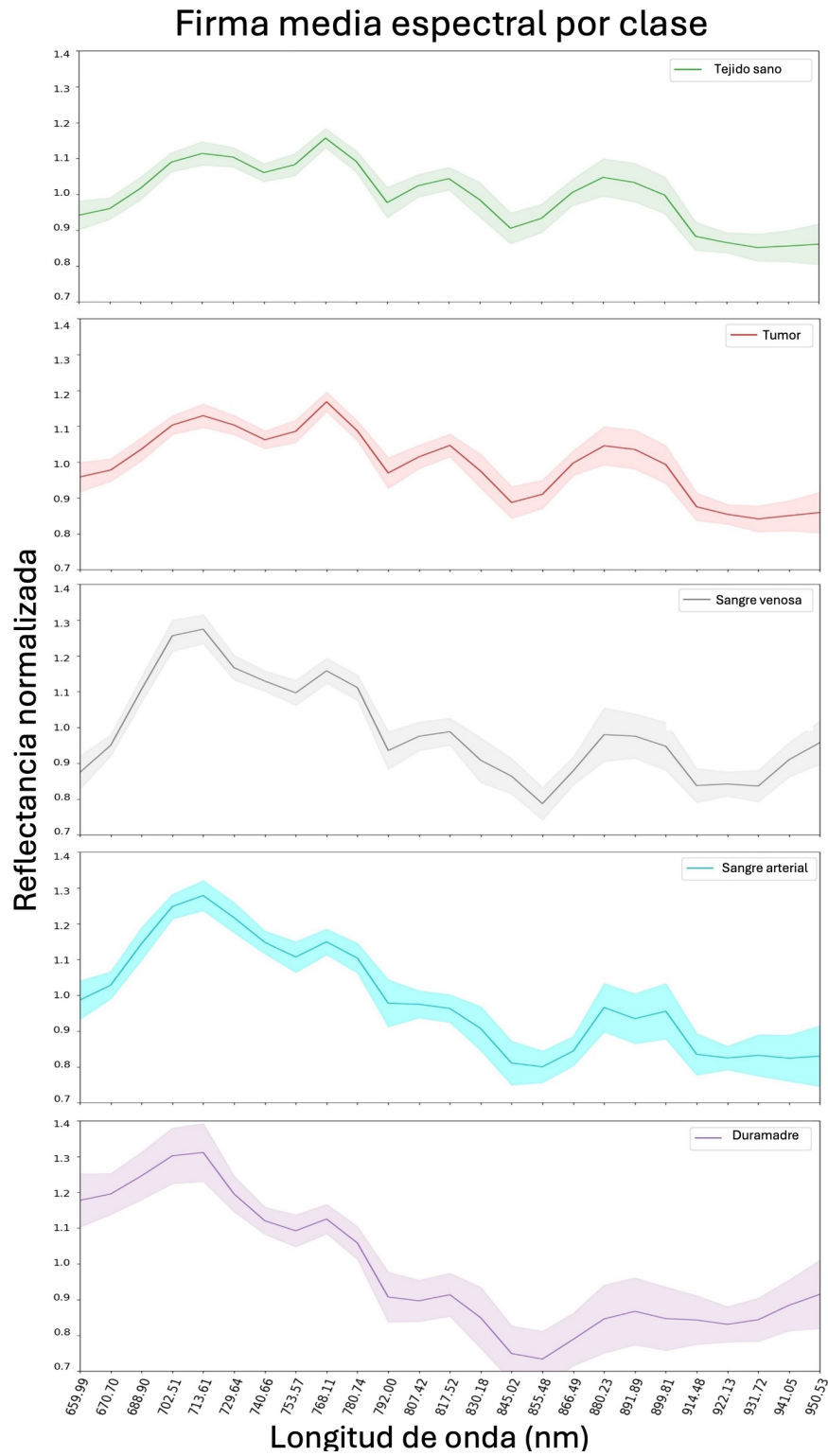


Figura 4.9: Firma media espectral por clase, representada de manera independiente para cada tipo de tejido, incluyendo el intervalo de variabilidad (área sombreada).

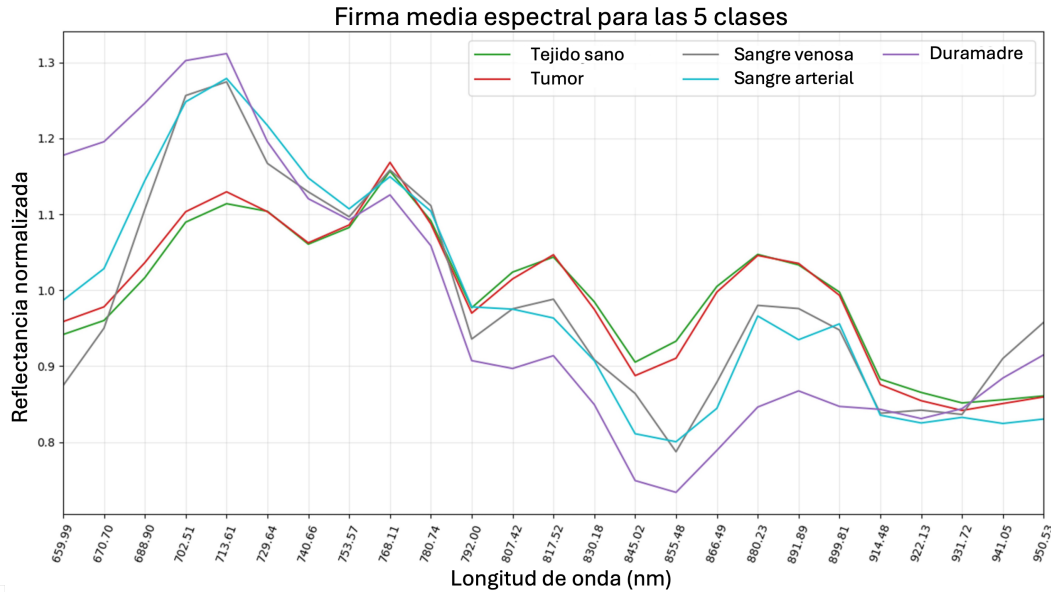


Figura 4.10: Comparación de la firma media espectral para las cinco clases en un mismo gráfico.

- Segmentación de materia gris y blanca: 2778 características combinadas.

Se ha analizado la correlación entre el conjunto de características completas que incluye variables clínicas y radiómicas extraídas de tres regiones de interés segmentadas automáticamente (hemorragia del Experimento 2, hemorragia segmentada con SMCCH y parénquima cerebral) y los principales desenlaces clínicos (resultado funcional, mortalidad, vasoespasma e hidrocefalia).

La Tabla 4.9 resume las variables que muestran mayor correlación con los distintos desenlaces en las tres cohortes evaluadas. De manera consistente, la escala clínica WFNS al ingreso se identifica como una de las variables más fuertemente asociadas con el resultado funcional y la mortalidad. En las regiones hemorrágicas (Experimento 2 y SMCCH) destacan además numerosos descriptores texturales derivados de transformadas wavelet y filtros laplacianos (GLRLM, GLSZM, GLDM), especialmente aquellos relacionados con la no uniformidad de niveles de gris, que se asocian significativamente con desenlaces funcionales y mortalidad.

En el parénquima cerebral también se observan asociaciones relevantes, en particular con el resultado funcional y la mortalidad. Parámetros texturales de secuencias filtradas (GLRLM en escala log-sigma) muestran correlaciones de magnitud comparable a las obtenidas en las regiones hemorrágicas.

Para hidrocefalia, las correlaciones son más modestas y heterogéneas, involucrando variables de primer orden y dependencias de nivel de gris en las tres cohortes, sin un patrón tan marcado como en los desenlaces funcionales. En el caso del vasoespasma, destacan asociaciones con características de alta frecuencia (wavelet-HHH) en la hemorragia y con descriptores de forma y textura en el parénquima.

En conjunto, los hallazgos reflejan que, además de los factores clínicos clásicos, existen firmas radiómicas que muestran correlaciones significativas con los principales eventos clínicos, lo

que abre la puerta a su posible integración en modelos predictivos multimodales.

Además, la Figura 4.11 muestra un análisis de Bland–Altman para evaluar la concordancia en la estimación del volumen de la hemorragia entre el modelo propuesto (Experimento 2) y el método de referencia descrito en la literatura (SMCCH). El volumen se calcula a partir de la variable radiómica `original_shape_Volume`, que cuantifica directamente el tamaño tridimensional de la región segmentada.

En el eje horizontal del gráfico se representa la media de los volúmenes obtenidos por ambos métodos (Experimento 2 y SMCCH), mientras que en el eje vertical se muestra la diferencia entre ellos (Experimento 2 – SMCCH). La línea roja discontinua indica la diferencia media entre los dos métodos, y las líneas grises delimitan los límites de concordancia (media $\pm 1,96$ DE), dentro de los cuales se espera que se sitúe el 95 % de las diferencias individuales.

El gráfico evidencia un sesgo sistemático negativo, ya que la mayoría de los puntos se sitúan por debajo de la línea de cero, lo que indica que la hemorragia obtenida del Experimento 2 tiende a reportar volúmenes inferiores a los de SMCCH. Además, se observa que este sesgo aumenta a medida que crece la magnitud de las medidas (efecto proporcional), con diferencias especialmente marcadas en los casos de mayor volumen.

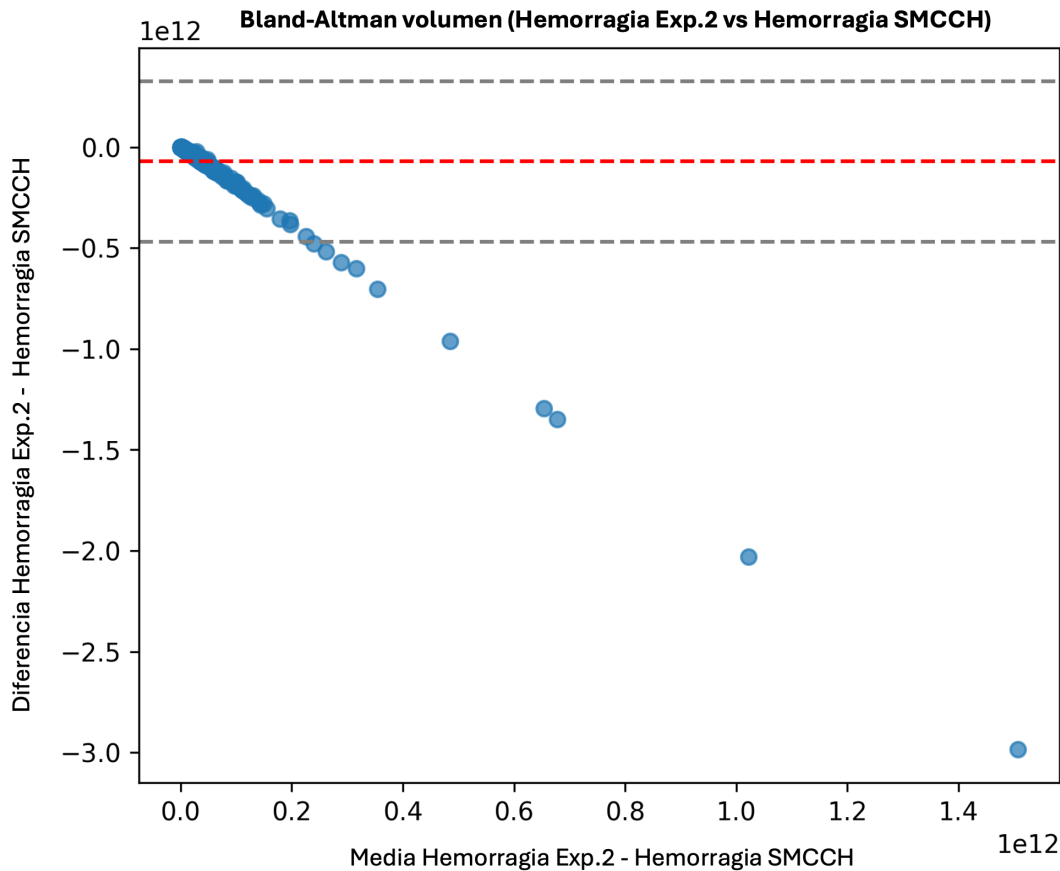


Figura 4.11: Gráfico de Bland–Altman global comparando las mediciones de volumen de hemorragia obtenidas con los algoritmos automáticos nuestro Experimento 2 y con SMCCH.

ROI	Desenlace	Característica	Coefficiente	p-valor
HSA del Experimento 2	Resultado Funcional	wfns_al_ingreso	-0,57	9.53e-43
		wavelet-HLH_glrmlm_RunLengthNonUniformity	-0,52	2.16e-34
		wavelet-LHH_glrmlm_RunLengthNonUniformity	-0,52	3.15e-34
		wavelet-LHH_gldm_DependenceNonUniformity	-0,51	8.97e-34
	Hidrocefalia	log-sigma-0-5-mm-3D_gldm_DependenceNonUniformity	-0,51	1.02e-33
		log-sigma-0-5-mm-3D_glszm_LargeAreaLowGrayLevelEmphasis	0,19	1.00e-03
		original_gldm_GrayLevelNonUniformity	0,18	1.03e-03
		wavelet-LLL_glszm_ZoneVariance	0,18	1.19e-03
		wavelet-LLL_glszm_LargeAreaEmphasis	0,17	1.22e-03
	Mortalidad	log-sigma-0-5-mm-3D_glrmlm_GrayLevelNonUniformityNormalized	-0,18	1.23e-03
		wfns_al_ingreso	-0,50	4.74e-32
		wavelet-LLH_glrmlm_RunLengthNonUniformity	0,48	5.36e-29
		wavelet-LLL_glrmlm_RunLengthNonUniformity	0,48	7.86e-29
		wavelet-HLH_glrmlm_RunLengthNonUniformity	0,48	1.52e-28
	Vasoespasmio	wavelet-LHH_glrmlm_RunLengthNonUniformity	0,48	1.58e-28
		wavelet-HHH_gldm_LargeDependenceLowGrayLevelEmphasis	0,25	9.52e-06
		wavelet-HHH_glcm_Imc2	-0,24	1.66e-05
		log-sigma-3-mm-3D_gldm_LargeDependenceEmphasis	0,24	3.12e-05
		wavelet-HHH_glcm_JointEnergy	-0,24	3.19e-05
HSA de SMCCCH	Resultado Funcional	wavelet-HHH_gldm_SmallDependenceEmphasis	-0,23	3.35e-05
		wfns_al_ingreso	-0,57	6.49e-44
		wavelet-HLH_glszm_GrayLevelNonUniformity	-0,41	7.68e-22
		wavelet-HHH_glszm_GrayLevelNonUniformity	-0,40	1.99e-20
		wavelet-LHH_glszm_GrayLevelNonUniformity	-0,40	2.88e-20
	Hidrocefalia	wavelet-HLL_glszm_GrayLevelNonUniformity	-0,39	9.67e-20
		log-sigma-0-5-mm-3D_firstorder_Kurtosis	0,18	8.29e-04
		wavelet-LHL_glcm_Correlation	0,17	2.10e-03
		logarithm_glcm_Idn	0,17	2.26e-03
		wfns_al_ingreso	-0,17	2.39e-03
	Mortalidad	original_gldm_SmallDependenceLowGrayLevelEmphasis	-0,16	3.13e-03
		wfns_al_ingreso	-0,49	4.41e-31
		wavelet-HLH_glszm_GrayLevelNonUniformity	0,40	1.22e-20
		wavelet-LHH_glszm_GrayLevelNonUniformity	0,39	1.43e-19
		wavelet-HHH_glszm_GrayLevelNonUniformity	0,39	2.22e-19
	Vasoespasmio	wavelet-LLH_glszm_SizeZoneNonUniformity	0,39	2.50e-19
		log-sigma-0-5-mm-3D_glcm_Imc2	-0,25	7.12e-06
		log-sigma-0-5-mm-3D_glcm_Imc1	0,23	2.76e-05
		original_firstorder_Minimum	0,23	4.55e-05
wavelet-HHH_glcm_Correlation		0,22	1.02e-04	
Parénquima	Resultado Funcional	log-sigma-0-5-mm-3D_firstorder_Mean	0,21	1.65e-04
		wfns_al_ingreso	-0,57	2.01e-43
		log-sigma-0-5-mm-3D_glrmlm_HighGrayLevelRunEmphasis_GREY	-0,43	8.30e-24
		log-sigma-0-5-mm-3D_glrmlm_GrayLevelNonUniformityNormalized_GREY	0,43	1.30e-23
		log-sigma-0-5-mm-3D_glrmlm_LowGrayLevelRunEmphasis_GREY	0,43	2.41e-23
	Hidrocefalia	log-sigma-0-5-mm-3D_glrmlm_GrayLevelVariance_GREY	-0,42	8.65e-23
		wavelet-HHL_firstorder_Median_GREY	0,20	2.69e-04
		logarithm_firstorder_Kurtosis_GREY	-0,19	6.79e-04
		logarithm_glszm_ZoneEntropy_WHITE	-0,18	8.65e-04
		wavelet-HLH_firstorder_Kurtosis_GREY	0,18	8.71e-04
	Mortalidad	log-sigma-3-mm-3D_firstorder_Maximum_WHITE	-0,18	9.02e-04
		wfns_al_ingreso	-0,49	3.77e-31
		log-sigma-0-5-mm-3D_glrmlm_GrayLevelNonUniformityNormalized_GREY	-0,42	4.72e-23
		log-sigma-0-5-mm-3D_glrmlm_HighGrayLevelRunEmphasis_GREY	0,42	2.46e-22
		log-sigma-0-5-mm-3D_glrmlm_LowGrayLevelRunEmphasis_GREY	-0,41	6.62e-22
	Vasoespasmio	log-sigma-0-5-mm-3D_glrmlm_GrayLevelVariance_GREY	0,41	3.19e-21
		log-sigma-5-mm-3D_firstorder_Skewness_GREY	0,29	9.12e-08
		original_shape_Flatness_WHITE	0,26	1.84e-06
		wavelet-LHH_glszm_SmallAreaEmphasis_WHITE	0,26	1.99e-06
wavelet-LHH_glszm_SizeZoneNonUniformityNormalized_WHITE		0,26	2.14e-06	
		wavelet-LHH_glszm_SmallAreaHighGrayLevelEmphasis_WHITE	0,26	2.53e-06

Tabla 4.9: Características radiómicas más correlacionadas con cada desenlace en las diferentes cohortes (coeficiente de correlación de Spearman y p-valor).

4.3 Modelado y evaluación

La presente sección constituye el núcleo del análisis cuantitativo de esta tesis doctoral, centrado en el modelado y la evaluación de los métodos propuestos. Tras la preparación de los datos y la extracción de características, resulta esencial establecer estrategias de modelado que permitan no solo clasificar y segmentar la información de manera robusta, sino también evaluar su capacidad predictiva y compararla con los estándares clínicos de referencia.

En primer lugar, se presentan los experimentos de clasificación tisular en GBM mediante HSI, donde se explora el potencial de esta modalidad para diferenciar tejidos tumorales en un contexto intraoperatorio (Sección 4.3.1).

A continuación, se muestran los resultados de la segmentación automática de hemorragia en TC (Sección 4.3.2), comparando los algoritmos desarrollados con métodos descritos en la literatura y con la valoración de expertos. Posteriormente, se analizan los modelos predictivos radiómicos (Sección 4.3.3), entrenados a partir de la cohorte de pacientes del H12O con HSAa para estimar desenlaces clínicamente relevantes, como mortalidad, resultado funcional, desarrollo de hidrocefalia o vasoespasmos. Con el fin de favorecer su aplicabilidad, se incluye un apartado específico dedicado a la interpretabilidad de los modelos, donde se identifican las características que más contribuyen a las predicciones y se examina su coherencia con el conocimiento clínico existente. Finalmente, se lleva a cabo una comparación con escalas clínicas clásicas, lo que permite contextualizar el valor añadido de los modelos basados en radiómica y evaluar su potencial como complemento a la práctica clínica habitual.

4.3.1 Clasificación tisular en glioblastoma (Imagen Hiperespectral)

Antes de presentar los resultados, conviene recordar que se han planteado dos experimentos con objetivos complementarios (véase Sección 3.4.1). El **Experimento A** aborda la clasificación *intrapaciente*, entendida como la capacidad del modelo para reconocer correctamente los distintos tipos de tejido dentro de un mismo paciente. Para ello, en cada caso individual se reserva un 20% de los píxeles como **test**, mientras que el 80% restante se utiliza para entrenar y validar mediante una validación cruzada anidada triple. Este diseño permite estimar el rendimiento máximo alcanzable cuando el sistema usa para entrenar un porcentaje de datos del mismo paciente durante el entrenamiento.

En contraste, el **Experimento B** se plantea como clasificación *interpaciente*, es decir, evaluar si el modelo es capaz de generalizar cuando se enfrenta a pacientes completamente nuevos cuyos datos no se emplean en el entrenamiento. Este caso reproduce de forma más realista el escenario clínico en el que el modelo se aplicaría a sujetos no vistos previamente.

Los resultados de estos modelos parten de la base de datos descrita en la Sección 4.1.1.

Experimento A

En la Tabla 4.10 se muestra la métrica OACC obtenida por cada algoritmo ML tras la predicción individual de cada paciente, junto con sus valores medios y DE. Los resultados confirman que los tres modelos son capaces de aprender a partir del conjunto de imágenes seleccionadas, lo que respalda la viabilidad de esta base de datos, al menos para la fase de

entrenamiento del Experimento B.

Tal como se observa en la Figura 4.12, tanto RF como CNN alcanzan errores sistemáticamente menores que SVM en la mayoría de los pacientes. En RF, el error nunca supera el 10 %, con valores de OACC siempre superiores al 87,6 % y, en la mayoría de los casos, por encima del 95 %. Por su parte, CNN logra el OACC más alto en tres pacientes, mostrando además una distribución de errores muy similar a la de RF.

En contraste, SVM presenta un rendimiento más irregular, con valores de OACC inferiores al 80 % en la mayoría de los casos y un porcentaje de error sustancialmente mayor. Las únicas excepciones corresponden a los pacientes ID29 e ID35, donde los tres algoritmos muestran desempeños comparables.

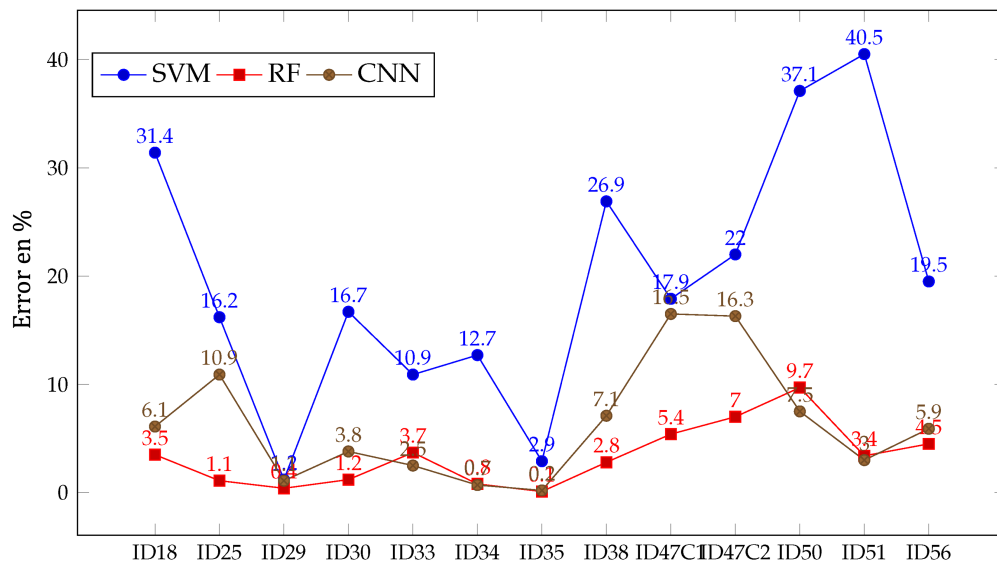


Figura 4.12: Error en % para SVM (azul), RF (rojo) y CNN (marrón) tras clasificar cada imagen de paciente en el experimento A.

Tras obtener las predicciones de los tres algoritmos, se calcula la ACC media para cada tipo de tejido considerando a todos los pacientes. Estos resultados se presentan en la Figura 4.13. En ella se aprecia que RF y CNN alcanzan valores de ACC muy similares en cada clase tisular, mostrando un desempeño consistente a lo largo de las predicciones. Sin embargo, SVM rinde claramente por debajo en tejido sano, tumor, sangre venosa y duramadre.

En términos absolutos, los valores medios de ACC obtenidos por RF y CNN superan el 90 % en todas las clases, con un comportamiento notablemente estable.

Desde el punto de vista clínico, durante la cirugía no resulta crítico que el neurocirujano distinga de manera precisa entre venas y arterias. Por este motivo, se ha llevado a cabo una evaluación adicional en la que ambas clases se fusionan en una única categoría de sangre. Para calcular la Sens de esta clase genérica de sangre, se aplica la siguiente ecuación:

$$\text{Sens}_{\text{sangre}} = \frac{TP_{\text{arterial}} + TP_{\text{venous}}}{(TP_{\text{arterial}} + FN_{\text{arterial}}) + (TP_{\text{venous}} + FN_{\text{venous}})}$$

Tabla 4.10: Resultados de OACC para los algoritmos SVM, RF y CNN en el Experimento A por paciente. Se incluyen la media y la DE.

Paciente	SVM	RF	CNN
ID18	0,668	0,953	0,916
ID25	0,731	0,968	0,843
ID29	0,947	0,988	0,980
ID30	0,787	0,984	0,956
ID33	0,855	0,960	0,968
ID34	0,831	0,987	0,981
ID35	0,961	0,998	0,996
ID38	0,644	0,949	0,910
ID47C1	0,710	0,927	0,788
ID47C2	0,631	0,885	0,812
ID50	0,562	0,876	0,898
ID51	0,573	0,959	0,961
ID56	0,797	0,964	0,938
Media	0,746	0,954	0,919
DE	0,125	0,035	0,062

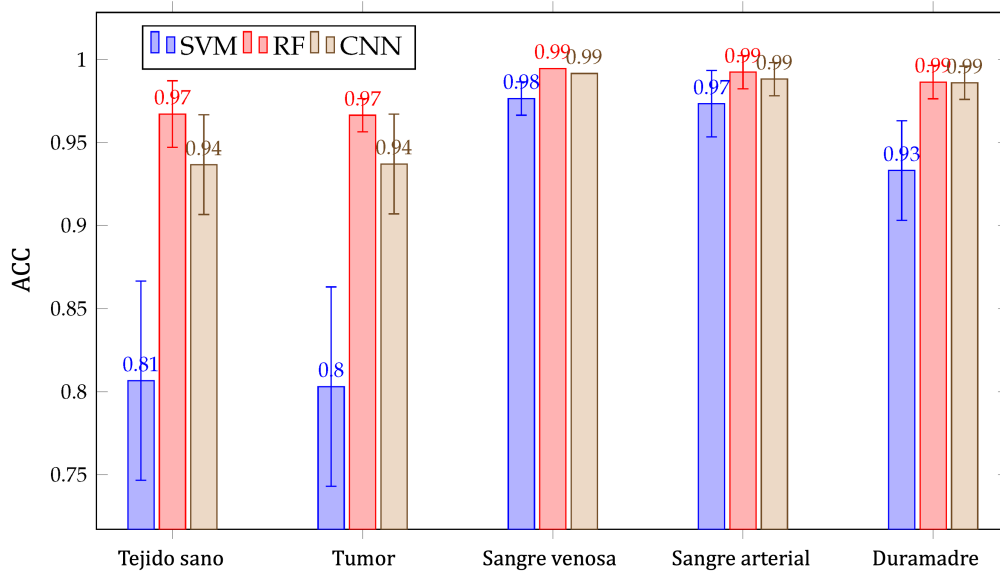


Figura 4.13: ACC medio para todas las predicciones de pacientes y todos los tejidos para los algoritmos SVM, RF y CNN obtenidos en el experimento A.

En la Figura 4.14 se presenta la Sens media de los tres algoritmos. Respecto a los resultados de la Sens en clasificación de los tejidos, los IC muestran un solapamiento considerable en todas las clases, lo que impide concluir de manera firme cuál de los modelos alcanza la mejor Sens global. No obstante, se aprecia que SVM obtiene sistemáticamente peores resultados que RF en las clases de tejido sano, tumor, sangre arterial, sangre venosa y duramadre. Además, la Figura 4.14 muestra que RF alcanza una Sens media superior al 97 % en la clase de sangre, mientras que SVM y CNN presentan valores similares, en torno al 87 %, aunque con diferencias relevantes en la dispersión. En particular, SVM exhibe un IC más amplio que CNN y RF, lo que refleja una mayor variabilidad interpaciente en sus resultados de Sens.

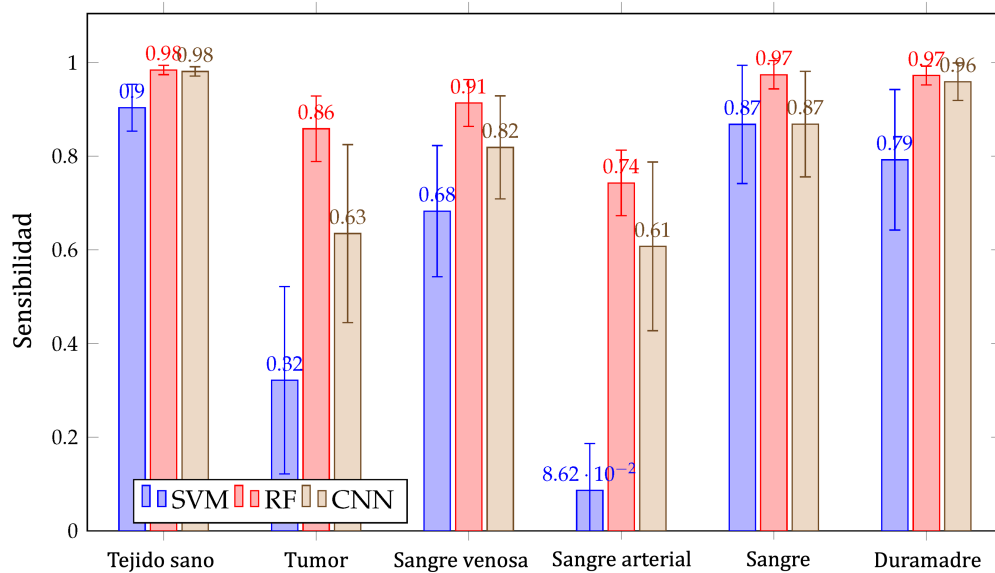


Figura 4.14: Sens medio para todas las predicciones de pacientes y todos los tejidos para los algoritmos SVM, RF y CNN obtenidos en el experimento A.

Por otro lado, la Figura 4.15 muestra los resultados de Spec. Los tres algoritmos alcanzan valores medios superiores al 90 % en las clases de tumor, sangre venosa, sangre arterial y duramadre. Sin embargo, en la clase tejido sano, SVM obtiene valores de Spec inferiores a los de RF, mientras que tanto CNN como RF presentan una variabilidad notable entre pacientes. Además, SVM vuelve a situarse por debajo de RF y CNN en la clase tumor, confirmando así su menor consistencia global frente a los otros dos modelos.

La Figura 4.16 presenta los mapas de clasificación obtenidos con RF, SVM y CNN durante el Experimento A para los pacientes ID33 e ID51, junto con sus correspondientes mapas RGB y GT sintéticos como referencia. En las representaciones, el verde corresponde al tejido sano, el rojo al tumor, el azul a la sangre venosa, el cian a la sangre arterial y el rosa a la duramadre.

En el paciente ID33, los tres algoritmos logran clasificar adecuadamente las regiones de sangre venosa y duramadre. Sin embargo, todos muestran dificultades en la distinción entre tejido sano y tumoral. Entre ellos, CNN reduce de manera más notable la cantidad de píxeles tumorales clasificados erróneamente.

En el paciente ID51, tanto RF como CNN logran identificar correctamente el tejido tumoral,

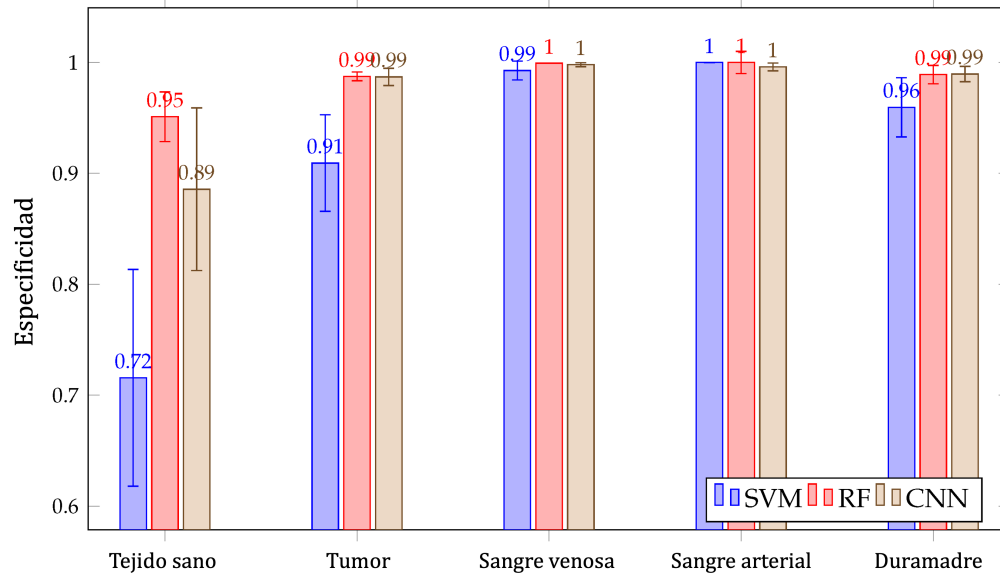


Figura 4.15: Spec medio para todas las predicciones de pacientes y todos los tejidos para los algoritmos SVM, RF y CNN obtenidos en el experimento A.

mientras que SVM clasifica la mayor parte de estas regiones como tejido sano.

Experimento B

La Figura 4.17 presenta la ACC media alcanzada por cada clase y algoritmo (SVM, RF y CNN). En términos generales, los tres métodos muestran un rendimiento comparable, con valores que en la mayoría de los casos superan el 60 %. Aunque SVM obtiene la media de ACC más alta, las diferencias entre algoritmos no resultan estadísticamente concluyentes debido a los amplios IC, que además se solapan.

En términos generales, la Tabla 4.11 muestra que SVM alcanza el mejor rendimiento en 7 de las 13 predicciones, seguido de CNN en 4 casos y RF en 2. Sin embargo, ninguno de los algoritmos supera un valor de OACC del 80 %, lo que indica un margen relevante de mejora.

En relación con los valores medios de Sens mostrados en la Figura 4.18, se observa que ninguno de los algoritmos logró discriminar adecuadamente las clases de tejido tumoral ni arterial. En el caso de la clase de sangre arterial, SVM no logró una clasificación adecuada en 10 de las 13 predicciones, alcanzando valores de Sens del 0 %.

Para los pacientes ID33 e ID51, los mapas de clasificación obtenidos mediante RF, SVM y CNN se muestran en la Figura 4.19. En ambos pacientes, los tres algoritmos logran una adecuada clasificación de la duramadre. En contraste, la clasificación del tejido tumoral resulta problemática para todos los métodos. En el paciente ID33, el mapa de clasificación de CNN ilustra cómo este algoritmo tiende a confundir el tejido sano con el tumoral, etiquetando gran parte de la superficie cerebral como tumor. Aunque RF y SVM también muestran errores de confusión entre tumor, tejido sano y sangre venosa, SVM parece incurrir en un menor número de falsos positivos en la clase de sangre venosa respecto a RF. Estos hallazgos coinciden con los bajos valores de Sens tumoral reportados en la Figura 4.18.

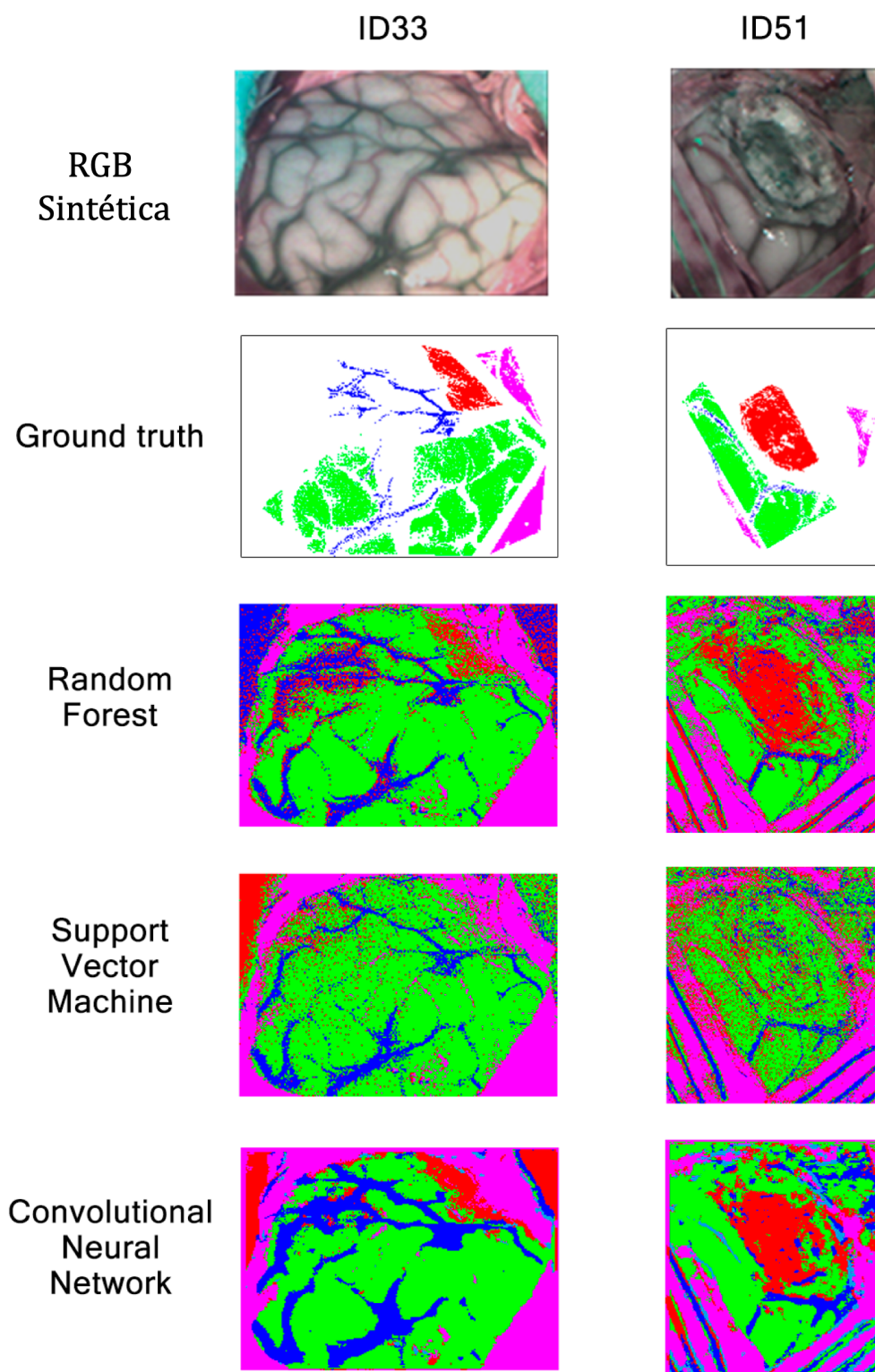


Figura 4.16: RGB sintéticos, *ground truths* y resultados de mapas de clasificación de los pacientes ID33 e ID51 obtenidos en el experimento A tras ser clasificados por RF, SVM y CNN.

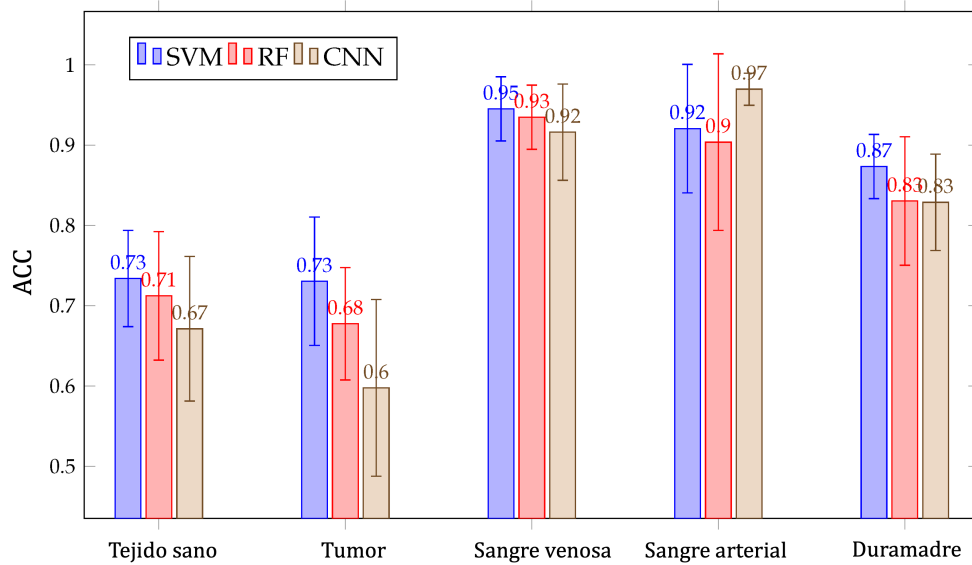


Figura 4.17: ACC medio para todas las predicciones de pacientes y todos los tejidos para los algoritmos SVM, RF y CNN obtenidos en el Experimento B.

Tabla 4.11: Resultados de OACC para los algoritmos SVM, RF y CNN tras clasificar cada imagen de paciente en el Experimento B.

Paciente	SVM	RF	CNN
ID18	0,412	0,287	0,425
ID25	0,680	0,549	0,593
ID29	0,786	0,685	0,835
ID30	0,683	0,635	0,385
ID33	0,666	0,526	0,239
ID34	0,798	0,466	0,615
ID35	0,154	0,113	0,256
ID38	0,552	0,584	0,244
ID47C1	0,701	0,669	0,496
ID47C2	0,623	0,609	0,536
ID50	0,529	0,457	0,474
ID51	0,527	0,560	0,541
ID56	0,705	0,739	0,749
Media	0,601	0,529	0,491
DE	0,166	0,158	0,171

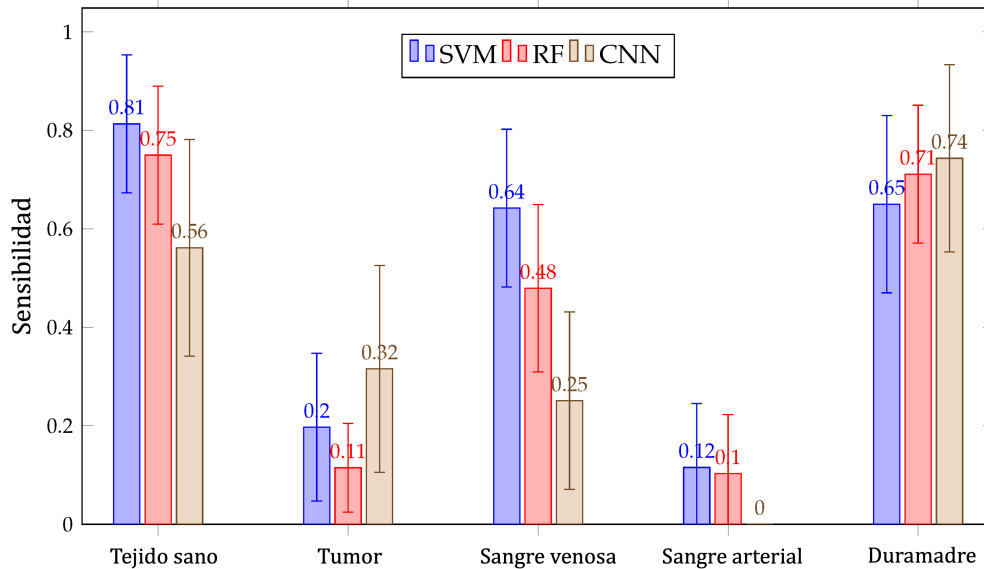


Figura 4.18: Sens medio para todas las predicciones de pacientes y todos los tejidos para los algoritmos SVM, RF y CNN obtenidos en el experimento B.

4.3.2 Segmentación automática en Hemorragia Subaracnoidea Aneurismática

En esta sección se presentan los resultados de los modelos de segmentación automática de hemorragia desarrollados desde dos perspectivas complementarias. En primer lugar, se ofrece una evaluación cuantitativa del rendimiento mediante métricas estándar de segmentación. En segundo lugar, se muestra una evaluación cualitativa basada en ejemplos representativos de casos favorables y desfavorables, así como los resultados de la evaluación cualitativa llevada a cabo por neurocirujanos. Como se ha mencionado en la Sección 3.4.2, en el **Experimento 1** se emplean los conjuntos H12O y HemSeg200 ($n = 445$), con 356 volúmenes destinados a entrenamiento/validación y 89 a test. En el **Experimento 2** se ha ampliado el entrenamiento a todos los conjuntos disponibles ($n = 722$), manteniendo inalterado el conjunto de test definido en el Experimento 1.

Los resultados de estos modelos parten de la base de datos descrita en la Sección 4.1.2.

Resultados cuantitativos

En la Tabla 4.12 se resumen las métricas de rendimiento obtenidas en la segmentación automática, comparadas con la segmentación manual como referencia, para los tres conjuntos de test: H12O, HemSeg200 y el conjunto combinado. Se incluyen tanto los algoritmos reportados en la bibliografía como los desarrollados en este estudio. Las métricas se expresan en términos de mediana e IQR ($Q1-Q3$), lo que permite una representación más robusta frente a valores extremos.

Los resultados muestran que los modelos propuestos en este trabajo consistentemente presentan valores medianos de rendimiento superiores a los algoritmos previos de la literatura de forma

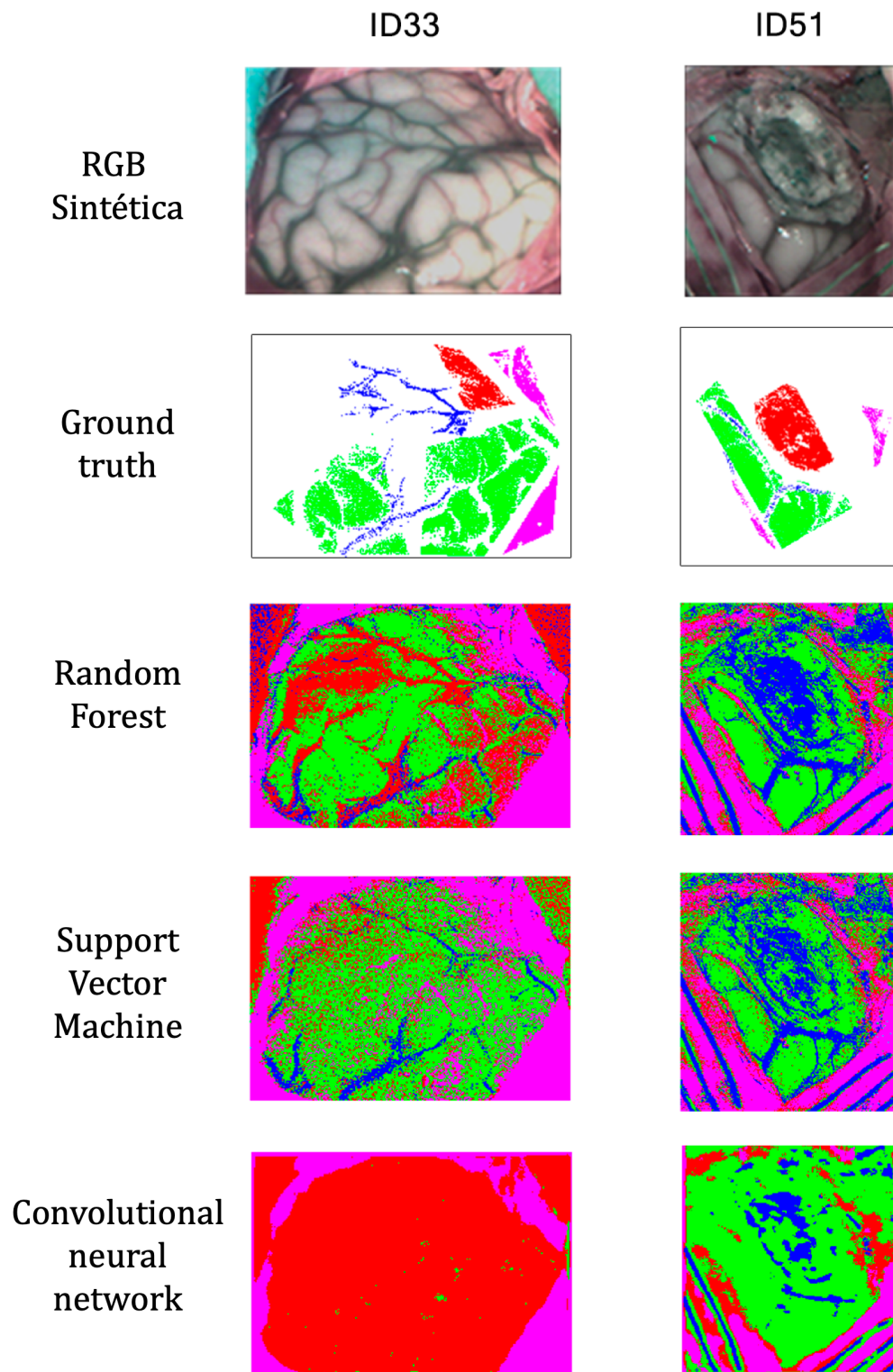


Figura 4.19: RGB sintéticos, verdades fundamentales y mapas de clasificación de los pacientes ID33 e ID51 obtenidos en el experimento B tras ser clasificados por RF, SVM y CNN.

consistente, tanto en el conjunto de test global como individualmente por cada conjunto de datos incluido en el test. En particular, el Experimento 2 alcanza el mejor rendimiento global en la mayoría de las métricas, con la excepción de la distancia de Hausdorff y la distancia media de la superficie, donde el Experimento 1 obtiene valores más favorables. Asimismo, se observa que las métricas de rendimiento son de forma sistemática más elevadas en el conjunto HemSeg200 que en el H12O, lo que sugiere que las características de este último plantean un mayor desafío para los modelos de segmentación.

Además, se ha realizado un análisis comparativo mediante IC95 % de la mediana del *Dice Score*. Bajo este enfoque, se considera que un modelo superaba a otro cuando sus IC95 % no se solapaban, cuyos resultados se representan en la Tabla 4.13. En particular, los experimentos realizados en este trabajo se sitúan dentro del grupo de mayor desempeño: presentando solapamiento de IC entre ellos (Experimento 1 y 2) y también con *ICH 2023* y *Viola-UNet v2*, lo que sugiere resultados equivalentes entre estas aproximaciones de alto rendimiento. Además, muestra IC *no* solapados frente a *Blast-CT* y *SMCCH*, indicando una ventaja clara.

De forma complementaria, *Blast-CT* queda claramente relegado, pues en todas sus comparaciones los IC no se solapan, indicando un rendimiento consistentemente inferior respecto al resto de métodos. Finalmente, el par *ICH 2023–Viola-UNet v2* presentó IC solapados, por lo que no se observan diferencias relevantes entre ambos. En conjunto, estos resultados muestran un conjunto de modelos con un Dice Score equivalente (*Experimento 1*, *Experimento 2*, *ICH 2023*, *Viola-UNet v2*) y un subconjunto claramente inferior (*Blast-CT* y, en menor medida, *SMCCH* frente a los anteriores).

Por otro lado, la Figura 4.20 presenta los gráficos de Bland–Altman que evalúan la concordancia entre los volúmenes de hemorragia (mL) estimados por los modelos del Experimento 1 y del Experimento 2, en comparación con las segmentaciones manuales de referencia. En estos gráficos se observa que el Experimento 2 ofrece una estimación volumétrica más estable, caracterizada por una menor dispersión de los puntos y una reducción en el número de valores atípicos respecto al Experimento 1. Este comportamiento sugiere una mayor precisión y un menor sesgo en las predicciones del Experimento 2, reforzando su utilidad como aproximación más fiable para la segmentación automática de la hemorragia.

Por último, en la Figura 4.21 se muestran los resultados obtenidos con el Experimento 2 del *Dice Score* en función del volumen de la hemorragia del conjunto de test para tres escenarios: el conjunto completo (HemSeg + H12O), únicamente el conjunto HemSeg y únicamente el conjunto H12O.

En el análisis global (panel izquierdo), que combina los casos de HemSeg y H12O, se observa una alta variabilidad en los volúmenes pequeños (<20 ml). En este rango existen casos con un desempeño excelente (Dice > 0,9), pero también otros con valores bajos cercanos a 0,5, lo que refleja inestabilidad del modelo en hemorragias de pequeño tamaño. A medida que el volumen aumenta por encima de los 40 ml, los valores de Dice tienden a concentrarse en un rango intermedio (0,7–0,8), lo que sugiere una mayor estabilidad en volúmenes grandes.

El conjunto HemSeg (panel central) muestra un comportamiento notablemente superior. La mayoría de los casos se encuentran por encima de 0,8, independientemente de su volumen, con numerosos casos que superan el 0,9 y solo un caso aislado con un volumen < 20 ml con

Tabla 4.12: Métricas de rendimiento en el conjunto de test (N=89) utilizando cinco algoritmos de la bibliografía y dos desarrollados en este trabajo, con respecto a la segmentación manual como referencia. Las métricas se presentan como la mediana y su IQR (Q1-Q3). En negrita se subrayan los valores medianos mas altos por métrica respecto a los modelos evaluados.

Métrica	Conjunto	Blast-CT	SMCCH	ISBI 2024	ICH 2023	Viola-UNet v2	Exp. 1	Exp. 2
Dice score	H12O	0,26 (0,14-0,38)	0,49 (0,32-0,58)	0,44 (0,18-0,60)	0,58 (0,38-0,70)	0,56 (0,46-0,65)	0,67 (0,59-0,73)	0,69 (0,57-0,75)
	HemSeg200	0,38 (0,26-0,55)	0,40 (0,10-0,65)	0,84 (0,76-0,93)	0,83 (0,71-0,91)	0,84 (0,76-0,88)	0,88 (0,76-0,90)	0,90 (0,81-0,92)
	Ambos	0,34 (0,22-0,51)	0,47 (0,17-0,61)	0,75 (0,51-0,90)	0,70 (0,52-0,86)	0,72 (0,57-0,85)	0,75 (0,67-0,88)	0,81 (0,72-0,90)
Sens	H12O	0,18 (0,04-0,35)	0,67 (0,52-0,80)	0,35 (0,10-0,55)	0,50 (0,34-0,71)	0,70 (0,57-0,81)	0,67 (0,54-0,75)	0,71 (0,59-0,82)
	HemSeg200	0,68 (0,40-0,86)	0,64 (0,31-0,80)	0,85 (0,73-0,94)	0,85 (0,73-0,93)	0,84 (0,76-0,88)	0,89 (0,81-0,95)	0,90 (0,82-0,94)
	Ambos	0,39 (0,08-0,71)	0,65 (0,45-0,81)	0,76 (0,39-0,91)	0,71 (0,46-0,88)	0,85 (0,69-0,92)	0,79 (0,64-0,89)	0,86 (0,72-0,92)
ICC3	H12O	0,26 (0,13-0,48)	0,79 (0,56-0,85)	0,71 (0,32-0,85)	0,80 (0,55-0,87)	0,82 (0,64-0,87)	0,83 (0,71-0,90)	0,86 (0,74-0,91)
	HemSeg200	0,41 (0,24-0,52)	0,80 (0,34-0,88)	0,94 (0,88-0,96)	0,93 (0,88-0,97)	0,92 (0,88-0,95)	0,94 (0,89-0,97)	0,95 (0,93-0,97)
	Ambos	0,36 (0,17-0,51)	0,79 (0,47-0,87)	0,89 (0,72-0,96)	0,88 (0,72-0,96)	0,88 (0,82-0,92)	0,90 (0,82-0,94)	0,92 (0,88-0,96)
Hausdorff (mm)	H12O	90,0 (70-114,3)	94,18 (67,04-102,85)	101,0 (70,55-130,95)	96,34 (59,47-119,86)	68,07 (51,5-88,32)	59,27 (44,52-78,23)	68,79 (51,39-93,3)
	HemSeg200	139,0 (111-163,6)	138,3 (85,26-168,40)	27,33 (6,59-89,55)	19,8 (8-96,65)	17,99 (6,99-94,8)	14,01 (6,87-61,8)	19,39 (6,59-90,6)
	Ambos	114,7 (83,7-148,2)	99,96 (79,35-148,7)	67,68 (13,6-124,9)	71,31 (21,21-115,6)	58,61 (12,25-90,26)	47,18 (13,78-72,01)	60,44 (13,24-91,8)
Distancia media Superficie (mm)	H12O	9,65 (5,46-16,24)	6,48 (3,93-9,67)	12,79 (4,28-31,75)	6,65 (3,16-13,7)	3,75 (2,28-7,36)	2,05 (1,6-5,02)	2,61 (1,8-5,34)
	HemSeg200	15,84 (6,91-29,89)	27,65 (11,49-73,2)	0,62 (0,25-3,48)	0,55 (0,3-3,08)	0,64 (0,41-1,93)	0,50 (0,3-0,88)	0,46 (0,34-2,84)
	Ambos	12,11 (5,73-21,49)	10,73 (4,69-35,4)	2,93 (0,34-13,54)	3,16 (0,49-11,6)	2,19 (0,57-5,26)	1,50 (0,48-4,21)	1,63 (0,42-3,79)
Volumen (mL)	H12O	9,38 (3,44-17,31)	68,52 (22,72-157,0)	7,88 (4,62-12,27)	3,98 (2,56-9,0)	1,95 (0,65-3,37)	2,52 (1,08-6,43)	1,92 (0,92-5,07)
	HemSeg200	12,22 (4,48-27,71)	4,62 (1,56-24,61)	1,38 (0,33-3,56)	1,72 (0,61-4,31)	6,51 (2,58-12,28)	1,72 (0,59-3,58)	1,24 (0,48-3,21)
	Ambos	10,68 (3,97-20,14)	23,58 (3,80-59,92)	3,06 (0,72-8,17)	2,97 (0,81-7,0)	3,17 (1,35-7,57)	1,91 (0,63-4,57)	1,40 (0,50-4,02)

Tabla 4.13: Matriz de comparación por pares según solapamiento de IC95 % de la mediana. “Equivalente” indica IC solapados (no diferencia concluyente); “Diferente” indica IC no solapados (diferencia relevante).

	Experimento 2	ISBI 2024	Blast-CT	Experimento 1	SMCCH	ICH 2023	Viola-UNet v2
Experimento 2	–	Equivalente	Diferente	Equivalente	Diferente	Equivalente	Equivalente
ISBI 2024	Equivalente	–	Diferente	Equivalente	Diferente	Equivalente	Equivalente
Blast-CT	Diferente	Diferente	–	Diferente	Diferente	Diferente	Diferente
Experimento 1	Equivalente	Equivalente	Diferente	–	Diferente	Equivalente	Equivalente
SMCCH	Diferente	Diferente	Diferente	Diferente	–	Diferente	Diferente
ICH 2023	Equivalente	Equivalente	Diferente	Equivalente	Diferente	–	Equivalente
Viola-UNet v2	Equivalente	Equivalente	Diferente	Equivalente	Diferente	Equivalente	–



Figura 4.20: Gráfico de Bland–Altman comparando los volúmenes hemorrágicos obtenidos mediante segmentación automática frente a la segmentación manual en el conjunto de test. A la izquierda se muestran los resultados del Experimento 1 y a la derecha los del Experimento 2.

un 0,6 de Dice Score. Por el contrario, en el conjunto H12O (panel derecho) el rendimiento es más variable. Parte de los casos se sitúan en un rango de Dice entre 0,5 y 0,7, aunque otros ejemplos que superan 0,8. Este comportamiento sobre todo se da en los volúmenes menores de 20 mL, e influyen a los resultados globales del primer panel.

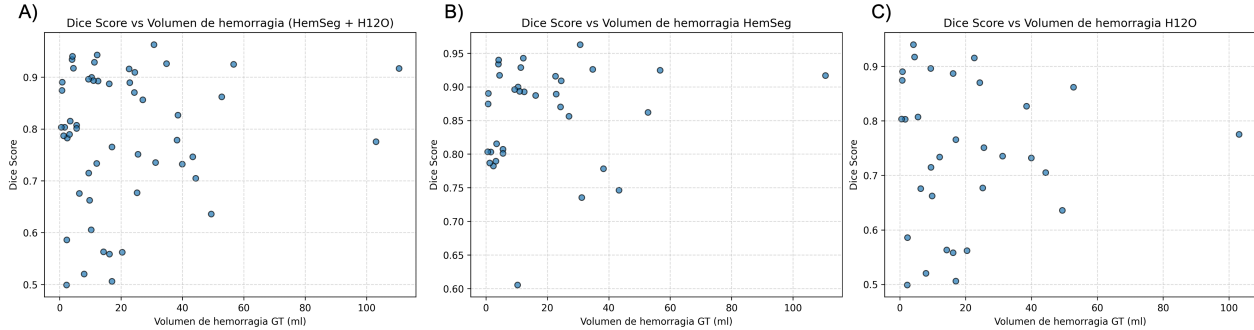


Figura 4.21: Diagrama de dispersión del Dice Score en función del volumen de la hemorragia en el conjunto de test. A) Conjunto completo (HemSeg+ H12O), B) Conjunto HemSeg, C) Conjunto H12O.

Resultados cualitativos

La Figura 4.22 ofrece una visión general del rendimiento de la segmentación usando diferentes modelos en los conjuntos de test de H12O y HemSeg200, mostrando ejemplos de casos favorables y desfavorables.

Por un lado, La Figura 4.22.A muestra casos representativos con un alto rendimiento (Dice > 85 %), en los que nuestros modelos muestran resultados consistentes tanto en el Experimento 1 como en el 2. En el conjunto de datos H12O, los modelos desarrollados en este trabajo superan a los métodos de la literatura: Blast-CT, SMCCH, ICH 2023, ISBI 2024, Viola-UNET v2 (García-García et al., 2024; Gong et al., 2024; Q. Liu et al., 2023; Monteiro et al., 2020; Rusche et al., 2023), especialmente en las regiones SAH, donde otros suelen producir falsos negativos. En HemSeg200, el rendimiento es comparable al de algoritmos de última generación como Viola-UNet v2, ICH 2023 e ISBI 2024, aunque SMCCH a menudo produce falsos positivos cerca del cráneo y Blast-CT tiende a sobresegmentar, generando resultados más grandes y con más ruido, comportamiento que coincide con las métricas cuantitativas obtenidas.

Por otro lado, La Figura 4.22.B muestra casos de bajo rendimiento (Dice < 67 %), que suelen implicar pequeños falsos positivos o falsos negativos en zonas de bajo contraste. A pesar de estos retos, los resultados de nuestros modelos parecen más precisos desde el punto de vista anatómico que los de otros enfoques, especialmente en casos de HSA, en los que algunos algoritmos no logran identificar ninguna región hemorrágica.

Por otro lado, la Figura 4.23 muestra las puntuaciones cualitativas medias asignadas por siete neurocirujanos a 15 casos no vistos previamente, es decir, que no han sido usados ni para entrenamiento ni para test de los modelos, ya que no cuentan con segmentación de hemorrágica.

Tal y como se explicó en la Sección 3.4.4, la evaluación se realizó en un estudio doble ciego en torno a cinco criterios: *cobertura anatómica*, *sobresegmentación*, *infrasegmentación*,

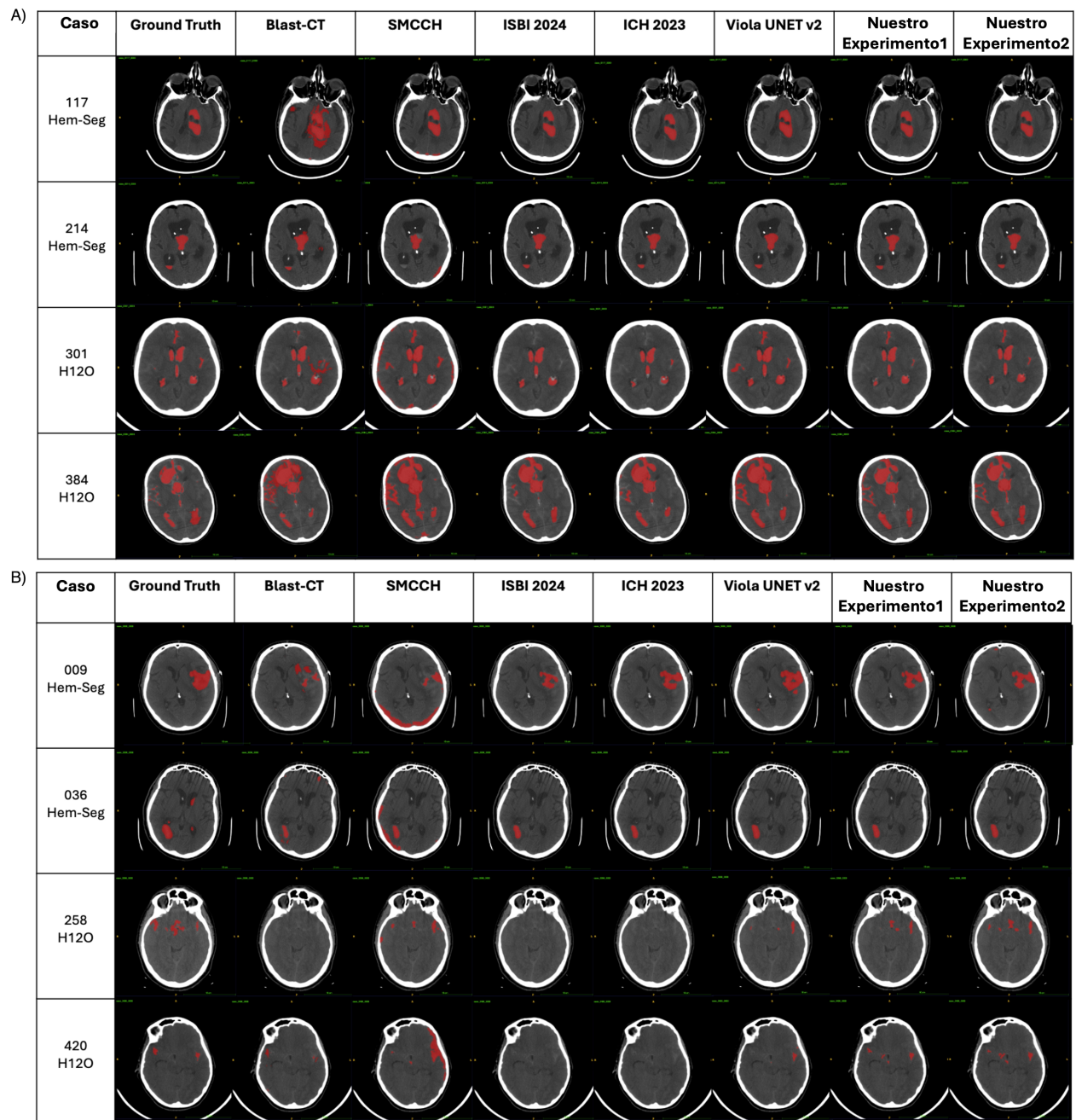


Figura 4.22: Segmentaciones obtenidas a partir del *ground truth*, modelos automáticos de la literatura y los modelos desarrollados en este trabajo: (A) casos favorables; (B) casos desfavorables.

consistencia entre cortes y precisión global, comparando los modelos desarrollados en los Experimentos 1 y 2 con cinco algoritmos de segmentación basados en IA. Las puntuaciones oscilaron entre 1 (deficiente) y 5 (excelente), representándose la media de las puntuaciones con sus correspondientes IC del 95 %.

- **Cobertura anatómica:** el modelo del Experimento 2 alcanza la puntuación media más alta (3,44), seguido de SMCCH (3,33), mientras que Blast-CT obtiene la más baja (3,04). Los únicos IC que no se solapan son respecto a Blast-CT y el Experimento 2, indicando mejor rendimiento de este último.
- **Sobresgmentación:** Viola-UNet v2 muestra el mejor desempeño (3,89), lo que sugiere una sobresgmentación mínima; en contraste, SMCCH obtiene la menor puntuación (3,30). Sin embargo, todos los IC se encuentran solapados.
- **Infrasegmentación:** SMCCH presenta el mejor resultado (2,94), mientras que Blast-CT evidencia la mayor infrasegmentación (2,42). No obstante, los únicos IC que no se solapan son respecto a Blast-CT y el Experimento 2, indicando mejor rendimiento estadístico de este último frente a Blast-CT.
- **Consistencia entre cortes:** el modelo del Experimento 1 y Viola-UNet v2 alcanzan la puntuación media más alta (3,40), mientras que Blast-CT vuelve a situarse en último lugar (3,17). Todos los IC se encuentran solapados.
- **Evaluación global:** el Experimento 2 obtiene la mayor puntuación media (2,99), reflejando un rendimiento equilibrado; Blast-CT obtiene la más baja (2,59). En este caso, los IC muestran que los modelos Viola-UNet v2, Experimento 1 y 2 muestran mejor resultado estadístico que Blast-CT.

Las clasificaciones finales de los modelos se obtienen aplicando el método de recuento de Borda, que integra las valoraciones de los expertos en los cinco criterios (véase Tabla 4.14). El Experimento 2 se posiciona en primer lugar, seguido de Viola-UNet v2, que destaca particularmente en sobresgmentación y precisión global. El Experimento 1 ocupa la tercera posición, evidenciando un rendimiento sólido y estable.

Los modelos restantes muestran un rendimiento global inferior, especialmente en los criterios más relevantes clínicamente, como la evaluación global. SMCCH e ISBI 2024 presentan fortalezas puntuales en métricas específicas, pero no logran superar de manera consistente a los tres modelos principales.

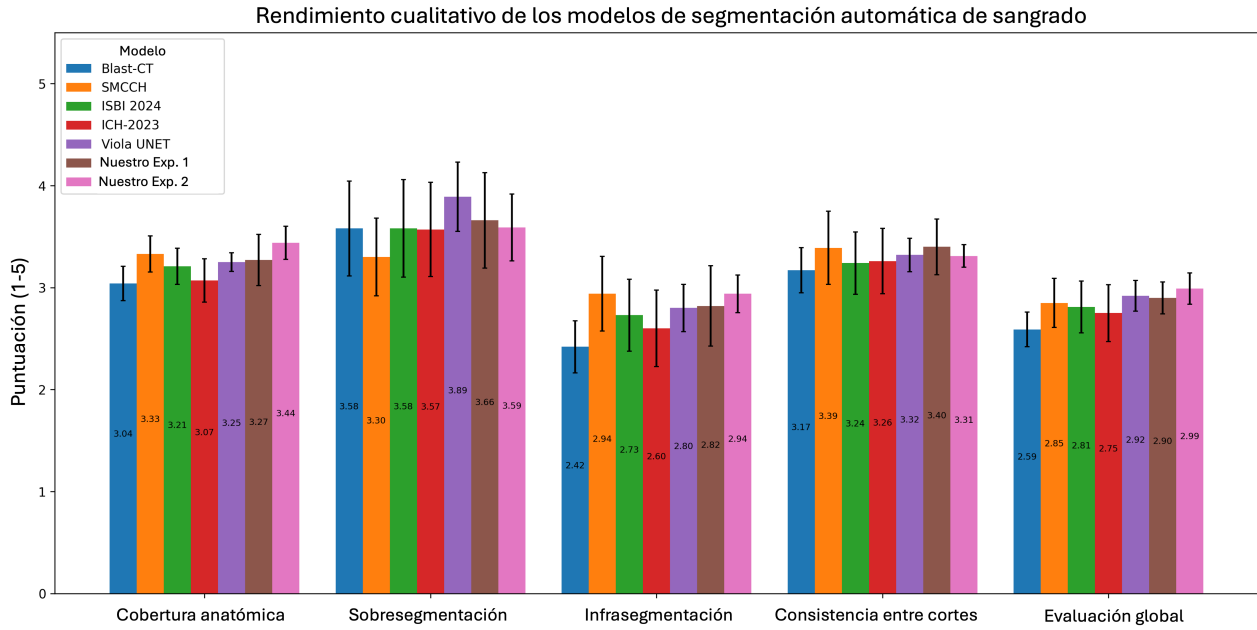


Figura 4.23: Media e IC95 % de las puntuaciones de los neurocirujanos según cinco criterios: cobertura anatómica, sobresegmentación, infrasegmentación, consistencia inter-corte y valoración global.

Tabla 4.14: Ranking de los modelos de segmentación de hemorragia en función de la suma de puntuaciones en los cinco criterios de evaluación (la posición en el ranking se indica entre paréntesis).

Modelo	Cobertura	Sobresseg.	Subseg.	Consistencia	Global
Experimento 2	378 (1)	404 (4)	343 (1)	369 (1)	342 (1)
Viola-UNet v2	351 (2)	413 (1)	310 (2)	361 (2)	324 (2)
Experimento 1	347 (3)	406 (3)	306 (3)	360 (3)	317 (3)
SMCCH	334 (4)	327 (7)	291 (4)	343 (4)	284 (4)
ISBI 2024	310 (5)	370 (6)	261 (5)	322 (7)	271 (5)
ICH 2023	305 (6)	401 (5)	258 (6)	334 (5)	269 (6)
Blast-CT	296 (7)	408 (2)	234 (7)	325 (6)	253 (7)

Eficiencia temporal: segmentación manual frente a automática

El tiempo medio ($\pm DE$) necesario para la segmentación manual es de aproximadamente 15 ± 4 minutos por caso, mientras que el método automático completa la misma tarea en una media de $23,4 \pm 2,1$ segundos.

Los tiempos de segmentación automática corresponden a la implementación con nnU-Net ejecutada en CPU (Intel Xeon Gold 6338, 2.0 GHz, 64 hilos; 256 GB RAM). No se empleó aceleración mediante GPU, por lo que los tiempos reportados constituyen una estimación conservadora.

La segmentación manual presenta una notable variabilidad en los tiempos de ejecución, con una DE que alcanza el 25 % del valor medio. Esta dispersión puede atribuirse tanto a la heterogeneidad anatómica entre pacientes como a la variación en el número de cortes afectados por la hemorragia en cada caso, factores que incrementan la carga de trabajo del observador. En contraste, la segmentación automática muestra una menor variabilidad, manteniendo tiempos de procesamiento constantes con independencia de la complejidad del caso o del volumen hemorrágico, aunque sí guardan relación con el número total de cortes a procesar en cada estudio. En la Tabla 4.16 se puede observar una comparativa del tiempo de segmentación manual y automático. En términos relativos, el método automático supone una reducción de tiempo cercana al **97,4 % por caso**, calculada como $\frac{900-23,4}{900}$.

Tabla 4.16: Comparación de la eficiencia temporal entre segmentación manual y automática

Método	Tiempo medio	Desv. estándar	Rango
Manual	15,2 min	3,8 min	9,5 – 21,4 min
Automática	23,4 s	2,1 s	19,6 – 27,5 s

El conjunto de casos que se segmentaron manualmente para este estudio ($n = 223$) representa una carga temporal aproximada de **55 h 45 min** ($223 \text{ casos} \times 15 \text{ min/caso}$). En contraste, la ejecución automática del algoritmo requiere únicamente alrededor de **1 h 27 min** ($223 \text{ casos} \times 23,4 \text{ s/caso}$).

Al ampliar la perspectiva a la *cohorte completa* ($n = 539$), el esfuerzo manual ascendería a **136 h 33 min** ($539 \times 15,2 \text{ min/caso}$), mientras que el método automático reduciría esta carga a **3 h 30 min** ($539 \times 23,4 \text{ s/caso}$).

Análisis comparativo con escalas clínicas convencionales

El objetivo de este análisis es comparar el valor clínico de tres aproximaciones diferentes: (i) los volúmenes hemorrágicos obtenidos mediante segmentación manual, (ii) los volúmenes derivados de la segmentación automática, obtenidas del modelo del Experimento 2, y (iii) la clasificación clínica mediante la escala de Fisher modificado, ampliamente utilizada en la práctica clínica. Para ello se han realizado análisis de regresión logística univariable y multivariable en la cohorte H12O ($N = 223$), para predecir el estado funcional del paciente a los seis meses, dicotomizado en desfavorable (GOS 1–3, $n = 119$) y favorable (GOS 4–5, $n = 104$). Los resultados detallados se muestran en la Tabla 4.17.

Tabla 4.17: Resultados de la regresión logística uni- y multivariable para la predicción de resultado funcional favorable (N=223).

	Multivariable			Univariable		
	Vol. manual	Vol. auto.	Fisher mod.	Vol. manual	Vol. auto.	Fisher mod.
AUC (IC95 %)	0,812 (0,749–0,861)	0,810 (0,748–0,862)	0,733 (0,650–0,816)	0,723 (0,637–0,802)	0,719 (0,635–0,802)	0,593 (0,516–0,668)
F1-Score (IC95 %)	0,759 (0,697–0,813)	0,778 (0,717–0,832)	0,705 (0,627–0,780)	0,735 (0,659–0,804)	0,771 (0,699–0,833)	0,698 (0,619–0,766)
Sens (IC95 %)	0,773 (0,698–0,846)	0,798 (0,726–0,866)	0,724 (0,617–0,825)	0,820 (0,737–0,902)	0,856 (0,772–0,927)	0,728 (0,641–0,808)
Pseudo R ²	0,221	0,224	0,122	0,097	0,098	0,031
Log-Likelihood	–116,94	–116,62	–87,62	–90,11	–90,01	–96,74
Volumen (coef.)	–0,0248 (p=0,003)	–0,0325 (p=0,003)	–0,199 (p=0,571)	–0,037 (p<0,001)	–0,044 (p<0,001)	–0,744 (p=0,017)
WFNS (coef.)	–0,5189 (p<0,001)	–0,4861 (p<0,001)	–0,467 (p<0,001)	—	—	—
Edad (coef.)	–0,0086 (p=0,383)	–0,0104 (p=0,285)	–0,022 (p=0,089)	—	—	—
Var. explicada (%)	Vol.: 15,5, WFNS: 37,6, Edad: 1.1	Vol.: 16,3, WFNS: 30,2, Edad: 1.7	Fisher mod.: 4,2, 71,3, 1.5	—	—	—

En los modelos univariables, tanto los volúmenes manuales como los automáticos se comportan como predictores significativos del desenlace funcional ($p < 0,001$), alcanzando AUC medias de 0,723 y 0,719, respectivamente. Por comparación, la escala de Fisher modificado también muestra asociación significativa ($p = 0,017$), aunque con una capacidad discriminatoria claramente inferior ($AUC = 0,593$).

En los análisis multivariables, ajustados por edad y puntuación WFNS al ingreso, los volúmenes hemorrágicos (manuales y automáticos) mantienen una asociación significativa con el pronóstico funcional (manual: coeficiente = $-0,0248$, $p = 0,003$; automático: coeficiente = $-0,0325$, $p = 0,003$). La escala de Fisher modificado, en cambio, pierde significación estadística ($p = 0,571$). En este escenario, el modelo basado en volumen automático alcanza un AUC medio de 0,810, prácticamente idéntico al obtenido con el volumen manual ($AUC = 0,812$). Asimismo, las métricas de Sens y F1-Score medias son muy similares entre el volumen manual y hemorrágico (Sens: 0,773-0,798, F1-Score: 0,759-0,778). En contraste, la escala de Fisher modificado muestra un rendimiento medio más bajo ($AUC = 0,733$; $F1 = 0,705$; $Sens = 0,724$).

Finalmente, el análisis de la varianza explicada confirma que la WFNS constituye el principal determinante del poder predictivo en todos los modelos. No obstante, el volumen automático explica de forma consistente una proporción de varianza superior a la de la escala de Fisher modificado (16,3 % frente a 4,2 %), lo que muestra la utilidad de los métodos volumétricos cuantitativos frente a las escalas cualitativas tradicionales en la predicción de resultado funcional (GOS).

Finalmente, la Figura 4.24 muestra la distribución de los resultado funcional (GOS dicoto-

mizado) en función del volumen de hemorragia segmentado manual y automáticamente. Se observa una tendencia clara: los volúmenes inferiores a 30 cc se asocian predominantemente con buen resultado funcional (>60–80%), mientras que la probabilidad de mal resultado aumenta de forma marcada más allá de este umbral.

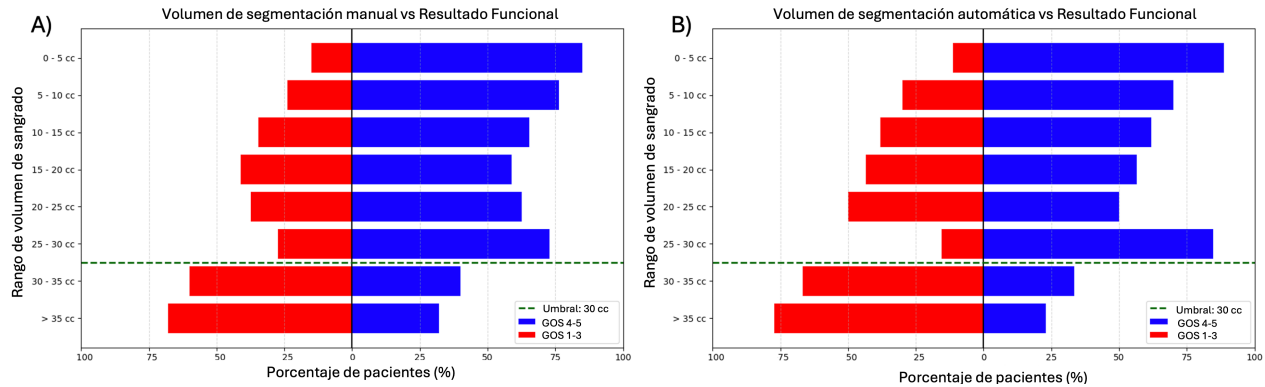


Figura 4.24: Distribución del resultado funcional en función del volumen de hemorragia. A) Volumen obtenido con segmentación manual, B) con segmentación automática

4.3.3 Modelos predictivos radiómicos

Como se ha comentado en la Sección 4.1.3, en estos modelos se incluyen 403 pacientes con HSAa confirmado para el desarrollo de los modelos y 41 en un test externo, evaluando desenlace funcional y mortalidad a 6 meses, así como vasoespasma e hidrocefalia. Para estos dos últimos desenlaces se han excluido los pacientes fallecidos en la primera semana sin haberlos desarrollado. Con esta cohorte se han desarrollado modelos de predicción basados en tres enfoques: únicamente con variables clínicas, únicamente con descriptores radiómicos, y con la combinación de ambas fuentes de información. La extracción radiómica se realiza tanto a partir del parénquima (sustancia gris y blanca) como de la hemorragia. Para esta última, se utilizan dos segmentaciones automáticas: la generada por un modelo de la literatura entrenado para HSAa (SMCCH) y la obtenida con el modelo desarrollado en este trabajo en el **Experimento 2**, que es el mejor valorado por los neurocirujanos y alcanza los valores medianos más altos en las métricas cuantitativas.

La Figura 4.25 presenta los resultados de los modelos de predicción pronóstica basados en radiómica y variables clínicas. En ella se muestra la comparación entre cuatro familias de modelos predictivos: un modelo basado exclusivamente en variables clínicas, uno de radiómica del parénquima cerebral, otro de radiómica centrado en la hemorragia a partir del algoritmo de la literatura SMCCH y, finalmente, nuestro modelo propuesto (Experimento 2). Se evalúan los desenlaces de mortalidad, resultado funcional (GOS), vasoespasma e hidrocefalia, considerando el rendimiento en términos de AUC, BalAcc, Sens y Spec. Se muestran resultados de test del desarrollo de la validación cruzada anidada estratificada en 5 folds con 3 semillas diferentes.

En relación con la **mortalidad**, el modelo clínico alcanza la mayor AUC media (0.84), ligeramente superior a la de los modelos radiómicos de parénquima y del *Experimento 2* (ambos con 0.82), mientras que el modelo **SMCCH** obtiene un rendimiento inferior (0.76). En cuanto a la BalAcc, el modelo clínico y el de parénquima comparten el mejor resultado

(0.75), seguidos de cerca por el *Experimento 2* (0.74). En las métricas de discriminación, el modelo de parénquima muestra la mayor Sens (0.79), mientras que el *Experimento 2* destaca por su Spec (0.79). No obstante, los IC se solapan en la mayoría de los casos, salvo en la Sens, donde el *Experimento 2* se sitúa por debajo del modelo clínico y del radiómico de parénquima.

Para el **resultado funcional (GOS)**, el *Experimento 2* obtiene la mayor AUC media (0.82) y la mayor Spec (0.76). El modelo clínico presenta valores muy próximos en BalAcc (0.75) y una Sens media algo superior (0.76). En cualquier caso, los IC de todos los modelos se solapan ampliamente.

El **vasoespasma** es el desenlace con mayor dificultad predictiva, con AUCs modestas (0.66–0.70) y amplios IC tanto en Sens como Spec. Las diferencias entre modelos son reducidas: la radiómica de parénquima alcanza la Sens media más alta (0.73), mientras que el *Experimento 2* presenta una Spec ligeramente superior (0.62). Nuevamente, todos los IC se encuentran solapados.

En cuanto a la **hidrocefalia**, la radiómica de parénquima muestra un rendimiento consistentemente superior en todas las métricas, destacando por su AUC (0.79), BalAcc (0.67), Sens (0.82) y Spec (0.63). En contraste, el modelo **SMCCH** ofrece los resultados más bajos de media. Sin embargo, también en este desenlace los IC permanecen solapados.

En términos globales, los modelos clínicos y radiómicos muestran rendimientos **similares** en la mayoría de desenlaces evaluados, con IC solapados en la mayoría de los casos. Esto indica que las diferencias entre modelos no son estadísticamente significativas y que tanto los modelos clínicos como los radiómicos ofrecen rendimientos comparables.

La Figura 4.26 extiende la comparativa de los modelos puramente radiómicos con los modelos que combinan las variables radiómicas y clínicas por cada una de las ROIs consideradas en este estudio. De nuevo, se muestran resultados en el test de la validación cruzada estratificada en 5 folds con 3 semillas diferentes en cada uno de los desenlaces.

En **mortalidad**, la combinación clínica+radiómica mejora la AUC media en parénquima (0,82→0,87) y SMCCH (0,76→0,83) aunque solapando los IC, mientras que en Experimento 2 también se elevan las métricas (0,83→0,87) con unos IC más pequeños y no solapados. La BalAcc sigue el mismo patrón, alcanzando hasta 0,79 en Experimento 2 combinado. El aumento se explica fundamentalmente por un incremento en la **Spec**, que sube de 0,71 a 0,76 en parénquima, de 0,68 a 0,76 en SMCCH y de 0,76 a 0,82 en Experimento 2, manteniéndose Sens elevadas (0,71–0,79). Además, en el Experimento 2 se observa una subida significativa (0,68→0,76), sin solapamientos en IC.

En el **GOS funcional**, los modelos combinados muestran también ganancias en las métricas medias. La AUC del parénquima asciende de 0,76 a 0,85, la de SMCCH de 0,75 a 0,79 y la de Experimento 2 de 0,82 a 0,87. Respecto a los IC en AUC los modelos que combinan parénquima+clínica y Experimento2+clínica son mejores que los modelos radiómicos puros de la hemorragia extraída con SMCCH. En BalAcc, los valores alcanzan un 0,79 para los modelos combinados, frente a 0,75 en sus equivalentes puros. El efecto más consistente es el aumento de **Spec**, que en el parénquima pasa de 0,63 a 0,75 y en Experimento 2 de 0,62 a 0,78, en ambos reduciendo sus IC en los modelos combinados. Las Sens de todos los modelos

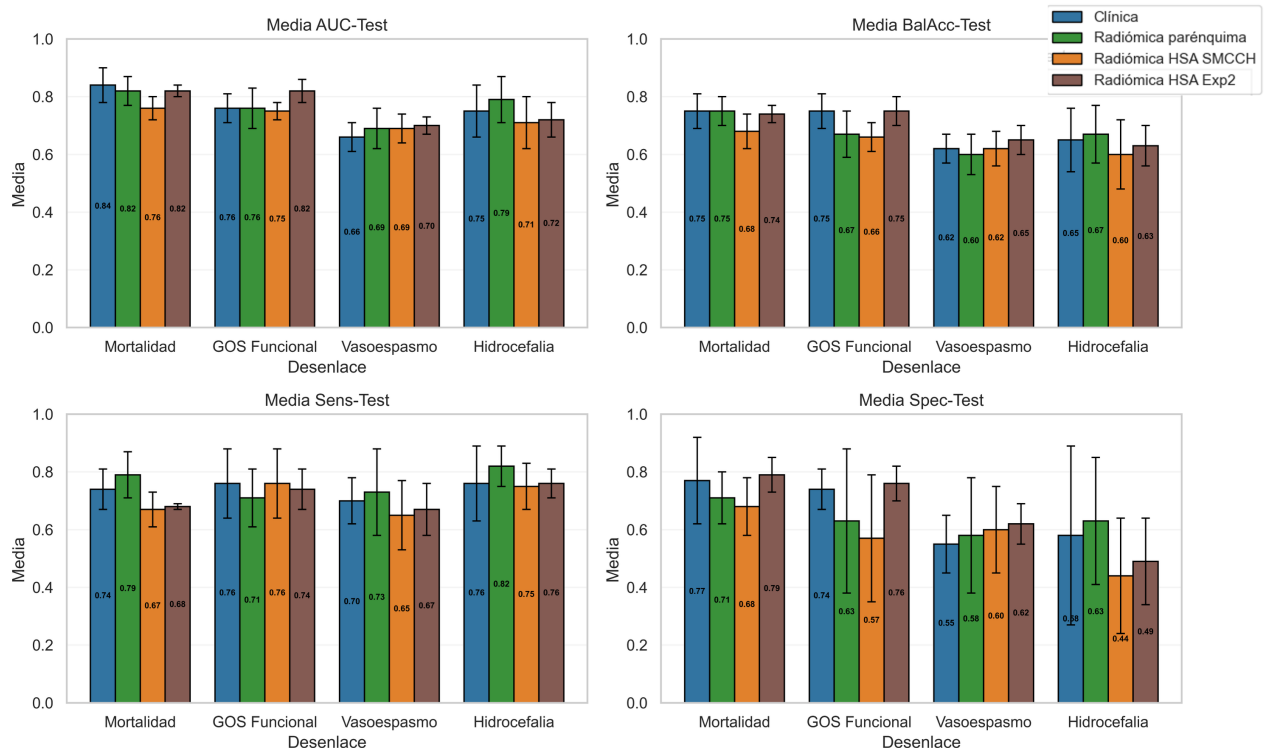


Figura 4.25: Comparación del rendimiento entre modelos clínicos y modelos radiómicos derivados de distintas ROIs para los diferentes desenlaces: azul para los modelos clínicos, verde para radiómica del parénquima, naranja para radiómica del sangrado obtenida con SMCCH y marrón para radiómica del sangrado correspondiente al Experimento 2.

son estables con valores medios entre 0,71-0,80.

En **vasoespasma**, el desenlace más complejo, las mejoras son más discretas, mostrando ligeros incrementos en los valores medios de las métricas en general. Sin embargo, en todos los modelos se puede apreciar que las métricas medias suben ligeramente o se mantienen en los modelos combinados respecto a los radiómicos puros, solapando los IC de todos los modelos.

En **hidrocefalia**, el modelo parénquima pasa de una AUC media de 0,79 a 0,86, acompañado de un aumento en BalAcc medio (0,60→0,74) y en Spec (0,53→0,68), manteniendo la Sens elevada (0,81). SMCCH también mejora (AUC 0,71→0,75; BalAcc 0,60→0,64), aunque con valores más modestos. En Experimento 2, la combinación produce un incremento de BalAcc (0,63→0,68). Sin embargo, los IC entre modelos vuelven a solaparse.

En términos generales, la combinación de variables clínicas con radiómica mejora sistemáticamente el rendimiento medio de los modelos en todas las métricas, pero especialmente en Spec.

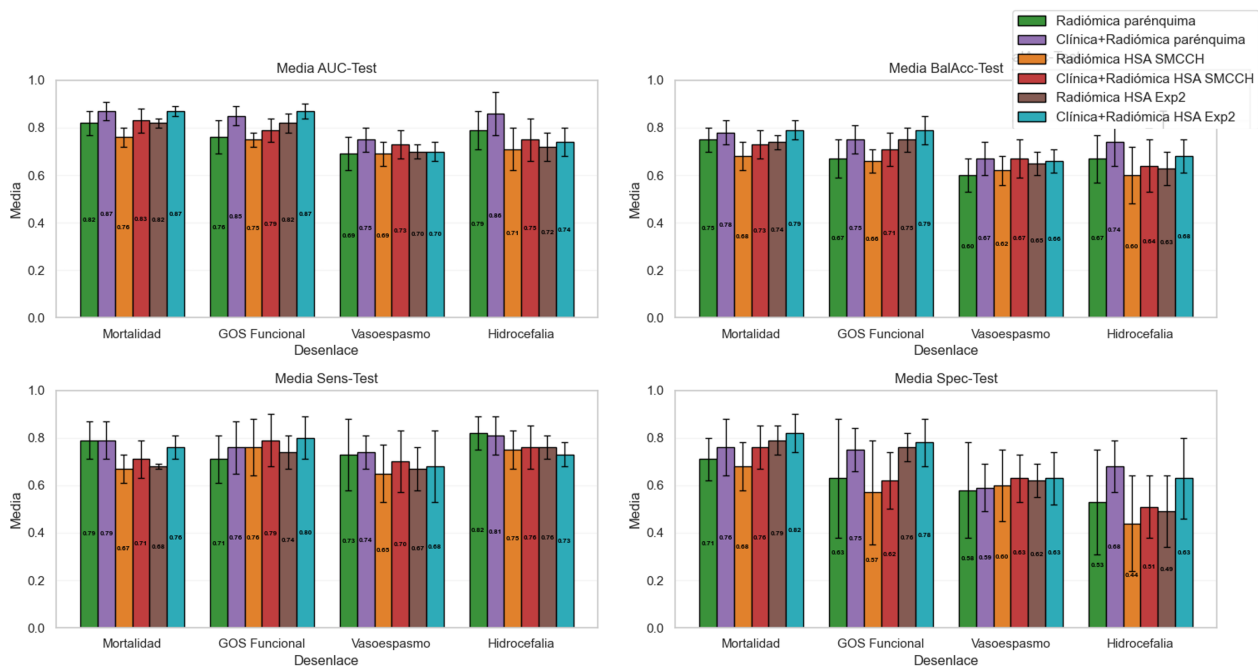


Figura 4.26: Comparación del rendimiento de modelos radiómicos y radiómico-clínicos en distintas ROIs: verde (radiómica parénquima), morado (radiómica+clínica parénquima), naranja (radiómica sangrado SMCCH), rojo (radiómica+clínica sangrado SMCCH), marrón (radiómica sangrado Exp. 2) y cian (radiómica+clínica sangrado Exp. 2).

Posteriormente, se han evaluado los modelos que integran características radiómicas con variables clínicas en el conjunto de test independiente (Figura 4.27).

En **mortalidad**, el modelo clínica+parénquima alcanza el mejor desempeño global, con AUC y BalAcc media de 0,89 y una Spec muy elevada (0,98), manteniendo además buena Sens (0,78). Los modelos SMCCH y Experimento 2 muestran un rendimiento inferior (AUC 0,77 y 0,71 respectivamente), especialmente penalizados por una Sens reducida. Estadísticamente, los modelos de clínica+parénquima mejoran los resultados del Experimento 2, y solo solapan

los IC en Spec de SMCCH.

En el **GOS funcional**, los tres enfoques ofrecen rendimientos competitivos, si bien con perfiles diferentes. El modelo clínica+parénquima es el más robusto (AUC 0,87) presentando un equilibrio adecuado, con Spec de 0,82 y Sens de 0,88. El modelo que combina clínica+hemorragia del Experimento 2 alcanza la segunda mayor AUC media (0,82), mientras que el modelo de clínica + hemorragia de SMCCH alcanza la Sens media más alta (0,91), aunque con menor Spec (0,53). Estadísticamente, los modelos de clínica+parénquima mejoran los resultados de hemorragia en AUC y BalAcc, solapando IC con el modelo de clínica+hemorragia de SMCCH en Sens, y con clínica+hemorragia del Experimento 2 en Spec.

En **vasoespasmio**, el desenlace más desafiante, los rendimientos son limitados (AUC 0,58–0,66). Clínica+parénquima obtiene los mejores valores medios (AUC 0,66, Sens 0,62), con una Sens y Spec equilibrada (0,62-0,63). Los otros dos modelos se sitúan por debajo, destacando la Sens muy baja del SMCCH (0,24), que compromete su aplicabilidad clínica.

En **hidrocefalia**, el rendimiento es más heterogéneo. Los modelos de clínica+parénquima alcanzan la AUC más alta (0,72), con una Sens adecuada (0,75) y Spec aceptable (0,69). En contraste, los modelos basados en hemorragia (SMCCH y Experimento 2) presentan peor rendimiento, con desviaciones entre Sens y Spec, que limita su utilidad en práctica clínica. Estadísticamente, los modelos basados en clínica+parénquima mejoran los de hemorragia en AUC, BalAcc y Spec.

En términos generales, en el conjunto de test externo independiente, la **radiómica del parénquima combinada con clínica emerge como la estrategia más robusta**, liderando en mortalidad, GOS funcional e hidrocefalia.

Respecto a la configuración de los parámetros de entrenamiento, los modelos con mejor rendimiento constantemente evitan la eliminación de valores atípicos, aplican el balanceo de datos y utilizan algoritmos ExtraTrees, RF o XGBoost. Los modelos tanto radiómicos como clínicos suelen seleccionar entre 5 y 20 características. La configuración específica usada por los modelos en cada fold los modelos que combinan clínica y radiómica de parénquima y del Experimento 2 se encuentra en el Anexo B (véase Sección B.5). Además, las matrices de confusión medias normalizadas de los resultados de test de la validación cruzada anidada que combinan radiómica y clínica se encuentran en el Anexo B, en la Figura B.4.

El rendimiento de los modelos varía según la edad del paciente (véase Anexo B, Figura B.2): los pacientes más jóvenes (<70 años) muestran métricas más altas e IC más estrechos, mientras que en mayores de 70 años el rendimiento desciende y la variabilidad aumenta, especialmente en los modelos basados en radiómica del parénquima.

En el Anexo B, la Tabla B.7 muestra el rendimiento del modelo según el grado clínico (WFNS 1-3 frente a 4-5). Los modelos tienden a identificar con algo más de exactitud a los pacientes con grados favorables, aunque la magnitud de esta diferencia no está muy marcada.

Por último, en el Anexo B, la Figura B.3 muestra las curvas de calibración de los conjuntos de test en los 5 folds de validación cruzada para los modelos usando la máscara de parénquima utilizando datos clínicos, radiómica o ambos.

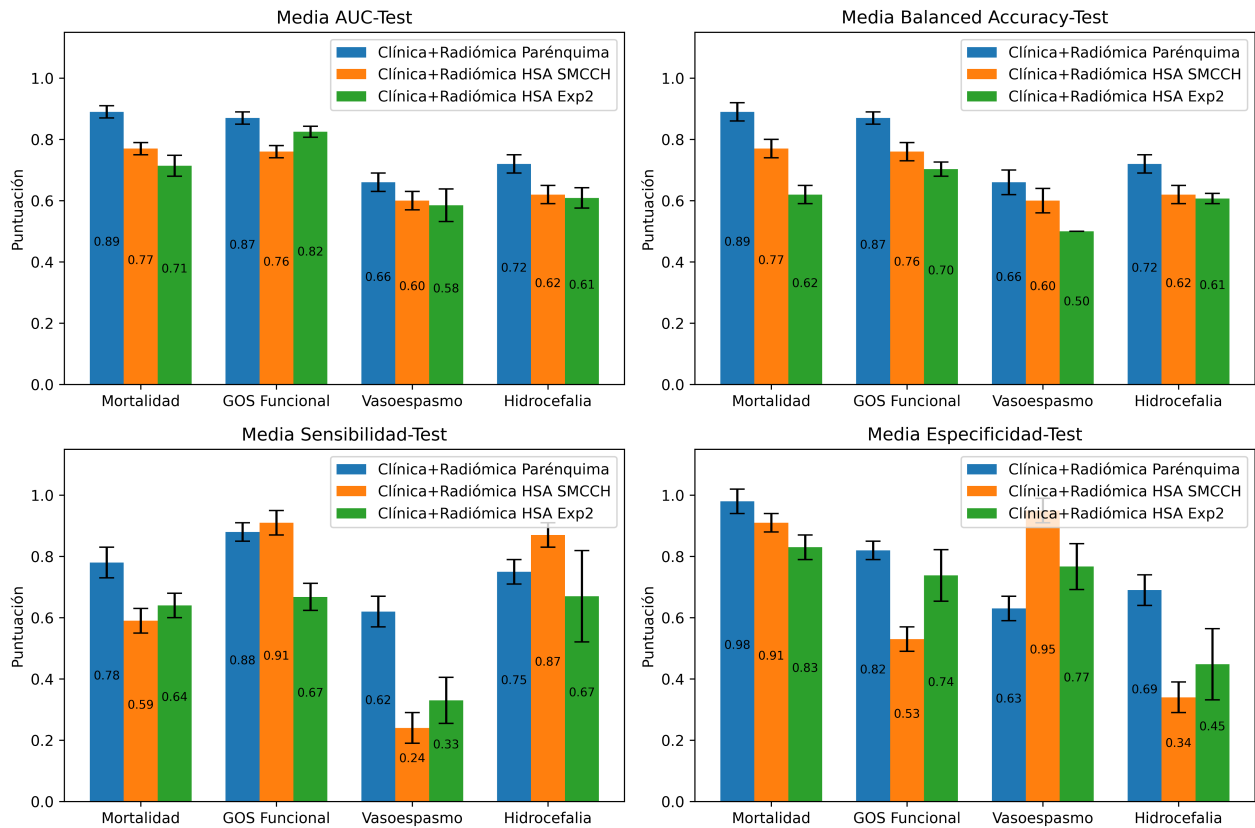


Figura 4.27: Rendimiento de los modelos en el conjunto de test externo basados en clínica + radiómica usando las distintas ROIs. En azul, radiómica extraída del parénquima, en naranja del sangrado con SMCCCH y en verde del sangrado con el Experimento 2.

Interpretabilidad de los modelos

A partir de los resultados obtenidos, surge el interés de conocer la interpretabilidad de los modelos de parénquima, ya que además de ofrecer unos resultados de predicción robustos, la radiómica puede ser más interpretable en las estructuras cerebrales que en la hemorragia. Por ello, la Figura 4.28 muestra los gráficos de valores SHAP para el modelo de mortalidad (desenlace con mayor rendimiento medio) basado en la segmentación de la parénquima cerebral. Las características se clasifican según su impacto, siendo las más relevantes las que aparecen en la parte superior.

La Figura 4.28 representa los valores SHAP de los modelos clínico (A), radiómico (B) y combinado clínica + radiómica (C) entrenados sobre el parénquima. Cada punto corresponde a un paciente, situándose en el eje X según el impacto de esa variable sobre la predicción (positivo implica mayor riesgo, negativo menor riesgo). El color indica la magnitud relativa de la variable (de valores bajos en azul a valores altos en rojo), lo que permite observar no solo la importancia global de cada característica, sino también la dirección de su efecto. Así, los gráficos permiten visualizar simultáneamente la relevancia, el sentido y la variabilidad del efecto de cada predictor.

En el modelo clínico (Figura 4.28), destaca la *WFNS* al ingreso como principal determinante del riesgo, con un desplazamiento consistente hacia peores desenlaces a medida que aumenta su valor. Variables como la dislipemia/obesidad, la glucosa basal y la escala de Fisher modificado muestran contribuciones secundarias, mientras que el sexo presenta un efecto más limitado y variable. En conjunto, el modelo se mantiene altamente interpretable, sustentado en un número reducido de predictores que reflejan de forma coherente el estado clínico inicial del paciente.

El modelo radiómico (Figura 4.28, en contraste, se basa en 20 descriptores de textura e intensidad derivados del parénquima cerebral. No se aprecia una única variable claramente determinante, sino un patrón más distribuido en el que diferentes atributos de co-ocurrencia (*GLCM*), de longitudes de corrida (*GLRLM*) y medidas de primer orden (p. ej. rango, mediana o desviación robusta) capturan la heterogeneidad estructural.

Finalmente, en el modelo combinado (Figura 4.28 se integran predictores clínicos y radiómicos, observándose que tanto la *WFNS al ingreso* como la glucosa mantienen un peso relevante, pero junto a ellos emergen múltiples variables radiómicas que complementan la señal clínica.

En el Anexo B, en la Figura B.5 se representan las gráficas SHAP para cada desenlace predicho en este trabajo (mortalidad, desenlace funcional, hidrocefalia y vasoespasm) usando la radiómica del cerebro.

Además, en el Anexo B, la Figura B.6 muestra el top 10 de variables más frecuentes en los cinco folds de la validación cruzada, con el objetivo de evaluar la estabilidad de las características seleccionadas en el modelo de GOS funcional. La figura evidencia que, en los modelos clínico, radiómico y combinado, algunas variables aparecen de forma consistente en varios folds, sugiriendo patrones de selección robustos. Este análisis complementa los valores SHAP al mostrar no solo la importancia relativa de las variables, sino también su persistencia durante la validación.

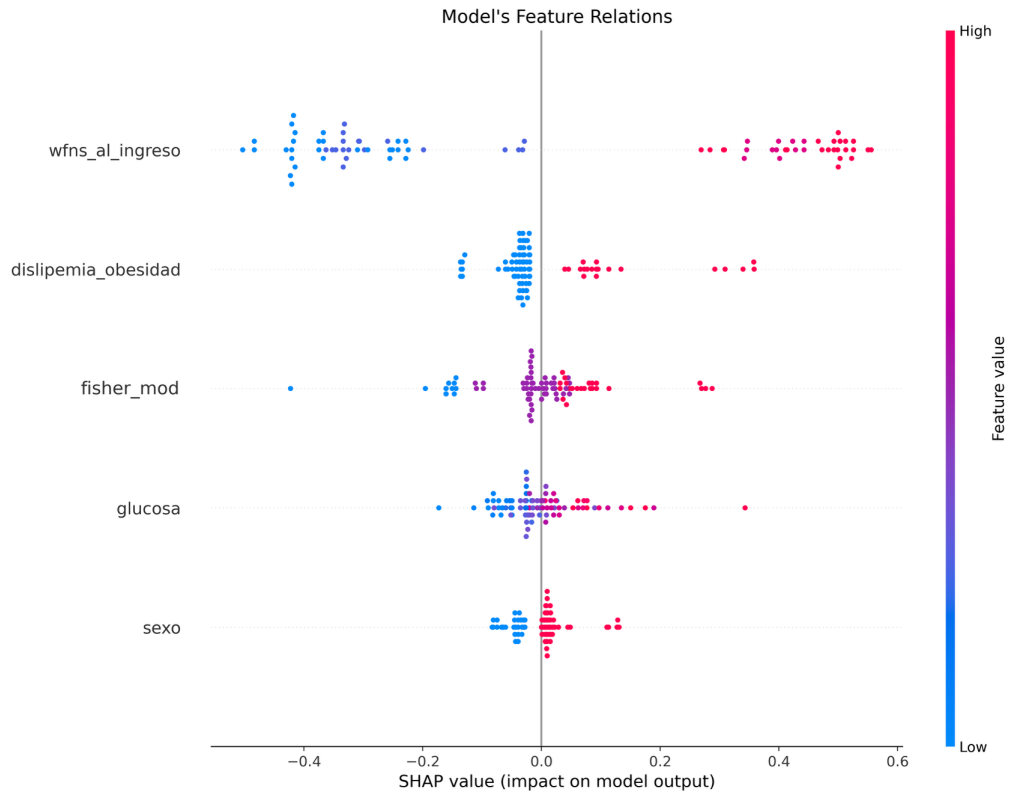


Figura 4.28: (a) Valores SHAP en modelo clínico

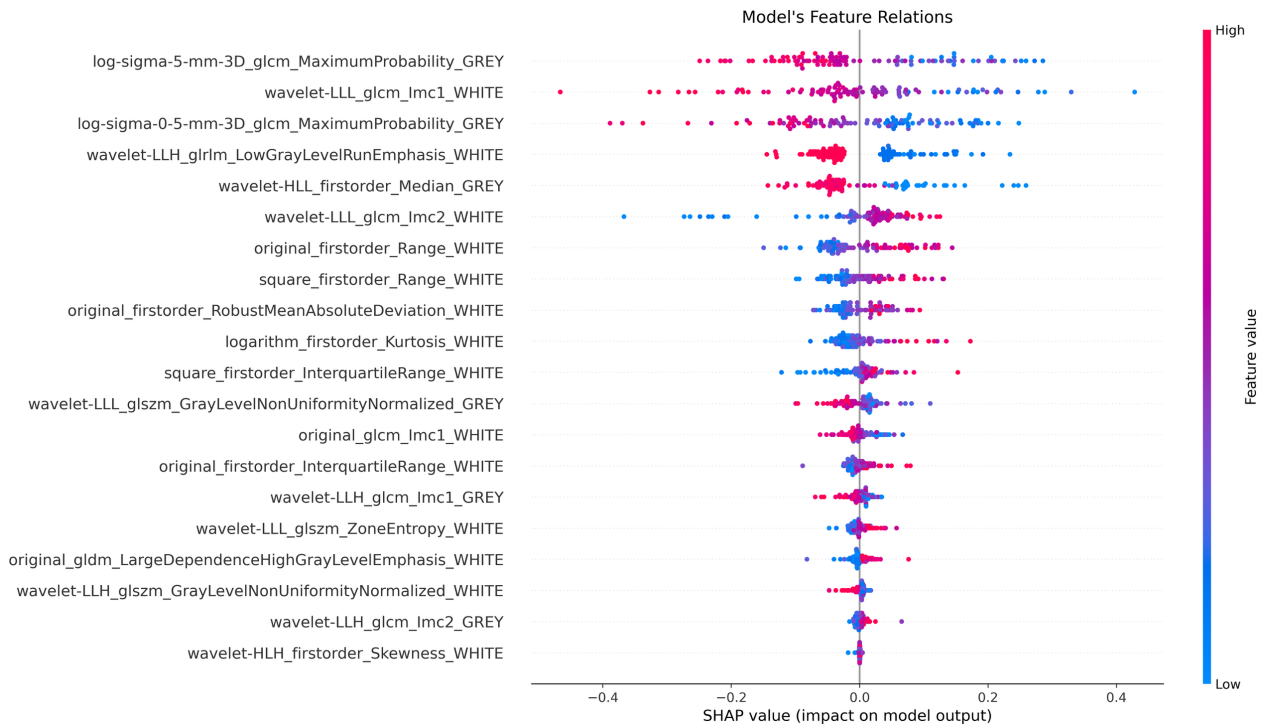


Figura 4.28: (b) Valores SHAP en modelo radiómico

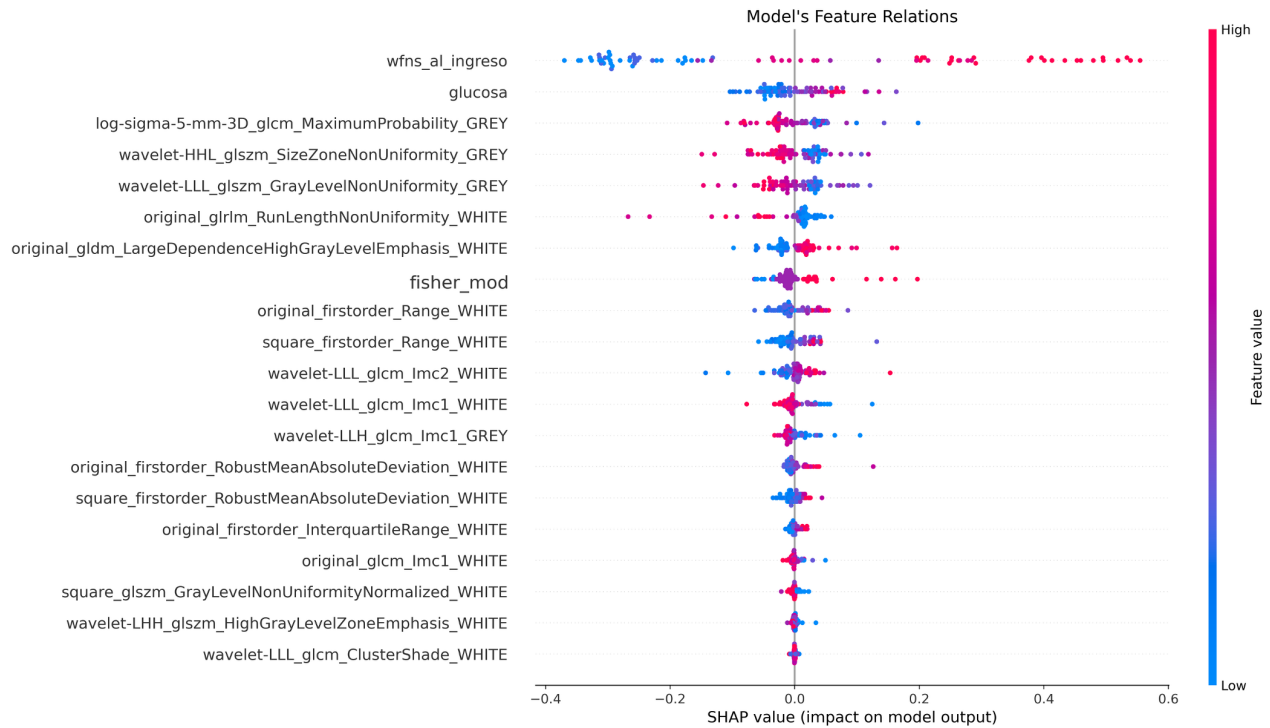


Figura 4.28: (c) Valores SHAP en modelo radiómico + clínico

Figura 4.28: Gráficos de valores SHAP de los modelos entrenados sobre el parénquima: A) clínico , B) radiómico y C) combinado clínica + radiómica..

Comparación con escalas clínicas

Para contextualizar el rendimiento de los modelos radiómicos desarrollados, se ha comparado con escalas clínicas utilizadas en la práctica asistencial como predictores de pronóstico. Entre ellas, se han evaluado la *ictWFNS*, reportada en la literatura como marcador temprano de gravedad clínica, y el *SEBES* y *SEBES6c*, una escala radiológica de edema cerebral ampliamente utilizada. La evaluación de ambas frente a los modelos automáticos de radiómica obtenida del parénquima ofrece un punto de referencia clínico frente al que valorar el rendimiento de las aproximaciones automáticas.

Capacidad predictiva de la escala clínica *ictWFNS* frente a modelos radiómicos

Sobre el conjunto de test externo (n=41), un neurocirujano experto calculó la puntuación individual para cada paciente. El cálculo de la *ictWFNS* integra las puntuaciones *SEBES*, *Hijdra* y *LeRoux*, cuya distribución se resume en la Tabla B.17 suplementaria. Posteriormente se aplicó una regresión logística univariable sobre cada desenlace para evaluar su capacidad predictiva. La Tabla 4.19 resume sus resultados frente a los obtenidos por los modelos radiómicos del parénquima cerebral.

En términos globales, la *ictWFNS* alcanza un valor de AUC de 0,70 tanto para la predicción de mortalidad como de mal pronóstico funcional, lo que refleja un rendimiento aceptable pero

claramente limitado para tareas de clasificación individual. En cuanto a la discriminación, la Sens se mantiene en un rango moderado (0,46 en mortalidad y 0,62 en pronóstico funcional). La Spec, en cambio, muestra un perfil desigual: muy alta para la predicción de mortalidad (0,95), pero considerablemente menor para el pronóstico funcional (0,64).

En los desenlaces secundarios, como el vasoespasma y la hidrocefalia, la capacidad discriminativa de la *ictWFNS* es aún más reducida. El AUC desciende a 0,60 para vasoespasma y a 0,56 para hidrocefalia, con una Sens prácticamente nula en este último caso (0,00).

En contraste, los modelos radiómicos demuestran un rendimiento consistentemente superior en todos los desenlaces evaluados. Para mortalidad y pronóstico funcional alcanzan valores de AUC de 0,89 y 0,87 respectivamente, con Sens y Spec equilibradas en torno al 80–90%. Incluso en los desenlaces más complejos de predecir, como vasoespasma (AUC=0,66) e hidrocefalia (AUC=0,72), los modelos de radiómica superan a la escala clínica.

Tabla 4.19: Rendimiento predictivo de la *ictWFNS* y de los modelos radiómicos cerebrales en la cohorte de test (n=41).

Desenlace	Modelo	AUC	Sens.	Spec.	p-valor
Mortalidad	<i>ictWFNS</i>	0,696	0,455	0,947	0,049
	Modelo radiómico cerebral	0,890	0,780	0,980	—
Resultado clínico	<i>ictWFNS</i>	0,699	0,625	0,643	0,044
	Modelo radiómico cerebral	0,870	0,880	0,820	—
Vasoespasma	<i>ictWFNS</i>	0,596	0,500	0,615	0,243
	Modelo radiómico cerebral	0,660	0,660	0,620	—
Hidrocefalia	<i>ictWFNS</i>	0,560	0,000	1,000	0,726
	Modelo radiómico cerebral	0,720	0,750	0,690	—

* El valor de p corresponde a la regresión logística. No aplicable a los modelos radiómicos no paramétricos.

Capacidad predictiva de las escalas clínica **SEBES** y **SEBES6c** frente a modelos radiómicos

Otra escala clínica ampliamente utilizada es el **SEBES** (y su variante **SEBES6c**), calculada por neurocirujanos sobre la misma cohorte de pacientes empleada en los modelos radiómicos. Para comparar su capacidad predictiva con la radiómica, se ajustan modelos de regresión logística: univariantes (**SEBES** o **SEBES6c**) y multivariantes, incorporando variables clínicas relevantes (*WFNS* al ingreso, Fisher modificada, edad, obesidad, hipertensión y tabaquismo). Los resultados se expresan mediante los *odds ratios* (OR) de cada variable, mientras que la interpretación de los modelos radiómicos se apoya en los valores SHAP, representados en la Figura B.5 del Anexo B.

La Tabla 4.20 presenta un resumen comparativo del rendimiento predictivo de los modelos basados en **SEBES/SEBES6c** frente a los modelos radiómicos del parénquima para los cuatro desenlaces evaluados (mortalidad, resultado funcional, vasoespasma e hidrocefalia). Se comparan seis aproximaciones: modelos univariantes con **SEBES** o **SEBES6c**, modelos

multivariantes que combinan estas escalas con variables clínicas, modelos puramente radiómicos y modelos que integran radiómica con información clínica. Los resultados muestran que los modelos univariantes basados únicamente en escalas SEBES/SEBES6c alcanzan rendimientos cercanos al azar (AUC 0,50–0,60), mientras que la adición de variables clínicas mejora sustancialmente su capacidad discriminativa (AUC \approx 0,60–0,84). Sin embargo, los modelos radiómicos, especialmente cuando se combinan con variables clínicas, superan consistentemente a todas las aproximaciones basadas en escalas clínicas en todos los desenlaces evaluados, destacando el valor añadido de los biomarcadores derivados de imagen.

Para los detalles completos de estas comparaciones, incluyendo modelos univariantes y multivariantes con desglose de contribuciones de variables, véanse las Tablas B.18, B.19, B.20 y B.21 en el Anexo B.

Tabla 4.20: Resumen comparativo del rendimiento predictivo (AUC y BalAcc con IC 95 %) de modelos basados en SEBES/SEBES6c y radiómica del parénquima para los cuatro desenlaces evaluados. Los mejores valores para cada desenlace aparecen en negrita.

Desenlace	SEBES Univ.	SEBES6c Univ.	SEBES +Clín.	SEBES6c +Clín.	Radio-mica	Radio-mica +Clín.
AUC (IC 95 %)						
Mortalidad	0,500	0,528	0,799	0,794	0,82 (0,77–0,83)	0,87 (0,83–0,88)
GOS funcional	0,517	0,544	0,796	0,795	0,76 (0,72–0,80)	0,85 (0,81–0,89)
Vasoespasmio	0,532	0,519	0,596	0,597	0,69 (0,65–0,73)	0,75 (0,71–0,79)
Hidrocefalia	0,570	0,593	0,840	0,840	0,79 (0,75–0,83)	0,86 (0,82–0,90)
BalAcc (IC 95 %)						
Mortalidad	0,50	0,50	0,662	0,680	0,75 (0,70–0,80)	0,78 (0,76–0,82)
GOS funcional	0,500	0,544	0,723	0,717	0,67 (0,63–0,71)	0,75 (0,71–0,79)
Vasoespasmio	0,500	0,500	0,546	0,539	0,60 (0,56–0,64)	0,67 (0,63–0,71)
Hidrocefalia	0,570	0,593	0,772	0,758	0,67 (0,63–0,71)	0,74 (0,70–0,78)

Con el objetivo de investigar si las variables radiómicas pueden reproducir la evaluación visual que realiza el neurocirujano al puntuar la escala SEBES sobre la imagen, se han realizado distintos análisis.

Con el objetivo de investigar si las variables radiómicas pueden reproducir la evaluación visual que realiza el neurocirujano al puntuar la escala SEBES sobre la imagen, se realizan dos análisis complementarios.

En primer lugar, se evalúa la correlación entre cada característica radiómica y la puntuación SEBES mediante el coeficiente de Spearman. Aunque la mayoría de las correlaciones son débiles, se identifica un subconjunto de variables con correlaciones negativas moderadas y alta significación estadística. Esto indica que estas características radiómicas capturan patrones de imagen relacionados con el edema cerebral que el neurocirujano evalúa visualmente al puntuar SEBES.

La Tabla 4.21 muestra las cinco características con mayor correlación con SEBES y SEBES6c. Cuatro de ellas coinciden en ambas escalas (*Uniformity, Maximum Probability, Joint Energy*

y *Run Variance*), todas derivadas de imágenes filtradas mediante wavelets (wavelet-LLH o log-sigma). Estas variables, procedentes principalmente de tejido gris transformado por wavelets, capturan menor uniformidad de intensidad y mayor heterogeneidad conforme aumentan las puntuaciones SEBES. Dado que SEBES y SEBES6c muestran patrones de correlación muy similares (lo que sugiere información redundante), los análisis posteriores se centran únicamente en SEBES.

Tabla 4.21: Top 5 características radiómicas correlacionadas con SEBES y SEBES6c. Nota: FO = first order; GLCM = gray-level co-occurrence matrix; GLRLM = gray-level run length matrix; σ = log-sigma.

Rango	SEBES – Característica	ρ	p	SEBES6c – Característica	ρ	p
1	LLH_FO_Uniformity	-0,325	<0,001	LLH_FO_Uniformity	-0,303	<0,001
2	LLH_GLCM_MaxProb	-0,316	<0,001	LLH_GLCM_JointEnergy	-0,294	<0,001
3	LLH_GLCM_JointEnergy	-0,294	<0,001	$\sigma_{0,5}$ _GLRLM_LRLGE	-0,293	<0,001
4	LLH_GLRLM_GLNUN	-0,280	<0,001	LLH_GLCM_MaxProb	-0,291	<0,001
5	σ_3 _GLRLM_RunVar	-0,278	<0,001	σ_3 _GLRLM_RunVar	-0,290	<0,001

En segundo lugar, se construye un modelo “SEBES radiómico” mediante selección LASSO incremental, con el objetivo de predecir la puntuación SEBES únicamente a partir de variables radiómicas. La Figura 4.29 muestra los resultados de este modelo. En el panel izquierdo se presenta la evolución del rendimiento (AUC) en función del número de características incorporadas: el modelo alcanza un AUC de 0,76 con solo 5 variables y su máximo rendimiento con 35 características (AUC 0,89). Este resultado confirma que las variables radiómicas pueden reproducir eficazmente la evaluación visual del neurocirujano al puntuar SEBES sobre la imagen.

En la evaluación por categorías (panel derecho de la Figura 4.29), las puntuaciones SEBES 3–4 se predicen con Sens 0,85, Spec 0,84 y F1-score 0,90, mientras que SEBES 0–2 alcanza Sens 0,84, Spec 0,85 y F1-score 0,68, con una BalAcc de 0,85 en ambos grupos. Las 35 características seleccionadas por LASSO y sus coeficientes se detallan en el Anexo B (Tabla B.22).

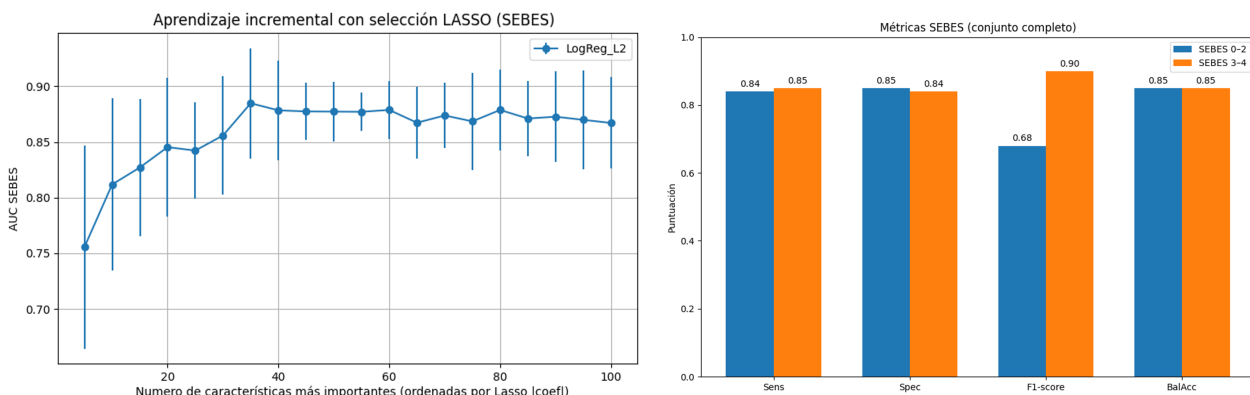


Figura 4.29: Visualización de la predicción de SEBES mediante LASSO incremental. Izquierda: curva del rendimiento del modelo en función del número de características radiómicas incluidas. Derecha: gráfico de barras con las métricas de clasificación (Sens, Spec, F1-score y balanced accuracy) diferenciadas para SEBES 0–2 y SEBES 3–4.

Una vez confirmado que las variables radiómicas pueden reproducir la evaluación visual del neurocirujano (SEBES), se ha explorado si estas mismas características también resultan útiles para predecir desenlaces clínicos. Para ello, se realizan dos análisis complementarios.

En primer lugar, se comparan las 35 características del modelo SEBES radiómico con aquellas identificadas como relevantes en los modelos radiómicos de cada desenlace mediante valores SHAP. El análisis de solapamiento (Tabla 4.22) revela que las variables utilizadas para cada propósito son prácticamente distintas: únicamente se identifican dos características comunes en total (una compartida entre SEBES e hidrocefalia, otra entre SEBES y vasoespasmio), mientras que para mortalidad y resultado funcional no se encuentra ninguna coincidencia.

En segundo lugar, se aplica el modelo SEBES radiómico directamente para predecir cada desenlace clínico mediante regresión logística. Los resultados (Tabla 4.23) muestran un rendimiento muy limitado en todos los casos: AUC de 0,65 para mortalidad, 0,67 para resultado funcional, 0,61 para vasoespasmio y 0,49 para hidrocefalia (sin capacidad discriminativa).

En conjunto, estos hallazgos indican que, aunque las variables radiómicas pueden reproducir eficazmente la evaluación visual del neurocirujano (SEBES), el perfil radiómico asociado al edema no resulta directamente útil como predictor de eventos clínicos. Este hallazgo sugiere que los biomarcadores de edema y los de pronóstico clínico capturan información diferente.

Tabla 4.22: Variables radiómicas comunes identificadas en los modelos de edema (SEBES) y los basados en radiómica para predecir desenlaces clínicos.

Desenlace	Nº solapadas	Características coincidentes
Mortalidad (6 meses)	0	–
Resultado funcional (6 meses)	0	–
Hidrocefalia	1	Original_shape_Elongation_GREY
Vasoespasmio	1	Wavelet-LLH_GLCM_MaximumProbability_GREY

Tabla 4.23: Rendimiento predictivo (AUC) del modelo de regresión logística entrenado con las características radiómicas seleccionadas para reproducir SEBES (“SEBES radiómico”) al aplicarse a los desenlaces clínicos.

Desenlace	AUC	Interpretación
Mortalidad (6 meses)	0,65	Capacidad discriminativa baja-moderada
Resultado funcional (6 meses)	0,67	Capacidad discriminativa baja-moderada
Vasoespasmio	0,61	Capacidad discriminativa pobre
Hidrocefalia	0,49	Sin capacidad discriminativa

Capítulo 5

Discusión

Los resultados obtenidos en este trabajo permiten abordar de manera crítica la aplicabilidad de modelos de segmentación, clasificación tisular y análisis radiómico en el ámbito de la neurociencia clínica. Más allá de las métricas alcanzadas, la discusión se centra en interpretar el significado de los hallazgos, contrastarlos con la literatura existente y valorar sus implicaciones tanto metodológicas como clínicas. La estructura de este apartado se organiza en torno a ejes temáticos que permiten integrar y contextualizar los diferentes experimentos: en primer lugar, el rendimiento técnico de los modelos propuestos y su comparación con métodos manuales y estudios previos; posteriormente, el valor clínico potencial de las medidas derivadas de imagen; y finalmente, las fortalezas y limitaciones metodológicas que condicionan la interpretación de los resultados.

5.1 Rendimiento técnico y comparación con la literatura

A continuación, se desarrollan los resultados de los modelos desarrollados en los distintos experimentos. En la Sección 5.1.1 se analizan los hallazgos de los dos experimentos que se han realizado para la clasificación tisular en GBM mediante imágenes HSI, junto con su comparación frente a otros enfoques descritos en la literatura.

La Sección 5.1.2 aborda la evaluación de los modelos de segmentación automática de hemorragia en TC, contrastándolos igualmente con trabajos previos.

Finalmente, en la Sección 5.1.3 se discuten los resultados de los modelos predictivos que se han basado en variables clínicas y radiómicas en pacientes con HSAa, así como la comparativa con otros modelos basados en radiómica de hemorragia cerebral de la literatura.

5.1.1 Clasificación tisular de glioblastoma en Imagen Hiperespectral

En este estudio se plantea una *prueba de concepto*, realizada en un entorno prospectivo, a partir de adquisiciones hiperespectrales recogidas durante intervenciones neuroquirúrgicas. El objetivo es comprobar la viabilidad técnica de esta tecnología y explorar si, a partir de dichas capturas, es posible desarrollar modelos efectivos de clasificación tisular. Este enfoque inicial

permite sentar las bases para una futura traslación al contexto intraoperatorio en tiempo real.

En primer lugar, la base de datos de *ground truths* obtenida a partir del etiquetado manual evidencia una notable variabilidad interpaciente en el número de píxeles anotados, reflejo directo de la heterogeneidad propia de los escenarios quirúrgicos reales, donde factores como el tamaño tumoral, la extensión de la craneotomía o la visibilidad del campo determinan la cantidad de información disponible. Este comportamiento, aunque introduce desequilibrios entre clases, también aporta un marco de validación más realista para los modelos, obligándolos a enfrentarse a proporciones tisulares desiguales y a clases minoritarias como los tejidos vasculares. La presencia consistente de duramadre y tejido sano garantiza una base sólida de entrenamiento para estas clases predominantes, mientras que la representación más limitada de sangre arterial y venosa plantea un reto añadido.

La caracterización de las firmas espectrales extraídas de los *ground truths* manuales y las capturas HSI ofrece información relevante sobre la diferenciación tisular, pero también plantea ciertas consideraciones (véase Figura 4.9). En primer lugar, la similitud entre las curvas de tejido sano y tumor pone de manifiesto la dificultad intrínseca de discriminar entre ambos únicamente a partir de su reflectancia promedio, lo que refuerza la necesidad de emplear algoritmos de clasificación avanzados capaces de captar patrones sutiles y multidimensionales más allá de la inspección visual de las curvas. En contraste, la sangre arterial y venosa exhiben picos espectrales característicos, mientras que la duramadre muestra una tendencia descendente en las longitudes de onda más altas que la separa del resto de tejidos, sugiriendo un perfil espectral más distintivo. No obstante, la interpretación de estas firmas debe realizarse considerando la distribución desigual de píxeles entre clases (Tabla 4.1). El predominio de tejido sano y tumor proporciona curvas más estables y con menor dispersión, mientras que las clases minoritarias, como las vasculares, presentan una mayor variabilidad relativa que puede estar influenciada no solo por sus propiedades ópticas, sino también por el reducido tamaño muestral. Este desequilibrio constituye una limitación relevante, ya que puede condicionar tanto la consistencia de las firmas promedio como el rendimiento posterior de los modelos de clasificación.

El **Experimento A** constituye una primera aproximación orientada a validar la viabilidad de la base de datos y a explorar la capacidad de diferentes algoritmos para aprender patrones discriminativos a partir de las HSI. Los resultados (Tabla 4.10, Figura 4.12) muestran que RF y CNN superan claramente a SVM, con OACC por encima del 87,6 % en todos los pacientes y medias superiores al 95 % en la mayoría. Los errores de RF y CNN son sistemáticamente < 10 % y, a menudo, < 5 %. Además, CNN alcanza el OACC más alto en tres pacientes, con distribución de errores muy similar a RF. Esta consistencia interpaciente refuerza la capacidad de ambos modelos para ajustarse a la variabilidad individual y sugiere que el conjunto de datos contiene una señal espectral suficientemente rica como para ser explotada en tareas de clasificación tisular.

Sin embargo, este comportamiento tan estable también sugiere riesgo de sobreajuste, especialmente considerando el tamaño limitado de la cohorte y el elevado número de características disponibles en el dominio espectral. La posibilidad de que el modelo esté aprendiendo no solo las propiedades discriminativas de los tejidos, sino también peculiaridades específicas de cada paciente, constituye una preocupación que será abordada de manera más rigurosa en el

Experimento B. Aunque el sobreajuste constituye una limitación habitual en el desarrollo de modelos de aprendizaje automático, en este contexto también señala una oportunidad para RF, dado que combina alta ACC y elevada Sens en la clasificación de los cinco tipos tisulares considerados.

En contraste, SVM muestra un rendimiento más irregular con OACC por debajo del 80 % en la mayoría de los casos y errores más elevados en comparación con los otros algoritmos, lo que sugiere que la complejidad espectral de los datos puede sobrepasar la capacidad del clasificador lineal utilizado para SVM para separar adecuadamente las clases.

Como se observa en la Figura 4.13, los algoritmos RF y CNN alcanzan valores medios de ACC superiores al 90 % en todas las clases, mostrando un rendimiento consistente. Sin embargo, SVM presenta una variabilidad interpaciente mucho mayor en la clase tumor.

La evaluación de la Sens (Figura 4.14) aporta matices relevantes. RF destaca por ofrecer valores elevados y estables, manteniéndose siempre por encima de en la media de ambos, y solapándose con CNN. Sin embargo, aunque la mayor dispersión de CNN limita la consistencia global de este modelo. En cambio, SVM muestra el patrón más inestable, con caídas drásticas en la clase tumoral. Esta fragilidad refleja la incapacidad de los límites de decisión lineales para capturar la complejidad espectral de los datos, especialmente en clases desbalanceadas como la sangre arterial.

El análisis de la Spec (Figura 4.15) refuerza las tendencias descritas. Aunque los tres algoritmos superan el 90 % medio en la mayoría de clases, SVM vuelve a situarse sistemáticamente por debajo de RF y CNN en tejido sano y tumor, confirmando su menor solidez global. Este aspecto es relevante en términos clínicos: los errores que confunden tumor con tejido sano no tienen el mismo impacto que los cometidos en clases menores, pues pueden condicionar directamente la extensión de la resección y la preservación de tejido funcional.

Los mapas de clasificación de la Figura 4.16 ofrecen una visión cualitativa complementaria de estos hallazgos. En el paciente ID33, CNN reduce el número de FN en tumor respecto a RF y SVM, alcanzando una Sens tumoral un 10 % superior a RF y casi un 70 % mayor que SVM, lo que sugiere una mayor capacidad para detectar tejido tumoral en contextos complejos. En ID51, tanto RF como CNN logran identificar correctamente el tumor, mientras que SVM clasifica la mayor parte como tejido sano, reflejando un error clínicamente crítico y en línea con su baja Sens tumoral (25 % frente al 95 % de RF y CNN).

El **Experimento B** se plantea como un paso necesario para validar la capacidad de generalización de los modelos ante pacientes no vistos. Los tres clasificadores (SVM, RF y CNN) muestran un rendimiento global comparable, con ACC medias por clase que en la mayoría de los casos superan el 60 % (Figura 4.17).

Aunque SVM obtiene la media de ACC más alta, las diferencias entre algoritmos no son estadísticamente concluyentes por los amplios IC y su solapamiento, lo que sugiere una inestabilidad de los modelos. Esta amplitud puede atribuirse tanto a la variabilidad interpaciente de las firmas espectrales como a la falta de datos para el entrenamiento, que favorece la varianza del modelo. La comparación entre SVM, RF y CNN confirma que el comportamiento de los algoritmos varía en función tanto de la clase como del paciente. Ningún modelo supera el 80 %

de OACC (Tabla 4.11), lo que indica un margen de mejora y la variabilidad de resultados apunta a que el problema podría estar condicionado por factores específicos de cada paciente.

Aunque RF y SVM también presentan errores de confusión entre tumor, tejido sano y sangre venosa, SVM incurre en un número menor de FP en la clase venosa. Este comportamiento es coherente con una firma espectral más distintiva de la sangre venosa respecto a otras clases (Figura 4.9). En cambio, la baja sensibilidad tumoral del modelo SVM se atribuye a la escasa diferenciación espectral entre tejido tumoral y parénquima sano en el rango de 660–950 nm, lo que complica su separación en el espacio de características, así como al desequilibrio de clases en el conjunto de entrenamiento, que reduce la capacidad del modelo para reconocer correctamente las muestras tumorales.

Este problema se refleja en la Figura 4.18, donde las Sens tumorales del **Experimento B** es variante, y baja en algunos casos. Esto contrasta con el Experimento A, donde tanto RF como CNN alcanzan sensibilidades cercanas al 90 % y logran detectar de manera consistente regiones tumorales. Este contraste pone de manifiesto la diferencia entre ambos esquemas experimentales: mientras el Experimento A (intrapaciente) se beneficia de la homogeneidad espectral dentro de cada sujeto y permite al modelo aprender patrones específicos del paciente, el Experimento B (interpaciente) exige una mayor capacidad de generalización ante la variabilidad biológica y técnica entre individuos. Esta heterogeneidad interpaciente explica la mayor variabilidad y la menor sensibilidad observada en algunos casos, al dificultar la discriminación espectral entre tejido tumoral y sano.

A nivel de clases, ninguno de los modelos consigue discriminar de forma adecuada ni el tumor ni la sangre arterial (Figura 4.18). El caso de la clase arterial es especialmente extremo: CNN alcanza una Sens del 0 % en todas las predicciones. Este resultado se explica por el fuerte desequilibrio de datos: la clase arterial apenas suma 1,256 píxeles y, tras la doble validación cruzada, solo ~ 64 % de ellos se emplean en el ajuste y calibración (80 % para entrenamiento y, de este, un 80 % para calibración). Dicho volumen es insuficiente para aprender límites de decisión robustos. Un efecto análogo se observa en la sangre venosa.

El análisis cualitativo de los mapas de clasificación respalda estas observaciones (Figura 4.19). La duramadre se identifica de forma consistente por los tres algoritmos, probablemente por su firma espectral más distintiva. En cambio, el tejido tumoral presenta confusiones frecuentes con parénquima sano y sangre venosa. En ID33, CNN tendió a sobredetectar tumor, clasificando extensas áreas de superficie cerebral; RF y SVM también incurren en errores, aunque SVM muestra menos FP en sangre venosa que RF. En ID51, los tres métodos clasifican adecuadamente los píxeles tumorales, pero los mapas de RF y SVM resultan más nítidos e interpretables que los de CNN. En conjunto, los resultados indican que no existe un clasificador claramente superior de forma consistente entre clases y pacientes.

Conviene recordar que la metodología propuesta se basa en una cámara hiperespectral instantánea capaz de capturar 25 bandas en el rango del rojo y el infrarrojo cercano. Aunque este número de bandas es considerablemente inferior al de los sistemas empleados en otros estudios, y pese a la similitud entre algunas firmas espectrales de los tejidos, en el Experimento A la OACC llega al 95 % utilizando RF, mientras que en el Experimento B se alcanza un 60 % con SVM.

Si se sitúan estos hallazgos en el contexto del estado del arte, se observa que las mejores cifras reportadas corresponden a un enfoque híbrido 3D-2D CNN, con una OACC del 80 %, mientras que con SVM se alcanza un 76 % (Manni et al., 2020). A primera vista, el rendimiento del algoritmo SVM en este trabajo (60 %) parece significativamente inferior, con una diferencia de alrededor del 20 % respecto al mejor enfoque híbrido publicado. Sin embargo, esta comparación debe matizarse: los sistemas previos se apoyan en cámaras HSI con capacidades muy superiores, llegando a registrar hasta 128 (Cruz-Guerrero et al., 2020; Manni et al., 2020) o incluso 826 bandas (Fabelo et al., 2016), frente a las 25 bandas capturadas por la cámara snapshot utilizada en nuestro estudio.

Para evaluar de manera más equitativa la relación entre rendimiento y complejidad espectral, se introduce una métrica ajustada (OACC/bandas), que expresa la precisión media obtenida por cada banda disponible. La Tabla 5.1 muestra que, bajo esta métrica, la metodología aquí propuesta resulta claramente más eficiente. En concreto, la mejor relación alcanzada con la cámara snapshot es del 2,4 % (SVM propuesto), lo que multiplica por 3,8 el valor reportado por el mejor enfoque previo (0,63 % con CNN 3D-2D (Manni et al., 2020)). Esta diferencia muestra la eficacia relativa de la propuesta, capaz de extraer información discriminativa competitiva a partir de un espectro mucho más reducido.

Tabla 5.1: Comparación entre los resultados de OACC de enfoques llevados a cabo en el estado del arte y los propuestos en este estudio.

Enfoque	OACC original (%)	N.º de bandas	OACC/Bandas (% por banda)
CNN 3D-2D (Manni et al., 2020)	80	12	0,63
3D-2D-CNN + SVM (Manni et al., 2020)	75	12	0,59
SVM (Manni et al., 2020)	76	128	0,59
CNN 2D (Manni et al., 2020)	72	12	0,56
CNN 1D (Manni et al., 2020)	78	12	0,61
ANN (Fabelo et al., 2016)	96,7	826	0,11
RF (Fabelo et al., 2016)	99,46	826	0,12
Separación espectral (Cruz-Guerrero et al., 2020)	76	128	0,59
SVM propuesto	60	25	2,40
RF propuesto	53	25	2,12
3D-CNN propuesto	49	25	1,96

Otro aspecto importante que condiciona la comparación con el estado del arte es el número de clases consideradas. Mientras que algunos estudios emplean cuatro categorías (tejido normal,

tumor, vasos sanguíneos y fondo) (Cruz-Guerrero et al., 2020; Manni et al., 2020) y otros únicamente tres (normal, tumor y otros) (Fabelo et al., 2016), en nuestro caso la clasificación se realiza sobre cinco clases distintas. Este incremento en la complejidad de la tarea tiende a reducir el rendimiento de los modelos, lo que confiere un valor adicional a los resultados obtenidos.

En conjunto, aunque los valores absolutos de OACC puedan parecer más modestos respecto a los mejores reportes de la literatura, la eficiencia espectral alcanzada, unida a la mayor dificultad del escenario de clasificación planteado, avalan la solidez de la metodología propuesta. Además, la naturaleza snapshot de las cámaras HSI, junto con la baja dimensionalidad espectral y la rapidez de los modelos de ML empleados, favorece su potencial aplicación en entornos intraoperatorios en tiempo real, donde la rapidez de decisión resulta crítica. Estos hallazgos sugieren que las cámaras HSI snapshot de bajo número de bandas, combinadas con algoritmos de ML, constituyen una alternativa viable y escalable para la clasificación intraoperatoria de tumores cerebrales.

Además, los experimentos realizados en este trabajo y el propio procedimiento de adquisición de imágenes hiperespectrales sirven como punto de partida para el desarrollo de una base de datos más amplia y estandarizada: SLIMBRAIN (SpectraL Imaging of the human BRAIn for INtraoperative applications) (Martín-Pérez et al., 2025). Esta iniciativa surge como una extensión natural del protocolo de captura empleado en este estudio, utilizando la misma cámara hiperespectral y configuración óptica, pero ampliando el número de pacientes, la diversidad de tejidos y las condiciones quirúrgicas incluidas.

5.1.2 Segmentación automática de Hemorragia Subaracnoidea Aneurismática en TC

En primer lugar, se ha desarrollado un sistema de **segmentación automática** del sangrado cerebral en imágenes de TC basado en CNNs. El conjunto de datos principal procede de una cohorte del H12O, compuesta por 223 pacientes con diagnóstico confirmado de HSAa. Con el fin de mejorar la capacidad de generalización del algoritmo frente a datos de distinta procedencia, se han incorporado también bases de datos públicas durante las fases de entrenamiento y validación.

Se han diseñado dos experimentos diferenciados. En el *Experimento 1*, el modelo se ha entrenado con la cohorte del H12O y con la base pública HemSeg200, que incluye segmentaciones de HIP y HIV, potencialmente relacionadas con la HSAa. En el *Experimento 2*, se han añadido al entrenamiento dos bases públicas de HIC secundarias a traumatismos craneoencefálicos, una patología distinta. En este caso, dichas bases de datos se emplean exclusivamente como estrategia de aumento de datos, sin ser consideradas en la validación final.

En ambos experimentos se observan diferencias notables tanto en los parámetros de adquisición como en las segmentaciones manuales disponibles, representadas en la Figura 4.4 para la cohorte del H12O y, de manera global, en la Tabla 4.3 y en la Figura 4.5. Esta heterogeneidad refleja la práctica clínica real, donde los protocolos de imagen y los equipos varían entre centros y momentos de adquisición. Aunque estas diferencias pueden constituir una fuente de sesgo (al introducir variabilidad técnica o de anotación no relacionada con la patología) también

aportan valor, ya que exponen al modelo a escenarios diversos y favorecen su capacidad de generalización.

Además, la variabilidad tiene un doble efecto. Por un lado, entrenar con datos heterogéneos puede actuar como un mecanismo de regularización, reduciendo el riesgo de sobreajuste a un único contexto. Pero, por otro, ciertas discrepancias (como las asociadas al tamaño de píxel o al grosor de corte) influyen en la resolución espacial y en la anisotropía de los volúmenes, afectando a la detección de hemorragias pequeñas. A esto se suma la desigualdad en el número de cortes por exploración y en la proporción de cortes segmentados, que puede introducir un desbalance entre regiones con y sin hemorragia que puede dificultar la identificación de patrones hemorrágicos.

El marco `nnU-Net` contribuye a mitigar en gran medida estas limitaciones gracias a su preprocesamiento estandarizado, que incluye el resamplado a espacio isótropo y la normalización de intensidades en UH. Además, durante el entrenamiento implementa un muestreo inteligente de parches (o *patches*): aproximadamente la mitad de los parches extraídos contienen de forma forzada algún vóxel de la estructura de interés, mientras que la otra mitad se selecciona de manera aleatoria en todo el volumen. Este procedimiento automático equilibra la proporción entre fondo y lesión, evitando que la red quede sesgada hacia el predominio de regiones sin información clínica relevante y facilitando un aprendizaje más robusto de las clases minoritarias. Estas estrategias permiten un aprendizaje más robusto y consistente. Sin embargo, no eliminan por completo la incertidumbre cuando el modelo se enfrenta a datos obtenidos en condiciones muy distintas de las contempladas en el entrenamiento, algo que conviene tener presente al interpretar los resultados y pensar en su aplicación futura.

Los modelos desarrollados se han evaluado tanto de forma cuantitativa como cualitativa, evaluando sobre un conjunto de test los modelos desarrollados en este trabajo como otros cinco existentes en la literatura (Viola-UNet v2, ICH 2023, ISBI 2024, Blast-CT, y SMCCH) para la segmentación de HIC.

Respecto a la evaluación cuantitativa (Tabla 4.12), los hallazgos de este estudio confirman que los modelos desarrollados se sitúan en el grupo de mayor rendimiento dentro del campo de la segmentación automática de hemorragias en TC. Tanto el Experimento 1 como el Experimento 2 alcanzan valores medianos sistemáticamente superiores y un desempeño general equiparable a modelos de referencia recientes, como ICH 2023 o Viola-UNet v2. Los modelos han sido evaluados de manera independiente en la cohorte pública HemSeg200 y en la cohorte hospitalaria del H12O, manteniendo en ambos casos un rendimiento competitivo y consistente frente a los algoritmos previamente descritos en la literatura. No obstante, se observa de manera sistemática un mejor desempeño en HemSeg200 que en H12O. Esta diferencia puede atribuirse a la mayor complejidad que plantea la HSAa: distribución difusa, morfología menos predecible y una heterogeneidad clínica superior frente a las HIC (HIV o HIP) incluidas mayoritariamente en HemSeg200. En este sentido, la HSAa representa un escenario particularmente desafiante para la segmentación automática, lo que realza el valor de haber obtenido resultados sólidos en dicha patología. La importancia de este hallazgo es aún mayor si se considera que la cohorte principal de este trabajo corresponde a HSAa, cuyas características radiológicas (dispersión espacial, focos satélite de pequeño tamaño y bordes difusos) difieren notablemente de las observadas en otras bases de datos públicas.

Esta diferencia puede apreciarse desde la construcción del *ground truth* (Figura 4.2), donde las segmentaciones de H12O muestran hemorragias más diseminadas y con límites menos definidos que las contenidas en HemSeg200.

El análisis comparativo entre los dos experimentos aporta matices que muestran dilemas clásicos en el uso de DL para la segmentación de imágenes médicas. El Experimento 2, que incorpora datos adicionales de HIC traumáticas como estrategia de aumento de datos en el entrenamiento, alcanza el mejor rendimiento global y la menor desviación volumétrica. Esto sugiere que entrenar con datos diversos hace a los modelos más robustos frente a la variabilidad entre centros y mejora su capacidad de generalización, un aspecto esencial cuando se piensa en aplicaciones multicéntricas. Sin embargo, este enfoque también tiene un coste: se pierde parte de la precisión en el trazado de los bordes. Así lo muestran métricas sensibles a la calidad de los contornos, como la distancia de Hausdorff o la distancia media de superficie, en las que el Experimento 1 obtiene mejores resultados. En definitiva, los dos experimentos reflejan un equilibrio entre generalización y precisión espacial, un debate ampliamente presente en la literatura sobre segmentación de lesiones en neuroimagen. Mientras que entrenar con datos homogéneos permite un ajuste más fino de los bordes, la heterogeneidad ofrece cobertura a un abanico mayor de escenarios clínicos, aunque a costa de sacrificar detalle morfológico.

Este hallazgo plantea un debate metodológico relevante: en qué medida conviene priorizar la robustez global frente a la precisión local. En un contexto clínico, las implicaciones difieren según el uso previsto de la segmentación. Para la cuantificación volumétrica y la monitorización longitudinal, la estabilidad y ausencia de sesgo sistemático, como la mostrada por el Experimento 2, resultan más valiosas que el ajuste a contornos finos. Por el contrario, en escenarios donde la localización exacta de los límites de la hemorragia condiciona decisiones terapéuticas (por ejemplo, en procedimientos endovasculares complejos o en el cálculo de índices de desplazamiento de estructuras), un modelo con mayor fidelidad espacial, como el Experimento 1, podría ser preferible.

La interpretación conjunta de las gráficas de dispersión (Figura 4.21) y del análisis de Bland–Altman (Figura 4.20) permite evaluar también las fortalezas y limitaciones de los modelos de segmentación. La relación entre el Dice Score y el volumen hemorrágico muestra que los casos de pequeño volumen se segmentan con alta precisión y baja variabilidad, mientras que, a medida que aumenta la extensión de la hemorragia, emergen discrepancias mayores. El análisis de Bland–Altman aporta una dimensión adicional al evidenciar cómo esas discrepancias se traducen en sesgos sistemáticos de cuantificación. En el Experimento 1 se observa una tendencia mayor a la sobreestimación de volúmenes, sin embargo, el Experimento 2 presenta un sesgo cercano a cero y una dispersión más acotada, lo que indica una mayor estabilidad y robustez en la estimación volumétrica. La Figura 4.21, muestra que en el conjunto de HemSeg200 los resultados suelen ser mejores que en H12O, sobre todo en volúmenes menores a 20 mL, esto puede ser debido a que la coexistencia de focos satélite y configuraciones irregulares particulares a la enfermedad de HSAA tiende a generar falsos positivos que deterioran la métrica global.

Respecto a los resultados cualitativos, la Figura 4.22 muestra resultados visuales de las segmentaciones obtenidas tanto con los modelos desarrollados como en la literatura, y la Figura 4.23, junto con la Tabla 4.14, permiten una valoración amplia del rendimiento de

los modelos desde el punto de vista cualitativo de siete neurocirujanos (posibles futuros usuarios). En conjunto, los resultados indican que los modelos propuestos alcanzan un rendimiento comparable al de las aproximaciones actuales del estado del arte, mostrando un comportamiento equilibrado y potencialmente útil en contextos clínicos.

La Figura 4.22 ilustra con claridad la diferencia entre casos favorables y desfavorables. En los ejemplos con Dice superiores al 85 %, los modelos desarrollados en el Experimento 1 y 2 ofrecen resultados robustos y estables, incluso en comparación con métodos consolidados como Viola-UNet v2 o ICH-2023. Destaca especialmente su capacidad para identificar de manera fiable regiones de HSA, donde otros algoritmos tienden a mostrar falsos negativos. Sin embargo, algunos algoritmos como SMCCH (modelo entrenado para la segmentación de HSAa) presentaba patrones de falsos positivos, mientras que Blast-CT mostraba segmentaciones ruidosas. Esto es algo que se muestra también en las métricas cuantitativas, donde por ejemplo, en SMCCH la diferencia de volumen entre segmentaciones es la más alta que el resto de modelos, llegando a una diferencia media de 23.58 mL.

En los ejemplos con Dice Score inferiores al 67 %, las dificultades se han concentrado en focos hemorrágicos pequeños o regiones de bajo contraste. Aunque estos escenarios siguen suponiendo un reto, nuestros modelos han conservado una coherencia anatómica comparable o superior a la de los otros algoritmos de la literatura. Resulta particularmente relevante que, incluso en casos adversos, los modelos propuestos han identificado regiones hemorrágicas que otros enfoques pasaron por alto, lo que subraya su utilidad potencial en contextos clínicos donde los falsos negativos pueden tener mayor impacto.

El estudio cualitativo con siete neurocirujanos (Figura 4.23) añade una dimensión clave a la evaluación. En términos de cobertura anatómica e infrasegmentación, el modelo del Experimento 2 se sitúa como el mejor valorado de media, superando de manera estadísticamente significativa a Blast-CT. Aunque en sobresegmentación las diferencias entre métodos son menos claras (presentando una mayor media Viola-UNet v2), los modelos desarrollados en este trabajo muestran un comportamiento equilibrado, sin excesos de ruido ni pérdidas anatómicas graves.

La síntesis de estas valoraciones mediante el método de Borda (Tabla 4.14) ha situado al Experimento 2 en primera posición, seguido de Viola-UNet v2 y del Experimento 1, que han ocupado la segunda y tercera posición respectivamente. Este resultado confirma que los modelos desarrollados no solo son competitivos en métricas cuantitativas, sino que también son percibidos por expertos clínicos como más fiables que varias de las alternativas disponibles.

5.1.3 Modelos predictivos radiómicos en Hemorragia Subaracnoidea Aneurismática

A partir de las bases de datos de TC de pacientes con HSAa del H12O, se han desarrollado **modelos predictivos de desenlaces clínicos** centrados en cuatro eventos de alta relevancia pronóstica: mortalidad, resultado funcional (GOS), vasoespasmo e hidrocefalia. Para ello, se han construido modelos basados en tres fuentes de información: (i) variables clínicas disponibles en el momento del ingreso, (ii) características radiómicas extraídas del parénquima cerebral (sustancia gris y blanca) segmentado automáticamente y (iii) características radiómicas

derivadas de las regiones hemorrágicas segmentadas automáticamente mediante dos enfoques diferentes: el modelo SMCCCH, previamente descrito en la literatura y entrenado en cohortes de HSAa, y el modelo desarrollado en este trabajo mejor valorado por los expertos y a nivel de métricas globales, correspondiente al Experimento 2.

Tal y como se muestra en la Figura 4.11, el análisis Bland–Altman de los volúmenes hemorrágicos obtenidos por ambos modelos evidencia un sesgo sistemático: SMCCCH tiende a sobreestimar el volumen respecto al modelo del Experimento 2, con discrepancias que aumentan conforme crece el tamaño de la hemorragia. Este hallazgo resulta coherente con lo observado en la Sección 4.3.2, donde SMCCCH es el método que muestra mayores diferencias de volumen respecto a la referencia manual de los neurocirujanos, además de generar extensas áreas de FP en regiones adyacentes al cráneo. Considerando que ambos algoritmos son específicamente entrenados en HSAa, pero difieren de manera significativa en la segmentación, se ha decidido extraer radiómica de ambos con el fin de evaluar comparativamente el impacto de la definición de la zona hemorrágica en el rendimiento de los modelos predictivos.

Una aportación novedosa de esta tesis doctoral es la incorporación del **análisis radiómico del parénquima**, más allá de las regiones hemorrágicas. La hipótesis planteada se fundamenta en la evidencia de que las HSA inducen cambios estructurales y de textura en la sustancia blanca y gris, relacionados con procesos como edema, isquemia o disfunción axonal (Lee et al., 2020; Ru et al., 2021). Por tanto, explorar si estos patrones pueden ser captados de manera objetiva mediante radiómica ha representado un paso adicional para enriquecer los modelos predictivos y aproximarnos a una caracterización más integral del daño cerebral secundario a la HSAa.

La radiómica depende de forma crítica de la calidad y homogeneidad de los estudios de imagen empleados. En este trabajo se observa que el tamaño de píxel axial se mantiene relativamente estable entre las cohortes usadas para el desarrollo de modelos y test independiente definido con una temporalidad posterior a la de entrenamiento, (reproduce un entorno prospectivo para evaluar la generalización temporal del modelo), con valores concentrados en el rango de 0,47–0,49 mm, lo que probablemente ha favorecido la robustez de las características extraídas. En contraste, el grosor de corte ha presentado una mayor variabilidad en la cohorte de desarrollo de modelos (predominio de 1.5 mm en el 74 % y de 3 mm en el 23 % de los estudios), mientras que en la cohorte de test independiente todos los casos son adquiridos con cortes uniformes de 1.5 mm.

Esta diferencia es relevante porque el grosor de corte puede modificar de forma sustancial la estabilidad de la radiómica, al suavizar bordes y reducir la complejidad textural en cortes más gruesos. Para mitigar este problema, en el presente estudio se han aplicado estrategias de preprocesado orientadas a reducir la heterogeneidad, como la interpolación a un vóxel isotrópico de $1 \times 1 \times 1$ mm y la discretización de intensidades con binning fijo en unidades Hounsfield. Estas medidas no eliminan completamente el efecto de las diferencias de adquisición, pero constituyen un estándar recomendado por las guías IBSI para garantizar la comparabilidad de los descriptores (Zwanenburg et al., 2020).

En este sentido, cabe destacar que la uniformidad del conjunto de test independiente (cortes de 1.5 mm en el 100 % de los casos) puede haber favorecido un rendimiento más estable

en la validación del test independiente, mientras que la mayor variabilidad del conjunto de desarrollo de modelos refleja mejor la práctica clínica habitual, en la que la radiómica debe demostrar estabilidad frente a condiciones no estandarizadas.

Más allá de las diferencias técnicas en la adquisición de imagen, otro de los factores que puede condicionar el rendimiento de los modelos predictivos es la divergencia en los perfiles clínicos entre el conjunto de desarrollo de modelos y el de test independiente.

En la cohorte para el desarrollo de modelos, los pacientes presentan una edad media más baja, una menor carga de comorbilidades cardiovasculares y, en general, un perfil clínico algo más favorable. Por el contrario, la cohorte de test independiente ha incluido pacientes de mayor edad, con mayor prevalencia de factores de riesgo vascular (hipertensión, tabaquismo, dislipemia/obesidad) y con parámetros analíticos indicativos de mayor fragilidad (hiperglucemia significativa y plaquetopenia marcada). Asimismo, la severidad al ingreso es sustancialmente más elevada en el grupo de test independiente, con mayor proporción de WFNS ≥ 4 y Fisher modificado ≥ 3 . Estas diferencias no son menores: los modelos se entrenan en una cohorte compuesta por pacientes más jóvenes y con menor gravedad inicial, pero son evaluados en un conjunto de test caracterizado por un perfil clínico más complejo y desfavorable. En consecuencia, el test independiente con otra temporalidad representa un escenario más exigente para los predictores, lo que convierte sus resultados en una prueba especialmente valiosa de la capacidad de los modelos para generalizar a contextos clínicos distintos y más adversos.

Posteriormente, se ha llevado a cabo un análisis de correlación entre las variables clínicas y radiómicas y los diferentes desenlaces estudiados. La Tabla 4.9 resume aquellas que muestran mayor grado de asociación. La identificación de la WFNS como principal predictor clínico confirma su papel consolidado en la literatura como marcador determinante de la gravedad inicial de la HSAa (CG, 1988). No obstante, entre las variables con mayor correlación también han emergido descriptores radiómicos, lo que sugiere un valor añadido frente a las escalas tradicionales, aportando información cuantitativa y objetiva que resulta difícil de captar mediante la simple inspección clínica o radiológica.

En este análisis se observa que las asociaciones radiómicas involucran patrones complejos de textura tanto en la región de la parénquima como en la hemorragia. En este sentido, la radiómica aporta un nivel de granularidad que escapa a los descriptores clásicos como el Fisher modificado, que solo codifica la extensión espacial de la sangre. En los desenlaces primarios relacionados con el resultado funcional y la mortalidad, las asociaciones de las variables son más fuertes que en los desenlaces secundarios (vasoespasmos e hidrocefalia). Las asociaciones radiómicas observadas para hidrocefalia son modestas y sin un patrón consistente, lo que limita su interpretación y sugiere que la capacidad predictiva en este evento es, por el momento, menos clara. En el caso del vasoespasmos, aunque las correlaciones son de menor magnitud que en el resultado funcional o la mortalidad, sí se han identificado algunas variables radiómicas con correlaciones algo más altas que podrían influir en la susceptibilidad a fenómenos de vasoconstricción arterial.

Respecto a los resultados de la test de la validación cruzada anidada estratificada usando modelos puramente clínicos o radiómicos (Figura 4.25), los modelos clínicos se sitúan entre

los más consistentes en mortalidad y resultado funcional (AUC 0,84 y 0,76, respectivamente). Sin embargo, los modelos puramente radiómicos del parénquima y del Experimento 2 ofrecen rendimientos muy próximos (AUC 0,82 en mortalidad y 0,76/0,82 en resultado funcional), lo que evidencia que la información contenida en la imagen basal es pronósticamente comparable a la de las escalas clínicas. Cada aproximación muestra además puntos fuertes diferenciales: el parénquima maximiza Sens (0,79), mientras que el Experimento 2 alcanza mayor Spec (0,79), lo que sugiere que capturan facetas complementarias del riesgo. En contraste, los modelos basados en SMCCH quedan sistemáticamente por debajo en valores medios (AUC 0,75–0,76), reflejando cómo los errores de segmentación afectan de forma negativa al valor predictivo.

En los desenlaces secundarios, los resultados son más modestos y deben interpretarse con cautela. El vasoespasma se confirma como el desenlace más desafiante, con AUC limitadas (0,66–0,70) y amplios IC. Esta dificultad es coherente con la propia práctica clínica, donde el diagnóstico de vasoespasma carece de un estándar único y puede basarse en criterios heterogéneos (clínicos o radiológicos), lo que introduce un importante grado de variabilidad y ruido en las etiquetas. Incluso bajo estas condiciones, los modelos logran Sens aceptables en el parénquima (0,73), lo que sugiere que la imagen basal contiene cierta señal relacionada con la susceptibilidad a desarrollar esta complicación.

En hidrocefalia, la base de datos ha estado marcada por un fuerte desbalance entre positivos y negativos (90% de casos negativos), lo que en general dificulta la construcción de modelos discriminativos estables. A pesar de ello, la radiómica del parénquima alcanza el rendimiento más prometedor (AUC 0,79; BalAcc 0,67), con una Sens elevada (0,82) y una Spec moderada (0,63). Estos resultados indican que, incluso en condiciones adversas de prevalencia y con un número reducido de eventos positivos, podrían existir patrones cuantitativos en la textura y morfología del parénquima que capturan de manera indirecta procesos asociados al desarrollo de hidrocefalia. Aunque las diferencias con respecto a otros modelos no alcanzan significación estadística, el hecho de que en todos los desenlaces el parénquima se mantenga como una fuente informativa refuerza su papel como región clave para integrar en futuros desarrollos predictivos.

La comparación entre modelos radiómicos puros y aquellos combinados con clínica (Figura 4.26) muestra la utilidad de los enfoques multimodales: en mortalidad y resultado funcional, la integración ha producido mejoras consistentes en AUC (hasta 0,87), BalAcc (0,79) y Spec (incrementos de 0,08–0,13), sin pérdida de Sens. Aunque los IC se solapan en varios casos, el patrón global apunta a que la fusión de información clínica y de imagen contribuye a una mejor discriminación y, sobre todo, a reducir FP en la mayoría de desenlaces.

Por otro lado, la evaluación en el test independiente temporal evidencia que la combinación clínica+radiómica del parénquima constituye la estrategia más robusta y consistente, alcanzando las AUC más altas tanto en mortalidad ($AUC \approx 0,89$) como en resultado funcional ($AUC \approx 0,87$), con un perfil equilibrado de Sens y Spec (0,88–0,82) y una ganancia notable en Spec para mortalidad. Este comportamiento es relevante si recordamos que el conjunto de pacientes para el desarrollo de modelos y test independiente presentaba diferencias significativas en la gravedad del cuadro clínico de los pacientes y su edad, lo que sugiere que las características derivadas de la parénquima capturan información pronóstica relativamente independiente de la heterogeneidad clínica. En contraste, los modelos clínica+hemorragia

(SMCCH y Experimento 2) ofrecen un rendimiento útil, pero menos estable: en resultado funcional SMCCH tiende a priorizar la Sens a costa de la Spec, mientras que el Experimento 2 muestra un rendimiento intermedio. Este comportamiento sugiere que los modelos basados en radiómica de la hemorragia pueden estar sobreajustados al conjunto de entrenamiento, mientras que los modelos basados en parénquima son capaces de generalizar mejor. Esto resulta fisiopatológicamente coherente, dado que las texturas del parénquima pueden capturar procesos difusos que las texturas de la hemorragia no logran reflejar.

Aunque los modelos combinados mantienen un alto rendimiento en los desenlaces primarios del test independiente, las métricas descienden en los secundarios. En vasoespasma la caída fue de aproximadamente un 9%, y en hidrocefalia de un 12%, lo que probablemente refleja discrepancias en las definiciones de los casos y variaciones entre cohortes. En el vasoespasma esta diferencia ha sido particularmente marcada: todos los casos del test se han establecido por criterios radiológicos, mientras que en el entrenamiento coexistían 96 clínicos y solo 29 radiológicos, generando un desajuste en la representación del fenómeno. A ello se suma la heterogeneidad de los protocolos de imagen entre tipos de escáner utilizado (escáner Philips Brilliance 6 mayoritariamente en el conjunto de entrenamiento y desarrollo de modelos y GE Revolution EVO mayoritariamente utilizado en el conjunto de test independiente), que puede haber amplificado las dificultades de generalización en desenlaces con menor tamaño muestral o mayor variabilidad diagnóstica.

Sin embargo, los resultados concretos de los desenlaces secundarios mantienen cierta coherencia con lo observado en el test de la validación cruzada anidada. En vasoespasma, el rendimiento es moderado, con clínica+parénquima como la aproximación más equilibrada ($AUC = 0,66$, $Sens = 0,62$, $Spec = 0,63$). En hidrocefalia, pese al acusado desbalanceo de clases, clínica+parénquima vuelve a ofrecer el mejor compromiso ($AUC = 0,72$, $Sens = 0,75$, $Spec = 0,69$), mientras que los modelos centrados en hemorragia muestran un comportamiento más inestable.

En conjunto, los resultados de los modelos ML en el test independiente (Figura 4.27) refuerzan la idea de que la radiómica del parénquima, al capturar procesos difusos y heterogéneos de daño cerebral, ofrece un valor pronóstico más consistente que la radiómica de la hemorragia, cuya utilidad parece más sensible a la composición del dataset y a la definición de las cohortes de entrenamiento. Esto pone en valor esta idea novedosa planteada en este trabajo, ya que hasta ahora sólo se habían desarrollado modelos para predecir desenlaces de HIC usando la radiómica de la hemorragia (L. Chen et al., 2025; Shan et al., 2023). En la Tabla 2.8 se aprecia que los estudios previos reportan resultados comparables a los obtenidos en nuestro trabajo ($AUC \geq 0,80$). No obstante, existen diferencias relevantes en la definición del desenlace funcional: mientras que (Dierksen et al., 2024; Huang et al., 2022; Shan et al., 2023) evalúan el resultado al alta o a los tres meses, nuestro estudio se centra en la evolución a los seis meses. De todos ellos, únicamente (Shan et al., 2023) se ha llevado a cabo específicamente en pacientes con HSA, alcanzando un AUC de 0,82 en el test de validación cruzada, valor muy próximo al obtenido en nuestra cohorte cuando se empleó exclusivamente radiómica de la hemorragia ($AUC \approx 0,82$), a pesar del tiempo al desenlace más prolongado.

Aunque los resultados no son estrictamente comparables, ya que los conjuntos de entrenamiento, validación y test difieren entre el problema de segmentación automática de hemorragia y los

modelos predictivos basados en radiómica, la comparación relativa para predecir resultado funcional a los 6 meses resulta ilustrativa. Los modelos basados en ML que integran radiómica de la parénquima y variables clínicas alcanzan en el test independiente un AUC medio de 0,87, con una Sens de 0,88. Por su parte, la regresión logística multivariable que incluye únicamente variables clínicas junto con el volumen hemorrágico del Experimento 2 muestra un rendimiento inferior, con un área bajo la curva de 0,81 y una Sens de 0,80. Este resultado demuestra que los descriptores radiómicos del parénquima tienen un valor predictivo robusto, situándose como posibles biomarcadores emergentes que podrían complementar y enriquecer el volumen hemorrágico, hasta ahora el referente más consolidado en la práctica clínica (Lagares et al., 2015).

Respecto a la configuración de los parámetros de entrenamiento, los modelos que obtienen mejores resultados siguen unos patrones que ponen de relieve varios aspectos metodológicos de interés (Anexo B, Sección B.5). En primer lugar, la decisión de mantener los valores atípicos parece haber sido beneficiosa para preservar la variabilidad real de los datos clínicos y de imagen. De igual forma, el balanceo de clases resulta beneficioso para mitigar los sesgos derivados del desbalanceo de datos en los datasets. En cuanto a los algoritmos, se observa que la distribución de los modelos (ExtraTrees, RF y XGBoost) ha sido relativamente homogénea en los distintos escenarios, lo que refleja que ninguno de los tres algoritmos domina de manera exclusiva en todas las condiciones. Sin embargo, se han identificado ligeras preferencias: RF y ExtraTrees muestran mayor presencia en mortalidad y resultado funcional, mientras que XGBoost se ha utilizado con mayor frecuencia en vasoespasma e hidrocefalia. Esta diversidad sugiere que el comportamiento de los clasificadores depende en gran medida del tipo de desenlace y de la naturaleza de la señal contenida en las variables clínicas y radiómicas.

Además, el hecho de que la mayoría de los modelos hayan seleccionado entre 5 y 20 características sugiere que el valor predictivo se concentra en un subconjunto reducido de variables, lo que facilita su interpretación clínica y sugiere que hay un balance entre capacidad predictiva y riesgo de sobreajuste.

5.2 Implicaciones clínicas

En esta sección se discuten las implicaciones clínicas derivadas de los tres ámbitos de estudio. El objetivo es valorar hasta qué punto los hallazgos pueden trasladarse a la práctica neuroclínica, considerando su utilidad potencial para la toma de decisiones, la planificación terapéutica y la estratificación pronóstica de los pacientes.

En la Sección 5.2.1 se analizan tanto la posible interpretabilidad biológica como las implicaciones clínicas del sistema de clasificación tisular en GBM mediante HSI.

La Sección 5.2.2 aborda las potenciales ventajas clínicas derivadas de la integración de la segmentación automática de HSAa en TC, destacando la viabilidad de emplear el volumen hemorrágico obtenido como biomarcador con mayor capacidad que las escalas clínicas tradicionales.

Finalmente, en la Sección 5.2.3 se examinan la influencia de la edad en los modelos predictivos basados en radiómica y variables clínicas en pacientes con HSAa, así como la interpretabilidad

de dichos modelos y su comparación con escalas clínicas previamente reportadas en la literatura.

5.2.1 Clasificación tisular de glioblastoma en Imagen Hiperespectral

La incorporación de técnicas de HSI a la neurocirugía abre la posibilidad de disponer, en tiempo real, de mapas tisulares objetivos que complementen la apreciación visual del cirujano. En el caso del GBM, contar con herramientas que discriminen entre tejido tumoral, sano y estructuras críticas representaría un avance potencialmente transformador. No obstante, trasladar los resultados experimentales a un contexto clínico real exige considerar no solo la capacidad técnica de los modelos, sino también las particularidades biológicas del tumor y las limitaciones inherentes al propio parénquima cerebral.

Desde un punto de vista clínico, la similitud entre las firmas espectrales del tejido sano y el canceroso podría significar que lo que se etiqueta como cerebro sano podría no ser tejido completamente sano. Los GBM son muy infiltrantes, por lo que las áreas que rodean el tumor pueden presentar alteraciones fisiológicas. La falta de divergencia entre las firmas espectrales puede ser problemática para la generalización de los modelos de clasificación, principalmente porque la información espectral es su principal fuente a la hora de diferenciar un tejido de otro.

Este hallazgo tiene una relevancia clínica directa. La dificultad para separar tumor de parénquima sano plantea que los algoritmos basados en HSI no deberían entenderse únicamente como un mecanismo para delimitar bordes de resección, sino también como una herramienta complementaria que puede alertar al cirujano de zonas “grises” o de transición. En la práctica, estas áreas corresponden al margen infiltrante, donde se concentran células tumorales residuales responsables de la recurrencia precoz. La posibilidad de señalar dichas regiones podría ayudar a optimizar el balance entre extensión de la resección y preservación de tejido funcional, un dilema permanente en la neurocirugía oncológica.

Por otra parte, la capacidad de identificar estructuras vasculares con alta Sens cobra especial importancia clínica. En la cirugía de GBM, donde la resección suele realizarse en campos muy vascularizados, la detección precisa de vasos podría reducir el riesgo de complicaciones hemorrágicas. No obstante, en nuestros experimentos se observa una clara diferencia según el contexto: mientras que en el Experimento A se alcanzan Sens elevadas en la identificación de tejido vascular, en el Experimento B los resultados son más modestos, especialmente en la sangre arterial. Esto indica que, aunque la mera capacidad de reconocer tejido vascular y diferenciarlo de tumor o duramadre representa un avance relevante en términos de seguridad quirúrgica, todavía es necesario reforzar la robustez de los modelos para garantizar un rendimiento consistente en escenarios clínicos reales.

En este estudio se ha demostrado la viabilidad de emplear una cámara HSI de tipo *snapshot*, capaz de adquirir de forma instantánea 25 bandas espectrales en el rango visible e infrarrojo cercano. Esta configuración presenta dos ventajas clave para la práctica quirúrgica: por un lado, permite la captura de imágenes en tiempo real sin necesidad de movimientos de barrido o tiempos de integración prolongados; por otro, posibilita su implementación en quirófano sin

interferir en el flujo operatorio ni comprometer la esterilidad del campo quirúrgico.

Este planteamiento contrasta con otros trabajos publicados en el ámbito de la neurocirugía, donde se han empleado cámaras HSI de tipo *pushbroom* o de escaneo lineal, que requieren movimientos de barrido y tiempos de adquisición más largos, poco compatibles con un entorno intraoperatorio dinámico. Por ejemplo, estudios pioneros como el de (Fabelo, Ortega, Lazcano et al., 2018) utilizaron cámaras pushbroom con 172 y 826 bandas espectrales, obteniendo buenos resultados en la clasificación tisular ($OACC \approx 0,80$) (Fabelo et al., 2019), aunque con procesos de captura más lentos, lo que supone limitaciones prácticas para su integración en quirófano. En contraste, el presente estudio aporta evidencia de que una reducción sustancial en el número de bandas (25 frente a >100) puede ser viable para lograr rendimientos competitivos, al tiempo que facilita de manera decisiva su viabilidad intraoperatoria.

5.2.2 Segmentación automática de Hemorragia Subaracnoidea Aneurismática en TC

La HSAa constituye un escenario especialmente desafiante para la segmentación automática en TC. A diferencia de otras lesiones intracraneales más localizadas, como la HIP, la HSAa se caracteriza por una distribución difusa, irregular y con focos satélite de pequeño tamaño, lo que complica tanto la anotación manual como el entrenamiento de modelos supervisados. La segmentación precisa del sangrado en HSAa no debe entenderse únicamente como un desafío técnico, sino como un elemento clave para avanzar en investigación traslacional. Disponer de volúmenes y distribuciones hemorrágicas obtenidas de forma reproducible permite correlacionarlas con desenlaces clínicos, validar escalas pronósticas existentes y explorar nuevas métricas cuantitativas con valor predictivo.

Desde un punto de vista crítico, los resultados de los modelos de segmentación automática estudiados ponen de relieve tanto el potencial del trabajo como los retos pendientes. Por un lado, la integración de análisis cuantitativo y cualitativo confirma que los modelos propuestos aportan mejoras significativas en aspectos concretos, como la cobertura anatómica y la detección de hemorragia en regiones subaracnoideas, donde la precisión resulta clave. Por otro, persisten limitaciones ligadas a la sobresegmentación o detección de focos pequeños.

La comparación entre los tiempos de segmentación manual y automática muestra una diferencia sustancial en términos de eficiencia y escalabilidad. Mientras que la segmentación manual requiere una media de 15 minutos por caso, con una variabilidad considerable atribuida a la heterogeneidad anatómica y a la extensión de la hemorragia, el método automático es capaz de completar la tarea en menos de 30 segundos de forma prácticamente constante. Esta reducción del 97% en el tiempo necesario no es únicamente un avance técnico, sino que tiene implicaciones directas en la práctica clínica y en la investigación en neurociencia aplicada.

Desde la perspectiva asistencial, disponer de segmentaciones rápidas y reproducibles puede aliviar de forma significativa la carga de trabajo de neurorradiólogos y neurocirujanos. En patologías como la HSAa, donde el tiempo es un recurso crítico, contar en menos de un minuto con una estimación objetiva y cuantitativa del volumen hemorrágico y su distribución podría contribuir a una estratificación pronóstica más precisa y a una planificación terapéutica más ágil. Del mismo modo, la homogeneidad de estas medidas favorece su integración en protocolos

clínicos estandarizados y su uso en la monitorización seriada de pacientes.

En el ámbito de la investigación en neurociencia clínica, la automatización podría transformar un cuello de botella tradicional: la anotación manual. Lo que supondría semanas de trabajo intensivo para generar cohortes segmentadas puede resolverse en horas mediante un modelo de IA, facilitando el desarrollo de bases de datos amplias y heterogéneas. Al proporcionar un criterio uniforme, se reduce de forma sustancial la variabilidad introducida por distintos anotadores y se incrementa la consistencia de los datos, lo que acelera la validación de biomarcadores de imagen y favorece la colaboración entre centros, contribuyendo a la construcción de estudios más sólidos y representativos.

Por último, se ha realizado un análisis comparativo del poder predictivo de los volúmenes hemorrágicos (obtenidos tanto mediante segmentación manual como automática) frente a la escala de Fisher modificada, con el objetivo de valorar su utilidad como biomarcadores pronósticos a seis meses (Tabla 4.17). Dado que la escala de Fisher modificada constituye una referencia habitual en la práctica clínica, la comparación permite explorar el potencial añadido de las métricas volumétricas frente a las aproximaciones cualitativas tradicionales. Los modelos univariados muestran que tanto el volumen manual como el automático son predictores significativos del resultado funcional, con una capacidad discriminativa ($AUC \approx 0,72$) claramente superior a la obtenida con la escala de Fisher modificada ($AUC \approx 0,59$). Este hallazgo muestra una limitación inherente a las escalas cualitativas: la simplificación de la complejidad del daño hemorrágico en categorías amplias puede que solo capturen parcialmente su variabilidad morfológica y pronóstica.

El valor añadido de la cuantificación volumétrica se hace aún más evidente en los modelos multivariados. Ajustando por edad y escala WFNS al ingreso, los volúmenes mantienen una asociación estadísticamente significativa con el resultado funcional, tanto si proceden de segmentación manual como automática. En contraste, la escala de Fisher modificada pierde significación, lo que sugiere que su capacidad explicativa se diluye cuando se incorporan otros predictores. Además, los volúmenes automáticos ofrecen métricas de discriminación y ajuste prácticamente idénticas a las obtenidas con segmentación manual, lo que respalda la validez de los métodos automáticos como posible sustituto en contextos clínicos y de investigación. El análisis de la varianza explicada aporta una conclusión relevante: aunque la escala WFNS es el determinante principal del pronóstico, la incorporación de medidas volumétricas objetivas complementa de manera significativa la información clínica disponible, enriqueciendo la estratificación pronóstica.

Por último, la distribución de los resultados funcionales en función del volumen hemorrágico (Figura 4.24) muestra un patrón clínicamente intuitivo y consistente: volúmenes inferiores a 30 cc se asocian de forma predominante con buen pronóstico, mientras que a partir de este umbral la probabilidad de desenlace desfavorable aumenta de manera marcada. Este hallazgo no solo valida la coherencia del marcador, sino que abre la posibilidad de emplear puntos de corte volumétricos como herramientas de apoyo a la toma de decisiones clínicas, aportando un grado de objetividad y reproducibilidad difícilmente alcanzable con escalas puramente cualitativas.

5.2.3 Modelos predictivos radiómicos en Hemorragia Subaracnoidea Aneurismática

La búsqueda de herramientas pronósticas en la HSAa ha estado tradicionalmente dominada por escalas clínicas y radiológicas semicuantitativas, como WFNS o la escala de Fisher. Sin embargo, estas aproximaciones presentan limitaciones en cuanto a subjetividad, resolución y reproducibilidad. En este contexto, la radiómica se plantea como una alternativa capaz de extraer información cuantitativa y reproducible a partir de las TC iniciales, con el potencial de identificar patrones de daño cerebral que trascienden la simple valoración visual. En este trabajo se ha explorado tanto la radiómica del volumen hemorrágico segmentado como la del parénquima cerebral, lo que permite comparar su poder pronóstico y valorar hasta qué punto cada enfoque refleja distintos mecanismos fisiopatológicos de la enfermedad.

Interpretabilidad y factores que condicionan el rendimiento de los modelos

La Figura B.2 muestra el impacto de la edad en el rendimiento de los modelos combinados clínica+radiómica. Dado que la edad constituye un factor pronóstico bien establecido en la HSAa (Ohkuma et al., 2017; Said et al., 2021), se estratificaron los pacientes con un punto de corte en 70 años. Los resultados muestran que en los individuos más jóvenes los modelos alcanzan AUC más elevadas y consistentes, con IC estrechos. Este hallazgo indica que tanto las características clínicas como las radiómicas conservan mayor valor predictivo en este grupo.

En contraste, en los pacientes de edad avanzada (≥ 70 años) el rendimiento de los modelos se reduce, con métricas medias más bajas y mayor variabilidad en los IC. Este deterioro puede explicarse por la heterogeneidad clínica y anatómica de este grupo, la presencia de comorbilidades y los cambios estructurales cerebrales asociados al envejecimiento (atrofia, dilataciones ventriculares), que pueden interferir con la firma radiómica. El caso de la hidrocefalia resulta paradigmático: en pacientes mayores la mayoría de los modelos muestran un perfil más sensible pero menos específico, con incremento de falsos positivos, posiblemente debido a confundir dilataciones ventriculares por envejecimiento con hallazgos patológicos.

En conjunto, estos hallazgos sugieren que la edad condiciona no solo el pronóstico, sino también la capacidad de los modelos predictivos para capturar patrones clínicos y radiómicos de forma estable. Mientras que en los pacientes jóvenes el valor predictivo es elevado y reproducible, en los mayores se reduce la robustez.

Por otro lado, la exploración de la interpretabilidad mediante valores SHAP en el parénquima cerebral aporta información particularmente valiosa para entender los mecanismos internos de los modelos y sus posibles implicaciones clínicas. A diferencia de la hemorragia, donde la señal radiómica está dominada por un contraste muy intenso y homogéneo que limita la interpretabilidad de la radiómica extraída en esa zona, el parénquima cerebral presenta una mayor variabilidad en términos de textura, densidad y organización tisular. Esta heterogeneidad puede vincularse con fenómenos fisiopatológicos conocidos como la presencia de edema, la isquemia o la pérdida de integridad de la sustancia blanca. En consecuencia, las características radiómicas pueden resultar más interpretables, por lo que se estudió la interpretabilidad de los modelos puramente clínicos, radiómicos del parénquima y combinados. La Figura 4.28 ilustra que, en el modelo clínico, las predicciones se apoyan en un número reducido de variables, con

la escala WFNS al ingreso como principal predictor, lo cual resulta coherente con la práctica clínica y refuerza la validez del modelo (Lagares et al., 2002; Teasdale et al., 1988). Variables adicionales, como la obesidad, glucosa basal o la escala de Fisher modificada, muestran un efecto secundario pero consistente, mientras que otras, como el sexo, ejercen una influencia limitada. Esta configuración confiere al modelo clínico una estructura sencilla y altamente interpretable, variables que además coinciden con estudios anteriores, lo cual refleja la solidez de este estudio (Kawate et al., 2017; Kruyt et al., 2010; Shi et al., 2021).

En contraste, el modelo radiómico se fundamenta en un conjunto más amplio y distribuido de descriptores de textura e intensidad del parénquima (tanto sustancia blanca como gris). No se identifica una única variable determinante, sino que diferentes características GLCM, GLRLM y medidas de primer orden contribuyen conjuntamente a la predicción. Este patrón refleja la capacidad de la radiómica para capturar la heterogeneidad microestructural del cerebro, posiblemente relacionadas con edema cerebral o los patrones hemorrágicos complejos (Connolly Jr et al., 2012; Starke et al., 2009). Sin embargo, estas interpretaciones siguen siendo hipotéticas y requieren una validación adicional.

El modelo combinado representa un escenario intermedio y probablemente el más útil desde el punto de vista clínico. La escala WFNS y la glucosa basal se mantienen como predictores relevantes, pero se ven complementados por un conjunto de variables radiómicas que aportan matices adicionales. En este sentido, los resultados sugieren que la radiómica del parénquima no sustituye a los predictores clásicos, sino que los refuerza y afina, permitiendo una estratificación más precisa. Este equilibrio puede ser clave para la traslación de los modelos a la práctica clínica, ya que favorece tanto la confianza del clínico como la capacidad del algoritmo para identificar patrones sutiles con valor pronóstico.

Comparación con escalas clínicas tradicionales

La comparación con escalas clínicas basadas en hallazgos radiológicos, ampliamente reconocidas como predictores pronósticos en la literatura, permite valorar de forma más precisa las fortalezas y limitaciones de los modelos radiómicos frente a estas referencias clínicas consolidadas. Primero, a partir de la *ictWFNS* en el conjunto de test externo se desarrollan modelos de regresión logística donde en los desenlaces principales (mortalidad y resultado funcional), la escala alcanza valores de AUC en torno a 0,70. Su Spec es alta en mortalidad (0,95), lo que confirma que la *ictWFNS* es útil para identificar con fiabilidad a los pacientes con bajo riesgo de fallecimiento. Sin embargo, la Sens se mantiene claramente limitada (0,46), lo que se traduce en un número importante de FN y en una escasa capacidad para detectar a aquellos pacientes que evolucionarán hacia un mal desenlace. Estas limitaciones se amplifican en los desenlaces secundarios. En vasoespasma, la capacidad discriminativa apenas alcanza un AUC de 0,60, y en hidrocefalia desciende a 0,56, con una Sens nula.

En contraste, los modelos radiómicos del parénquima cerebral demuestran un rendimiento significativamente superior en todos los desenlaces evaluados. Para mortalidad y resultado funcional, alcanzan AUC de 0,89 y 0,87 respectivamente, con Sens y Spec equilibradas en torno al 80–90 %, lo que refleja una mayor utilidad práctica para la clasificación individual. Incluso en escenarios de mayor complejidad, como vasoespasma (AUC 0,66) e hidrocefalia (AUC 0,72), el rendimiento es claramente más favorable que el de la *ictWFNS*.

No obstante, estos resultados deben interpretarse con cautela, ya que el tamaño de la cohorte de prueba externa ha sido reducido ($n = 41$), lo que limita la potencia estadística y la precisión de las estimaciones. Por ello, resulta imprescindible validar estos hallazgos en poblaciones más amplias y multicéntricas, que permitan confirmar la generalización de los modelos y su utilidad en diferentes entornos clínicos.

Para contextualizar el rendimiento de nuestros modelos, la Tabla 5.2 presenta una comparación directa con lo reportado en estudios previos usando escalas clínicas tradicionales, como WFNS, su versión modificada (*hWFNS*), la escala de Fisher y una puntuación que mide la extensión de la HIV. En este análisis comparativo se observa que los modelos radiómicos aplicados al parénquima cerebral alcanzan rendimientos predictivos de moderados a altos, superando en la mayoría de los casos a los sistemas clásicos de estratificación.

Tabla 5.2: Rendimiento comparativo de los modelos radiómicos frente a escalas clínicas tradicionales en la predicción de resultados en HSAa

Desenlace	Métrica	Radiómica	WFNS/ Fisher	hWFNS/ Punt. HIV	Referencia
Mortalidad	Sens/Spec	0,75 / 0,76 (6 meses)	WFNS 0,89 / 0,19 (6 meses)	hWFNS 0,44 / 0,93 (6 meses)	(Raabe et al., 2022)
Resultado funcional	AUC	0,85 (6 meses)	WFNS ~0,837 (3 meses)	WFNS mod. ~0,839 (3 meses)	(Hofmann et al., 2022 ; Nguyen et al., 2023)
Vasoespasmio	AUC	0,75 (6 meses)	Fisher/ Fisher mod. ~0,65–0,70 (3 meses)	–	(Couret et al., 2024)
Hidrocefalia	AUC	0,86 (6 meses)	Fisher mod. ~0,81 (1 mes)	Punt. HIV ~0,85 (3 meses)	(Couret et al., 2024 ; D. Rao et al., 2024)

En la predicción de mortalidad a 6 meses, nuestros modelos radiómicos muestran un perfil equilibrado (Sens 0,75, Spec 0,76). En contraste, la WFNS tradicional se caracteriza por una elevada Sens (0,89) pero muy baja Spec (0,19), mientras que la *hWFNS* mejora la Spec (0,93) a costa de una Sens reducida (0,44) (Raabe et al., [2022](#)).

En cuanto al resultado funcional, los modelos presentan un AUC de 0,85, alcanzando niveles similares las escalas clínicas (WFNS AUC ~0,837; modWFNS AUC ~0,839) (Hofmann et al., [2022](#); Nguyen et al., [2023](#)).

Respecto a las complicaciones secundarias, la predicción de vasoespasmó alcanza un AUC de 0,75, superior al rango descrito para la escala de Fisher y su versión modificada (AUC 0,65–0,70) (Couret et al., 2024). En hidrocefalia, los modelos radiómicos obtienen un AUC de 0,86, comparable a los valores alcanzados por la puntuación de HIV (AUC 0,85) y el Fisher modificado (AUC 0,81) (Couret et al., 2024; D. Rao et al., 2024).

Un aspecto clave a destacar es que, en estas últimas comparaciones, los modelos de la literatura para predecir resultado funcional y vasoespasmó son a los 3 meses e hidrocefalia al mes. Nuestros modelos desarrollados con radiómica usan ventanas de predicción más amplias (6 meses). Esto incrementa su valor potencial como soporte a la toma de decisiones clínicas a largo plazo, permitiendo ajustar tratamientos y estrategias de monitorización a lo largo del tiempo.

Por último, se ha evaluado la capacidad predictiva de los modelos radiómicos frente a la escala SEBES y SEBES6c (Tablas B.18 - B.21). Nuestros modelos radiómicos demuestran un rendimiento superior de forma consistente frente a SEBES, tanto evaluados de manera independiente como en combinación con variables clínicas, en todos los desenlaces analizados. Este hallazgo confirma que la radiómica es capaz de extraer información más compleja y detallada que esta escala basada en la inspección visual.

En el análisis correlacional, observamos una relación negativa entre las variables radiómicas y la puntuación SEBES y SEBES6c: una menor definición en los vóxeles de la sustancia gris se correlaciona con puntuaciones más elevadas en la escala, lo que probablemente corresponde al borramiento de surcos identificado visualmente por el observador (Tabla 4.21). No obstante, la correlación es solo moderada ($\approx -0,3$). Tanto en SEBES como en SEBES6c, las características radiómicas con mayor correlación han resultado muy similares, lo que refuerza la consistencia de estos descriptores como posibles marcadores objetivos de la severidad del edema cerebral. Sin embargo, se muestra un hallazgo de especial interés: SEBES6c se asocia de manera independiente con la aparición de hidrocefalia, incluso en los modelos multivariados incluyendo otras variables clínicas como WFNS. Hasta la fecha, solo (Fang et al., 2019) habían planteado una relación entre el edema cerebral precoz y la hidrocefalia, mediada por biomarcadores inflamatorios en el LCR. No obstante, esta asociación podría estar parcialmente condicionada por la mayor proporción de casos graves de HSAa en los grupos con SEBES6c elevado. Por ello, SEBES6c debe interpretarse con cautela como marcador temprano de hidrocefalia.

Con el objetivo de explorar hasta qué punto la radiómica puede reproducir la valoración visual del clínico, se ha desarrollado un “SEBES radiómico”, un concepto novedoso que no había sido previamente descrito en la literatura. Este enfoque consiste en un modelo de regresión logística regularizado mediante LASSO, entrenado a partir de los descriptores radiómicos más informativos, con la finalidad de replicar la puntuación de la escala visual SEBES. El modelo reproduce la puntuación con elevada fidelidad (AUC = 0,89) usando 35 características radiómicas (Figura 4.29), demostrando que la radiómica puede capturar objetivamente el borramiento de surcos y la pérdida de definición cortical que aprecia el ojo humano. Este resultado confirma que la radiómica puede emular con éxito la escala visual, ofreciendo la posibilidad de estandarizar la interpretación radiológica y agilizar su aplicación, especialmente en entornos con menor experiencia.

Sin embargo, el análisis comparativo de las variables seleccionadas para el “SEBES radiómico” frente a aquellas identificadas como relevantes por SHAP en los modelos predictivos específicos (mortalidad, resultado funcional, hidrocefalia, vasoespasma) revela una coincidencia mínima. Ninguna característica se comparte con el resultado funcional o mortalidad; solo una coincide en hidrocefalia (`original_shape_Elongation_GREY`) y otra distinta en vasoespasma (`wavelet-LLH_glcm_MaximumProbability_GREY`).

Asimismo, cuando el “SEBES radiómico” se aplica directamente a la predicción de desenlaces clínicos, su rendimiento cae de manera marcada ($AUC \leq 0,67$ para GOS; $\leq 0,61$ para vasoespasma; $0,49$ para hidrocefalia). Esto demuestra que las características relacionadas con el edema no resultan suficientes por sí solas para anticipar el pronóstico funcional ni las principales complicaciones tras la HSAa. Por el contrario, los modelos radiómicos entrenados de manera específica para cada desenlace ofrecen un rendimiento sustancialmente superior, confirmando que los descriptores asociados al edema no constituyen predictores clínicos precisos, mientras que la radiómica dirigida a desenlaces concretos sí logra capturar la señal relevante.

En conjunto, el análisis con escalas clínicas subraya que, aunque estas escalas podrían mantener un papel clínico relevante como herramienta de referencia (especialmente por su sencillez y aplicabilidad inmediata), en algunos casos (como el SEBES, SEBES6c o ictWFNS) sus limitaciones son evidentes en el ámbito de la predicción individualizada y en la anticipación de complicaciones secundarias. Los modelos radiómicos, por su parte, muestran un mayor poder discriminativo, lo que respalda su integración como complemento a las escalas clínicas tradicionales. Este enfoque combinado podría ofrecer la doble ventaja de preservar la interpretabilidad clínica y, al mismo tiempo, mejorar la capacidad de estratificación y predicción en la práctica real.

Un aspecto adicional que merece ser destacado es el compromiso temporal asociado al proceso de extracción de características radiómicas según la región de interés analizada. En el caso de la radiómica aplicada al parénquima cerebral, el procedimiento resulta considerablemente más costoso en términos computacionales y de preprocesamiento, debido a la necesidad de segmentar volúmenes amplios y heterogéneos, lo que incrementa tanto el tiempo de ejecución como la complejidad técnica. Por el contrario, la radiómica centrada en el sangrado intracraneal se beneficia de un volumen de interés más acotado, lo que permite una extracción de descriptores más rápida y eficiente.

A partir de estos resultados, conviene plantearse con mayor claridad el papel de la radiómica como biomarcador en la HSAa. La comparación entre modelos clínicos y radiómicos ha mostrado un rendimiento predictivo similar, lo que pone de relieve tanto las fortalezas como las limitaciones inherentes a cada enfoque. Las escalas clínicas clásicas, como WFNS o Fisher, se basan en décadas de experiencia asistencial y mantienen su vigencia como herramientas rápidas, sencillas y ampliamente utilizadas en la práctica rutinaria. Sin embargo, su simplicidad es también su principal limitación: se basan en categorías discretas, sujetas a variabilidad interobservador y poco sensibles a la complejidad del sangrado o al grado de afectación cerebral.

La radiómica, en cambio, ofrece un enfoque cuantitativo, reproducible y potencialmente auto-

matizable, con la capacidad de identificar patrones invisibles al ojo humano. Esta característica resulta especialmente valiosa en escenarios donde la información clínica es incompleta o poco fiable, como en pacientes intubados, en contextos de urgencia o en estudios retrospectivos multicéntricos. Su naturaleza objetiva y escalable la convierte en una posible alternativa para sistemas de apoyo a la decisión clínica basados en IA.

5.3 Limitaciones

Finalmente, se abordan de manera sistemática las limitaciones metodológicas identificadas en cada uno de los tres experimentos que se han llevado a cabo. Primero, en la Sección 5.3.1 se discuten las limitaciones encontradas en la clasificación tisular en GBM mediante HSI; en la Sección 5.3.2 las limitaciones de los modelos de segmentación automática de HSAa en TC; y por último, en la Sección 5.2.3 se discuten las limitaciones de los modelos predictivos basados en radiómica. Este apartado busca examinar de forma crítica el alcance de los resultados, identificando las limitaciones y consideraciones que definen su aplicabilidad.

5.3.1 Clasificación tisular de glioblastoma en Imagen Hiperespectral

Como se ha comentado en secciones anteriores, este trabajo debe entenderse como una *prueba de concepto* desarrollada en un entorno retrospectivo. El objetivo principal ha sido demostrar la viabilidad técnica de emplear cámaras HSI *snapshot* en neurocirugía y explorar si, a partir de estas capturas, es posible entrenar modelos de clasificación tisular. Sin embargo, este planteamiento inicial conlleva una serie de limitaciones que condicionan tanto la interpretación de los resultados como su potencial traslación clínica.

En primer lugar, el tamaño muestral es reducido y los datos proceden de un único centro, lo que limita la generalización de los hallazgos. Además, la base de datos de *ground truths* se ha construido a partir de anotaciones manuales procedentes de distintos neurocirujanos, sujetas a la variabilidad del observador. También debe destacarse el fuerte desbalance entre clases: mientras que la duramadre y el parénquima cerebral cuentan con un número abundante de píxeles, las clases vasculares apenas están representadas, lo que condiciona tanto la estabilidad de las firmas espectrales como el rendimiento de los algoritmos.

Desde el punto de vista técnico, la cámara utilizada captura únicamente 25 bandas en el rango visible e infrarrojo cercano. Aunque este diseño *snapshot* tiene claras ventajas operativas (rapidez, compatibilidad con la dinámica quirúrgica, no interferencia en el campo estéril), también supone una restricción frente a otros sistemas con mayor cobertura espectral, capaces de captar información fisiológica adicional en bandas más largas. Por otra parte, las adquisiciones se han realizado en condiciones variables de iluminación, presencia de sangre o irrigación, lo que puede afectar la reflectancia registrada y añadir variabilidad no controlada.

En cuanto a los modelos empleados, los resultados del Experimento A muestran rendimientos muy elevados, pero con un riesgo evidente de sobreajuste, dado el número limitado de pacientes y la gran cantidad de características espectrales disponibles. Cuando los clasificadores se evalúan en pacientes no vistos (Experimento B), el rendimiento desciende de forma notable y

la variabilidad interpaciente se hace mucho más evidente. Esto refleja tanto la dificultad de discriminar entre tumor y tejido sano (debido a su gran similitud espectral y a la infiltración característica del GBM) como la falta de datos suficientes para entrenar modelos más robustos.

Otra limitación reside en que el estudio no ha abordado la viabilidad en tiempo real. Aunque la cámara permite capturas instantáneas, no se ha medido de manera sistemática la latencia completa del proceso (adquisición, preprocesado, inferencia y visualización). Tampoco se ha evaluado la integración práctica en el flujo quirúrgico, por ejemplo mediante superposición en el microscopio o interacción directa con el cirujano.

En conjunto, este trabajo demuestra la viabilidad técnica del enfoque y establece las bases metodológicas para el desarrollo futuro de sistemas de clasificación hiperespectral intraoperatoria. Los resultados obtenidos confirman que las cámaras HSI *snapshot* pueden integrarse en el entorno quirúrgico y generar información cuantitativa útil para la discriminación tisular. Aunque las limitaciones identificadas en este estudio inicial requieren de investigaciones posteriores con cohortes más amplias y validación externa, los hallazgos obtenidos representan un avance significativo hacia la aplicación clínica de esta tecnología, abriendo nuevas posibilidades para mejorar la precisión y objetividad en la cirugía neuroquirúrgica. En este sentido, resulta alentador constatar que la HSI está experimentando una adopción creciente en la comunidad científica, como lo evidencian los trabajos presentados en conferencias internacionales de referencia como MICCAI, donde se están desarrollando enfoques avanzados de segmentación y clasificación hiperespectral para aplicaciones médicas (Roddan et al., 2025), lo que confirma el interés y potencial de esta tecnología en el ámbito de la imagen médica.

5.3.2 Segmentación automática de Hemorragia Subaracnoidea Aneurismática en TC

En este trabajo se alcanza un rendimiento satisfactorio, donde tanto cuantitativamente como cualitativamente se logran rendimientos equiparables o superiores a otros modelos de segmentación automática de hemorragia cerebral desarrollados en la literatura.

No obstante, este trabajo presenta varias limitaciones que conviene señalar. La primera se relaciona con la necesidad de una validación externa en cohortes independientes de HSAa, requisito indispensable para confirmar la capacidad de generalización de los modelos más allá de los datos aquí analizados. Del mismo modo, la aplicabilidad a otros subtipos de HIC (HSD o HED) permanece sin explorar, lo que limita el alcance inmediato a otras patologías.

En segundo lugar, no se evalúa de manera sistemática el impacto de los artefactos de imagen (como los derivados de implantes dentales, material de clipaje o movimientos del paciente) sobre la precisión de la segmentación y el rendimiento predictivo. Estos factores, frecuentes en la práctica clínica, podrían condicionar la robustez de los modelos en escenarios reales.

Asimismo, persisten limitaciones intrínsecas a la segmentación de pequeños focos hemorrágicos o regiones de bajo contraste, donde los FP y FN impactan de manera desproporcionada en las métricas globales. Aunque su relevancia clínica es distinta de la masa hemorrágica principal, estos errores pueden limitar la utilidad de los modelos en escenarios de monitorización fina o en estudios fisiopatológicos donde las pequeñas hemorragias pueden tener interés específico.

Conviene señalar también que los modelos no se han evaluado en sujetos sanos o en pacientes sin hemorragia, por lo que podrían aparecer FP en este contexto, un aspecto que debería investigarse en futuros estudios para garantizar la especificidad del sistema.

Desde un punto de vista metodológico, la heterogeneidad de los protocolos de adquisición entre centros (distintas marcas y modelos de escáner, variaciones en el grosor de corte, tamaño del píxel) introduce variabilidad que puede afectar a la segmentación.

A nivel de integración clínica, los modelos se han evaluado de forma retrospectiva y *offline*, sin incluir pruebas prospectivas ni en tiempo real. Esto limita la capacidad de valorar su utilidad inmediata en un flujo asistencial donde factores como el tiempo de procesamiento, la interoperabilidad con sistemas hospitalarios o la aceptación del usuario final resultan determinantes.

En conjunto, estas limitaciones ponen de relieve que, aunque los resultados obtenidos son prometedores, la traslación clínica de los modelos de segmentación automática requerirá validar su robustez frente a cohortes externas, artefactos de imagen y escenarios en tiempo real.

5.3.3 Modelos predictivos radiómicos en Hemorragia Subaracnoidea Aneurismática

En este estudio se han desarrollado modelos predictivos basados en radiómica extraída de ROIs cerebrales que alcanzan métricas prometedoras en la predicción de distintos desenlaces clínicos de la HSAa, que en muchos casos mejoran las escalas clínicas tradicionales hasta ahora utilizados en la práctica clínica e investigación en neurociencia.

Sin embargo, el estudio presenta varias limitaciones que deben ser consideradas al interpretar los hallazgos. En primer lugar, su diseño retrospectivo y unicéntrico introduce un riesgo inherente de sesgo de selección y limita la generalización de los resultados a otros entornos clínicos. El tamaño muestral junto con el notable desequilibrio de clases en algunos desenlaces, compromete la potencia estadística y puede afectar a la calibración de los modelos, incluso a pesar de haber aplicado técnicas de sobremuestreo y validación cruzada. Además, la definición heterogénea del desenlace del vasoespasma introduce variabilidad adicional difícil de controlar.

Por otro lado, la alta dimensionalidad de los datos radiómicos conlleva un riesgo de sobreajuste. Para evitar este riesgo de sobreajuste se ha utilizado MRMR, que ha permitido seleccionar un subconjunto óptimo de predictores con alta relevancia y baja redundancia. Aun así, sería conveniente verificar la consistencia de este proceso frente a diferentes configuraciones o cohortes, a fin de garantizar su robustez ante variaciones en los datos. De manera adicional, la cohorte de test independiente ha sido limitada en términos absolutos y particularmente reducida en ciertos subgrupos clínicamente relevantes, como los pacientes con vasoespasma clínico, lo que restringe la precisión de las estimaciones de rendimiento.

En relación con la segmentación, se han empleado volúmenes derivados de segmentaciones automáticas, lo cual introduce sus propias fuentes de error. En el caso de SMCCH, las predicciones han tendido a sobrestimar el volumen hemorrágico (más de 20 mL de media), mientras que en el Experimento 2, aunque mejor valorado por los clínicos, no puede descartarse

la presencia de fallos en focos pequeños o de bajo contraste. Estas imprecisiones pueden repercutir directamente en la calidad de los descriptores radiómicos extraídos y, por tanto, en el rendimiento de los modelos predictivos.

Pese a los resultados prometedores expuestos, la radiómica aún enfrenta retos considerables para consolidarse como biomarcador de uso clínico. La dependencia de múltiples pasos técnicos (segmentación, normalización, extracción de descriptores), la susceptibilidad a la heterogeneidad en la adquisición y el riesgo de sobreajuste derivados del elevado número de variables en relación con el tamaño de las cohortes, limitan su generalización. Por este motivo, la equivalencia observada con los modelos clínicos no debe interpretarse como una sustitución inmediata, sino más bien como evidencia de que ambas aproximaciones pueden capturar información complementaria. De hecho, en la mayoría de los desenlaces la combinación de información clínica y radiómica ha mostrado un efecto aditivo, lo que sugiere que la integración de ambos mundos representa una estrategia más sólida y realista.

Capítulo 6

Conclusiones y líneas futuras

El hilo conductor de esta tesis se centra en explorar hasta qué punto los modelos supervisados pueden contribuir a una caracterización más objetiva del cerebro y de las lesiones que lo afectan en el contexto de la neurociencia clínica, utilizando para ello distintas modalidades de imagen con finalidades complementarias: discriminar tejidos intraoperatoriamente, segmentar lesiones agudas en TC y anticipar la evolución clínica mediante radiómica. Considerados de manera conjunta, estos experimentos muestran que no se trata de problemas aislados, sino de distintas manifestaciones de un mismo reto: traducir señales heterogéneas y complejas del cerebro en biomarcadores cuantitativos estandarizados, reproducibles y clínicamente útiles.

El valor de esta aproximación común reside en tres aspectos. Primero, en evidenciar que la automatización puede aliviar cuellos de botella clásicos de la práctica neuroclínica y de la investigación traslacional (como los tiempos prolongados de anotación, la dependencia del observador o la subjetividad de las escalas), generando información en plazos compatibles con la urgencia y la toma de decisiones en patologías cerebrales críticas. Segundo, en mostrar que el rendimiento técnico de un modelo no puede evaluarse al margen de su contexto de uso: el impacto de los errores sobre la preservación de tejido sano o la identificación de lesiones, la heterogeneidad de los protocolos de adquisición o la definición de los desenlaces clínicos condicionan tanto o más que las métricas globales. Y tercero, en subrayar que la integración de fuentes (imagen multimodal, radiómica y clínica) resulta más prometedora que la búsqueda de un sustituto único de las escalas existentes, ya que permite capturar dimensiones complementarias de los procesos cerebrales, desde la infiltración tumoral hasta el edema o la isquemia secundaria.

Desde esta perspectiva, este trabajo no solo aporta modelos y resultados comparativos, sino también contribuciones aplicables a la neurociencia clínica, que sientan bases para pensar en cómo deben desarrollarse, evaluarse y validarse estas herramientas si aspiran a ser útiles más allá del laboratorio. El objetivo común es, en definitiva, avanzar hacia sistemas que equilibren precisión técnica, robustez metodológica y relevancia clínica, y que lo hagan bajo condiciones que reflejen la complejidad intrínseca del cerebro humano y la variabilidad inherente a su enfermedad.

En esta sección se presentan, en primer lugar, las conclusiones principales del trabajo,

organizadas en función de los objetivos planteados en la Sección 1.2. Además, se resalta su contribución al avance del conocimiento en el campo de la caracterización automática del tejido cerebral y las lesiones asociadas mediante imagen médica multimodal.

A continuación, se exponen las líneas futuras de investigación que se desprenden de los hallazgos obtenidos, estructuradas de manera paralela a los diferentes experimentos desarrollados.

Finalmente, se incluye una subsección dedicada al **impacto científico**, donde se sintetizan las principales aportaciones derivadas de esta tesis en términos de publicaciones, presentaciones en congresos y colaboraciones en proyectos de investigación.

6.1 Conclusiones

El objetivo principal de esta tesis doctoral consiste en desarrollar y validar un sistema automatizado basado en técnicas avanzadas de procesamiento de imágenes médicas y algoritmos de aprendizaje automático, con el fin de mejorar la precisión diagnóstica, predecir la evolución clínica y apoyar la toma de decisiones en el ámbito de la neurociencia clínica. Este propósito se ha alcanzado en gran medida a través de los distintos objetivos específicos planteados.

Las conclusiones presentadas en esta sección se estructuran siguiendo la organización de los objetivos específicos definidos en la Sección 1.2, lo que permite establecer una correspondencia directa entre los resultados obtenidos y los propósitos inicialmente planteados. En primer lugar, el **OBJ-S1** aborda la clasificación y segmentación automática de lesiones cerebrales, incluyendo la clasificación tisular intraoperatoria con imagen hiperespectral (OBJ-S1.1) y la segmentación de hemorragias en TC (OBJ-S1.2). El **OBJ-S2** se centra en la integración de biomarcadores avanzados en modelos predictivos, mediante el desarrollo de firmas espectrales (OBJ-S2.1) y la extracción de características radiómicas para la predicción de desenlaces clínicos (OBJ-S2.2). Finalmente, el **OBJ-S3** aborda la validación clínica del sistema integrado, evaluando la aplicabilidad de la imagen hiperespectral en cirugía de glioblastoma y el valor pronóstico de los biomarcadores en hemorragia subaracnoidea aneurismática.

6.1.1 OBJ-S1: Clasificación y segmentación automática de lesiones cerebrales

En relación con el **OBJ-S1**, se abordan dos ámbitos principales: la clasificación tisular intraoperatoria (OBJ-S1.1) y la segmentación automática de hemorragias cerebrales en TC (OBJ-S1.2).

Clasificación tisular intraoperatoria mediante HSI (OBJ-S1.1)

Los **Experimentos A y B** demuestran la viabilidad técnica y metodológica de aplicar técnicas de aprendizaje automático a imágenes hiperespectrales adquiridas con cámaras *snapshot* de 25 bandas en un entorno quirúrgico real. En el escenario intrapaciente (Experimento A), que simula condiciones ideales de clasificación tisular, los modelos basados en Random Forest (RF) y redes neuronales convolucionales (CNN) alcanzan OACC superiores al 95 % y errores de clasificación por debajo del 10 %, validando así su capacidad de discriminar múltiples clases tisulares (tumor, tejido sano, vasos, duramadre) en un entorno controlado con variabilidad

espectral mínima.

En contraste, el escenario más exigente y clínicamente relevante de clasificación interpaciente (Experimento B), que evalúa la capacidad de generalización entre sujetos distintos, muestra un descenso en el rendimiento (OACC $\sim 60\%$), evidenciando la dificultad intrínseca de trasladar patrones espectrales entre individuos con variaciones en las condiciones de adquisición, la anatomía o la composición tisular. Aun así, los resultados obtenidos confirman la viabilidad del enfoque en un contexto realista de variabilidad espectral y establecen una línea base sólida sobre la cual seguir optimizando la metodología.

En comparación con trabajos previos que emplean cámaras hiperespectrales de 128–826 bandas, el presente estudio demuestra que, al normalizar el rendimiento por número de bandas, las 25 bandas utilizadas alcanzan una eficiencia espectral superior, lo que refuerza el potencial de los sistemas HSI de menor complejidad para aplicaciones clínicas.

Hallazgo principal: Las cámaras HSI *snapshot* de 25 bandas logran rendimientos superiores al 95 % en clasificación intrapaciente y del 60 % en clasificación interpaciente. Su elevada eficiencia espectral, superior a la de sistemas de 128–826 bandas, respalda el desarrollo de soluciones HSI más portátiles, rápidas y escalables para entornos quirúrgicos en tiempo real.

Segmentación automática de hemorragias cerebrales en TC (OBJ-S1.2)

En lo relativo a la segmentación de hemorragias en TC, los modelos **nnU-Net** del **Experimento 1**, entrenados exclusivamente con cohortes locales que incluyen casos de HSA y hemorragias secundarias (HIV e HIP), alcanzan una mediana de Dice de 0,75 y un ASD mediano de 1,50,mm, generando contornos precisos y espacialmente coherentes.

El **Experimento 2** amplía el entrenamiento mediante la incorporación de dos bases de datos públicas con mayor diversidad de lesiones (HSA, HIV, HIP, HED y HSD). En la evaluación sobre el mismo conjunto que el Experimento 1, se observa un ligero aumento en la métrica de distancia (ASD mediano 1,63,mm), junto con una mejora en los valores de Dice (0,81 de mediana), lo que sugiere una mejor adaptación a distintos patrones hemorrágicos.

Ambos modelos funcionan correctamente tanto en la cohorte institucional, con casos de HSA, HIP y HIV, como en una cohorte externa que incluye lesiones del mismo tipo pero con mayor variabilidad en la procedencia de los datos. En términos generales, el rendimiento es algo superior en la cohorte externa, probablemente debido a que los casos institucionales de HSAa presentan una mayor complejidad clínica y morfológica.

Los modelos desarrollados en este trabajo muestran un rendimiento superior, en términos de valores medianos, frente a cinco modelos de segmentación de hemorragia reportados previamente en la literatura. Esta solidez técnica cuantitativa es corroborada por una evaluación cualitativa exhaustiva realizada por neurocirujanos con experiencia en evaluación de segmentaciones, quienes, tras una revisión doble ciega y aleatoria de los siete modelos comparados (incluyendo los dos propuestos y cinco de la literatura), manifiestan una preferencia global significativa por el modelo correspondiente al Experimento 2, destacando especialmente su capacidad para detectar hemorragias en regiones subaracnoideas y mantener coherencia anatómica.

Hallazgo principal: Los modelos desarrollados segmentación automática de hemorragia cerebral para HSAa alcanzan un equilibrio entre robustez y fidelidad espacial, superando a modelos previos de la literatura a nivel cuantitativo y cualitativo. Los tiempos de inferencia se mantienen por debajo de 30 s en CPU, representando una reducción aproximada del 97 % respecto al tiempo requerido para segmentaciones manuales.

6.1.2 OBJ-S2: Integración de biomarcadores avanzados en modelos predictivos

El **OBJ-S2** se centra en la integración de biomarcadores avanzados en modelos de apoyo diagnóstico y pronóstico. Este objetivo se desarrolla en dos líneas principales: por un lado, la caracterización de firmas espectrales en imagen hiperespectral para clasificación tisular (OBJ-S2.1), y por otro, la extracción de características radiómicas en hemorragia subaracnoidea aneurismática para la predicción de desenlaces clínicos (OBJ-S2.2).

Firmas espectrales en HSI para clasificación tisular (OBJ-S2.1)

En el ámbito de HSI (**OBJ-S2.1**) se demuestra que las firmas espectrales de 25 bandas, cubriendo la transición entre el extremo rojo del visible (620–740 nm) y el VNIR (740–970 nm), pueden aportar información discriminativa relevante entre múltiples tipos de tejido (tumor, tejido sano, vasos, duramadre). Este hallazgo confirma la viabilidad de emplear dispositivos *snapshot* de adquisición rápida en entornos quirúrgicos, al tiempo que establece el potencial de la información espectral como biomarcador cuantitativo.

Hallazgo principal: Las firmas espectrales de 25 bandas en el rango visible-VNIR permiten capturar información discriminativa entre múltiples tipos de tejido cerebral. Este hallazgo establece las bases para el uso de la información espectral como biomarcador cuantitativo y objetivo en neurocirugía.

Características radiómicas en HSAa para predicción de desenlaces (OBJ-S2.2)

En el caso de la HSAa (**OBJ-S2.2**), se extraen descriptores radiómicos estandarizados (según la guía IBSI) de las regiones hemorrágicas y del parénquima cerebral adyacente, incluyendo características de primer a tercer orden. El análisis de correlación con los desenlaces de primarios (mortalidad, resultado funcional (GOS) a los seis meses), y secundarios (vasoespasmos e hidrocefalia) muestra asociaciones significativas con ambas regiones, apoyando su valor como biomarcadores complementarios con potencial pronóstico independiente.

Se comparan cuatro estrategias de modelización mediante validación cruzada anidada: (1) modelos basados en variables clínicas y de laboratorio, (2) modelos radiómicos de la hemorragia, (3) modelos radiómicos del parénquima y (4) modelos combinados de radiómica y clínica. Los resultados indican que los modelos puramente clínicos y los radiómicos alcanzan rendimientos similares, confirmando la relevancia de la radiómica como fuente predictiva independiente. No obstante, la integración de variables clínicas y radiómicas mejora el desempeño, especialmente en especificidad, reduciendo falsos positivos y evidenciando una sinergia entre ambas fuentes de información.

En la validación independiente, los modelos que combinan la radiómica del parénquima

con variables clínicas obtienen los resultados más consistentes (AUC medio de 0.89 para mortalidad y 0.87 para resultado funcional), superando tanto a las escalas tradicionales como a los modelos basados solo en la hemorragia. Esto sugiere que el parénquima cerebral aporta información pronóstica adicional, posiblemente relacionada con edema, daño difuso o alteraciones microestructurales tempranas.

Para los desenlaces secundarios, más heterogéneos, se alcanzan valores moderados pero clínicamente relevantes (AUC = 0.66 en vasoespasmo y 0.72 en hidrocefalia), en concordancia con lo descrito en la literatura.

Finalmente, los análisis de interpretabilidad mediante SHAP y correlación con escalas clínicas permiten identificar las variables más influyentes, destacando la WFNS al ingreso junto con descriptores texturales del parénquima, lo que refuerza la coherencia clínica y el valor complementario de los biomarcadores radiómicos.

Metodológicamente, este trabajo busca aportar un mayor rigor respecto a estudios previos mediante la inclusión de radiómica del parénquima junto con la hemorragia segmentada por dos algoritmos distintos, la aplicación de validación cruzada anidada con múltiples semillas, la comparación sistemática de cuatro estrategias de modelización y la evaluación del rendimiento en un test independiente con distinta temporalidad.

Hallazgo principal: La radiómica del parénquima cerebral, en combinación con variables clínicas, alcanza rendimientos superiores (AUC 0,89 para mortalidad, AUC 0,87 para resultado funcional a los seis meses) a escalas clínicas tradicionales. La combinación de información clínica y radiómica reduce los falsos positivos y mejora la media de rendimiento, demostrando complementariedad entre ambas fuentes de información.

6.1.3 OBJ-S3: Validación clínica del sistema integrado

El **OBJ-S3**, orientado a la validación clínica del sistema propuesto, integra de manera conjunta los resultados de los tres ámbitos metodológicos desarrollados: la clasificación tisular mediante HSI, la segmentación automática de hemorragias y el modelado pronóstico radiómico. El objetivo trasciende la simple comparación de métricas y busca demostrar la aplicabilidad, robustez y valor clínico de estas aproximaciones en entornos reales y heterogéneos, validando su potencial traslacional. Esta validación se articula en tres ejes: la aplicabilidad de la HSI en cirugía de GBM, el valor pronóstico del volumen hemorrágico en HSAA y la caracterización de biomarcadores radiómicos para la predicción de desenlaces clínicos.

Aplicabilidad clínica de la HSI en cirugía de GBM

En el contexto de la cirugía de GBM, los resultados obtenidos muestran que la adquisición con cámaras hiperespectrales (*HSI*) de tipo *snapshot* permite generar mapas tisulares intraoperatorios, compatibles con el flujo quirúrgico estándar y sin interferir con la dinámica operatoria. Estos mapas han demostrado un **potencial relevante** para discriminar entre tumor y tejido sano, así como para identificar estructuras anatómicas críticas como vasos y duramadre. Este comportamiento sugiere una **posible utilidad complementaria** de la HSI en la planificación de la resección y en la mejora de la seguridad quirúrgica, al aportar información objetiva que podría reducir la dependencia de la evaluación subjetiva del neurocirujano.

Hallazgo principal: La HSI intraoperatoria con cámaras *snapshot* muestra un potencial prometedor para generar mapas tisulares en tiempo real, manteniendo la compatibilidad con el flujo quirúrgico habitual. Los resultados sugieren su capacidad para diferenciar entre tumor y tejido sano e identificar estructuras anatómicas relevantes (vasos, duramadre), lo que podría contribuir a mejorar la seguridad de la resección y apoyar la toma de decisiones intraoperatorias.

Biomarcadores pronósticos en HSAa: volumen hemorrágico

En el caso de la HSAa, se confirma de manera robusta que el volumen hemorrágico, ya sea obtenido de forma manual por expertos o automática mediante modelos de segmentación, constituye un biomarcador pronóstico robusto e independiente para el resultado funcional del paciente a seis meses. El análisis comparativo demuestra que el rendimiento pronóstico del volumen hemorrágico automático es incluso superior al de escalas ampliamente utilizadas y consolidadas como la de Fisher modificada, estableciendo así el valor de la cuantificación automática como herramienta complementaria o alternativa a las escalas visuales tradicionales. La alta concordancia práctica entre segmentación manual de referencia y automática refuerza sólidamente la viabilidad de una automatización traslacional de este proceso mediante modelos de IA, con el potencial de estandarizar y acelerar la evaluación pronóstica.

Hallazgo principal: El volumen hemorrágico automático constituye un biomarcador pronóstico robusto e independiente, con rendimiento superior a la escala de Fisher modificada. La alta concordancia entre segmentación manual y automática valida la viabilidad de automatización mediante IA, con potencial para estandarizar y acelerar la evaluación pronóstica en HSAa.

Biomarcadores pronósticos en HSAa: radiómica

Los modelos de radiómica del parénquima, en combinación con variables clínicas, muestran un desempeño medio superior en la mayoría de los desenlaces evaluados frente a escalas consolidadas como *ictWFNS* y SEBES/SEBES6c, tanto en los desenlaces primarios (mortalidad y pronóstico funcional a los seis meses) como el vasoespasmo e hidrocefalia. De forma complementaria, el desarrollo del *SEBES radiómico*, un modelo diseñado específicamente para reproducir la escala visual tradicional SEBES a partir de características radiómicas, permite reproducir de manera objetiva y cuantitativa la evaluación subjetiva del neurocirujano, demostrando que es posible capturar mediante algoritmos la información que el experto visualmente evalúa. No obstante, este modelo no logra reemplazar el valor añadido de los modelos pronósticos desarrollados específicamente para cada desenlace clínico, sugiriendo que la optimización directa hacia el objetivo pronóstico supera a la aproximación indirecta de replicar escalas visuales.

Hallazgo principal: Los modelos de radiómica del parénquima combinados con variables clínicas superan de manera consistente a escalas consolidadas (*ictWFNS*, SEBES/SEBES6c) en la predicción de mortalidad y pronóstico funcional. El desarrollo del *SEBES radiómico* demuestra que es posible reproducir objetivamente la evaluación visual del neurocirujano, aunque la optimización directa hacia desenlaces clínicos supera a la replicación de escalas visuales.

6.1.4 Marco metodológico y contribución global

Los resultados de este trabajo muestran que la innovación lograda no radica únicamente en alcanzar métricas competitivas frente a métodos de referencia, sino en haber desarrollado un conjunto de estrategias reproducibles y escalables que permiten su aplicación en entornos clínicos reales. La estandarización de los procesos de preprocesamiento y la integración de técnicas de interpretabilidad facilitan la traducción de la información derivada de la imagen a un lenguaje comprensible y clínicamente útil.

Contribución principal de la tesis: Esta tesis presenta un marco unificado para la caracterización automática del tejido cerebral y sus lesiones mediante modelos supervisados aplicados a imagen médica multimodal. Este enfoque combina rigor técnico, interpretabilidad y relevancia clínica, y sienta las bases para el desarrollo de herramientas automáticas de apoyo al diagnóstico aplicables en neurociencia, tanto en el ámbito quirúrgico como hospitalario.

6.1.5 Contribuciones a la caracterización automática del tejido cerebral y lesiones asociadas mediante modelos supervisados en imagen médica multimodal para diagnóstico asistido

Este trabajo articula un marco unificado para la caracterización automática del cerebro y sus lesiones, combinando (i) clasificación tisular intraoperatoria con HSI en GBM, (ii) segmentación automática de HSAa en TC, y (iii) modelización predictiva basada en radiómica. Más allá de resultados aislados, los hallazgos convergen en una tesis central: *los modelos supervisados, cuando se alimentan de descriptores de imagen estandarizados y se evalúan con diseños robustos, son capaces de capturar señales clínicas relevantes a diferentes escalas y de forma complementaria a las escalas tradicionales.*

1. **Eficiencia espectral en HSI con recursos limitados.** Con sólo 25 bandas, los clasificadores obtienen una relación rendimiento/recursos (OACC por banda) claramente superior a enfoques previos con cámaras de 128–826 bandas. Este resultado sugiere que una adecuada definición de clases, junto con estrategias de normalización y validación rigurosas, puede favorecer el desarrollo de sistemas HSI portátiles, a tiempo real y escalables en quirófano.
2. **Contribuciones a la base de datos *SLIMBRAIN* (Martín-Pérez et al., 2025).** Se establece una colección hiperespectral abierta centrada en neurocirugía, generada a partir de adquisiciones intraoperatorias en pacientes con glioblastoma. Esta base de datos, que amplía los experimentos realizados en este trabajo, busca promover la validación multicéntrica, la comparabilidad metodológica y la reproducibilidad futura en el campo del análisis espectral cerebral.
3. **Segmentación de HSAa robusta y clínicamente útil.** Los modelos basados en **nnU-Net** no sólo alcanzan un equilibrio entre *robustez inter-centro* y *fidelidad espacial*, sino que además superan de forma consistente los resultados previamente publicados en la literatura. A nivel cuantitativo, se obtienen métricas de solapamiento y error espacial superiores, mientras que a nivel cualitativo, las segmentaciones son valoradas

por neurocirujanos como más precisas que otros modelos de segmentación de hemorragia existentes en la literatura.

4. **Radiómica del parénquima como biomarcador emergente.** Los modelos que incorporan texturas del parénquima en TC y variables clínicas alcanzan rendimientos superiores a enfoques centrados exclusivamente en la hemorragia. La señal pronóstica del parénquima, más estable que la radiómica del sangrado, sugiere que el daño difuso temprano (edema, desorganización tisular) es clínicamente explotable.
5. **Complementariedad clínica–imagen.** La combinación de variables clínicas con descriptores radiómicos reduce los falsos positivos y mejora las métricas de rendimiento en la predicción de desenlaces. La fusión de ambas fuentes demuestra que los modelos pueden beneficiarse de la estabilidad y claridad de las variables clínicas, sin renunciar al poder discriminativo y la riqueza cuantitativa de la radiómica.
6. **Validez traslacional y eficiencia operativa.** Inferencias de aproximadamente un minuto en CPU, junto con la automatización de la segmentación, abren la puerta a despliegues hospitalarios sin necesidad de infraestructura GPU. Esto permite reducir los cuellos de botella derivados de la anotación manual y facilita estudios multicéntricos con criterios reproducibles.
7. **De la escala al biomarcador cuantitativo.** El volumen hemorrágico calculado automáticamente, junto con las características texturales del parénquima, proporcionan un valor pronóstico superior y complementario al de escalas clínicas tradicionales como WFNS, Fisher, SEBES y sus variantes, abriendo la puerta a biomarcadores objetivos y reproducibles.

6.2 Líneas futuras

En esta sección se abordan las líneas futuras derivadas de los estudios presentados en esta tesis doctoral, con el objetivo de delinear las direcciones de investigación que pueden consolidar y ampliar los hallazgos obtenidos. La Sección 6.2.1 explora los próximos pasos en el desarrollo de la HSI aplicada a la neurocirugía. La Sección 6.2.2 recoge las perspectivas de mejora en los modelos de segmentación automática de la HSAa en TC. A continuación, la Sección 6.2.3 examina las oportunidades y desafíos que se abren en el campo de los modelos predictivos basados en radiómica para la HSAa. Finalmente, en la Sección 6.2.4 se sintetizan una serie de retos y oportunidades de carácter transversal, comunes a las tres áreas, que resultan estratégicos para su futura traslación clínica.

6.2.1 Imagen hiperespectral aplicada a neurocirugía

La incorporación de la HSI en la práctica neuroquirúrgica constituye un campo emergente con gran potencial, pero todavía enfrenta limitaciones que condicionan su traslación clínica. A partir de los hallazgos de este trabajo, pueden delinearse varias líneas estratégicas para avanzar hacia su consolidación como herramienta intraoperatoria.

En primer lugar, uno de los principales retos radica en alcanzar un equilibrio adecuado

entre resolución espectral, espacial y temporal. Los resultados de este estudio han puesto de manifiesto que las firmas espectrales del tejido tumoral y del tejido sano muestran una notable similitud, lo que dificulta su discriminación automática. Esta similitud podría deberse, al menos en parte, a la resolución espectral empleada, lo que sugiere la necesidad de explorar configuraciones instrumentales con mayor granularidad capaces de capturar diferencias más sutiles entre ambos tejidos. Sin embargo, incrementar la resolución espectral, aunque potencia la capacidad discriminativa, también conlleva un aumento de los tiempos de adquisición y procesamiento, comprometiendo la aplicabilidad en un entorno quirúrgico dinámico. De ahí que resulte prioritario avanzar hacia dispositivos de captura y algoritmos de análisis que generen mapas de clasificación tisular en tiempo real. Este objetivo requiere optimizaciones simultáneas tanto en hardware como en software, desde el desarrollo de sensores de alto rendimiento hasta la implementación de modelos de inferencia ultrarrápida, apoyados en técnicas de selección de bandas y en estrategias de pérdidas coste-sensibles que permitan maximizar la separabilidad espectral sin comprometer la velocidad operativa.

En segundo lugar, la estandarización del procesamiento de datos constituye un desafío crítico. Las firmas espectrales están condicionadas por múltiples factores, como la iluminación, la geometría de adquisición, la presencia de sangre o líquido cefalorraquídeo y la variabilidad interpaciente. En este sentido, la armonización de datos, ya sea mediante métodos estadísticos (p. ej., *ComBat* (Orlhac et al., 2021)) o a través de modelos generativos (GANs, modelos de difusión) podría ayudar a mitigar el impacto de la heterogeneidad de las adquisiciones y favorecer la reproducibilidad (Haberl et al., 2024; M. Wu et al., 2024). Complementariamente, la generación de datos sintéticos y la aplicación de técnicas de aumento de datos pueden contribuir a paliar la escasez de grandes cohortes multicéntricas, un aspecto indispensable para entrenar modelos robustos y generalizables.

Desde el punto de vista del modelado, se abren varias líneas de mejora que podrían potenciar tanto la precisión como la aplicabilidad clínica de los sistemas desarrollados. Conviene señalar, además, que los algoritmos de ML empleados (SVM, RF y una CNN relativamente sencilla) representan aproximaciones básicas en comparación con el repertorio actual disponible en la literatura. Se pueden explorar técnicas más avanzadas de optimización, como la búsqueda sistemática de hiperparámetros mediante frameworks automatizados, la reducción de dimensionalidad mediante métodos no lineales (tales como autoencoders o técnicas de manifold learning), la detección y eliminación de *outliers* espectrales mediante métodos de detección de anomalías, o el uso de estrategias de balanceo de clases avanzadas que podrían mejorar la robustez de los modelos. Asimismo, se pueden aplicar enfoques de regularización avanzada, aprendizaje auto-supervisado para aprovechar datos no etiquetados, arquitecturas profundas con mecanismos de atención espectral-espacial o *transformers*, que en otros ámbitos han demostrado mayor capacidad de generalización (Ahmad et al., 2025; Che et al., 2025).

Una primera vía de mejora pasa por aplicar técnicas de optimización de hiperparámetros sistemáticas, junto con estrategias específicas para detectar y gestionar valores atípicos, equilibrar las clases y, en general, mejorar la robustez del entrenamiento. De igual manera, la incorporación de información anatómica previa, por ejemplo a partir de segmentaciones estructurales derivadas de RM preoperatoria y registradas espacialmente con la adquisición hiperespectral intraoperatoria, podría ayudar a restringir las predicciones a regiones plausibles.

Esta estrategia, combinada con técnicas de fine-tuning adaptadas a cada paciente o cohorte, favorecería un mejor ajuste de los modelos a la variabilidad biológica real. A ello se suman enfoques orientados a la fase de predicción, como el postprocesado espacial para reducir falsos positivos o la combinación de salidas mediante *ensembles* de modelos, que contribuirían a incrementar la estabilidad y consistencia de los resultados.

En paralelo, cobra especial importancia avanzar en la interpretabilidad: comprender qué bandas o regiones espectrales aportan un mayor valor diagnóstico no solo permitiría aumentar la confianza del clínico en las decisiones asistidas por el algoritmo, sino que también orientaría el diseño de futuros dispositivos, más compactos, eficientes y dirigidos a las necesidades reales del quirófano. Esto podría realizarse mediante el uso de técnicas de atribución como los valores SHAP, capaces de cuantificar la contribución de cada banda a la predicción, o a través de mapas de atención y análisis de importancia espectral, que faciliten la identificación de las regiones del espectro con mayor relevancia clínica.

Otro ámbito futuro es la integración de la HSI en el flujo quirúrgico real. Para que esta tecnología sea viable, debe cumplir requisitos estrictos: portabilidad, compatibilidad con el instrumental existente, adquisición rápida y sin contacto, interfaces intuitivas, robustez frente a condiciones quirúrgicas cambiantes y cumplimiento de normativas sanitarias. Una línea especialmente prometedora consiste en acoplar la cámara HSI a un microscopio quirúrgico, lo que permitiría capturar imágenes de forma directa durante las fases críticas de la resección tumoral, sin interferir en el campo estéril ni alterar el flujo de trabajo habitual del neurocirujano.

En cuanto al diseño de las clases de interés, desde la perspectiva clínica resulta relevante explorar enfoques que simplifiquen la interpretación. Por ejemplo, la consideración de una clase vascular combinada podría ser más útil intraoperatoriamente que la distinción detallada entre subtipos vasculares, dado que lo crucial para el cirujano es reconocer de manera inequívoca el tejido vascular y preservarlo.

Otra línea futura de investigación podría orientarse a ampliar el uso de la HSI a otros tipos de tumores cerebrales distintos del GBM. Aunque en este trabajo se ha demostrado su viabilidad para discriminar entre tejido tumoral y sano en este contexto, sería interesante evaluar hasta qué punto la técnica mantiene esa capacidad en lesiones con distinta naturaleza histológica, como astrocitomas, oligodendrogliomas o metástasis de distintos grados. A partir de ahí, cabría incluso explorar la posibilidad de extender su uso hacia la clasificación de subtipos tumorales intraoperatoriamente, lo que abriría un campo prometedor para una estratificación quirúrgica más precisa y para una integración más estrecha entre la práctica neuroquirúrgica y el análisis histopatológico.

Una línea de trabajo complementaria y estratégica consiste en ampliar la base de datos pública SLIMBRAIN con estas capturas hiperespectrales. Esta base de datos, que ya incluye datos multimodales de tumores cerebrales, representa una plataforma valiosa para compartir y estandarizar datos de HSI intraoperatoria, permitiendo ampliar el conocimiento en este campo y facilitar la validación multicéntrica de algoritmos. La incorporación de datos hiperespectrales a SLIMBRAIN no solo contribuiría a incrementar el tamaño muestral disponible para entrenar modelos más robustos y generalizables, sino que también permitiría establecer estándares de

adquisición, anotación y preprocesamiento que favorezcan la reproducibilidad y comparabilidad entre diferentes grupos de investigación.

Finalmente, la HSI ofrece oportunidades adicionales en el ámbito formativo y de estandarización. Para neurocirujanos en formación, disponer de mapas espectrales en tiempo real podría acelerar el aprendizaje y reducir la dependencia de la experiencia subjetiva, mientras que para equipos experimentados proporcionaría un soporte cuantitativo adicional que minimice la variabilidad interobservador.

6.2.2 Segmentación automática de Hemorragia Subaracnoidea Aneurismática en TC

A pesar de los resultados prometedores presentados en este trabajo, existen múltiples direcciones de investigación orientadas a fortalecer la robustez y la aplicabilidad clínica de los modelos desarrollados de segmentación automática de HSAa en TC.

Por un lado, la integración de análisis cuantitativo y cualitativo confirma que los modelos propuestos aportan mejoras significativas en aspectos concretos, como la cobertura anatómica y la detección de hemorragia en regiones subaracnoideas, donde la precisión resulta especialmente relevante para el pronóstico. Por otro, persisten limitaciones ligadas a la sobresegmentación y a la detección de focos hemorrágicos de pequeño tamaño, que impactan de manera directa en las métricas globales. Estas limitaciones señalan líneas claras de trabajo futuro, orientadas a la reducción de FP, el refinamiento en escenarios complejos y una ampliación de la validación prospectiva. En este sentido, estrategias como el entrenamiento de modelos con casos sanos para mejorar la capacidad de distinguir entre patrones normales y patológicos, o el desarrollo de modelos secundarios específicos para detectar y eliminar FP, podrían resultar especialmente útiles. Asimismo, el diseño de arquitecturas híbridas, capaces de combinar la predicción volumétrica global con módulos refinadores especializados en pequeñas lesiones, podría contribuir de forma significativa a este objetivo.

Un hallazgo destacado es que la segmentación ha mostrado un rendimiento comparativamente más sólido en la base de datos externa, que en la cohorte interna de casos puramente de HSAa. Esta diferencia puede atribuirse a la naturaleza más difusa y de menor contraste de la HSA, así como a la presencia de estructuras óseas y LCR que dificultan la diferenciación tisular. Para superar esta limitación, podrían explorarse varias estrategias. Una posibilidad es incorporar canales adicionales de información durante el entrenamiento, como mapas derivados de TC angiográfica, perfusión o incluso segmentaciones previas del espacio subaracnoideo, que aporten contexto anatómico y fisiopatológico adicional. Otras aproximaciones pasan por el desarrollo de modelos especializados en subcompartimentos, entrenados de manera focalizada en HSA, o la implementación de módulos de atención espacial capaces de resaltar señales sutiles en las cisternas y surcos.

Otra dirección relevante sería testar el rendimiento de los modelos desarrollados en otras enfermedades neurológicas asociadas a hemorragia, como el TCE, donde las lesiones son multifocales y altamente heterogéneas. De igual modo, explorar su aplicabilidad en HSD y HED, y contrastar los resultados con los modelos existentes en la literatura, permitirá valorar hasta qué punto los algoritmos desarrollados mantienen su eficacia en escenarios más amplios

y clínicamente frecuentes.

En paralelo, será fundamental avanzar hacia la segmentación subregional, diferenciando de manera explícita los distintos compartimentos hemorrágicos (HSA, HIP, HIV, HSD, HED). Este nivel de detalle permitiría construir biomarcadores de imagen más alineados con escalas clínicas de uso habitual, como la Fisher modificada, y valorar si los volúmenes individuales aportan un poder pronóstico superior al volumen global. De forma complementaria, otra línea de trabajo consistiría en desarrollar algoritmos de clasificación automática capaces de identificar el subtipo de hemorragia, o bien entrenar directamente modelos de segmentación multiclase que combinen ambas tareas (detección y clasificación) en un único proceso.

Desde un plano metodológico, resulta prioritario explorar arquitecturas más allá de **nnU-Net**. Modelos basados en transformers visuales (como Swin UNETR o nnFormer), aproximaciones CNN–transformer híbridas y mecanismos avanzados de atención han demostrado mejoras en tareas de segmentación médica y podrían aumentar la capacidad de generalización (Hatamizadeh et al., 2021; H.-Y. Zhou et al., 2021).

Además, en futuros estudios multicéntricos podría explorarse si los umbrales óptimos de edad y volumen difieren entre centros o subpoblaciones, ya que es posible que los pacientes de mayor edad toleren volúmenes hemorrágicos mayores, mientras que los pacientes más jóvenes presenten peor evolución con volúmenes menores.

Asimismo, una línea de investigación relevante consistiría en evaluar el volumen obtenido a partir de la segmentación automática no solo como predictor del resultado funcional a los seis meses, sino también como biomarcador de otros desenlaces clínicamente significativos, como la mortalidad, la aparición de vasoespasma o el desarrollo de hidrocefalia. Estudiar estas asociaciones permitiría ampliar el valor clínico de la segmentación automática y consolidar su utilidad como herramienta integral de estratificación pronóstica.

Por último, la validación prospectiva y en condiciones de uso ágil y con respuesta rápida constituye un paso imprescindible para la traslación clínica. Este reto no se limita a optimizar la eficiencia de la inferencia mediante técnicas de pruning (Blalock et al., 2020; Han et al., 2015), cuantización (Banner et al., 2018; Jacob et al., 2018) o conversión a entornos de ejecución acelerados como ONNX Runtime (Contributors, 2025) o implementaciones en C++ (Zunin, 2021), sino que exige también asegurar una integración fluida en el flujo asistencial. La interoperabilidad con los sistemas hospitalarios y la aceptación por parte de los profesionales son factores determinantes para su implantación. Del mismo modo, el desarrollo de interfaces intuitivas, acompañadas de métricas de confianza y mapas de incertidumbre que informen del grado de fiabilidad de cada predicción, puede aumentar la transparencia y favorecer la confianza del usuario final, facilitando así su incorporación en la práctica clínica diaria.

6.2.3 Modelos predictivos basados en radiómica para Hemorragia Subaracnoidea Aneurismática en TC

A partir de los modelos ML desarrollados para predecir desenlaces en pacientes con HSAA, se han identificado líneas futuras orientadas a transformar la radiómica en un biomarcador clínico sólido, fiable y con verdadero impacto en la toma de decisiones asistenciales.

En primer lugar, la validación externa y prospectiva resulta un paso indispensable. El carácter retrospectivo y unicéntrico de los datos utilizados limita de manera natural la generalización de los resultados. Por ello, resulta prioritario diseñar estudios multicéntricos que reúnan cohortes amplias, heterogéneas y representativas de la diversidad clínica y demográfica de la HSAa. Dichos estudios deberían incluir un tamaño muestral suficiente para abordar de forma robusta desenlaces como el vasoespasma e hidrocefalia y garantizar definiciones homogéneas de los mismos, minimizando así la variabilidad interinstitucional.

Un segundo eje de desarrollo se relaciona con la armonización de los datos de imagen y la robustez de los descriptores radiómicos. Los parámetros de adquisición, las diferencias entre escáneres y la presencia de artefactos pueden alterar la radiómica. Para abordar este reto, se podrían aplicar técnicas de armonización como ComBat (Orlhac et al., 2021) para minimizar diferencias entre protocolos o escáneres, y explorar estrategias más recientes basadas en modelos generativos que permitan reducir la variabilidad técnica sin comprometer la señal biológica (Mali et al., 2021). Del mismo modo, conviene evaluar explícitamente el impacto de artefactos comunes en TC, como los derivados de movimiento, clipajes quirúrgicos o implantes dentales, ya que representan situaciones habituales en la práctica clínica que los modelos deberán aprender a manejar con robustez.

Otro aspecto de gran interés se refiere a la segmentación y su relación con la radiómica. En este trabajo se emplearon volúmenes derivados de segmentaciones automáticas, lo que introduce inevitablemente errores que se trasladan a los descriptores. De cara al futuro, será necesario avanzar hacia segmentaciones más conscientes de la incertidumbre, incorporando ensembles y mapas de incertidumbre que informen sobre la fiabilidad de cada región segmentada. Asimismo, se abre la posibilidad de extraer radiómica no solo del volumen global de la hemorragia, sino también de compartimentos específicos: HSA, HIV y HIP. Este análisis subregional podría revelar patrones diferenciales con valor pronóstico añadido y facilitar la construcción de biomarcadores más robustos e interpretables.

En cuanto al modelado estadístico y de ML, conviene explorar técnicas más sofisticadas que vayan más allá de MRMR para la selección de variables. Enfoques como LASSO, Boruta o la selección basada en estabilidad podrían aumentar la robustez frente al sobreajuste y mejorar la reproducibilidad (Kursa y Rudnicki, 2010; Meinshausen y Bühlmann, 2010; Tibshirani, 1996). De igual manera, la implementación de estrategias de optimización más avanzadas, como aplicar el framework de Optuna (Akiba et al., 2019), contribuiría a encontrar configuraciones más ajustadas al problema.

Desde la perspectiva de los desenlaces clínicos, una línea futura de investigación consistiría en ampliar el espectro de desenlaces asociados al volumen obtenido mediante segmentación automática. Se ha explorado su valor como biomarcador de resultado funcional, mortalidad, vasoespasma e hidrocefalia; sin embargo, podrían evaluarse también otros eventos de gran relevancia clínica, como el resangrado o la isquemia cerebral diferida. Además, será necesario ir más allá de los modelos binarios de clasificación y avanzar hacia enfoques tiempo-a-evento y de riesgos competitivos, que reflejen con mayor fidelidad la evolución clínica de la HSAa y sus complicaciones. Asimismo, integrar variables de tratamiento (como el tipo de clipado o embolización, el uso de drenajes ventriculares o la estancia en UCI) podría permitir reducir la confusión y captar mejor la interacción entre enfermedad, intervención y desenlaces. Incluso

puede plantearse la aplicación de metodologías de inferencia causal, orientadas a distinguir correlaciones espurias de verdaderos efectos fisiopatológicos.

Como línea futura, si se quisiera comparar de forma directa el rendimiento entre desenlaces, sería necesario trabajar con un subconjunto común de pacientes y asegurar que todos los modelos emplean las mismas particiones de entrenamiento y prueba. Solo así podrían atribuirse las diferencias al desenlace y no a variaciones en la cohorte. En este estudio, en cambio, se ha optado por utilizar el máximo número de pacientes disponible para cada outcome, priorizando la solidez y la especificidad de cada análisis (mortalidad, funcionalidad, hidrocefalia y vasoespasma). Este enfoque permite evaluar cada dominio con la mayor potencia posible, aunque impide comparaciones directas entre ellos, que se plantean como una línea de trabajo futura.

Por último, un bloque de líneas futuras se relaciona con la traslación tecnológica y la implementación clínica. La radiómica, para convertirse en una herramienta útil, no puede quedar restringida a entornos experimentales. Es necesario avanzar hacia prototipos que puedan desplegarse en entornos hospitalarios, midiendo de manera explícita la latencia, la carga computacional y la usabilidad en tiempo ágil. De igual modo, el diseño de interfaces clínicas intuitivas, que combinen las predicciones con métricas de confianza fácilmente interpretables, resultará determinante para la aceptación por parte de los profesionales.

6.2.4 Líneas futuras transversales

Más allá de las particularidades metodológicas y técnicas de cada modalidad (HSI, segmentación en TC o radiómica), existen retos y oportunidades de carácter estratégico que trascienden las fronteras entre disciplinas y que requieren enfoques integrados e interdisciplinarios para su resolución. Estos aspectos transversales se refieren a la integración de múltiples fuentes de información, la evaluación de equidad y sesgos, el diseño de sistemas de evaluación continua, y la creación de marcos regulatorios y de certificación.

Validación multicéntrica y prospectiva.

Un aspecto crítico compartido por todas las aproximaciones presentadas es la necesidad de validar los modelos en entornos multicéntricos, prospectivos y con cohortes heterogéneas. En el caso de la clasificación tisular mediante HSI, la validación multicéntrica permitiría evaluar la robustez de los clasificadores ante variaciones en los sistemas de adquisición, protocolos quirúrgicos y condiciones de iluminación entre diferentes centros. Para la segmentación automática de hemorragias en TC, la validación prospectiva en múltiples hospitales con distintos escáneres y protocolos de adquisición permitiría confirmar que los modelos de **nnU-Net** mantienen su rendimiento en condiciones reales de uso clínico. En el ámbito de la radiómica para predicción de desenlaces en HSAa, la validación multicéntrica es especialmente relevante, ya que las diferencias en parámetros de adquisición entre escáneres pueden afectar la reproducibilidad de las características radiómicas. La dependencia de bases de datos retrospectivas y unicéntricas, aunque necesaria en fases iniciales, limita de manera natural la generalización de los resultados. Estudios colaborativos entre instituciones permitirían no solo ampliar el tamaño muestral y la diversidad demográfica, sino también someter a los modelos a la variabilidad técnica derivada de distintos escáneres, protocolos y entornos asistenciales,

lo que constituye una condición indispensable para garantizar su robustez y transferibilidad a la práctica clínica.

Integración multimodal y fusión de información.

Una línea estratégica transversal consiste en explorar cómo los tres enfoques desarrollados (clasificación tisular HSI, segmentación de hemorragias y radiómica) pueden integrarse en sistemas de apoyo a la decisión unificados que aprovechen la sinergia entre modalidades. Por ejemplo, en un paciente con HSAa que requiere cirugía de clipado, podría combinarse la información radiómica extraída de la TC preoperatoria (para estratificación pronóstica) con la segmentación automática de la hemorragia (para cuantificación volumétrica) y, potencialmente en el futuro, con mapas hiperespectrales intraoperatorios (para guía durante la resección del aneurisma o evaluación del tejido isquémico). Esta integración multimodal requeriría no solo la sincronización técnica de las diferentes modalidades, sino también el desarrollo de arquitecturas de fusión que combinen decisiones de forma inteligente, considerando la incertidumbre asociada a cada fuente de información y los contextos clínicos específicos en los que cada modalidad aporta mayor valor.

Evaluación de equidad y sesgos poblacionales.

Otra línea transversal de gran relevancia es la evaluación de la equidad de los modelos. La estratificación por sexo, edad, raza u otros factores sociodemográficos resulta esencial para identificar posibles sesgos de representación y rendimiento diferencial entre subgrupos poblacionales. La creciente literatura en IA aplicada a la salud subraya la importancia de evitar que los algoritmos refuercen desigualdades preexistentes, y de asegurar que las predicciones sean consistentes y justas para todos los pacientes, independientemente de sus características individuales (Chinta et al., 2024; Hasanzadeh et al., 2025). Este enfoque no solo es éticamente necesario, sino que además aumenta la fiabilidad y la aceptación de las herramientas desarrolladas.

Sistemas de evaluación continua y monitorización de deriva.

Un desafío transversal poco explorado es el desarrollo de sistemas que permitan evaluar de forma continua el rendimiento de los modelos una vez desplegados en la práctica clínica. En clasificación tisular mediante HSI, los modelos pueden sufrir *data drift* si se actualiza el sistema de captura, cambian las condiciones de iluminación quirúrgica o se modifica la configuración de la cámara, alterando las firmas espectrales y degradando el rendimiento sin que el usuario lo perciba. Para la segmentación automática de hemorragias en TC, el *data drift* puede ocurrir cuando un hospital actualiza sus escáneres o cambia los protocolos de adquisición. El *concept drift* es especialmente relevante para los modelos radiómicos de predicción de desenlaces en HSAa: si mejoran los protocolos de tratamiento del vasoespasmio o se introducen nuevas estrategias terapéuticas, la relación entre características radiómicas del parénquima y desenlace funcional puede cambiar, haciendo que un modelo entrenado con datos históricos pierda precisión. No obstante, esta discrepancia también puede tener una lectura clínica útil: si un modelo entrenado con cohortes antiguas tiende a sobrestimar la mortalidad o la morbilidad en pacientes tratados con protocolos actuales, esto puede indicar que las nuevas intervenciones están siendo más efectivas y que el pronóstico real de los

pacientes ha mejorado respecto a lo esperado. Implementar sistemas de monitorización que detecten automáticamente estos cambios mediante comparación de distribuciones de variables de entrada, seguimiento de métricas de rendimiento y alertas cuando sea necesario reentrenar o recalibrar resulta esencial para mantener la confiabilidad a largo plazo.

Marcos regulatorios y certificación para IA en imagen médica.

Finalmente, todos estos enfoques comparten el reto de navegar por un panorama regulatorio en evolución. Para la HSI intraoperatoria como dispositivo de apoyo a la decisión quirúrgica, será necesario demostrar seguridad y eficacia conforme al Reglamento de Dispositivos Médicos (MDR) en Europa, incluyendo estudios de usabilidad que validen que los mapas tisulares mejoran la precisión del neurocirujano. Los modelos de segmentación automática de hemorragias en TC, si se utilizan para cuantificación volumétrica que informe decisiones terapéuticas, pueden requerir certificación como Software como Dispositivo Médico (SaMD) clase II, lo que implica validación clínica prospectiva y procesos de calidad. Para los modelos radiómicos de predicción de desenlaces en HSAA, la clasificación regulatoria dependerá de su uso previsto. En todos los casos, será necesario establecer procesos de trazabilidad que documenten qué versión del modelo generó cada predicción y cómo se integra la incertidumbre algorítmica en la decisión clínica final. La creación de estándares de certificación específicos para biomarcadores de imagen basados en IA será clave para facilitar la traslación y asegurar seguridad y calidad para uso clínico.

6.3 Impacto

En esta sección se presenta la producción científica derivada de la presente tesis doctoral, que comprende la participación en congresos y conferencias internacionales, las publicaciones en revistas científicas, así como los manuscritos actualmente en proceso de revisión. Asimismo, se recogen los trabajos co-supervisados y los proyectos de investigación en los que se ha colaborado de manera activa.

6.3.1 Conferencias y congresos internacionales

1. L. Ruiz, G. Urbanos. *NEMESIS-3D-CM: Aplicación de la imagen hiperespectral en la diferenciación de lesiones cerebrales*. 19º Congreso Sociedad de Neurocirugía de la Comunidad Autónoma de Madrid (SONCAM), Online, 23/10/2020. Ponencia.
2. G. Urbanos, E. Juárez, A. Lagares. *Optical properties of tissues: HSI application*. 2nd PhD Poster Competition on Design of Circuits and Integrated Systems (DCIS 2020), Segovia, 18–20/11/2020. Póster.
3. J. Sancho, G. Urbanos, L. Ruiz, M. Villanueva, G. Rosa, A. Martín, M. Villa, M. Chavarrías, A. Lagares, R. Salvador, E. Juárez, C. Sanz. *Towards GPU Accelerated HyperSpectral Depth Estimation in Medical Applications*. XXXV Conference on Design of Circuits and Integrated Systems (DCIS 2020), Segovia, 18–20/11/2020. Ponencia.
4. G. Rosa, M. Villanueva, J. Sancho, G. Urbanos, L. Ruiz, M. Villa, A. Martín, E. Juárez, L. Jiménez, M. Chavarrías, A. Lagares, C. Sanz. *Hyperspectral Images Acquisition:*

- an Efficient Capture and Processing Stitching Procedure for Medical Environments*. XXXV Conference on Design of Circuits and Integrated Systems (DCIS 2020), Segovia, 18–20/11/2020. Ponencia.
5. L. Ruiz, A. Martín, G. Urbanos, M. Villanueva, J. Sancho, G. Rosa, M. Villa, M. Chavarrías, A. Pérez, E. Juárez, A. Lagares, C. Sanz. *Multiclass Brain Tumor Classification Using Hyperspectral Imaging and Supervised Machine Learning*. XXXV Conference on Design of Circuits and Integrated Systems (DCIS 2020), Segovia, 18–20/11/2020. Ponencia.
 6. M. Villa, J. Sancho, M. Villanueva, G. Urbanos, P. Sutradhar, G. Rosa, G. Vázquez, A. Martín, M. Chavarrías, L. Pérez, A. Nuñez, A. Lagares, E. Juárez, C. Sanz. *Stitching technique based on SURF for Hyperspectral Pushbroom Linescan Cameras*. XXXVI Conference on Design of Circuits and Integrated Systems (DCIS 2021), Oporto, 24–26/11/2021. Ponencia.
 7. J. Sancho, M. Villa, G. Urbanos, M. Villanueva, P. Sutradhar, G. Rosa, A. Martín, G. Vázquez, M. Chavarrías, R. Salvador, L. Pérez, A. Nuñez, A. Lagares, E. Juárez, C. Sanz. *An Embedded GPU Accelerated Hyperspectral Video Classification System in Real-Time*. XXXVI Conference on Design of Circuits and Integrated Systems (DCIS 2021), Oporto, 24–26/11/2021. Ponencia.
 8. M. Villa, J. Sancho, G. Vázquez, G. Rosa, G. Urbanos, A. Martín-Perez, P. Sutradhar, R. Salvador, M. Chavarrías, A. Lagares, E. Juárez, C. Sanz. *Data-Type Assessment for Real-Time Hyperspectral Classification in Medical Imaging*. DASIP 2022: Workshop on Design and Architectures for Signal and Image Processing, Budapest, 20–22/06/2022. Ponencia.
 9. G. Urbanos, G. Vázquez, A. Martín, M. Villa, G. Rosa, J. Sancho, P. Sutradhar, M. Chavarrías, L. Jiménez-Roldán, A. Pérez-Nuñez, A. Lagares, E. Juárez, C. Sanz. *Intraoperative brain tumor hyperspectral imaging database*. XXV Congreso de la Sociedad Española de Neurocirugía (SENEC), Oviedo, 17–20/05/2022. Ponencia.

6.3.2 Publicaciones en revistas

1. Urbanos, G., Martín, A., Vázquez, G., Villanueva, M., Villa, M., Jiménez-Roldán, L., ... & Sanz, C. *Supervised Machine Learning Methods and Hyperspectral Imaging Techniques Jointly Applied for Brain Cancer Classification*. Sensors, 21(11):3827, 31 May 2021. Categoría JCR: Instruments & Instrumentation (SCIE), Q2.
2. Martín-Pérez, A., Villa, M., Rosa Olmeda, G., Sancho, J., Vázquez, G., Urbanos, G., ... & Sanz, C. *SLIMBRAIN database: A multimodal image database of in vivo human brains for tumour detection*. Scientific Data (Nature Portfolio), 12: Article nº 836, 21 May 2025. Categoría JCR: Multidisciplinary Sciences, Q1.
3. Urbanos, G., Castaño-León, A. M., Maldonado-Luna, M., Salvador, E., Ramos, A., Lechuga, C., ... & Lagares, A. *Comprehensive predictive modeling in subarachnoid hemorrhage: integrating radiomics and clinical variables*. Neurosurgical Review (Springer), 48(1):1–15, 24 June 2025. Categoría JCR: Surgery, Q1.

6.3.3 Otras publicaciones

- Jimenez-Pastor, A., & Urbanos-García, G. *How to Extract Radiomic Features from Imaging*. En: *Basics of Image Processing: The Facts and Challenges of Data Harmonization to Improve Radiomics Reproducibility*. Springer International Publishing, Cham, pp. 61–78, 25 February 2024. Tipo: Capítulo de libro (sin JCR).

6.3.4 Premios y reconocimientos

- Primer Premio en el 24^o Congreso de la Sociedad de Neurocirugía de la Comunidad de Madrid (SONCAM) 2025 por la comunicación titulada “Radiomics vs the Human Eye: Rethinking SEBES for Prognostic Stratification in Aneurysmal Subarachnoid Hemorrhage”. Mónica Maldonado-Luna*, Gemma Urbanos*, Ana M. Castaño-León, Andreea E. Baciu, Luis Miguel Moreno-Gómez, Guillermo García-Posadas, Leandro Tosi, Carlos Loynaz-Cardona, Alfonso Lagares, *Both authors contributed equally to this work.

Trabajos en revisión

No todos los trabajos desarrollados en el marco de esta tesis doctoral han alcanzado todavía la fase de publicación definitiva. En la actualidad, varios manuscritos derivados de la investigación se encuentran en proceso de revisión por revistas internacionales indexadas, lo que refleja la continuidad y proyección científica del trabajo realizado. Entre ellos destacan:

- *Deep Learning-Based Segmentation of Aneurysmal Subarachnoid Hemorrhage: Toward Accurate and Scalable Prognostic Imaging*.
- *Radiomics vs the Human Eye: Rethinking SEBES for Prognostic Stratification in Aneurysmal Subarachnoid Hemorrhage*.

6.3.5 Trabajos cosupervisados

- Cáceres Bordallo, J. (2021). Sistema de reconocimiento de gestos utilizando técnicas de Machine Learning en un entorno quirúrgico.

6.3.6 Proyectos de investigación

- **NEMESIS-3D-CM** (clasificación intraoperatoria de tumores cerebrales mediante modelos InnerSivos). Este proyecto se ha desarrollado en colaboración con la Fundación para la Investigación Biomédica del Hospital Universitario 12 de Octubre, integra la investigación tecnológica en procesamiento de imágenes hiperespectrales, computación de alto rendimiento y técnicas de aprendizaje automático con la experiencia clínica y la práctica diaria de los neurocirujanos. Esta colaboración genera sinergias entre la ingeniería y la medicina, con el propósito de diseñar un sistema médico basado en arquitecturas multiprocesador heterogéneas, capaz de dar soporte a la parte sistemática del diagnóstico intraoperatorio en tiempo real. El objetivo último es proporcionar a los neurocirujanos información complementaria que aumente la precisión en la toma de decisiones críticas durante la resección de tumores cerebrales.

Referencias

- Acerbi, F., Cavallo, C., Broggi, M., Cordella, R., Anghileri, E., Eoli, M., Schiariti, M., Broggi, G., & Ferroli, P. (2014). Fluorescein-guided surgery for malignant gliomas: a review. *Neurosurgical review*, 37(4), 547-557.
- Adamczyk, B., Węgrzyn, K., Wilczyński, T., Maciarz, J., Morawiec, N., & Adamczyk-Sowa, M. (2021). The most common lesions detected by neuroimaging as causes of epilepsy. *Medicina*, 57(3), 294.
- Adduru, V., Baum, S., Zhang, C., Helguera, M., Zand, R., Lichtenstein, M., Griessenauer, C., & Michael, A. (2020). A method to estimate brain volume from head CT images and application to detect brain atrophy in Alzheimer disease. *American Journal of Neuroradiology*, 41(2), 224-230.
- Aerts, H. J., Velazquez, E. R., Leijenaar, R. T., Parmar, C., Grossmann, P., Carvalho, S., Bussink, J., Monshouwer, R., Haibe-Kains, B., Rietveld, D., et al. (2014). Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature communications*, 5(1), 4006.
- Aftab, K., Aamir, F. B., Mallick, S., Mubarak, F., Pope, W. B., Mikkelsen, T., Rock, J. P., & Enam, S. A. (2022). Radiomics for precision medicine in glioblastoma. *Journal of neuro-oncology*, 156(2), 217-231.
- Agresti, A., et al. (1996). An introduction to categorical data analysis.
- Ahmad, M., Distefano, S., Khan, A. M., Mazzara, M., Li, C., Li, H., Aryal, J., Ding, Y., Vivone, G., & Hong, D. (2025). A comprehensive survey for hyperspectral image classification: The evolution from conventional to transformers and mamba models. *Neurocomputing*, 130428.
- Ahn, S.-H., Savarraj, J. P., Pervez, M., Jones, W., Park, J., Jeon, S.-B., Kwon, S. U., Chang, T. R., Lee, K., Kim, D. H., et al. (2018). The subarachnoid hemorrhage early brain edema score predicts delayed cerebral ischemia and clinical outcomes. *Neurosurgery*, 83(1), 137-145.
- Akbari, H., Halig, L. V., Zhang, H., Wang, D., Chen, Z. G., & Fei, B. (2012). Detection of cancer metastasis using a novel macroscopic hyperspectral method. *Proceedings of SPIE*, 8317, 831711.
- Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2623-2631.

- Algethami, H., Lam, F. C., Rojas, R., & Kasper, E. M. (2021). Pre-Surgical and Surgical Planning in Neurosurgical Oncology-A Case-Based Approach to Maximal. *Frontiers in Clinical Neurosurgery*, 35.
- Ali, H. (2022). AI in neurodegenerative disease research: Early detection, cognitive decline prediction, and brain imaging biomarker identification. *Int J Eng Technol Res Manag*, 6(10), 71.
- Anderson, G. B., Ashforth, R., Steinke, D. E., & Findlay, J. M. (2000). CT angiography for the detection of cerebral vasospasm in patients with acute subarachnoid hemorrhage. *American Journal of Neuroradiology*, 21(6), 1011-1015.
- Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B. A., Litjens, G., Menze, B., Ronneberger, O., Summers, R. M., et al. (2022). The medical segmentation decathlon. *Nature communications*, 13(1), 4128.
- Arévalo-Sáenz, A., Rodríguez-Boto Amago, G., & Pedrosa Sánchez, M. (2022). High-grade glioma and solitary metastasis: Differentiation by spectroscopy and advanced magnetic resonance techniques. *Egyptian Journal of Neurosurgery*, 37(1), 34.
- Aryasinghe, L., Kazim, Y., Obeid, H. F., & Hashim, H. (2016). The hyperdense basilar artery sign: a case of locked-in syndrome. *International Journal of Emergency Medicine*, 9(1), 12.
- Ashburner, J., Barnes, G., Chen, C.-C., Daunizeau, J., Flandin, G., Friston, K., Kiebel, S., Kilner, J., Litvak, V., Moran, R., et al. (2014). SPM12 manual. *Wellcome Trust Centre for Neuroimaging, London, UK*, 2464(4), 53.
- Ashraf, M., Choudhary, N., Hussain, S. S., Kamboh, U. A., & Ashraf, N. (2020). Role of intraoperative computed tomography scanner in modern neurosurgery—An early experience. *Surgical neurology international*, 11, 247.
- Banner, R., Hubara, I., Hoffer, E., & Soudry, D. (2018). Scalable methods for 8-bit training of neural networks. *Advances in neural information processing systems*, 31.
- Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8), 1798-1828.
- Benjamin, C. F., Dhingra, I., Li, A. X., Blumenfeld, H., Alkawadri, R., Bickel, S., Helmstaedter, C., Meletti, S., Bronen, R. A., Warfield, S. K., et al. (2018). Presurgical language fMRI: technical practices in epilepsy surgical planning. *Human brain mapping*, 39(10), 4032-4042.
- Bhatele, K. R., & Bhadauria, S. S. (2022). Machine learning application in glioma classification: review and comparison analysis. *Archives of Computational Methods in Engineering*, 29(1), 247-274.
- Bin-Alamer, O., Abou-Al-Shaar, H., Gersey, Z. C., Huq, S., Kallos, J. A., McCarthy, D. J., Head, J. R., Andrews, E., Zhang, X., & Hadjipanayis, C. G. (2023). Intraoperative imaging and optical visualization techniques for brain tumor resection: a narrative review. *Cancers*, 15(19), 4890.
- Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4). Springer.
- Black, D., Byrne, D., Walke, A., Liu, S., Di Ieva, A., Kaneko, S., Stummer, W., Salcudean, T., & Suero Molina, E. (2024). Towards machine learning-based quantitative hyperspectral image guidance for brain tumor resection. *Communications Medicine*, 4(1), 131.

- Blalock, D., Gonzalez Ortiz, J. J., Frankle, J., & Gutttag, J. (2020). What is the state of neural network pruning? *Proceedings of machine learning and systems*, 2, 129-146.
- Bonmatí, L. M., Alberich-Bayarri, A., García-Martí, G., Requena, R. S., Castillo, C. P., Sierra, J. C., & Herrera, J. M. (2012). Biomarcadores de imagen, imagen cuantitativa y bioingeniería. *Radiologia*, 54(3), 269-278.
- Boone, J. M. (2006). Multidetector CT: opportunities, challenges, and concerns associated with scanners with 64 or more detector rows. *Radiology*, 241(2), 334-337.
- Boss, A., Bisdas, S., Kolb, A., Hofmann, M., Ernemann, U., Claussen, C. D., Pfannenber, C., Pichler, B. J., Reimold, M., & Stegger, L. (2010). Hybrid PET/MRI of intracranial masses: initial experiences and comparison to PET/CT. *Journal of Nuclear Medicine*, 51(8), 1198-1205.
- Bot, M., Van Den Munckhof, P., Bakay, R., Stebbins, G., & Verhagen Metman, L. (2017). Accuracy of intraoperative computed tomography during deep brain stimulation procedures: comparison with postoperative magnetic resonance imaging. *Stereotactic and Functional Neurosurgery*, 95(3), 183-188.
- Bravo-Garrido, G., Vargas-Lopez, A., Fernández-Gómez, M., Gomar-Alba, M., Urreta-Juárez, G., Amaya-Pascasio, L., & Martínez-Sánchez, P. (2025). Volumetric Blood Cutoff for Outcome Prediction in Aneurysmal Subarachnoid Hemorrhage Using Semiautomated Quantification.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Brodal, P. (2010). *The central nervous system*. oxford university Press.
- Brown, T. J., Brennan, M. C., Li, M., Church, E. W., Brandmeir, N. J., Rakszawski, K. L., Patel, A. S., Rizk, E. B., Suki, D., Sawaya, R., et al. (2016). Association of the extent of resection with survival in glioblastoma: a systematic review and meta-analysis. *JAMA oncology*, 2(11), 1460-1469.
- Brudfors, M. (2020). *Generative Models for Preprocessing of Hospital Brain Scans* [Tesis doctoral, UCL (University College London)].
- Brudfors, M., Balbastre, Y., Flandin, G., Nachev, P., & Ashburner, J. (2020). Flexible Bayesian Modelling for Nonlinear Image Registration. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 253-263.
- Bzdok, D., & Meyer-Lindenberg, A. (2018). Machine learning for precision psychiatry: opportunities and challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3(3), 223-230.
- Cai, J. C., Akkus, Z., Philbrick, K. A., Boonrod, A., Hoodeshenas, S., Weston, A. D., Rouzrokh, P., Conte, G. M., Zeinoddini, A., Vogelsang, D. C., et al. (2020). Fully automated segmentation of head CT neuroanatomy using deep learning. *Radiology: Artificial Intelligence*, 2(5), e190183.
- Calin, M. A., Parasca, S. V., Savastru, D., & Manea, D. (2014). Hyperspectral imaging in the medical field: Present and future. *Applied Spectroscopy Reviews*, 49(6), 435-447.
- Caplan, L. R. (2015). Intracerebral hemorrhage. *Intracerebral Hemorrhage*, 221.
- Castellano, A., Cirillo, S., Bello, L., Riva, M., & Falini, A. (2017). Functional MRI for surgery of gliomas. *Current treatment options in neurology*, 19, 1-23.
- Catapano, G., Sgulò, F. G., Seneca, V., Lepore, G., Columbano, L., & di Nuzzo, G. (2017). Fluorescein-guided surgery for high-grade glioma resection: an intraoperative “contrast-enhancer”. *World neurosurgery*, 104, 239-247.

- Cattarinussi, G., Grimaldi, D. A., & Sambataro, F. (2023). Spontaneous brain activity alterations in first-episode psychosis: a meta-analysis of functional Magnetic Resonance Imaging studies. *Schizophrenia bulletin*, *49*(6), 1494-1507.
- Celaya, A., Riviere, B., & Fuentes, D. (2023). A generalized surface loss for reducing the hausdorff distance in medical imaging segmentation. *arXiv preprint arXiv:2302.03868*.
- CG, D. (1988). Report of World Federation of Neurological Surgeons Committee on a universal subarachnoid hemorrhage grading scale. *J neurosurg*, *68*, 985-986.
- Chapleau, M., Iaccarino, L., Soleimani-Meigooni, D., & Rabinovici, G. D. (2022). The role of amyloid PET in imaging neurodegenerative disorders: a review. *Journal of Nuclear Medicine*, *63*(Supplement 1), 13S-19S.
- Che, J., Sun, M., & Wang, Y. (2025). Deep learning-based time series prediction in multispectral and hyperspectral imaging for cancer detection. *Frontiers in Medicine*, *12*, 1605865.
- Chen, L., Wang, X., Wang, S., Zhao, X., Yan, Y., Yuan, M., & Sun, S. (2025). Development of a non-contrast CT-based radiomics nomogram for early prediction of delayed cerebral ischemia in aneurysmal subarachnoid hemorrhage. *BMC Medical Imaging*, *25*(1), 1-12.
- Chen, S., Luo, J., Reis, C., Manaenko, A., & Zhang, J. (2017). Hydrocephalus after subarachnoid hemorrhage: pathophysiology, diagnosis, and treatment. *BioMed research international*, *2017*(1), 8584753.
- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785-794.
- Chen, W. (2007). Clinical applications of PET in brain tumors. *Journal of nuclear medicine*, *48*(9), 1468-1481.
- Cheplygina, V., De Bruijne, M., & Pluim, J. P. (2019). Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Medical image analysis*, *54*, 280-296.
- Chételat, G., Arbizu, J., Barthel, H., Garibotto, V., Law, I., Morbelli, S., van de Giessen, E., Agosta, F., Barkhof, F., Brooks, D. J., et al. (2020). Amyloid-PET and 18F-FDG-PET in the diagnostic investigation of Alzheimer's disease and other dementias. *The Lancet Neurology*, *19*(11), 951-962.
- Chicco, D. (2021). Siamese neural networks: An overview. *Artificial neural networks*, 73-94.
- Chinta, S. V., Wang, Z., Palikhe, A., Zhang, X., Kashif, A., Smith, M. A., Liu, J., & Zhang, W. (2024). AI-Driven Healthcare: A Review on Ensuring Fairness and Mitigating Bias. *arXiv preprint arXiv:2407.19655*.
- Cleveland Clinic. (2024). Brain: Anatomy, Function, Conditions [Accessed: 16 August 2025].
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. routledge.
- Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. (2015a). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Journal of British Surgery*, *102*(3), 148-158.
- Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. (2015b). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Journal of British Surgery*, *102*(3), 148-158.
- Connolly Jr, E. S., Rabinstein, A. A., Carhuapoma, J. R., Derdeyn, C. P., Dion, J., Higashida, R. T., Hoh, B. L., Kirkness, C. J., Naidech, A. M., Ogilvy, C. S., et al. (2012).

- Guidelines for the management of aneurysmal subarachnoid hemorrhage: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*, 43(6), 1711-1737.
- Contributors, O. (2025). ONNX: Open standard for machine learning interoperability [[Online; accessed 13-September-2025]].
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.
- Couret, D., Boussen, S., Cardoso, D., Alonzo, A., Madec, S., Reyre, A., Brunel, H., Girard, N., Graillon, T., Dufour, H., et al. (2024). Comparison of scales for the evaluation of aneurysmal subarachnoid haemorrhage: a retrospective cohort study. *European Radiology*, 34(11), 7526-7536.
- Cramér, H. (1999). *Mathematical methods of statistics* (Vol. 9). Princeton university press.
- Cruz-Guerrero, I. A., Leon, R., Campos-Delgado, D. U., Ortega, S., Fabelo, H., & Callico, G. M. (2020). Classification of hyperspectral in vivo brain tissue based on linear unmixing. *Applied Sciences*, 10(16), 5686.
- Cutler, D. R., Edwards Jr, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, 88(11), 2783-2792.
- Cyran, K. A., Kawulok, J., Kawulok, M., Stawarz, M., Michalak, M., Pietrowska, M., Widłak, P., & Polańska, J. (2013). Support vector machines in biomedical and biometrical applications. En *Emerging paradigms in machine learning* (pp. 379-417). Springer.
- Dierksen, F., Sommer, J. K., Tran, A. T., Lin, H., Haider, S. P., Maier, I. L., Aneja, S., Sanelli, P. C., Malhotra, A., Qureshi, A. I., et al. (2024). Machine Learning Models for 3-Month Outcome Prediction Using Radiomics of Intracerebral Hemorrhage and Perihematomal Edema from Admission Head Computed Tomography (CT). *Diagnostics*, 14(24), 2827.
- Dietrich, O., Raya, J. G., Reeder, S. B., Reiser, M. F., & Schoenberg, S. O. (2007). Measurement of signal-to-noise ratios in MR images: influence of multichannel coils, parallel imaging, and reconstruction filters. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 26(2), 375-385.
- Dixon, L., Lim, A., Grech-Sollars, M., Nandi, D., & Camp, S. (2022). Intraoperative ultrasound in brain tumor surgery: A review and implementation guide. *Neurosurgical Review*, 45(4), 2503-2515.
- Doniselli, F. M., Pascuzzo, R., Mazzi, F., Padelli, F., Moscatelli, M., Akinci D'Antonoli, T., Cuocolo, R., Aquino, D., Cuccarini, V., & Sconfienza, L. M. (2024). Quality assessment of the MRI-radiomics studies for MGMT promoter methylation prediction in glioma: a systematic review and meta-analysis. *European Radiology*, 34(9), 5802-5815.
- Dos Santos, P., Scoczynski, M., Amorim Nogueira, S., Maffei Loureiro, R., Gonçalves, C., & Pacheco Calixto, W. (2024). A novel unsupervised segmentation approach for brain computed tomography employing hyperparameter optimization. *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, 2091-2094.
- Dupont, S. A., Wijdicks, E. F., Manno, E. M., Lanzino, G., & Rabinstein, A. A. (2009). Prediction of angiographic vasospasm after aneurysmal subarachnoid hemorrhage: value of the Hijdra sum scoring system. *Neurocritical care*, 11(2), 172-176.
- Eghbali, R., Nedelec, P., Weiss, D., Bhalerao, R., Xie, L., Rudie, J. D., Liu, C., Sugrue, L. P., & Rauschecker, A. M. (2025). Automated Lesion and Feature Extraction Pipeline for Brain MRIs with Interpretability. *Neuroinformatics*, 23(1), 2.

- Eibach, M., Won, S.-Y., Bruder, M., Keil, F., Herrmann, E., Berkefeld, J., Seifert, V., & Konczalla, J. (2020). Age dependency and modification of the subarachnoid hemorrhage early brain edema score. *Journal of Neurosurgery*, *134*(3), 946-952.
- Eisenhauer, E. A., Therasse, P., Bogaerts, J., Schwartz, L. H., Sargent, D., Ford, R., Dancey, J., Arbuck, S., Gwyther, S., Mooney, M., et al. (2009). New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *European journal of cancer*, *45*(2), 228-247.
- Equipo de la Torre. (2021, 2 de noviembre). *Las siete funciones vitales del líquido cefalorraquídeo* [Accessed: 16 August 2025]. <https://www.neurocirugiaequipodelatorre.es/las-siete-funciones-vitales-del-liquido-cefalorraquideo>
- Fabelo, H., Halicek, M., Ortega, S., Shahedi, M., Szolna, A., Piñeiro, J. F., Sosa, C., O'Shanahan, A. J., Bisshopp, S., Espino, C., et al. (2019). Deep learning-based framework for in vivo identification of glioblastoma tumor using hyperspectral images of human brain. *Sensors*, *19*(4), 920.
- Fabelo, H., Ortega, S., Kabwama, S., Callico, G. M., Bulters, D., Szolna, A., Pineiro, J. F., & Sarmiento, R. (2016). HELICoiD project: A new use of hyperspectral imaging for brain cancer detection in real-time during neurosurgical operations. *Hyperspectral Imaging Sensors: Innovative Applications and Sensor Standards 2016*, 9860, 986002.
- Fabelo, H., Ortega, S., Lazcano, R., Madroñal, D., M. Callicó, G., Juárez, E., Salvador, R., Bulters, D., Bulstrode, H., Szolna, A., et al. (2018). An intraoperative visualization system using hyperspectral imaging to aid in brain tumor delineation. *Sensors*, *18*(2), 430.
- Fabelo, H., Ortega, S., Ravi, D., Kiran, B. R., Sosa, C., Bulters, D., Callicó, G. M., Bulstrode, H., Szolna, A., Piñeiro, J. F., et al. (2018). Spatio-spectral classification of hyperspectral images for brain cancer detection during surgical operations. *PloS one*, *13*(3), e0193721.
- Fang, Y.-J., Mei, S.-H., Lu, J.-N., Chen, Y.-K., Chai, Z.-H., Dong, X., Araujo, C., Reis, C., Zhang, J.-M., & Chen, S. (2019). New risk score of the early period after spontaneous subarachnoid hemorrhage: for the prediction of delayed cerebral ischemia. *CNS neuroscience & therapeutics*, *25*(10), 1173-1181.
- Feghali, J., Yang, W., & Huang, J. (2020). Updates in chronic subdural hematoma: epidemiology, etiology, pathogenesis, treatment, and outcome. *World neurosurgery*, *141*, 339-345.
- Feigin, V. L., Norrving, B., & Mensah, G. A. (2017). Global burden of stroke. *Circulation research*, *120*(3), 439-448.
- Feng, Q., & Ding, Z. (2020). MRI radiomics classification and prediction in Alzheimer's disease and mild cognitive impairment: a review. *Current Alzheimer Research*, *17*(3), 297-309.
- Figueroa, J., & DeVine, J. G. (2017). Spontaneous spinal epidural hematoma: literature review. *Journal of Spine Surgery*, *3*(1), 58.
- Fink, J. R., Muzi, M., Peck, M., & Krohn, K. A. (2015). Multimodality brain tumor imaging: MR imaging, PET, and PET/MR imaging. *Journal of Nuclear Medicine*, *56*(10), 1554-1561.
- Fisher, C., Kistler, J., & Davis, J. (1980). Relation of cerebral vasospasm to subarachnoid hemorrhage visualized by computerized tomographic scanning. *Neurosurgery*, *6*(1), 1-9.

- Fonov, V., Evans, A. C., McKinstry, R. C., Almlie, C. R., & Collins, D. L. (2011). Unbiased average age-appropriate atlases for pediatric studies [Referencia estándar para el espacio anatómico MNI152.]. *NeuroImage*, *54*(1), 313-327. <https://doi.org/10.1016/j.neuroimage.2010.07.033>
- Fortin, J.-P., Cullen, N., Sheline, Y. I., Taylor, W. D., Aselcioglu, I., Cook, P. A., Adams, P., Cooper, C., Fava, M., McGrath, P. J., et al. (2018). Harmonization of cortical thickness measurements across scanners and sites. *Neuroimage*, *167*, 104-120.
- Frade, H. C., Wilson, S. E., Beckwith, A., & Powers, W. J. (2022). Comparison of outcomes of ischemic stroke initially imaged with cranial computed tomography alone vs computed tomography plus magnetic resonance imaging. *JAMA Network Open*, *5*(7), e2219416-e2219416.
- Frontera, J. A., Claassen, J., Schmidt, J. M., Wartenberg, K. E., Temes, R., Connolly, E. S., Macdonald, R. L., & Mayer, S. A. (2006). Prediction of symptomatic vasospasm after subarachnoid hemorrhage: the modified fisher scale. *Neurosurgery*, *59*(1), 21-27.
- Ganau, M., Ligarotti, G. K., & Apostolopoulos, V. (2019). Real-time intraoperative ultrasound in brain surgery: neuronavigation and use of contrast-enhanced image fusion. *Quantitative imaging in medicine and surgery*, *9*(3), 350.
- García-García, S., Cepeda, S., Arrese, I., & Sarabia, R. (2024). A fully automated pipeline using swin transformers for deep learning-based blood segmentation on head computed tomography scans after aneurysmal subarachnoid hemorrhage. *World Neurosurgery*, *190*, e762-e773.
- Garibotto, V., Heinzer, S., Vulliemoz, S., Guignard, R., Wissmeyer, M., Seeck, M., Lovblad, K.-O., Zaidi, H., Ratib, O., & Vargas, M.-I. (2013). Clinical applications of hybrid PET/MRI in neuroimaging. *Clinical nuclear medicine*, *38*(1), e13-e18.
- Gebel, J. M., & Broderick, J. P. (2000). Intracerebral hemorrhage. *Neurologic clinics*, *18*(2), 419-438.
- Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine learning*, *63*(1), 3-42.
- Giannantonio, T., Alperovich, A., Semeraro, P., Atzori, M., Zhang, X., Hauger, C., Freytag, A., Luthman, S., Vandebriel, R., Jayapala, M., et al. (2023). Intra-operative brain tumor detection with deep learning-optimized hyperspectral imaging. *Optical Biopsy XXI: Toward Real-Time Spectroscopic Imaging and Diagnosis*, *12373*, 80-98.
- Giavarina, D. (2015). Understanding bland altman analysis. *Biochemia medica*, *25*(2), 141-151.
- Gillies, R. J., Kinahan, P. E., & Hricak, H. (2016). Radiomics: images are more than pictures, they are data. *Radiology*, *278*(2), 563-577.
- Gong, S., Zhong, Y., Gong, Y., Chan, N. Y., Ma, W., Mak, C. H.-K., Abrigo, J., & Dou, Q. (2024). Segmentation of Tiny Intracranial Hemorrhage Via Learning-to-Rank Local Feature Enhancement. *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, 1-5.
- Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1). MIT press Cambridge.
- Granziera, C., Wuerfel, J., Barkhof, F., Calabrese, M., De Stefano, N., Enzinger, C., Evangelou, N., Filippi, M., Geurts, J. J., Reich, D. S., et al. (2021). Quantitative magnetic resonance imaging towards clinical application in multiple sclerosis. *Brain*, *144*(5), 1296-1311.

- Grasso, G. (2016). Extent of resection and survival in glioblastoma multiforme. *JAMA oncology*, 2(11), 1508-1509.
- Grefkes, C., & Fink, G. R. (2011). Reorganization of cerebral networks after stroke: new insights from neuroimaging with connectivity approaches. *Brain*, 134(5), 1264-1276.
- Group, B. D. W., Atkinson Jr, A. J., Colburn, W. A., DeGruttola, V. G., DeMets, D. L., Downing, G. J., Hoth, D. F., Oates, J. A., Peck, C. C., Schooley, R. T., et al. (2001). Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clinical pharmacology & therapeutics*, 69(3), 89-95.
- Gupta, V., Ambrosius, W., Qian, G., Blazejewski, A., Kazmierski, R., Urbanik, A., & Nowinski, W. L. (2010). Automatic segmentation of cerebrospinal fluid, white and gray matter in unenhanced computed tomography images. *Academic radiology*, 17(11), 1350-1358.
- Haberl, D., Spielvogel, C. P., Jiang, Z., Orlhac, F., Iommi, D., Carrió, I., Buvat, I., Haug, A. R., & Papp, L. (2024). Multicenter PET image harmonization using generative adversarial networks. *European Journal of Nuclear Medicine and Molecular Imaging*, 51(9), 2532-2546.
- Hadjipanayis, C. G., & Stummer, W. (2019). 5-ALA and FDA approval for glioma surgery. *Journal of neuro-oncology*, 141, 479-486.
- Hadjipanayis, C. G., Widhalm, G., & Stummer, W. (2015). What is the surgical benefit of utilizing 5-aminolevulinic acid for fluorescence-guided surgery of malignant gliomas? *Neurosurgery*, 77(5), 663-673.
- Halicek, M., Little, J. V., Wang, X., Patel, M., Griffith, C. C., Chen, A. Y., & Fei, B. (2018). Tumor margin classification of head and neck cancer using hyperspectral imaging and convolutional neural networks. *Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling*, 10576, 17-27.
- Halleivi, H., Albright, K., Aronowski, J., Barreto, A., Martin-Schild, S., Khaja, A., Gonzales, N., Illoh, K., Noser, E., & Grotta, J. (2008). Intraventricular hemorrhage: anatomic relationships and clinical implications. *Neurology*, 70(11), 848-852.
- Han, S., Mao, H., & Dally, W. J. (2015). Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding. *arXiv preprint arXiv:1510.00149*.
- Hardigan, A. A., Jackson, J. D., & Patel, A. P. (2023). Surgical management and advances in the treatment of glioma. *Seminars in neurology*, 43(06), 810-824.
- Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O'Neal, L., McLeod, L., Delacqua, G., Delacqua, F., Kirby, J., et al. (2019). The REDCap consortium: building an international community of software platform partners. *Journal of biomedical informatics*, 95, 103208.
- Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of biomedical informatics*, 42(2), 377-381.
- Hasanzadeh, F., Josephson, C. B., Waters, G., Adedinsewo, D., Azizi, Z., & White, J. A. (2025). Bias recognition and mitigation strategies in artificial intelligence healthcare applications. *NPJ Digital Medicine*, 8(1), 154.

- Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H. R., & Xu, D. (2021). Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. *International MICCAI brainlesion workshop*, 272-284.
- He, H., Bai, Y., Garcia, E. A., & Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*, 1322-1328.
- Hefti, M., Siegner, A., Looser, H., et al. (2008). 5-aminolevulinic acid induced protoporphyrin IX fluorescence in high-grade glioma surgery. *Swiss medical weekly*, 138(1112), 180-180.
- Heller, N., Isensee, F., Maier-Hein, K. H., Hou, X., Xie, C., Li, F., Nan, Y., Mu, G., Lin, Z., Han, M., et al. (2021). The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging: Results of the KiTS19 challenge. *Medical image analysis*, 67, 101821.
- Heye, A. K., Culling, R. D., Hernández, M. d. C. V., Thrippleton, M. J., & Wardlaw, J. M. (2014). Assessment of blood–brain barrier disruption using dynamic contrast-enhanced MRI. A systematic review. *NeuroImage: Clinical*, 6, 262-274.
- Hijdra, A., Brouwers, P., Vermeulen, M., & Van Gijn, J. (1990). Grading the amount of blood on computed tomograms after subarachnoid hemorrhage. *Stroke*, 21(8), 1156-1161.
- Hill, C. S., Coleman, M. P., & Menon, D. K. (2016). Traumatic axonal injury: mechanisms and translational opportunities. *Trends in Neurosciences*, 39(5), 311-324.
- Hofmann, B. B., Fischer, I., Neyazi, M., Karadag, C., Donaldson, D. M., Abusabha, Y., Muhammad, S., Beseoglu, K., Cornelius, J. F., & Hänggi, D. (2022). Revisiting the WFNS score: native computed tomography imaging improves identification of patients with “false poor grade” aneurysmal subarachnoid hemorrhage. *Neurosurgery*, 10-1227.
- Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*. John Wiley & Sons.
- Hostettler, I. C., Seiffge, D. J., & Werring, D. J. (2019). Intracerebral hemorrhage: an update on diagnosis and treatment. *Expert review of neurotherapeutics*, 19(7), 679-694.
- Hssayeni, M., Croock, M., Salman, A., Al-khafaji, H., Yahya, Z., & Ghoraani, B. (2020). Computed tomography images for intracranial hemorrhage detection and segmentation. *Intracranial hemorrhage segmentation using a deep convolutional model. Data*, 5(1), 14.
- Hssayeni, M. D., Croock, M. S., Salman, A. D., Al-Khafaji, H. F., Yahya, Z. A., & Ghoraani, B. (2020). Intracranial hemorrhage segmentation using a deep convolutional model. *Data*, 5(1), 14.
- Hu, P., Zhou, H., Yan, T., Miu, H., Xiao, F., Zhu, X., Shu, L., Yang, S., Jin, R., Dou, W., et al. (2023). Deep learning-assisted identification and quantification of aneurysmal subarachnoid hemorrhage in non-contrast CT scans: development and external validation of Hybrid 2D/3D UNet. *Neuroimage*, 279, 120321.
- Huang, X., Wang, D., Zhang, Q., Ma, Y., Zhao, H., Li, S., Deng, J., Ren, J., Yang, J., Zhao, Z., et al. (2022). Radiomics for prediction of intracerebral hemorrhage outcomes: A retrospective multicenter study. *NeuroImage: Clinical*, 36, 103242.
- Iglesias, J. E., & et al. (2021). Joint super-resolution and synthesis of 1 mm isotropic MP-RAGE volumes from clinical MRI exams with scans of different orientation, resolution, and contrast. *NeuroImage*, 237, 118206.

- Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., & Maier-Hein, K. H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, *18*(2), 203-211.
- Ishimaru, H., Morikawa, M., Iwanaga, S., Kaminogo, M., Ochi, M., & Hayashi, K. (2001). Differentiation between high-grade glioma and metastatic brain tumor using single-voxel proton MR spectroscopy. *European radiology*, *11*, 1784-1791.
- Itahashi, T., Yamada, T., Watanabe, H., Nakamura, M., Ohta, H., Kanai, C., Iwanami, A., Kato, N., & Hashimoto, R.-i. (2015). Alterations of local spontaneous brain activity and connectivity in adults with high-functioning autism spectrum disorder. *Molecular autism*, *6*, 1-14.
- Jack, C. R., Knopman, D. S., Jagust, W. J., Shaw, L. M., Aisen, P. S., Weiner, M. W., Petersen, R. C., & Trojanowski, J. Q. (2010). Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet Neurology*, *9*(1), 119-128.
- Jacob, B., Kligys, S., Chen, B., Zhu, M., Tang, M., Howard, A., Adam, H., & Kalenichenko, D. (2018). Quantization and training of neural networks for efficient integer-arithmetic-only inference. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2704-2713.
- Jansen, O., Szikora, I., Causin, F., Brückmann, H., & Lobotesis, K. (2017). Standards of practice in interventional neuroradiology. *Neuroradiology*, *59*, 541-544.
- Jansen-Winkeln, B., Germann, I., Köhler, H., Mehdorn, M., Maktabi, M., Sucher, R., Barberio, M., Chalopin, C., Diana, M., Moulla, Y., et al. (2021). Comparison of hyperspectral imaging and fluorescence angiography for the determination of the transection margin in colorectal resections—a comparative study. *International journal of colorectal disease*, *36*, 283-291.
- Jezzard, P., Chappell, M. A., & Okell, T. W. (2018). Arterial spin labeling for the measurement of cerebral perfusion and angiography. *Journal of Cerebral Blood Flow & Metabolism*, *38*(4), 603-626.
- Jiménez-Roldán, L., Alén, J. F., Gómez, P. A., Lobato, R. D., Ramos, A., Munarriz, P. M., & Lagares, A. (2013). Volumetric analysis of subarachnoid hemorrhage: assessment of the reliability of two computerized methods and their comparison with other radiographic scales. *Journal of neurosurgery*, *118*(1), 84-93.
- Jordan, L. C., Kleinman, J. T., & Hillis, A. E. (2009). Intracerebral hemorrhage volume predicts poor neurologic outcome in children. *Stroke*, *40*(5), 1666-1671.
- Kalra, A. (2017). Decoding the Bland–Altman plot: basic review. *Journal of the Practice of Cardiovascular Sciences*, *3*(1), 36-38.
- Kawate, N., Kayaba, K., Hara, M., Kotani, K., & Ishikawa, S. (2017). Body mass index and stroke incidence in Japanese community residents: The Jichi Medical School (JMS) Cohort Study. *Journal of epidemiology*, *27*(7), 325-330.
- Keator, D. B., Helmer, K., Steffener, J., Turner, J. A., Van Erp, T. G., Gadde, S., Ashish, N., Burns, G., & Nichols, B. N. (2013). Towards structured sharing of raw and derived neuroimaging data across existing resources. *Neuroimage*, *82*, 647-661.
- Kingsford, C., & Salzberg, S. L. (2008). What are decision trees? *Nature biotechnology*, *26*(9), 1011-1013.
- Kok, Y. E., Pszczolkowski, S., Law, Z. K., Ali, A., Krishnan, K., Bath, P. M., Sprigg, N., Dineen, R. A., & French, A. P. (2022). Semantic segmentation of spontaneous

- intracerebral hemorrhage, intraventricular hemorrhage, and associated edema on CT images using deep learning. *Radiology: Artificial Intelligence*, 4(6), e220096.
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, 15(2), 155-163.
- Koochaki, F., & Najafizadeh, L. (2024). A Siamese Convolutional Neural Network for Identifying Mild Traumatic Brain Injury and Predicting Recovery. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*.
- Kruse, F. A., Lefkoff, A. B., Boardman, J. W., Heidebrecht, K. B., Shapiro, A., Barloon, P., & Goetz, A. F. (1993). The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. *Remote sensing of environment*, 44(2-3), 145-163.
- Kruyt, N. D., Biessels, G. J., DeVries, J. H., Luitse, M. J., Vermeulen, M., Rinkel, G. J., Vandertop, W. P., & Roos, Y. B. (2010). Hyperglycemia in aneurysmal subarachnoid hemorrhage: a potentially modifiable risk factor for poor outcome. *Journal of cerebral blood flow & metabolism*, 30(9), 1577-1587.
- Kumar, V., Gu, Y., Basu, S., Berglund, A., Eschrich, S. A., Schabath, M. B., Forster, K., Aerts, H. J., Dekker, A., Fenstermacher, D., et al. (2012). Radiomics: the process and the challenges. *Magnetic resonance imaging*, 30(9), 1234-1248.
- Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the Boruta package. *Journal of statistical software*, 36, 1-13.
- Lagares, A., Gomez, P., Lobato, R., Alen, J., Alday, R., & Campollo, J. (2001). Prognostic factors on hospital admission after spontaneous subarachnoid haemorrhage. *Acta neurochirurgica*, 143, 665-672.
- Lagares, A., Gómez, P., Alén, J., Lobato, R., & Campollo, J. (2002). Global cerebral edema after subarachnoid hemorrhage. *Stroke*, 33(9), 2153-2154.
- Lagares, A., Jiménez-Roldán, L., Gomez, P. A., Munarriz, P. M., Castaño-León, A. M., Cepeda, S., & Alén, J. F. (2015). Prognostic value of the amount of bleeding after aneurysmal subarachnoid hemorrhage: a quantitative volumetric study. *Neurosurgery*, 77(6), 898-907.
- Lao, J., Chen, Y., Li, Z.-C., Li, Q., Zhang, J., Liu, J., & Zhai, G. (2017). A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. *Scientific reports*, 7(1), 10353.
- Larranaga, P., Calvo, B., Santana, R., Bielza, C., Galdiano, J., Inza, I., Lozano, J. A., Armananzas, R., Santafé, G., Pérez, A., et al. (2006). Machine learning in bioinformatics. *Briefings in bioinformatics*, 7(1), 86-112.
- Lasjaunias, P., Berenstein, A., & ter Brugge, K. G. (2013). *Clinical vascular anatomy and variations*. Springer Science & Business Media.
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (2002). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324.
- Lederer, J. (2021). Activation functions in artificial neural networks: A systematic overview. *arXiv preprint arXiv:2101.09957*.
- Lee, G. Y., Ryu, C.-W., Ko, H. C., & Jahng, G.-H. (2020). Correlation between gray matter volume loss followed by aneurysmal subarachnoid hemorrhage and subarachnoid hemorrhage volume. *Neuroradiology*, 62(11), 1401-1409.

- Leon, R., Fabelo, H., Ortega, S., Cruz-Guerrero, I. A., Campos-Delgado, D. U., Szolna, A., Piñeiro, J. F., Espino, C., O'Shanahan, A. J., Hernandez, M., et al. (2023). Hyperspectral imaging benchmark based on machine learning for intraoperative brain tumour detection. *NPJ Precision Oncology*, 7(1), 119.
- Lerch, J. P., Van Der Kouwe, A. J., Raznahan, A., Paus, T., Johansen-Berg, H., Miller, K. L., Smith, S. M., Fischl, B., & Sotiropoulos, S. N. (2017). Studying neuroanatomy using MRI. *Nature neuroscience*, 20(3), 314-326.
- LeRoux, P. D., Haglund, M. M., Newell, D. W., Grady, M. S., & Winn, H. R. (1992). Intraventricular hemorrhage in blunt head trauma: an analysis of 43 cases. *Neurosurgery*, 31(4), 678-685.
- Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, 42, 60-88.
- Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation forest. *2008 eighth IEEE international conference on data mining*, 413-422.
- Liu, Q., MacIntosh, B. J., Schellhorn, T., Skogen, K., Emblem, K. E., & Bjørnerud, A. (2023). Voxels intersecting along orthogonal levels attention u-net for intracerebral haemorrhage segmentation in head CT. *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, 1-5.
- Liu, X., Gao, K., Liu, B., Pan, C., Liang, K., Yan, L., Ma, J., He, F., Zhang, S., Pan, S., et al. (2021). Advances in deep learning-based medical image analysis. *Health Data Science, 2021*, 8786793.
- Liu, X., Li, Y., Qian, Z., Sun, Z., Xu, K., Wang, K., Liu, S., Fan, X., Li, S., Zhang, Z., et al. (2018). A radiomic signature as a non-invasive predictor of progression-free survival in patients with lower-grade gliomas. *NeuroImage: Clinical*, 20, 1070-1077.
- Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3431-3440.
- LoPresti, M. A., Bruce, S. S., Camacho, E., Kunchala, S., Dubois, B. G., Bruce, E., Appelboom, G., & Connolly Jr, E. S. (2014). Hematoma volume as the major determinant of outcomes after intracerebral hemorrhage. *Journal of the neurological sciences*, 345(1-2), 3-7.
- Lu, G., & Fei, B. (2014). Medical hyperspectral imaging: a review. *Journal of biomedical optics*, 19(1), 010901-010901.
- Luff, M. K., Khezri, N., Miralbes, S., Naravetla, B., Spiotta, A. M., Loehr, C., Martínez-Galdámez, M., McTaggart, R. A., Defreyne, L., Vega, P., et al. (2024). Hemorrhagic transformation in acute ischemic stroke: hemorrhagic subtypes and symptomatic intracranial hemorrhage. *Journal of NeuroInterventional Surgery*.
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.
- Lundervold, A. S., & Lundervold, A. (2019). An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift fuer medizinische Physik*, 29(2), 102-127.
- Madley-Dowd, P., Hughes, R., Tilling, K., & Heron, J. (2019). The proportion of missing data should not be used to guide decisions on multiple imputation. *Journal of clinical epidemiology*, 110, 63-73.

- Mali, S. A., Ibrahim, A., Woodruff, H. C., Andrearczyk, V., Müller, H., Primakov, S., Salahuddin, Z., Chatterjee, A., & Lambin, P. (2021). Making radiomics more reproducible across scanner and imaging protocol variations: a review of harmonization methods. *Journal of personalized medicine*, *11*(9), 842.
- Mandal, D. J., George, S., Pedersen, M., & Boust, C. (2021). Influence of acquisition parameters on pigment classification using hyperspectral imaging. *Color and Imaging Conference*, *29*, 334-346.
- Manni, F., van der Sommen, F., Fabelo, H., Zinger, S., Shan, C., Edström, E., Elmi-Terander, A., Ortega, S., Marrero Callicó, G., & de With, P. H. (2020). Hyperspectral imaging for glioblastoma surgery: improving tumor identification using a deep spectral-spatial approach. *Sensors*, *20*(23), 6955.
- Manniesing, R., Oei, M. T., Oostveen, L. J., Melendez, J., Smit, E. J., Platel, B., Sánchez, C. I., Meijer, F. J., Prokop, M., & van Ginneken, B. (2017). White matter and gray matter segmentation in 4D computed tomography. *Scientific Reports*, *7*(1), 119.
- Martín-Pérez, A., Villa, M., Rosa Olmeda, G., Sancho, J., Vazquez, G., Urbanos, G., Martínez de Ternero, A., Chavarrías, M., Jimenez-Roldan, L., Perez-Nuñez, A., et al. (2025). SLIMBRAIN database: A multimodal image database of in vivo human brains for tumour detection. *Scientific Data*, *12*(1), 836.
- Meinshausen, N., & Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *72*(4), 417-473.
- Miranda, L. B., Braxton, E., Hobbs, J., & Quigley, M. R. (2011). Chronic subdural hematoma in the elderly: not a benign disease. *Journal of neurosurgery*, *114*(1), 72-76.
- Monteiro, M., Newcombe, V. F., Mathieu, F., Adatia, K., Kamnitsas, K., Ferrante, E., Das, T., Whitehouse, D., Rueckert, D., Menon, D. K., et al. (2020). Multiclass semantic segmentation and quantification of traumatic brain injury lesions on head CT using deep learning: an algorithm development and multicentre validation study. *The Lancet Digital Health*, *2*(6), e314-e322.
- Mori, S., & Zhang, J. (2006). Principles of diffusion tensor imaging and its applications to basic neuroscience research. *Neuron*, *51*(5), 527-539.
- Mühle, R., Ernst, H., Sobottka, S. B., & Morgenstern, U. (2021). Workflow and hardware for intraoperative hyperspectral data acquisition in neurosurgery. *Biomedical Engineering/Biomedizinische Technik*, *66*(1), 31-42.
- Muir, K. W., Ford, G. A., Messow, C.-M., Ford, I., Murray, A., Clifton, A., Brown, M. M., Madigan, J., Lenthall, R., Robertson, F., et al. (2017). Endovascular therapy for acute ischaemic stroke: the Pragmatic Ischaemic Stroke Thrombectomy Evaluation (PISTE) randomised, controlled trial. *Journal of Neurology, Neurosurgery & Psychiatry*, *88*(1), 38-44.
- Müller, D., Soto-Rey, I., & Kramer, F. (2022). Towards a guideline for evaluation metrics in medical image segmentation. *BMC Research Notes*, *15*(1), 210.
- Neftel, C., Laffy, J., Filbin, M. G., Hara, T., Shore, M. E., Rahme, G. J., Richman, A. R., Silverbush, D., Shaw, M. L., Hebert, C. M., et al. (2019). An integrative model of cellular states, plasticity, and genetics for glioblastoma. *Cell*, *178*(4), 835-849.
- Nguyen, T. A., Mai, T. D., Vu, L. D., Dao, C. X., Ngo, H. M., Hoang, H. B., Tran, T. A., Pham, T. Q., Pham, D. T., Nguyen, M. H., et al. (2023). Validation of the accuracy of the modified World Federation of Neurosurgical Societies subarachnoid hemorrhage

- grading scale for predicting the outcomes of patients with aneurysmal subarachnoid hemorrhage. *PLOS ONE*, 18(8), e0289267.
- Nijiati, M., Tuersun, A., Zhang, Y., Yuan, Q., Gong, P., Abulizi, A., Tuoheti, A., Abulaiti, A., & Zou, X. (2022). A symmetric prior knowledge based deep learning model for intracerebral hemorrhage lesion segmentation. *Frontiers in Physiology*, 13, 977427.
- Nimsky, C., Ganslandt, O., Buchfelder, M., & Fahlbusch, R. (2006). Intraoperative visualization for resection of gliomas: the role of functional neuronavigation and intraoperative 1.5 T MRI. *Neurological research*, 28(5), 482-487.
- Ocaña-Tienda, B., Pérez-Beteta, J., Ortiz de Mendivil, A., Asenjo, B., Albillo, D., Pérez-Romasanta, L. A., LLorente, M., Carballo, N., Arana, E., & Pérez-García, V. M. (2024). Morphological MRI features as prognostic indicators in brain metastases. *Cancer Imaging*, 24(1), 111.
- O’connor, J. P., Aboagye, E. O., Adams, J. E., Aerts, H. J., Barrington, S. F., Beer, A. J., Boellaard, R., Bohndiek, S. E., Brady, M., Brown, G., et al. (2017). Imaging biomarker roadmap for cancer studies. *Nature reviews Clinical oncology*, 14(3), 169-186.
- Ohkuma, H., Shimamura, N., Naraoka, M., & Katagai, T. (2017). Aneurysmal subarachnoid hemorrhage in the elderly over age 75: a systematic review. *Neurologia medico-chirurgica*, 57(11), 575-583.
- Omar II, A. T., & Starreveld, Y. P. (2022). Meningiomas associated with subdural hematomas: a systematic review of clinical features and outcomes. *World Neurosurgery*, 158, e465-e475.
- Orlhac, F., Lecler, A., Savatovski, J., Goya-Outi, J., Nioche, C., Charbonneau, F., Ayache, N., Frouin, F., Duron, L., & Buvat, I. (2021). How can we combat multicenter variability in MR radiomics? Validation of a correction procedure. *European radiology*, 31(4), 2272-2280.
- Osgood, M. L., & Singer, R. J. (2021). Aneurysmal subarachnoid hemorrhage: review of the pathophysiology and management strategies. *Current Neurology and Neuroscience Reports*, 21(9), 50. <https://doi.org/10.1007/s11910-021-01136-9>
- Ostrom, Q. T., Price, M., Neff, C., Cioffi, G., Waite, K. A., Kruchko, C., & Barnholtz-Sloan, J. S. (2022). CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2015–2019. *Neuro-oncology*, 24(Supplement_5), v1-v95.
- Özcan, H., Emiroğlu, B. G., Sabuncuoğlu, H., Özdoğan, S., Soyer, A., & Saygı, T. (2021). A comparative study for glioma classification using deep convolutional neural networks. *Molecular Biology and Evolution*.
- Papacci, F., Pedicelli, A., & Montano, N. (2015). The role of preoperative angiography in the management of giant meningiomas associated to vascular malformation. *Surgical Neurology International*, 6, 114.
- Parmar, C., Grossmann, P., Bussink, J., & Lambin, P. (2017). Machine learning methods for quantitative radiomic biomarkers. *Machine learning applications for Radiomics*, 5, 125.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., . . . Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. En H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché-Buc, E. Fox & R. Garnett (Eds.), *Advances in*

- Neural Information Processing Systems (NeurIPS 2019)* (Vol. 32). Curran Associates, Inc.
- Pei, L., Fang, T., Xu, L., & Ni, C. (2024). A radiomics model based on CT images combined with multiple machine learning models to predict the prognosis of spontaneous intracerebral hemorrhage. *World Neurosurgery*, *181*, e856-e866.
- Pelletier, D., Garrison, K., & Henry, R. (2004). Measurement of whole-brain atrophy in multiple sclerosis. *Journal of Neuroimaging*, *14*, 11S-19S.
- Peng, H., Huo, J., Li, B., Cui, Y., Zhang, H., Zhang, L., & Ma, L. (2021). Predicting isocitrate dehydrogenase (IDH) mutation status in gliomas using multiparameter MRI radiomics features. *Journal of Magnetic Resonance Imaging*, *53*(5), 1399-1407.
- Petrella, J. R., Coleman, R. E., & Doraiswamy, P. M. (2003). Neuroimaging and early diagnosis of Alzheimer disease: a look to the future. *Radiology*, *226*(2), 315-336.
- Raabe, A., Beck, J., Goldberg, J., Z'Graggen, W. J., Branca, M., Marbacher, S., D'Alonzo, D., Fandino, J., Stienen, M. N., Neidert, M. C., et al. (2022). Herniation World Federation of Neurosurgical Societies scale improves prediction of outcome in patients with poor-grade aneurysmal subarachnoid hemorrhage. *Stroke*, *53*(7), 2346-2351.
- Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. *Nature medicine*, *28*(1), 31-38.
- Rao, D., Yang, L., Enxi, X., Siyuan, L., Yu, Q., Zheng, L., Zhou, Z., Yerong, C., Bo, C., Xiuhong, S., et al. (2024). A predictive model in patients with chronic hydrocephalus following aneurysmal subarachnoid hemorrhage: a retrospective cohort study. *Frontiers in Neurology*, *15*, 1366306.
- Rao, G., Gao, H., Wang, X., Zhang, J., Ye, M., & Rao, L. (2023). MRI measurements of brain hippocampus volume in relation to mild cognitive impairment and Alzheimer disease: A systematic review and meta-analysis. *Medicine*, *102*(36), e34997.
- Rawat, W., & Wang, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. *Neural computation*, *29*(9), 2352-2449.
- Reuter, M., Schmansky, N. J., Rosas, H. D., & Fischl, B. (2012). Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage*, *61*(4), 1402-1418.
- Roberts, D. W., Valdés, P. A., Harris, B. T., Fontaine, K. M., Hartov, A., Fan, X., Ji, S., Lollis, S. S., Pogue, B. W., Leblond, F., et al. (2011). Coregistered fluorescence-enhanced tumor resection of malignant glioma: relationships between δ -aminolevulinic acid-induced protoporphyrin IX fluorescence, magnetic resonance imaging enhancement, and neuropathological parameters. *Journal of neurosurgery*, *114*(3), 595-603.
- Roddan, A., Czempiel, T., Xu, C., Elson, D. S., & Giannarou, S. (2025). SAMSA: Segment Anything Model Enhanced with Spectral Angles for Hyperspectral Interactive Medical Image Segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 478-488.
- Rokach, L. (2010). Ensemble-based classifiers. *Artificial intelligence review*, *33*(1), 1-39.
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, 234-241.

- Ru, X., Gao, L., Zhou, J., Li, Q., Zuo, S., Chen, Y., Liu, Z., & Feng, H. (2021). Secondary white matter injury and therapeutic targets after subarachnoid hemorrhage. *Frontiers in Neurology*, *12*, 659740.
- Rubin, D. B. (2018). Multiple imputation. En *Flexible imputation of missing data, second edition* (pp. 29-62). Chapman; Hall/CRC.
- Ruiz, L., Martín, A., Urbanos, G., Villanueva, M., Sancho, J., Rosa, G., Villa, M., Chavarrías, M., Pérez, Á., Juárez, E., et al. (2020). Multiclass brain tumor classification using hyperspectral imaging and supervised machine learning. *2020 XXXV conference on design of circuits and integrated systems (DCIS)*, 1-6.
- Rusche, T., Wasserthal, J., Breit, H.-C., Fischer, U., Guzman, R., Fiehler, J., Psychogios, M.-N., & Sporns, P. B. (2023). Machine learning for onset prediction of patients with intracerebral hemorrhage. *Journal of Clinical Medicine*, *12*(7), 2631.
- Said, M., Gümüs, M., Herten, A., Dinger, T. F., Chihi, M., Darkwah Oppong, M., Deuschl, C., Wrede, K. H., Kleinschnitz, C., Sure, U., et al. (2021). Subarachnoid Hemorrhage Early Brain Edema Score (SEBES) as a radiographic marker of clinically relevant intracranial hypertension and unfavorable outcome after subarachnoid hemorrhage. *European Journal of Neurology*, *28*(12), 4051-4059.
- Salahuddin, Z., Woodruff, H. C., Chatterjee, A., & Lambin, P. (2022). Transparency of deep neural networks for medical image analysis: A review of interpretability methods. *Computers in biology and medicine*, *140*, 105111.
- San Diego Brain Injury Foundation. (n.d.). What's the difference between all the different head scans? [Accessed: 2025-08-03].
- Schölkopf, B., & Smola, A. J. (2002). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press.
- Senft, C., Bink, A., Franz, K., Vatter, H., Gasser, T., & Seifert, V. (2011). Intraoperative MRI guidance and extent of resection in glioma surgery: a randomised, controlled trial. *The lancet oncology*, *12*(11), 997-1003.
- Shan, D., Wang, J., Qi, P., Lu, J., & Wang, D. (2023). Non-contrasted CT radiomics for SAH prognosis prediction. *Bioengineering*, *10*(8), 967.
- Sheline, Y. I., Barch, D. M., Price, J. L., Rundle, M. M., Vaishnavi, S. N., Snyder, A. Z., Mintun, M. A., Wang, S., Coalson, R. S., & Raichle, M. E. (2009). The default mode network and self-referential processes in depression. *Proceedings of the National Academy of Sciences*, *106*(6), 1942-1947.
- Shen, D., Wu, G., & Suk, H.-I. (2017). Deep learning in medical image analysis. *Annual review of biomedical engineering*, *19*(1), 221-248.
- Shetty, P., Yeole, U., Singh, V., & Moiyadi, A. (2021). Navigated ultrasound-based image guidance during resection of gliomas: practical utility in intraoperative decision-making and outcomes. *Neurosurgical Focus*, *50*(1), E14.
- Shi, M., Yang, C., Tang, Q.-w., Xiao, L.-f., Chen, Z.-h., & Zhao, W.-y. (2021). The prognostic value of neutrophil-to-lymphocyte ratio in patients with aneurysmal subarachnoid hemorrhage: a systematic review and meta-analysis of observational studies. *Frontiers in Neurology*, *12*, 745560.
- Siddique, N., Sidike, P., Elkin, C., & Devabhaktuni, V. (2020). U-Net and its variants for medical image segmentation: theory and applications. *arXiv preprint arXiv:2011.01118*.

- Smith, E. J., Gohil, K., Thompson, C. M., Naik, A., & Hassaneen, W. (2021). Fluorescein-guided resection of high grade gliomas: a meta-analysis. *World Neurosurgery*, *155*, 181-188.
- Smits, M. (2021). MRI biomarkers in neuro-oncology. *Nature Reviews Neurology*, *17*(8), 486-500.
- Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, *14*(3), 199-222.
- Song, C., Zhao, Q., Li, J., Yue, X., Gao, R., Wang, Z., Gao, A., & Fu, G. (2024). HemSeg-200: A voxel-annotated dataset for intracerebral hemorrhages segmentation in brain CT scans. *2024 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 3376-3377.
- Spearman, C. (1987). The proof and measurement of association between two things. *The American journal of psychology*, *100*(3/4), 441-471.
- SphereOptics GmbH. (2024). Zenith Lite™ Reflectance Targets [Accessed: 2025-08-14].
- Starke, R. M., Komotar, R. J., Otten, M. L., Schmidt, J. M., Fernandez, L. D., Rincon, F., Gordon, E., Badjatia, N., Mayer, S. A., & Connolly, E. S. (2009). Predicting long-term outcome in poor grade aneurysmal subarachnoid haemorrhage patients utilising the Glasgow Coma Scale. *Journal of Clinical Neuroscience*, *16*(1), 26-31.
- Steyerberg, E. W., & Vergouwe, Y. (2014). Towards better clinical prediction models: seven steps for development and an ABCD for validation. *European heart journal*, *35*(29), 1925-1931.
- Student. (1908). The probable error of a mean. *Biometrika*, 1-25.
- Stummer, W., Pichlmeier, U., Meinel, T., Wiestler, O. D., Zanella, F., & Reulen, H.-J. (2006). Fluorescence-guided surgery with 5-aminolevulinic acid for resection of malignant glioma: a randomised controlled multicentre phase III trial. *The lancet oncology*, *7*(5), 392-401.
- Stupp, R., Mason, W. P., Van Den Bent, M. J., Weller, M., Fisher, B., Taphoorn, M. J., Belanger, K., Brandes, A. A., Marosi, C., Bogdahn, U., et al. (2005). Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *New England journal of medicine*, *352*(10), 987-996.
- Suero Molina, E., Black, D., Xie, A., Gill, J., Di Ieva, A., & Stummer, W. (2024). Machine and Deep Learning in Hyperspectral Fluorescence-Guided Brain Tumor Surgery. *Computational Neurosurgery*, 245-264.
- Sutton, P. A., van Dam, M. A., Cahill, R. A., Mieog, S., Polom, K., Vahrmeijer, A. L., & van der Vorst, J. (2023). Fluorescence-guided surgery: comprehensive review. *BJS open*, *7*(3), zrad049.
- Szandała, T. (2020). Review and comparison of commonly used activation functions for deep neural networks. En *Bio-inspired neurocomputing* (pp. 203-224). Springer.
- Tae, W.-S., Ham, B.-J., Pyun, S.-B., Kang, S.-H., & Kim, B.-J. (2018). Current clinical applications of diffusion-tensor imaging in neurological disorders. *Journal of clinical neurology (Seoul, Korea)*, *14*(2), 129.
- Teasdale, G., Drake, C., Hunt, W., Kassell, N., Sano, K., Pertuiset, B., & De Villiers, J. (1988). A universal subarachnoid hemorrhage scale: report of a committee of the World Federation of Neurosurgical Societies. *Journal of neurology, neurosurgery, and psychiatry*, *51*(11), 1457.

- Teasdale, G., & Jennett, B. (1974). Assessment of coma and impaired consciousness: a practical scale. *The Lancet*, *2*(7872), 81-84. [https://doi.org/10.1016/S0140-6736\(74\)91639-0](https://doi.org/10.1016/S0140-6736(74)91639-0)
- Thévenaz, P., Blu, T., & Unser, M. (2002). Interpolation revisited [medical images application]. *IEEE Transactions on medical imaging*, *19*(7), 739-758.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 267-288.
- Tonidandel, S., & LeBreton, J. M. (2011). Relative importance analysis: A useful supplement to regression analysis. *Journal of Business and Psychology*, *26*(1), 1-9.
- Tuan, P. M., Adel, M., Trung, N. L., Horowitz, T., Parlak, I. B., & Guedj, E. (2025). FDG-PET-based brain network analysis: a brief review of metabolic connectivity. *EJNMMI reports*, *9*(1), 4.
- Turc, G., Bhogal, P., Fischer, U., Khatri, P., Lobotesis, K., Mazighi, M., Schellinger, P. D., Toni, D., De Vries, J., White, P., et al. (2023). European Stroke Organisation (ESO)-European Society for Minimally Invasive Neurological Therapy (ESMINT) guidelines on mechanical thrombectomy in acute ischemic stroke. *Journal of neurointerventional surgery*, *15*(8), e8-e8.
- Urbanos, G., Castaño-León, A. M., Maldonado-Luna, M., Salvador, E., Ramos, A., Lechuga, C., Sanz, C., Juárez, E., & Lagares, A. (2025). Comprehensive predictive modeling in subarachnoid hemorrhage: integrating radiomics and clinical variables. *Neurosurgical Review*, *48*(1), 528.
- Urbanos, G., Martín, A., Vázquez, G., Villanueva, M., Villa, M., Jimenez-Roldan, L., Charvarrias, M., Lagares, A., Juárez, E., & Sanz, C. (2021). Supervised machine learning methods and hyperspectral imaging techniques jointly applied for brain cancer classification. *Sensors*, *21*(11), 3827.
- Valdés, P. A., Leblond, F., Kim, A., Harris, B. T., Wilson, B. C., Fan, X., Tosteson, T. D., Hartov, A., Ji, S., Erkmén, K., et al. (2011). Quantitative fluorescence in intracranial tumor: implications for ALA-induced PpIX as an intraoperative biomarker. *Journal of neurosurgery*, *115*(1), 11-17.
- van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*(3), 1-67. <https://doi.org/10.18637/jss.v045.i03>
- Van Calster, B., McLernon, D. J., Van Smeden, M., Wynants, L., Steyerberg, E. W., diagnostic tests, T. G. E., & prediction models' of the STRATOS initiative Bossuyt Patrick Collins Gary S. Macaskill Petra McLernon David J. Moons Karel GM Steyerberg Ewout W. Van Calster Ben van Smeden Maarten Vickers Andrew J. (2019). Calibration: the Achilles heel of predictive analytics. *BMC medicine*, *17*(1), 230.
- Van Gijn, J., Kerr, R. S., & Rinkel, G. J. (2007). Subarachnoid haemorrhage. *The Lancet*, *369*(9558), 306-318.
- Van Griethuysen, J. J., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R. G., Fillion-Robin, J.-C., Pieper, S., & Aerts, H. J. (2017). Computational radiomics system to decode the radiographic phenotype. *Cancer research*, *77*(21), e104-e107.
- Vande Vyvere, T., Pisciă, D., Wilms, G., Claes, L., Van Dyck, P., Snoeckx, A., van Den Hauwe, L., Pullens, P., Verheyden, J., Wintermark, M., et al. (2024). Imaging findings in acute traumatic brain injury: a national institute of neurological disorders and stroke

- common data element-based pictorial review and analysis of over 4000 admission brain computed tomography scans from the collaborative European NeuroTrauma effectiveness research in traumatic brain injury (CENTER-TBI) study. *Journal of Neurotrauma*, 41(19-20), 2248-2297.
- Velthuis, B. K., van Leeuwen, M. S., Witkamp, T. D., Ramos, L. M., van der Sprenkel, J. W. B., & Rinkel, G. J. (1999). Computerized tomography angiography in patients with subarachnoid hemorrhage: from aneurysm detection to treatment without conventional angiography. *Journal of neurosurgery*, 91(5), 761-767.
- Villa, M., Sancho, J., Rosa, G., Chavarrias, M., Juarez, E., & Sanz, C. (2024). HyperMRI: hyperspectral and magnetic resonance fusion methodology for neurosurgery applications. *International Journal of Computer Assisted Radiology and Surgery*, 19(7), 1367-1374.
- Villegas-Gomez, G. A., Ramos-Márquez, A., Hakim, F., Gómez-Amarillo, D. F., Ramón, J. F., & Ordóñez-Rubiano, E. G. (2022). Navigated Intraoperative Ultrasound Offers Effective and Efficient Real-Time Analysis of Intracranial Tumor Resection and Brain Shift. *Operative Neurosurgery*, 10-1227.
- Wach, J., Vychopen, M., Basaran, A. E., Güresir, A., Seidel, C., Kühnapfel, A., & Güresir, E. (2024). Efficacy and safety of intraoperative MRI in glioma surgery: a systematic review and meta-analysis of prospective randomized controlled trials. *Journal of Neurosurgery*, 1(aop), 1-12.
- Waldron, J. R., Crockett, M., & Phillips, T. (2018). Hyperacute intracranial aneurysm rebleed captured on planar and 3D digital subtraction angiography. *Radiology Case Reports*, 13(5), 1012-1017.
- Wang, H., Sun, Y., Ge, Y., Wu, P.-Y., Lin, J., Zhao, J., & Song, B. (2021). A clinical-radiomics nomogram for functional outcome predictions in ischemic stroke. *Neurology and Therapy*, 10(2), 819-832.
- Wang, Y., Liu, G., Hong, D., Chen, F., Ji, X., & Cao, G. (2016). White matter injury in ischemic stroke. *Progress in neurobiology*, 141, 45-60.
- Wasserthal, J., Breit, H.-C., Meyer, M. T., Pradella, M., Hinck, D., Sauter, A. W., Heye, T., Boll, D. T., Cyriac, J., Yang, S., et al. (2023). TotalSegmentator: robust segmentation of 104 anatomic structures in CT images. *Radiology: Artificial Intelligence*, 5(5), e230024.
- Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., Harvey, D., Jack Jr, C. R., Jagust, W., Morris, J. C., et al. (2017). The Alzheimer's Disease Neuroimaging Initiative 3: Continued innovation for clinical trial improvement. *Alzheimer's & Dementia*, 13(5), 561-571.
- Weisberg, S. (2005). *Applied linear regression* (Vol. 528). John Wiley & Sons.
- Wintermark, M. (2005). Brain perfusion-CT in acute stroke patients. *European Radiology Supplements*, 15, d28-d31.
- Wong, S. C., Nawawi, O., Ramli, N., & Abd Kadir, K. A. (2012). Benefits of 3D rotational DSA compared with 2D DSA in the evaluation of intracranial aneurysm. *Academic radiology*, 19(6), 701-707.
- Wu, B., Xie, Y., Zhang, Z., Ge, J., Yaxley, K., Bahadir, S., Wu, Q., Liu, Y., & To, M.-S. (2023). Bhsd: A 3d multi-class brain hemorrhage segmentation dataset. *International workshop on machine learning in medical imaging*, 147-156.

- Wu, M., Yu, M., Jing, S., Yap, P.-T., Zhang, Z., & Liu, M. (2024). Unpaired Volumetric Harmonization of Brain MRI with Conditional Latent Diffusion. *arXiv preprint arXiv:2408.09315*.
- XIMEA. (—). XIMEA xiSpec hyperspectral miniature camera (SM range 600–1000 nm, USB 3) [Accedido: 16 de agosto de 2025].
- XIMEA GmbH. (2017). *xiSpec01 Software Manual, Beta Version V01* [Accedido: 10 agosto 2025].
- Yeatts, S. D., & Martin, R. H. (2015). What is missing from my missing data plan? *Stroke*, *46*(6), e130-e132.
- Yelanchezian, M., Gonzalez-Prieto, C., Oulaghan, B., Yates, S., Morgan, C., Dobbie, G., Davis, D., & Cullum, S. (2025). The diagnostic accuracy of CTseg segmentation software for dementia in a New Zealand memory service. *Journal of Alzheimer's Disease Reports*, *9*, 25424823251332448.
- Yeo, M., Tahayori, B., Kok, H. K., Maingard, J., Kutaiba, N., Russell, J., Thijs, V., Jhamb, A., Chandra, R. V., Brooks, M., et al. (2021). Review of deep learning algorithms for the automatic detection of intracranial hemorrhages on computed tomography head imaging. *Journal of neurointerventional surgery*, *13*(4), 369-378.
- Young, H. P. (1974). An axiomatization of Borda's rule. *Journal of economic theory*, *9*(1), 43-52.
- Zhou, H.-Y., Guo, J., Zhang, Y., Yu, L., Wang, L., & Yu, Y. (2021). nnformer: Interleaved transformer for volumetric segmentation. *arXiv preprint arXiv:2109.03201*.
- Zhou, Z., Sodha, V., Rahman Siddiquee, M. M., Feng, R., Tajbakhsh, N., Gotway, M. B., & Liang, J. (2019). Models genesis: Generic autodidactic models for 3d medical image analysis. *International conference on medical image computing and computer-assisted intervention*, 384-393.
- Zunin, V. (2021). Intel openvino toolkit for computer vision: Object detection and semantic segmentation. *2021 International Russian Automation Conference (RusAutoCon)*, 847-851.
- Zwanenburg, A., Leger, S., Vallières, M., & Löck, S. (2016). Image biomarker standardisation initiative. *arXiv preprint arXiv:1612.07003*.
- Zwanenburg, A., Vallières, M., Abdalah, M. A., Aerts, H. J., Andrearczyk, V., Apte, A., Ashrafinia, S., Bakas, S., Beukinga, R. J., Boellaard, R., et al. (2020). The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*, *295*(2), 328-338.

Anexos

Anexo A Configuración técnica y experimental del sistema HSI en glioblastoma

A.1 Especificaciones técnicas del sistema HSI intraoperatorio

Tabla A.1: Estructura de archivos y contenido asociado a cada adquisición hiperespectral.

Elemento	Formato	Descripción
Cubo hiperespectral	.raw	Datos brutos en formato nativo
Referencia blanca	.raw	Imagen de calibración
Referencia negra	.raw	Captura sin iluminación para corrección de ruido
Metadatos técnicos	.json / .csv	Parámetros ópticos, geométricos y de iluminación
Información contextual	.csv / .txt	ID paciente, fase quirúrgica, operador

Tabla A.2: Resumen de especificaciones del sistema HSI intraoperatorio.

Parámetro	Especificación
Cámara	Ximea MQ022HG-IM-SM5X5-NIR (sensor VNIR, 25 bandas, 655–975 nm) XIMEA, —
Resolución espacial	217 x 409 píxeles
Lente	VIS/NIR 35 mm, baja distorsión
Filtro	Corte a 650 nm
Tiempo de exposición	70 ms (fijo)
Apertura	$f/4$ (fija)
Fuente de iluminación	Halógena 150 W, doble fibra lateral (180 cm, 1.2 cm diámetro)
Rango geométrico típico	Distancia cámara–superficie: 37–58 cm; incidencia: 25°–45°
Peso / dimensiones	32 g; 26 x 26 x 26 mm
Tasa máxima de adquisición	170 cubos/segundo (teórica)

A.2 Interfaz y visualización del sistema de anotación de las imágenes HSI intraoperatorias

La interfaz gráfica presenta tres paneles principales dispuestos horizontalmente, representada en la Figura A.1:

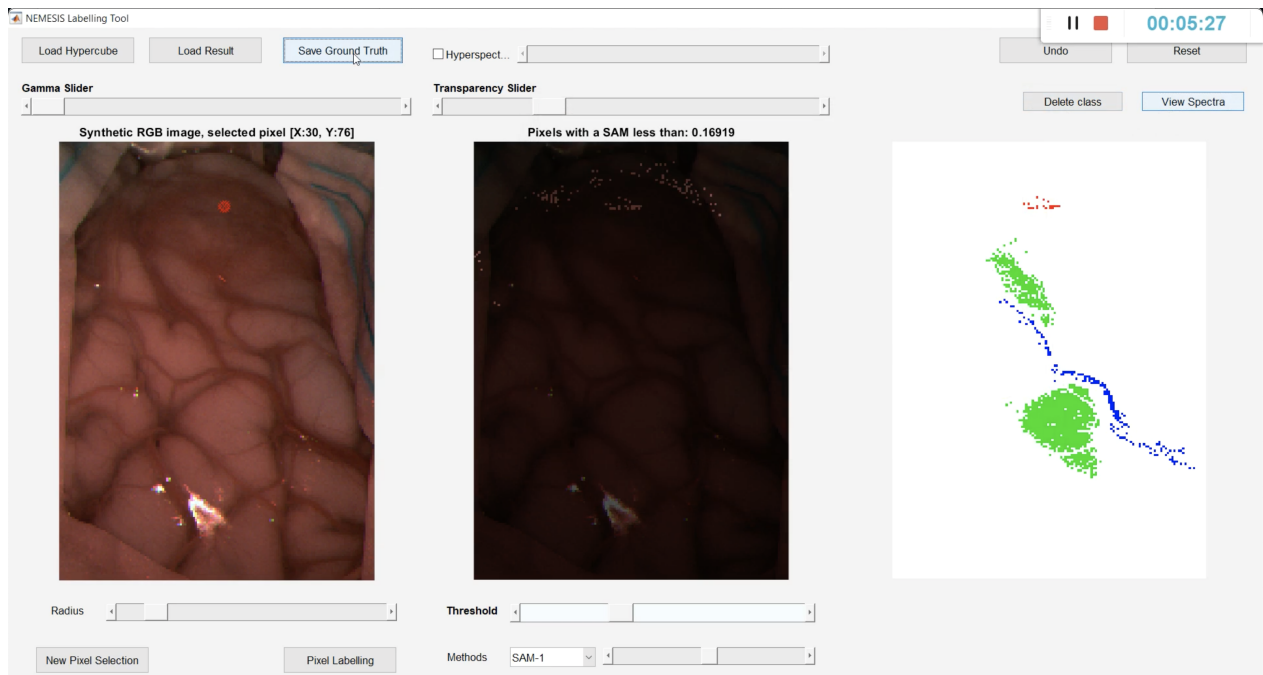


Figura A.1: Interfaz de la herramienta de etiquetado en HSI, mostrando los tres paneles principales: imagen RGB de referencia (izquierda), máscara de selección por similitud espectral (centro) y acumulación de etiquetas (*ground truth*) (derecha).

- **Izquierda:** imagen RGB utilizada para la selección del píxel o región de referencia. Para

configurar esta imagen RGB se ha realizado un renderizado con las siguientes longitudes de onda centrales: $R = 828,3$ nm, $G = 739,2$ nm y $B = 668,7$ nm. Estas longitudes de onda se eligen para cubrir las regiones alta, media y baja del espectro VNIR medido por el sensor (665–960 nm), maximizando el contraste visual entre parénquima, vasculatura y tumor.

- **Centro:** máscara de selección obtenida por similitud espectral, con opción de superponer la imagen RGB en modo semitransparente para facilitar la correspondencia anatómica.
- **Derecha:** acumulación de las máscaras ya etiquetadas, que constituyen el *ground truth* en construcción.

La herramienta incorpora controles deslizantes para ajustar el brillo, la opacidad y el umbral de similitud en tiempo real. Asimismo, permite visualizar las firmas espectrales medias de cada categoría etiquetada.

A.3 Configuración de hiperparámetros de los modelos de clasificación tisular

Tabla A.3: Configuración de los hiperparámetros empleados para cada algoritmo supervisado.

Modelo	Hiperparámetro	Valor
SVM (LinearSVC)	C	1.0
	Penalización	L_2
	<code>class_weight</code>	None
	<code>tol</code>	1×10^{-4}
Random Forest	$n_estimators$	100
	Criterio	<code>gini</code>
	<code>max_depth</code>	None
	<code>min_samples_split</code>	2
	<code>min_samples_leaf</code>	1
	<code>max_features</code>	<code>"sqrt"</code>
3D-CNN	Tamaño de entrada	$9 \times 9 \times 25$
	Filtros (<code>conv1</code> / <code>conv2</code>)	128 / 256
	Dropout	0.3 / 0.5
	Optimizador	AdamW
	Learning rate	5×10^{-5}
	Weight decay	0.01
	Épocas	2
	Batch size	1024

Anexo B Configuración y resultados complementarios de los modelos aplicados a Hemorragia Subaracnoidea Aneurismática en TC

B.1 Modelos de segmentación automática de hemorragia

Interfaz y funcionalidades de segmentación en *QUIBIM Precision*[®]

La plataforma *QUIBIM Precision*[®] se ha utilizado como entorno principal para la anotación y segmentación manual de las hemorragias intracraneales. Esta herramienta, orientada al análisis cuantitativo de imagen médica, incorpora funciones avanzadas de delineado y visualización que agilizan la creación del *ground truth*.

El sistema ofrecía dos modalidades de segmentación:

- **Manual mediante pincel digital**, que permite delinear con precisión milimétrica regiones de interés de morfología compleja.
- **Semiautomática por homogeneidad de intensidad**, en la que el algoritmo ajusta automáticamente los límites tras definir una región inicial, reduciendo el tiempo de delineado sin pérdida de exactitud.

Las máscaras resultantes se exportan en formato NIfTI para su análisis cuantitativo. Ejemplos del visor y de las segmentaciones se muestran en la Figura B.1.

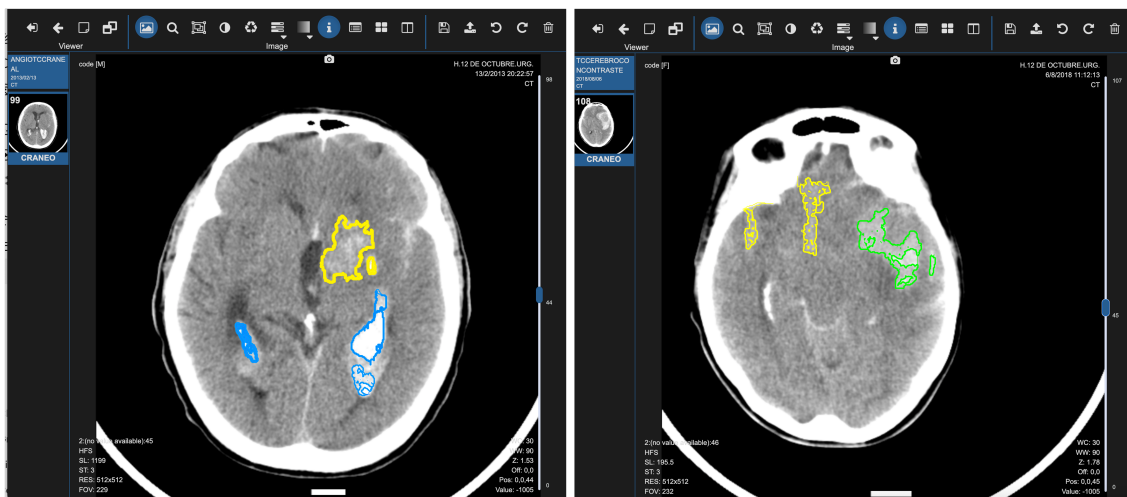


Figura B.1: Capturas del visor *QUIBIM Precision*[®] que muestran ejemplos de segmentaciones en dos pacientes distintos. A la izquierda se representan las regiones correspondientes a HSAa (amarillo) y HIV (azul); a la derecha, la HSAa (amarillo) y HIP (verde).

Hiperparámetros utilizados en el modelo de segmentación

Tabla B.1: Configuración detallada de los parámetros arquitectónicos y de entrenamiento utilizados en los Experimentos 1 y 2 con *nnU-Net*.

Parámetro	Experimento 1	Experimento 2
Datos		
Casos de entrenamiento	359	633
Casos de test externo	89	Ídem
Arquitectura		
Tamaño del <i>patch</i> de entrada	$224 \times 224 \times 40$	$192 \times 192 \times 64$
Espaciado final (mm)	$1,624 \times 0,819 \times 0,819$	$1,610 \times 0,811 \times 0,811$
Número base de <i>feature maps</i>	32	Ídem
Número de etapas / conv. por etapa	6 / 2	Ídem
Tamaño de <i>kernel</i> por etapa	[1, 3, 3]; [3, 3, 3]	resto Todos [3, 3, 3]
Entrenamiento		
<i>Batch size</i>	2	Ídem
Optimizador	SGD con momento de Nesterov	Ídem
<i>Momentum</i>	0.99	Ídem
<i>Weight decay</i>	3×10^{-5}	Ídem
Tasa de aprendizaje inicial	1×10^{-3}	Ídem
Épocas máximas	1000	Ídem
Criterio de parada temprana	Sin mejora en 50 épocas	Ídem
Aumento de datos (<i>data augmentation</i>)		
Rotación	$p = 0,2, \pm 30^\circ$	Ídem
Escalado	$p = 0,2, [0,7, 1,4]$	Ídem
Espejo	Ejes (0, 1, 2)	Ídem
Gamma	$p = 0,3, [0,7, 1,5]$	Ídem
Deformaciones elásticas	Activadas	Ídem
Brillo / contraste	Activadas	Ídem

B.2 Modelos basados en radiómica

Definición de las variables clínicas incluidas

Tabla B.2: Resumen de variables clínicas categóricas y continuas incluidas en el estudio, con sus valores o descripción correspondiente.

Variable	Valores / Descripción
Variables categóricas	
Sexo	0 = Hombre 1 = Mujer
Hipertensión	0 = No 1 = Sí
Tabaquismo	0 = No 1 = Sí
Diabetes	0 = No 1 = Sí
Dislipemia / obesidad	0 = No 1 = Sí
Alcoholismo	0 = No 1 = Sí
Otras drogas	0 = No 1 = Sí
Antecedentes familiares de HSAa	0 = No 1 = Sí
Arteriografía negativa	0 = No 1 = Sí tras una angiografía 2 = Sí tras dos angiografías
HSAa no filiada o idiopática	0 = No 1 = Sí
Patrón TAC	0 = Normal 1 = Perimesencefálico 2 = Aneurismático
WFNS al ingreso	1, 2, 3, 4, 5
Fisher	1, 2, 3, 4
Fisher Modificado	0, 1, 2, 3, 4
Aneurisma múltiple	0 = No 1 = Sí
MAV asociada	0 = No 1 = Sí
Circulación	0 = Anterior 1 = Posterior
WFNS pretratamiento	1, 2, 3, 4, 5
Tratamiento	0 = No tratamiento 1 = Quirúrgico 2 = Endovascular 3 = Mixto
Angiografía postratamiento	0 = No angio 1 = Exclusión completa 2 = Relleno de cuello 3 = Relleno de saco
Ingreso en UCI	0 = No 1 = Sí
Vasoespasmio	0 = No 1 = Angiográfico 2 = Clínico
Hidrocefalia	0 = No 1 = Temporal 2 = Requiere válvula definitiva
Resangrado	0 = No 1 = Sí
GOS al alta	1 = Muerte 2 = Estado vegetativo 3 = Discapacidad severa 4 = Discapacidad moderada 5 = Buena evolución
Variables continuas	
Record ID	Identificador único del registro
Fecha de nacimiento	Fecha de nacimiento del paciente
Fecha de HSAa	Fecha del evento hemorrágico
GOT	Valor sérico (U/L)
GPT	Valor sérico (U/L)
Proteína C reactiva	Concentración (mg/L)
Plaquetas	Recuento ($10^9/L$)
Leucocitos	Recuento ($10^9/L$)
Neutrófilos	Recuento ($10^9/L$)
Linfocitos	Recuento ($10^9/L$)
Monocitos	Recuento ($10^9/L$)
Eosinófilos	Recuento ($10^9/L$)
Basófilos	Recuento ($10^9/L$)
Fibrinógeno	Concentración plasmática (g/L)
Actividad de protrombina	Porcentaje de actividad
TTPA	Tiempo de tromboplastina parcial activada (s)
Glucosa	Concentración plasmática (mg/dL)
Potasio	Concentración sérica (mmol/L)
Sodio	Concentración sérica (mmol/L)

Tabla B.3: Valores fisiológicos de referencia de las variables clínicas incluidas.

Variable	Unidad	Rango normal
Glucosa (ayunas)	mg/dL	70–100
Plaquetas	$\times 10^9/L$	150–450
Sodio	mmol/L	135–145
Potasio	mmol/L	3,5–5,0
Leucocitos	$\times 10^9/L$	4–11
Neutrófilos	$\times 10^9/L$	2–7,5
Linfocitos	$\times 10^9/L$	1–4

Tabla B.4: Definiciones y frecuencias de las comorbilidades en la base de datos clínica.

Comorbilidad	Definición	Sí (%)	No (%)
Hipertensión arterial	Antecedente de presión arterial sistólica ≥ 130 mmHg, diastólica ≥ 80 mmHg o tratamiento antihipertensivo al ingreso.	41,69	58,31
Tabaquismo activo	Consumo de tabaco en el momento del ingreso.	28,54	71,46
Diabetes	Diagnóstico conocido de diabetes o tratamiento con insulina o antidiabéticos orales.	18,60	81,40
Dislipemia/obesidad	Antecedente de dislipemia (LDL ≥ 130 mg/dL) y/o IMC ≥ 30 kg/m ² .	17,86	82,14
Alcoholismo	Trastorno por consumo de alcohol documentado o consumo crónico excesivo.	5,71	94,29
Consumo de otras drogas	Consumo activo de drogas recreativas (p. ej., cocaína, cannabis, opioides).	3,22	96,78
Antecedentes familiares de HSAa	Familiar de primer grado con antecedente de hemorragia subaracnoidea.	1,49	98,51
MAV asociada	Diagnóstico de malformación arteriovenosa asociada a aneurisma.	1,40	98,60

Cumplimiento de la guía TRIPOD en la notificación de modelos predictivos

Sección / Tema	Ítem	Descripción de la lista de verificación	Secciones
Título y Resumen	1	Identificar el estudio como desarrollo y/o validación de un modelo multivariable de predicción, la población diana y el desenlace a predecir.	Título
	2	Proporcionar un resumen con objetivos, diseño del estudio, escenario, participantes, tamaño muestral, predictores, desenlace, análisis estadístico, resultados y conclusiones.	Resumen
Introducción	3a	Explicar el contexto médico (diagnóstico o pronóstico) y la justificación para desarrollar o validar el modelo, incluyendo referencias a modelos previos.	Cap. 1: Motivación; Cap.2: Estado del arte
	3b	Especificar los objetivos, indicando si el estudio describe desarrollo, validación o ambos.	Cap. 1: Objetivos
Métodos	4a	Describir el diseño del estudio o la fuente de datos (p. ej., cohorte, registro), separando los conjuntos de desarrollo y validación si procede.	Cap. 3: Sección 3.1; Cap. 4: Sección 4.1
	4b	Especificar las fechas del estudio: inicio/fin del reclutamiento y seguimiento.	Cap. 3: Sección 4.1
	5a	Describir el ámbito del estudio, número y localización de centros.	Cap. 3: Sección 3.1; Cap. 4: Sección 4.1
	5b	Detallar los criterios de elegibilidad de los participantes.	Cap. 3: Sección 3.1
	5c	Proporcionar información sobre los tratamientos recibidos, si es relevante.	No aplicable
	6a	Definir el desenlace predicho por el modelo, incluyendo cómo y cuándo fue evaluado.	Cap. 3: Sección 3.2
	6b	Informar de cualquier acción realizada para enmascarar la evaluación del desenlace.	No aplicable
	7a	Definir todos los predictores utilizados, incluyendo cómo y cuándo fueron medidos.	Cap. 3: Sección 3.2
	7b	Informar de cualquier acción realizada para enmascarar la evaluación de los predictores.	No aplicable

Continúa en la siguiente página

Sección Tema	/ Ítem	Descripción de la lista de verificación	Secciones
	8	Explicar cómo se determinó el tamaño muestral.	Cap. 3: Sección 3.1; Cap. 4: Sección 4.1
	9	Describir cómo se manejaron los datos perdidos, incluyendo el método de imputación.	Cap. 3: Sección 3.2
	10a	Describir cómo se trataron los predictores en el análisis.	Cap. 3: Sección 3.2 y Sección 3.3
	10b	Especificar el tipo de modelo, los procedimientos de construcción y la validación interna.	Cap. 3: Sección 3.4
	10d	Especificar las métricas utilizadas para evaluar el rendimiento del modelo y compararlo con otros.	Cap. 3: Sección 3.4; Anexo B: Sección C
	11	Describir cómo se crearon los grupos de riesgo, si corresponde.	No aplicable
Resultados	13a	Describir el flujo de participantes, número con/sin desenlace y tiempo de seguimiento. Puede incluirse un diagrama.	Cap. 4: Sección 4.3.3
	13b	Describir las características de los participantes, incluyendo los datos faltantes.	Cap. 4: Sección 4.1
	14a	Especificar el número de participantes y eventos de desenlace por análisis.	Cap. 4: Sección 4.1
	14b	Informar de las asociaciones no ajustadas entre predictores y desenlace (si se realizaron).	Cap. 4: Sección 4.2
	15a	Presentar el modelo completo	Anexo B: Sección 4.3
	15b	Explicar cómo utilizar el modelo.	Cap. 3: Sección 3.2 y 3.4; Capítulo 5. Sección 5.2; Anexo A: Sección A.3; Anexo B: Sección B.1; Anexo D: Detalles de implementación
	16	Informar de las medidas de rendimiento del modelo (con intervalos de confianza).	Cap. 4: Sección 4.3
Discusión	18	Discutir las limitaciones del estudio.	Cap. 5: Sección 5.3
	19b	Interpretar los resultados considerando los objetivos, limitaciones y otras evidencias.	Cap. 5: Sección 5.1
	20	Discutir el uso clínico del modelo y las implicaciones para futuras investigaciones.	Cap. 5: Sección 5.2

Continúa en la siguiente página

Sección Tema	/ Ítem	Descripción de la lista de verificación	Secciones
	21	Proporcionar información sobre recursos suplementarios.	Anexos A, B, D
	22	Indicar la fuente de financiación y el papel de los financiadores.	Cap. 6: Sección 6.3

Tabla B.5: Cumplimiento de la guía TRIPOD para la notificación transparente del desarrollo y evaluación de modelos predictivos.

Espacios de búsqueda y configuraciones de hiperparámetros

En esta sección se detallan los espacios de búsqueda empleados para la optimización de hiperparámetros de los modelos predictivos descritos en la Sección 3.4.1. El ajuste de parámetros se realizó exclusivamente dentro del bucle interno de la validación cruzada anidada, garantizando que ninguna información del conjunto de test externo (bucle externo) fuera utilizada durante la fase de optimización. El criterio de selección se basó en el AUC media obtenida en validación interna.

Random Forest (RF)

El modelo RF Breiman, 2001 se ha configurado con ponderación automática de clases (`class_weight='balanced'`) para mitigar el efecto del desbalanceo entre grupos. El espacio de búsqueda se representa en la Tabla B.5

Tabla B.5: Espacio de búsqueda de hiperparámetros para RF.

Hiperparámetro	Descripción	Valores explorados
<code>n_estimators</code>	Número de árboles en el bosque	{100, 200, 300}
<code>max_depth</code>	Profundidad máxima de los árboles	{2, 3}
<code>min_samples_split</code>	Mínimo de muestras para dividir un nodo	{3, 4, 5}
<code>min_samples_leaf</code>	Mínimo de muestras por hoja	{2, 3, 4}
<code>max_features</code>	Fracción de variables por árbol	{0.15, 0.3, 0.5}
<code>criterion</code>	Función de impureza	{gini, entropy}

Extremely Randomized Trees (ExtraTrees)

El algoritmo *Extra Trees* Geurts et al., 2006 comparte una estructura similar a RF, pero introduce una mayor aleatoriedad en las divisiones de los nodos, reduciendo la varianza a costa de un ligero incremento en el sesgo. Se activaron las opciones de `oob_score=True` y `bootstrap=True`.

Extreme Gradient Boosting (XGBoost)

El modelo *XGBoost* T. Chen y Guestrin, 2016 se ha configurado con función objetivo logística binaria (`objective='binary:logistic'`) y métrica de evaluación basada en AUC

Tabla B.6: Espacio de búsqueda de hiperparámetros para *Extra Trees*.

Hiperparámetro	Descripción	Valores explorados
<code>n_estimators</code>	Número de árboles	{100, 200, 300}
<code>max_depth</code>	Profundidad máxima de los árboles	{2, 3}
<code>min_samples_split</code>	Mínimo de muestras para dividir un nodo	{3, 4, 5}
<code>min_samples_leaf</code>	Mínimo de muestras por hoja	{2, 3, 4}
<code>max_features</code>	Fracción de variables por árbol	{0.15, 0.3, 0.5}
<code>criterion</code>	Función de impureza	{gini, entropy}

(`eval_metric='auc'`). El parámetro `scale_pos_weight` se ajusta dinámicamente en cada fold de acuerdo con la proporción entre clases positiva y negativa.

Tabla B.6: Espacio de búsqueda de hiperparámetros para *XGBoost*.

Hiperparámetro	Descripción	Valores explorados
<code>eta</code>	Tasa de aprendizaje	0.001–0.15 (10 valores)
<code>gamma</code>	Penalización del árbol	10-20 (5 valores)
<code>max_depth</code>	Profundidad máxima	{2, 3, 4}
<code>subsample</code>	Proporción de muestras	{0.3, 0.5, 0.7}
<code>colsample_bytree</code>	Proporción de características	{0.3, 0.5, 0.7}

Resultados de predicción basados en la edad

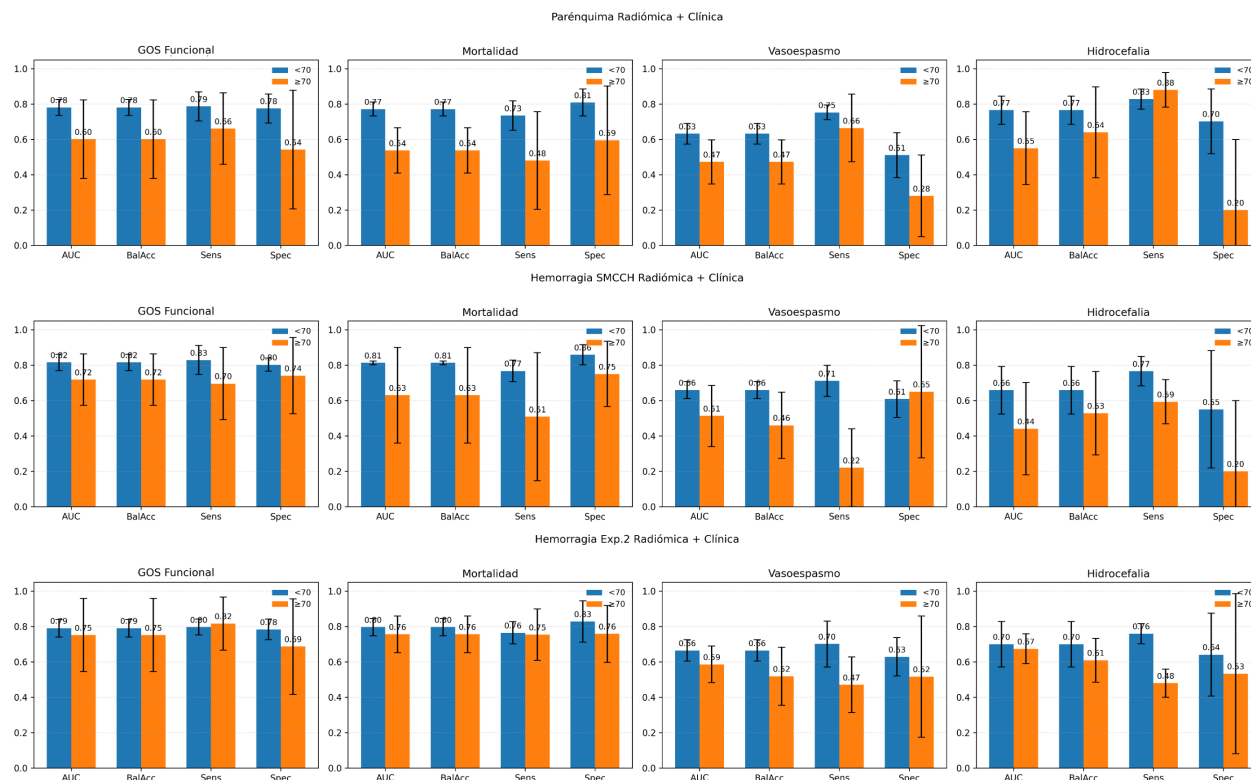


Figura B.2: Rendimiento de los modelos combinando clínica y radiómica a partir de distintas regiones, estratificado por grupos de edad (<70 y ≥70 años).

Resultados de predicción por grupos clínicos

Tabla B.7: Resultados de predicción por grupos clínicos en los distintos modelos y desenlaces.

Modelo	Grupo	AUC	BalAcc	Sens	Spec
Mortalidad					
Clínica	Buen desarrollo	0,80 ± 0,03	0,72 ± 0,05	0,75 ± 0,07	0,71 ± 0,20
	Mal desarrollo	0,76 ± 0,05	0,68 ± 0,08	0,67 ± 0,06	0,82 ± 0,21
Radiómica Parénquima	Buen desarrollo	0,65 ± 0,10	0,65 ± 0,10	0,79 ± 0,05	0,55 ± 0,12
	Mal desarrollo	0,63 ± 0,11	0,63 ± 0,11	0,74 ± 0,06	0,50 ± 0,15
Clínica + Radiómica Parénquima	Good clinic grade	0,65 ± 0,10	0,65 ± 0,10	0,79 ± 0,05	0,55 ± 0,12
	Poor clinic grade	0,63 ± 0,11	0,63 ± 0,11	0,74 ± 0,06	0,50 ± 0,15
Radiómica HSAa (SMCCH)	Buen desarrollo	0,68 ± 0,09	0,68 ± 0,09	0,85 ± 0,05	0,52 ± 0,17
	Mal desarrollo	0,66 ± 0,10	0,66 ± 0,10	0,80 ± 0,07	0,55 ± 0,20
Clínica + Radiómica HSAa (SMCCH)	Buen desarrollo	0,70 ± 0,05	0,70 ± 0,05	0,90 ± 0,03	0,50 ± 0,10
	Mal desarrollo	0,65 ± 0,07	0,65 ± 0,07	0,85 ± 0,05	0,45 ± 0,12
Radiómica HSAa Exp2	Buen desarrollo	0,78 ± 0,05	0,70 ± 0,05	0,70 ± 0,06	0,72 ± 0,15
	Mal desarrollo	0,73 ± 0,06	0,66 ± 0,06	0,68 ± 0,07	0,65 ± 0,18
Clínica + Radiómica HSAa Exp2	Buen desarrollo	0,80 ± 0,05	0,72 ± 0,05	0,74 ± 0,06	0,70 ± 0,14
	Mal desarrollo	0,75 ± 0,06	0,68 ± 0,06	0,70 ± 0,07	0,66 ± 0,16
Resultado funcional (GOS)					
Clínica	Buen desarrollo	0,80 ± 0,09	0,68 ± 0,09	0,26 ± 0,17	0,95 ± 0,04
	Mal desarrollo	0,62 ± 0,03	0,62 ± 0,03	0,70 ± 0,08	0,54 ± 0,12
Radiómica Parénquima	Buen desarrollo	0,60 ± 0,04	0,60 ± 0,04	0,67 ± 0,09	0,52 ± 0,11
	Mal desarrollo	0,60 ± 0,04	0,60 ± 0,04	0,67 ± 0,09	0,52 ± 0,11
Clínica + Radiómica Parénquima	Buen desarrollo	0,80 ± 0,09	0,68 ± 0,09	0,26 ± 0,17	0,95 ± 0,04
	Mal desarrollo	0,76 ± 0,05	0,63 ± 0,05	0,72 ± 0,12	0,66 ± 0,21
Radiómica HSAa (SMCCH)	Buen desarrollo	0,76 ± 0,05	0,63 ± 0,05	0,72 ± 0,12	0,66 ± 0,21
	Mal desarrollo	0,68 ± 0,09	0,68 ± 0,09	0,80 ± 0,04	0,56 ± 0,14
Clínica + Radiómica HSAa (SMCCH)	Buen desarrollo	0,63 ± 0,10	0,63 ± 0,10	0,75 ± 0,06	0,48 ± 0,15
	Mal desarrollo	0,68 ± 0,08	0,68 ± 0,08	0,71 ± 0,12	0,65 ± 0,21
Radiómica HSAa Exp2	Buen desarrollo	0,64 ± 0,06	0,64 ± 0,06	0,70 ± 0,10	0,60 ± 0,19
	Mal desarrollo	0,77 ± 0,06	0,65 ± 0,06	0,62 ± 0,10	0,74 ± 0,14
Clínica + Radiómica HSAa Exp2	Buen desarrollo	0,72 ± 0,07	0,61 ± 0,06	0,66 ± 0,11	0,62 ± 0,16
	Mal desarrollo	0,79 ± 0,06	0,67 ± 0,06	0,65 ± 0,10	0,74 ± 0,12
	Buen desarrollo	0,74 ± 0,06	0,63 ± 0,06	0,68 ± 0,11	0,62 ± 0,15
	Mal desarrollo	0,74 ± 0,06	0,63 ± 0,06	0,68 ± 0,11	0,62 ± 0,15
Vasoespasmó					
Clínica	Buen desarrollo	0,62 ± 0,08	0,62 ± 0,08	0,76 ± 0,09	0,50 ± 0,10
	Mal desarrollo	0,58 ± 0,06	0,58 ± 0,06	0,71 ± 0,07	0,45 ± 0,09
Radiómica Parénquima	Buen desarrollo	0,70 ± 0,02	0,67 ± 0,02	0,78 ± 0,08	0,69 ± 0,15
	Mal desarrollo	0,64 ± 0,07	0,54 ± 0,07	0,57 ± 0,11	0,56 ± 0,19
Clínica + Radiómica Parénquima	Buen desarrollo	0,70 ± 0,02	0,67 ± 0,02	0,78 ± 0,08	0,69 ± 0,15
	Mal desarrollo	0,64 ± 0,07	0,54 ± 0,07	0,57 ± 0,11	0,56 ± 0,19
Radiómica HSAa (SMCCH)	Buen desarrollo	0,64 ± 0,03	0,64 ± 0,03	0,76 ± 0,08	0,50 ± 0,12
	Mal desarrollo	0,58 ± 0,06	0,58 ± 0,06	0,70 ± 0,09	0,44 ± 0,10
Clínica + Radiómica HSAa (SMCCH)	Buen desarrollo	0,61 ± 0,01	0,61 ± 0,01	0,78 ± 0,08	0,45 ± 0,14
	Mal desarrollo	0,55 ± 0,05	0,55 ± 0,05	0,72 ± 0,12	0,38 ± 0,15
Radiómica HSAa Exp2	Buen desarrollo	0,66 ± 0,04	0,63 ± 0,05	0,66 ± 0,08	0,60 ± 0,12
	Mal desarrollo	0,61 ± 0,05	0,57 ± 0,05	0,60 ± 0,10	0,55 ± 0,12
Clínica + Radiómica HSAa Exp2	Buen desarrollo	0,68 ± 0,05	0,65 ± 0,05	0,67 ± 0,08	0,62 ± 0,11
	Mal desarrollo	0,63 ± 0,05	0,60 ± 0,05	0,61 ± 0,09	0,59 ± 0,12
Hidrocefalia					
Clínica	Buen desarrollo	0,75 ± 0,15	0,75 ± 0,15	0,80 ± 0,06	0,55 ± 0,20
	Mal desarrollo	0,70 ± 0,10	0,70 ± 0,10	0,78 ± 0,08	0,50 ± 0,15
Radiómica Parénquima	Buen desarrollo	0,80 ± 0,07	0,67 ± 0,17	0,73 ± 0,06	0,66 ± 0,30
	Mal desarrollo	0,73 ± 0,08	0,62 ± 0,08	0,65 ± 0,10	0,45 ± 0,20
Clínica + Radiómica Parénquima	Buen desarrollo	0,80 ± 0,07	0,67 ± 0,17	0,73 ± 0,06	0,66 ± 0,30
	Mal desarrollo	0,73 ± 0,08	0,62 ± 0,08	0,65 ± 0,10	0,45 ± 0,20
Radiómica HSAa (SMCCH)	Buen desarrollo	0,65 ± 0,14	0,65 ± 0,14	0,78 ± 0,07	0,52 ± 0,22

Continúa en la siguiente página

Modelo	Grupo	AUC	BalAcc	Sens	Spec
Clínica + Radiómica HSAa (SMCCH)	Mal desarrollo	0,61 ± 0,10	0,61 ± 0,10	0,75 ± 0,08	0,45 ± 0,20
	Buen desarrollo	0,62 ± 0,17	0,62 ± 0,17	0,82 ± 0,04	0,48 ± 0,23
Radiómica HSAa Exp2	Mal desarrollo	0,59 ± 0,09	0,59 ± 0,09	0,75 ± 0,11	0,42 ± 0,18
	Buen desarrollo	0,67 ± 0,05	0,64 ± 0,06	0,65 ± 0,09	0,62 ± 0,14
Clínica + Radiómica HSAa Exp2	Mal desarrollo	0,62 ± 0,07	0,58 ± 0,06	0,61 ± 0,11	0,56 ± 0,15
	Buen desarrollo	0,69 ± 0,06	0,65 ± 0,06	0,66 ± 0,09	0,63 ± 0,13
	Mal desarrollo	0,64 ± 0,06	0,60 ± 0,06	0,62 ± 0,10	0,58 ± 0,14

Curvas de calibración

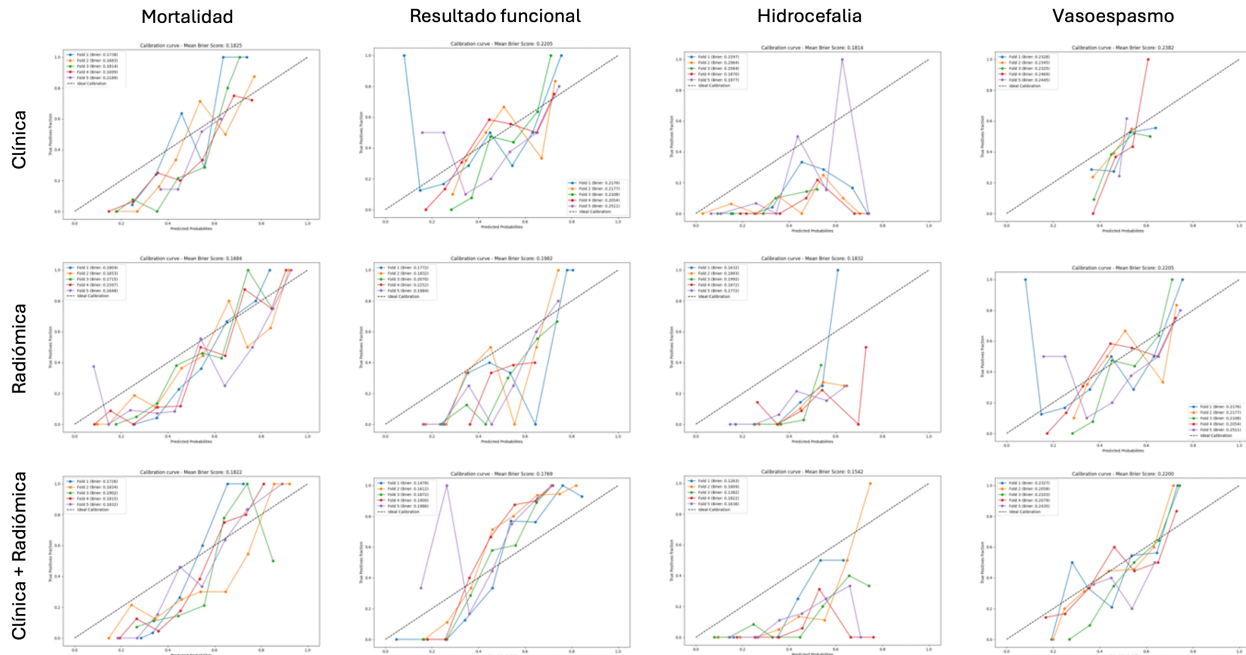


Figura B.3: Curvas de calibración de los modelos en los distintos folds de la validación cruzada por experimento: usando variables clínicas, radiómicas o su combinación. La radiómica fue extraída del parénquima cerebral.

Configuraciones seleccionadas de los modelos

En esta sección se resumen las configuraciones de los modelos que alcanzaron el mejor rendimiento en cada una de las tareas de predicción, tanto en el experimento 2 con HSAa como en los modelos radiómicos de parénquima. Para cada *fold* se especifica el algoritmo utilizado, la cantidad de características seleccionadas, y si se aplicaron técnicas de eliminación de valores atípicos o balanceo mediante *SMOTE*.

Las Tablas B.8–B.15 recogen los resultados de manera desglosada por outcome: mortalidad, GOS funcional, hidrocefalia y vasoespasmio.

Tabla B.8: Modelos seleccionados para Mortalidad (HSAa, Experimento 2).

Fold	Model	FS Method	Rmv Outliers	SMOTE
0	RF	MRMR_10	No	Yes
1	ExtraTrees	MRMR_5	No	Yes
2	XGBoost	MRMR_10	No	Yes
3	XGBoost	MRMR_5	No	Yes
4	ExtraTrees	MRMR_20	No	Yes

Tabla B.9: Modelos seleccionados para GOS funcional (HSAa, Experimento 2).

Fold	Model	FS Method	Rmv Outliers	SMOTE
0	ExtraTrees	MRMR_20	No	Yes
1	XGBoost	MRMR_10	No	Yes
2	ExtraTrees	MRMR_20	No	Yes
3	RF	MRMR_10	No	Yes
4	XGBoost	MRMR_5	No	Yes

Tabla B.10: Modelos seleccionados para Hidrocefalia (HSAa, Experimento 2).

Fold	Model	FS Method	Rmv Outliers	SMOTE
0	ExtraTrees	MRMR_20	No	Yes
1	RF	MRMR_15	No	Yes
2	ExtraTrees	MRMR_15	No	Yes
3	ExtraTrees	MRMR_20	No	Yes
4	ExtraTrees	MRMR_20	No	Yes

Tabla B.11: Modelos seleccionados para Vasoespasmo (HSAa, Experimento 2).

Fold	Model	FS Method	Rmv Outliers	SMOTE
0	ExtraTrees	MRMR_20	No	Yes
1	ExtraTrees	MRMR_10	No	Yes
2	ExtraTrees	MRMR_20	No	Yes
3	RF	MRMR_20	No	Yes
4	RF	MRMR_5	No	Yes

Tabla B.12: Modelos seleccionados para Mortalidad (Parénquima).

Fold	Model	FS Method	Rmv Outliers	SMOTE
0	ExtraTrees	MRMR_30	No	Yes
1	XGBoost	MRMR_20	No	Yes
2	ExtraTrees	MRMR_10	No	Yes
3	ExtraTrees	MRMR_20	No	Yes
4	ExtraTrees	MRMR_30	No	Yes

Tabla B.13: Modelos seleccionados para GOS funcional (Parénquima).

Fold	Model	FS Method	Rmv Outliers	SMOTE
0	RF	MRMR_15	No	No
1	RF	MRMR_10	No	No
2	ExtraTrees	MRMR_10	No	No
3	XGBoost	MRMR_10	No	No
4	ExtraTrees	MRMR_10	No	No

Tabla B.14: Modelos seleccionados para Hidrocefalia (Parénquima).

Fold	Model	FS Method	Rmv Outliers	SMOTE
0	RF	MRMR_20	No	Yes
1	ExtraTrees	MRMR_20	No	Yes
2	XGBoost	MRMR_10	No	Yes
3	ExtraTrees	MRMR_15	No	Yes
4	ExtraTrees	MRMR_5	No	Yes

Tabla B.15: Modelos seleccionados para Vasoespasmio (Parénquima).

Fold	Model	FS Method	Rmv Outliers	SMOTE
0	XGBoost	MRMR_10	No	Yes
1	XGBoost	MRMR_5	No	Yes
2	RF	MRMR_20	No	Yes
3	XGBoost	MRMR_20	No	Yes
4	ExtraTrees	MRMR_5	No	Yes

Matrices de confusión de los modelos predictivos

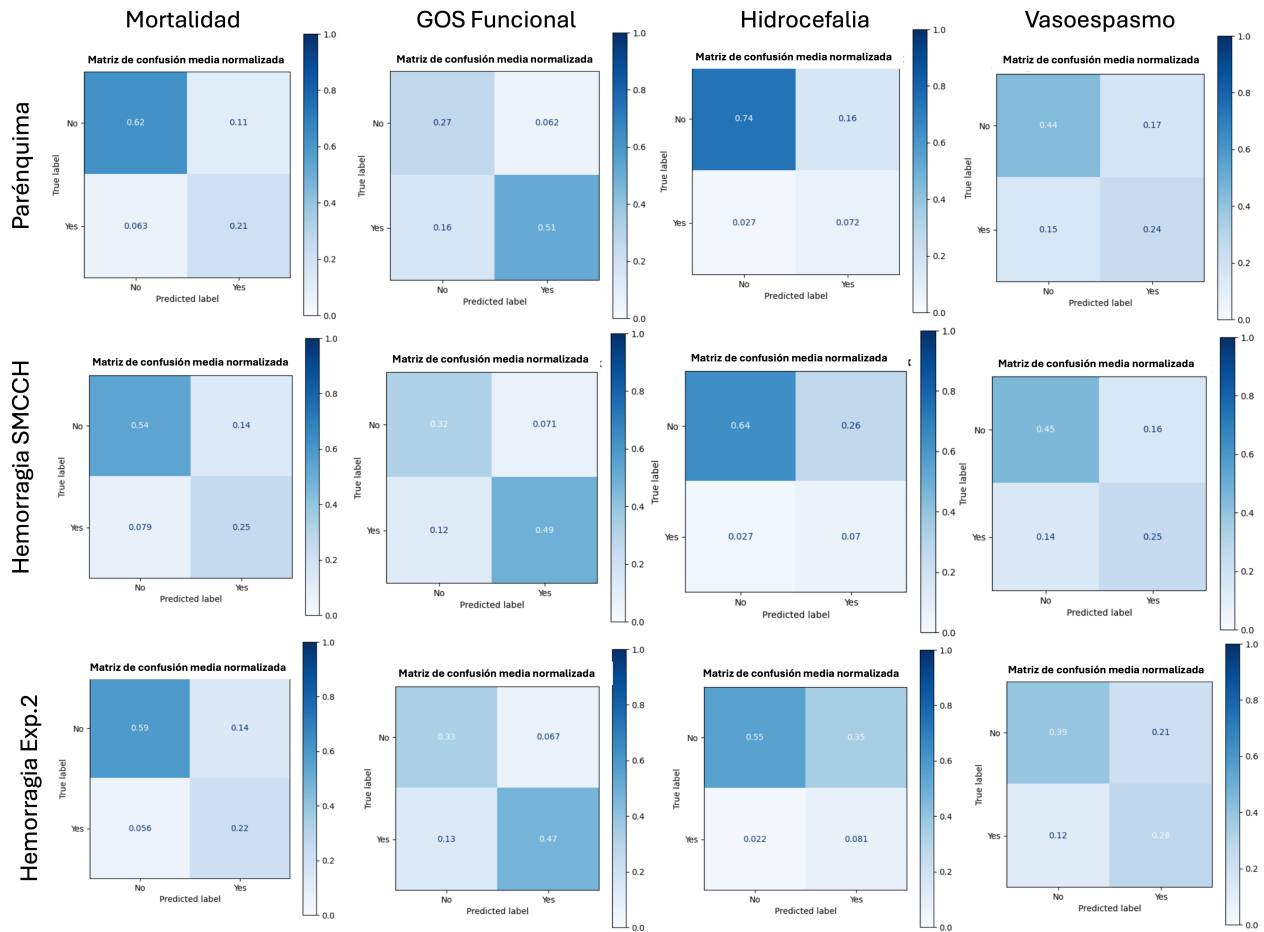


Figura B.4: Matrices de confusión normalizadas para los modelos que combinan radiómica y clínica por cada desenlace usando la radiómica extraída de la parénquima, hemorragia SMCCH y hemorragia del Experimento 2.

Visualización de importancia de variables con SHAP



Figura B.5: Gráficas SHAP obtenidas por cada desenlace clínico de los modelos que solo usan variables clínicas, radiómica de la parénquima o combinados (radiómica + clínica).

Top 10 variables más frecuentes (GOS funcional)

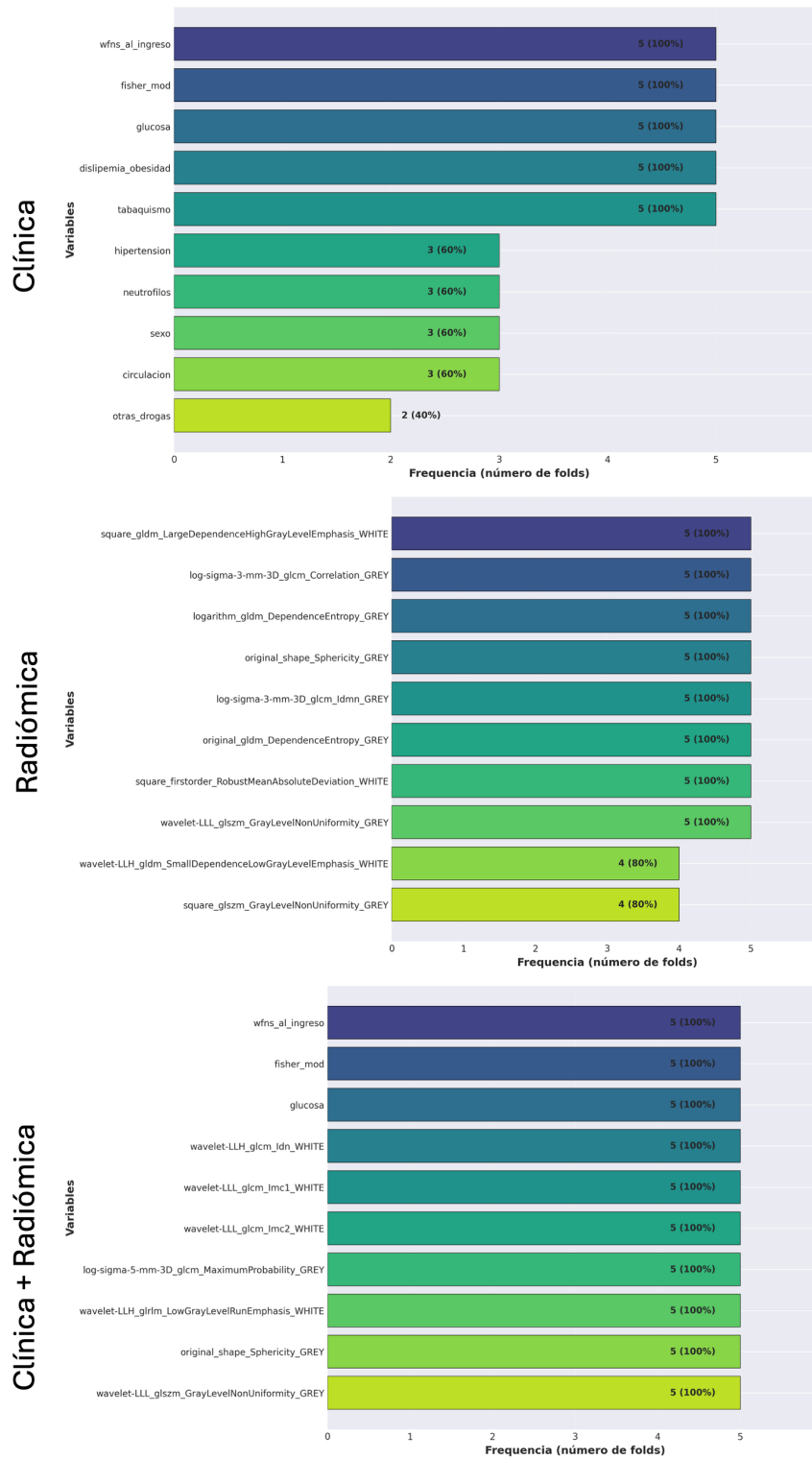


Figura B.6: Frecuencia de selección de variables en los cinco folds de validación cruzada en los modelos de GOS funcional que usan variables clínicas, radiómicas y combinadas.

Información adicional de las escalas clínicas evaluadas

Tabla B.16: Estadísticos descriptivos de las escalas radiológicas empleadas para calcular el ictWFNS en el conjunto de test ($n = 41$).

Escala	Media \pm DE	Mediana (IQR)	Rango
Hijdra	16.49 \pm 8.75	18 (10–24)	0–30
LeRoux	5.44 \pm 5.12	4 (1–9)	0–16

Las escalas de Hijdra y LeRoux cuantifican de forma cualitativa el volumen hemorrágico subaracnoideo e intraventricular, respectivamente. Los valores se presentan como media \pm desviación estándar, mediana (rango intercuartílico) y rango completo.

Tabla B.17: Distribución del SEBES en el conjunto de test ($n = 41$).

Puntuación SEBES	Frecuencia	Porcentaje (%)
1	2	5.38
2	3	7.32
3	5	12.20
4	31	75.61

El SEBES es una escala ordinal de 0–4 que refleja signos de edema cerebral precoz en la TC al ingreso. La distribución se muestra en valores absolutos y porcentajes.

Comparativa de modelos SEBES/SEBES6c y radiómicos por desenlace

Tabla B.18: Métricas comparativas de rendimiento de los modelos predictivos para mortalidad: aproximaciones basadas en SEBES frente a radiómica.

Métrica	SEBES Univariable	SEBES6c Univariable	SEBES+Clínica Multivariable	SEBES6c+Clínica Multivariable	Radiómica	Radiómica +Clínica
Pseudo R ²	0,0000001	0,0022	0,2137	0,2055	–	–
Log-Likelihood	-212.65	-212.18	-167.20	-168.94	–	–
Coef. SEBES	0,00 (<i>p</i> = 0, 97)	–	0,48 (<i>p</i> = 0, 03)	–	–	–
Coef. SEBES6c	–	0,22 (<i>p</i> = 0, 25)	–	0,71 (<i>p</i> = 0, 21)	–	–
AUC (IC 95 %)	0,500	0,528	0,799	0,794	0,82 (0,77–0,83)	0,87 (0,83–0,88)
Sensibilidad (IC 95 %)	0,00	0,00	0,491	0,536	0,79 (0,72–0,87)	0,79 (0,72–0,87)
Especificidad (IC 95 %)	1.00	1.00	0,833	0,824	0,71 (0,61–0,80)	0,76 (0,62–0,87)
BalAcc (IC 95 %)	0,50	0,50	0,662	0,680	0,75 (0,70–0,80)	0,78 (0,76–0,82)
Var. Explicada (%)	–	–	SEBES: 0,48 WFNS: 1.66 Edad: 1.01 Hipertensión: 1.15 Obes./Dislip.: 1.27 Tabaquismo: 0,75 Fisher mod.: 2.16	SEBES6c: 0,71 WFNS: 1.62 Edad: 1.01 Hipertensión: 1.22 Obes./Dislip.: 1.39 Tabaquismo: 0,76 Fisher mod.: 2.09	–	–

Tabla B.19: Métricas comparativas de rendimiento de los modelos predictivos para mal resultado funcional: aproximaciones basadas en SEBES frente a radiómica.

Métrica	SEBES Univariable	SEBES6c Univariable	SEBES+Clínica Multivariable	SEBES6c+Clínica Multivariable	Radiómica	Radiómica +Clínica
Pseudo R ²	0,001	0,006	0,214	0,208	–	–
Log-Likelihood	-230,41	-229.36	-181.25	-182.63	–	–
Coef. SEBES	-0,161 (<i>p</i> = 0, 500)	–	0,610 (<i>p</i> = 0, 050)	–	–	–
Coef. SEBES6c	–	-0,352 (<i>p</i> = 0, 110)	–	0,294 (<i>p</i> = 0, 287)	–	–
AUC (IC 95 %)	0,517 (0,471–0,563)	0,544 (0,493–0,595)	0,796 (0,751–0,843)	0,795 (0,750–0,843)	0,76 (0,72–0,80)	0,85 (0,81–0,89)
Sensibilidad (IC 95 %)	1.000 (0,96–1.00)	0,535 (0,43–0,63)	0,744 (0,65–0,83)	0,727 (0,64–0,82)	0,71 (0,63–0,79)	0,76 (0,68–0,84)
Especificidad (IC 95 %)	0,000 (0,00–0,00)	0,553 (0,475–0,629)	0,702 (0,631–0,770)	0,708 (0,643–0,781)	0,63 (0,51–0,75)	0,75 (0,65–0,85)
BalAcc (IC 95 %)	0,500 (0,500–0,500)	0,544 (0,493–0,595)	0,723 (0,678–0,771)	0,717 (0,672–0,768)	0,67 (0,63–0,71)	0,75 (0,71–0,79)
Var. Explicada (%)	–	–	SEBES: 0,9 WFNS: 9.6 Edad: 0,0 Obesidad: 0,0 Hipertensión: 0,0 Tabaquismo: 1.0 Fisher mod.: 3.4	SEBES6c: 0,2 WFNS: 8.9 Edad: 0,0 Obesidad: 0,1 Hipertensión: 0,0 Tabaquismo: 1.0 Fisher mod.: 3.1	–	–

Tabla B.20: Métricas comparativas de rendimiento de los modelos predictivos para vasoespasmo: aproximaciones basadas en SEBES frente a radiómica.

Métrica	SEBES Univariable	SEBES6c Univariable	SEBES+Clínica Multivariable	SEBES6c+Clínica Multivariable	Radiómica	Radiómica +Clínica
Pseudo R ²	0,003	0,001	0,017	0,018	-	-
Log-Likelihood	-159.96	-160,35	-157.75	-157.57	-	-
Coef. SEBES	0,296 (<i>p</i> = 0, 297)	-	0,210 (<i>p</i> = 0, 498)	-	-	-
Coef. SEBES6c	-	-0,150 (<i>p</i> = 0, 571)	-	-0,262 (<i>p</i> = 0, 361)	-	-
AUC	0,532	0,519	0,596	0,597	0,69	0,75
(IC 95 %)	(0,474–0,590)	(0,452–0,583)	(0,520–0,670)	(0,510–0,663)	(0,65–0,73)	(0,71–0,79)
Sensibilidad	0,000	0,000	0,202	0,212	0,73	0,74
(IC 95 %)	(0,00–0,05)	(0,00–0,05)	(0,13–0,30)	(0,17–0,32)	(0,67–0,79)	(0,68–0,80)
Especificidad	1.000	1.000	0,891	0,876	0,58	0,69
(IC 95 %)	(1.00–1.00)	(1.00–1.00)	(0,831–0,941)	(0,812–0,932)	(0,48–0,68)	(0,59–0,79)
BalAcc	0,500	0,500	0,546	0,539	0,60	0,67
(IC 95 %)	(0,500–0,500)	(0,500–0,500)	(0,496–0,592)	(0,489–0,586)	(0,56–0,64)	(0,63–0,71)
Var. Explicada (%)	-	-	SEBES: 0,2 WFNS: 0,3 Edad: 0,2 Obesidad: 0,0 Hipertensión: 0,4 Tabaquismo: 0,2 Fisher mod.: 0,2	SEBES6c: 0,3 WFNS: 0,4 Edad: 0,1 Obesidad: 0,0 Hipertensión: 0,4 Tabaquismo: 0,2 Fisher mod.: 0,4	-	-

Tabla B.21: Métricas comparativas de rendimiento de los modelos predictivos para hidrocefalia: aproximaciones basadas en SEBES frente a radiómica.

Métrica	SEBES Univariable	SEBES6c Univariable	SEBES+Clínica Multivariable	SEBES6c+Clínica Multivariable	Radiómica	Radiómica +Clínica
Pseudo R ²	0,016	0,025	0,266	0,267	-	-
Log-Likelihood	-165.87	-164.39	-123.75	-123.51	-	-
Coef. SEBES	0,652 (<i>p</i> = 0, 021)	-	0,110 (<i>p</i> = 0, 759)	-	-	-
Coef. SEBES6c	-	0,753 (<i>p</i> = 0, 004)	-	0,250 (<i>p</i> = 0, 447)	-	-
AUC	0,570	0,593	0,840	0,840	0,79	0,86
(IC 95 %)	(0,514–0,631)	(0,533–0,654)	(0,789–0,886)	(0,788–0,888)	(0,75–0,83)	(0,82–0,90)
Sensibilidad	0,754	0,570	0,711	0,737	0,82	0,81
(IC 95 %)	(0,64–0,84)	(0,46–0,68)	(0,61–0,82)	(0,68–0,79)	(0,78–0,86)	(0,77–0,85)
Especificidad	0,385	0,615	0,815	0,815	0,63	0,83
(IC 95 %)	(0,303–0,470)	(0,535–0,699)	(0,748–0,883)	(0,750–0,878)	(0,53–0,73)	(0,75–0,91)
BalAcc	0,570	0,593	0,772	0,758	0,67	0,74
(IC 95 %)	(0,514–0,631)	(0,533–0,654)	(0,718–0,826)	(0,705–0,811)	(0,63–0,71)	(0,70–0,78)
Var. Explicada (%)	-	-	SEBES: 0,0 WFNS: 12.0 Edad: 0,0 Obesidad: 0,0 Hipertensión: 0,1 Tabaquismo: 1.2 Fisher mod.: 3.9	SEBES6c: 0,2 WFNS: 11.3 Edad: 0,0 Obesidad: 0,0 Hipertensión: 0,1 Tabaquismo: 1.3 Fisher mod.: 3.9	-	-

Tabla B.22: Coeficientes obtenidos con regresión logística LASSO para las principales características radiómicas seleccionadas para predecir SEBES. Se muestran el nombre de la característica y el coeficiente estimado (*coef*).

Feature	Coef
wavelet-LLH_glcm_MaximumProbability_GREY	-0.33
log-sigma-3-mm-3D_glrmlm_RunVariance_GREY	-0.25
log-sigma-3-mm-3D_glrmlm_LongRunHighGrayLevelEmphasis_GREY	-0.24
log-sigma-5-mm-3D_glrmlm_RunLengthNonUniformity_WHITE	0.15
wavelet-LHL_glcm_ClusterShade_WHITE	0.15
log-sigma-0-5-mm-3D_glcm_MaximumProbability_GREY	-0.15
log-sigma-3-mm-3D_glrmlm_LongRunEmphasis_GREY	-0.13
log-sigma-3-mm-3D_glszm_GrayLevelVariance_WHITE	0.12
log-sigma-3-mm-3D_glcm_ClusterShade_WHITE	0.10
log-sigma-0-5-mm-3D_ngtdm_Busyness_GREY	-0.10
wavelet-LHL_firstorder_Mean_WHITE	-0.08
wavelet-LHL_firstorder_RootMeanSquared_WHITE	-0.08
wavelet-LLH_glszm_LargeAreaLowGrayLevelEmphasis_WHITE	0.07
log-sigma-5-mm-3D_firstorder_Maximum_WHITE	0.07
original_shape_MajorAxisLength_GREY	-0.07
square_glszm_LargeAreaHighGrayLevelEmphasis_WHITE	0.07
wavelet-LLH_glrmlm_RunLengthNonUniformity_WHITE	0.07
wavelet-LLL_glcm_Imc1_WHITE	0.07
log-sigma-3-mm-3D_glrmlm_RunEntropy_WHITE	0.06
log-sigma-5-mm-3D_glcm_Correlation_WHITE	0.06
wavelet-LHH_firstorder_Median_WHITE	0.06
logarithm_ngtdm_Busyness_GREY	-0.06
wavelet-LLH_glszm_LargeAreaHighGrayLevelEmphasis_GREY	-0.06
log-sigma-3-mm-3D_glszm_GrayLevelNonUniformityNormalized_WHITE	-0.05
logarithm_glszm_GrayLevelNonUniformity_WHITE	0.05
wavelet-LHL_firstorder_Skewness_GREY	0.05
wavelet-LLH_ngtdm_Busyness_WHITE	0.05
wavelet-HHH_firstorder_Skewness_GREY	-0.05
wavelet-LLL_glcm_Correlation_WHITE	-0.05
square_firstorder_Variance_WHITE	-0.05
exponential_glszm_SmallAreaEmphasis_GREY	0.05
wavelet-LHL_glszm_LowGrayLevelZoneEmphasis_WHITE	-0.05
log-sigma-5-mm-3D_glrmlm_RunEntropy_WHITE	0.04
log-sigma-3-mm-3D_glrmlm_LongRunEmphasis_WHITE	0.04
original_glszm_SmallAreaEmphasis_GREY	0.04

Anexo C Métricas de evaluación

Las métricas empleadas en este trabajo se agrupan en tres bloques: (i) métricas derivadas de la matriz de confusión, aplicables a tareas de clasificación y segmentación; (ii) métricas específicas para segmentación, centradas en solapamiento, distancia y volumen; y (iii) métricas estadísticas para modelos de regresión logística.

C.1 Métricas derivadas de la matriz de confusión

Sea A la predicción binaria del modelo y B la referencia (*ground truth*), con respecto a la **clase positiva**:

- **TP** (verdaderos positivos, de sus siglas en inglés *True Positives*): $A = 1, B = 1$
- **FP** (falsos positivos): $A = 1, B = 0$
- **TN** (verdaderos negativos, de sus siglas en inglés *True Negatives*): $A = 0, B = 0$
- **FN** (falsos negativos): $A = 0, B = 1$

	Referencia ($B = 1$)	Referencia ($B = 0$)
Predicción ($A = 1$)	TP	FP
Predicción ($A = 0$)	FN	TN

A partir de esta matriz se definen las siguientes métricas:

- **Sensibilidad (Sens)**: La sensibilidad, también denominada *recall* o Tasa de Verdaderos Positivos (TPR, de sus siglas en inglés *True Positive Rates*), mide la proporción de vóxeles de la región de interés presentes en la máscara de referencia que han sido correctamente identificados por el modelo como positivos (TP). Su valor oscila entre 0 y 1, donde 1 representa una detección completa de todos los positivos y 0 implica que no se ha reconocido ninguno.

$$\text{Sens} = \frac{TP}{TP + FN}$$

- **Especificidad (Spec)**: La especificidad, también llamada Tasa de Verdaderos Negativos (TNR, de sus siglas en inglés *True Negatives Rate*), cuantifica la proporción de vóxeles que no pertenecen a la región de interés en la máscara de referencia y que el modelo ha clasificado correctamente como negativos. Su rango es $[0, 1]$, siendo 1 la clasificación perfecta de todos los negativos y 0 la ausencia total de discriminación.

$$\text{Spec} = \frac{TN}{TN + FP}$$

- **Precisión global (OACC, por sus siglas en inglés *Overall Accuracy*)**: La precisión global o *overall accuracy* mide la fracción de predicciones correctas, tanto positivas como negativas, sobre el total de vóxeles evaluados. Representa una visión general del rendimiento del modelo. Sus valores posibles están entre 0 y 1, siendo 1 la clasificación perfecta.

$$\text{OACC} = \frac{TP + TN}{TP + TN + FP + FN}$$

- **Error global**: El error global refleja la proporción de vóxeles mal clasificados por el modelo, es decir, aquellos que corresponden a FP o FN. Constituye el complemento de

la precisión global. Su rango es $[0, 1]$, donde 0 indica ausencia de errores y 1 equivale a error absoluto.

$$\text{Error} = \frac{FP + FN}{TP + TN + FP + FN}$$

- **Exactitud (ACC, por sus siglas en inglés *Accuracy*)**: La exactitud o *accuracy* coincide con la definición de precisión global y corresponde a la métrica más empleada en problemas de clasificación binaria. Representa la proporción de vóxeles correctamente clasificados, tanto positivos como negativos, con valores que oscilan entre 0 (ninguna predicción correcta) y 1 (clasificación perfecta).

$$\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN}$$

- **Exactitud balanceada (BalAcc, por sus siglas en inglés *Balanced Accuracy*)**: La *balanced accuracy* o exactitud balanceada se obtiene como la media entre la sensibilidad y la especificidad. Este indicador resulta más robusto en contextos con clases desbalanceadas, ya que pondera por igual la detección de positivos y negativos. Sus valores oscilan entre 0 y 1, siendo 0,5 el rendimiento esperado por azar en un problema balanceado y 1 la clasificación perfecta.

$$\text{BalAcc} = \frac{\text{Sens} + \text{Spec}}{2}$$

- **Precisión**: La precisión mide la proporción de vóxeles clasificados por el modelo como pertenecientes a la región de interés que realmente lo son en la máscara de referencia. Evalúa, por tanto, la fiabilidad de las predicciones positivas. Su rango es $[0, 1]$, donde 1 significa que todas las predicciones positivas son correctas y 0 implica que ninguna lo es.

$$\text{Precisión} = \frac{TP}{TP + FP}$$

- **F1-score**: El F1-score es la media armónica entre sensibilidad y precisión, integrando en una única métrica la capacidad del modelo para identificar correctamente los positivos y evitar falsos positivos. Es especialmente útil en situaciones con clases desbalanceadas. Sus valores posibles se encuentran entre 0 y 1, donde 1 representa un equilibrio perfecto entre ambas métricas y 0 un desempeño nulo.

$$F_1 = 2 \cdot \frac{\text{Precisión} \cdot \text{Sens}}{\text{Precisión} + \text{Sens}}$$

- **Área bajo la curva ROC (AUC)**: El área bajo la curva ROC (AUC, *Area Under the Curve*) es una métrica global de rendimiento que cuantifica la capacidad de un clasificador para discriminar entre clases positivas y negativas en todos los posibles umbrales de decisión. Se obtiene como el área bajo la curva ROC, la cual representa la relación entre la tasa de verdaderos positivos (sensibilidad) y la tasa de falsos positivos ($1 - \text{especificidad}$). Matemáticamente se expresa como:

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR}) d(\text{FPR}), \quad \text{TPR} = \frac{TP}{TP + FN}, \quad \text{FPR} = \frac{FP}{FP + TN}.$$

C.2 Métricas específicas de segmentación

Métricas de solapamiento

- **Coefficiente de Dice (DSC, por sus siglas en inglés *Dice Similarity Coefficient*)** Müller et al., 2022: El coeficiente de Dice, también denominado *Dice Similarity Coefficient*, es una métrica ampliamente utilizada para evaluar la superposición entre dos segmentaciones. Se define como el doble de la intersección entre la máscara predicha y la de referencia, dividido por la suma de los vóxeles positivos en ambas. Su valor se encuentra en el rango $[0, 1]$, donde 0 indica ausencia total de solapamiento entre la predicción y la referencia, y 1 representa una coincidencia perfecta. Valores intermedios reflejan distintos grados de concordancia: cuanto más cercano a 1, mayor es la similitud entre ambas segmentaciones.

$$DSC = \frac{2TP}{2TP + FP + FN}$$

Métricas basadas en distancia

- **Distancia de Hausdorff (HD, por sus siglas en inglés *Hausdorff Distance*):** La distancia de Hausdorff mide la discrepancia máxima entre las superficies de dos segmentaciones. Se define como la mayor de todas las distancias mínimas desde un punto de una superficie a la superficie opuesta. De esta manera, refleja el peor caso de error en la localización de los contornos. El HD toma valores en el rango $[0, \infty)$, en unidades de distancia (habitualmente milímetros). Un valor de 0 indica coincidencia perfecta entre las superficies de referencia y predicción, mientras que valores elevados señalan discrepancias notables en los bordes.

$$HD(A, B) = \max \left\{ \max_{a \in S_A} \min_{b \in S_B} \|a - b\|, \max_{b \in S_B} \min_{a \in S_A} \|b - a\| \right\}$$

- **Distancia media de superficie (ASD, por sus siglas en inglés *Average Surface Distance*)** Celaya et al., 2023: La distancia media de superficie mide la discrepancia promedio entre las superficies de dos segmentaciones. Para cada punto de una superficie se calcula su distancia mínima a la superficie opuesta, y luego se obtiene la media considerando ambas direcciones (de S_A a S_B y de S_B a S_A). El ASD toma valores en el rango $[0, \infty)$, también expresados en unidades físicas (p. ej., milímetros). Un valor de 0 indica que las superficies coinciden exactamente, mientras que valores crecientes reflejan un mayor grado de separación media entre los contornos de la predicción y de la referencia.

$$ASD(A, B) = \frac{1}{|S_A| + |S_B|} \left(\sum_{a \in S_A} \min_{b \in S_B} \|a - b\| + \sum_{b \in S_B} \min_{a \in S_A} \|b - a\| \right)$$

Métricas basadas en volumen

- **Coefficiente de correlación intraclase** Koo y Li, 2016 El coeficiente de correlación intraclase (ICC, por sus siglas en inglés *Intraclass Correlation Coefficient*) evalúa el grado de concordancia entre dos métodos de medida, en este caso entre los volúmenes

segmentados automáticamente y los de referencia manual. La variante ICC(3,1) se emplea cuando se comparan siempre los mismos métodos de segmentación sobre distintos sujetos, valorando la consistencia entre ellos aunque exista un sesgo sistemático (por ejemplo, una ligera sobreestimación constante). Su valor varía entre 0 y 1, donde 0 indica ausencia de concordancia y 1 representa acuerdo perfecto.

$$ICC(3, 1) = \frac{MS_R - MS_E}{MS_R + (k - 1)MS_E}$$

donde MS_R representa la varianza entre sujetos, MS_E el error residual y k el número de medidas. Un valor de 0 corresponde a ausencia de concordancia y 1 a acuerdo perfecto.

- **Diferencia porcentual de volumen** La diferencia porcentual de volumen cuantifica el sesgo relativo entre el volumen segmentado por el modelo (V_{pred}) y el volumen de referencia (V_{ref}). Se calcula según la Ecuación (X):

$$\% \Delta V = \frac{V_{\text{pred}} - V_{\text{ref}}}{V_{\text{ref}}} \times 100$$

Un valor cercano a 0 indica buena concordancia en términos absolutos. Valores positivos implican sobrestimación del volumen por parte del modelo, mientras que valores negativos reflejan subestimación.

- **Gráfico de Bland–Altman** Giavarina, 2015 El análisis de Bland–Altman es una herramienta visual para evaluar la concordancia entre dos métodos de medida. Representa las diferencias de volumen entre predicción y referencia frente a su media, permitiendo identificar sesgos sistemáticos y estimar los límites de concordancia (media $\pm 1,96$ DE). La proximidad de los puntos al eje horizontal y la ausencia de tendencias sistemáticas son indicadores de buena concordancia.

C.3 Métricas estadísticas para regresión logística

- **Pseudo- R^2 de Nagelkerke** Weisberg, 2005 El pseudo- R^2 de Nagelkerke se deriva del estadístico de Cox–Snell, al que se le aplica un ajuste de escala para que su rango esté acotado entre 0 y 1, lo que facilita la interpretación. Este índice compara la verosimilitud del modelo ajustado (L_M) con la del modelo nulo (L_0), y se calcula de la siguiente forma:

$$R_{\text{Nag}}^2 = \frac{1 - \left(\frac{L_0}{L_M}\right)^{\frac{2}{n}}}{1 - L_0^n},$$

donde n es el tamaño muestral.

El valor de R_{Nag}^2 refleja la proporción de variabilidad en la variable dependiente que puede ser explicada por el modelo en comparación con el modelo nulo. Su rango teórico es $[0, 1]$: un valor cercano a 0 indica que el modelo no mejora sustancialmente la predicción respecto al modelo nulo, mientras que un valor cercano a 1 significa un ajuste casi perfecto a los datos.

- **Log-verosimilitud** Weisberg, 2005 La log-verosimilitud es una medida de qué tan bien se ajusta un modelo a los datos. En el caso de la regresión logística, indica con qué probabilidad el modelo es capaz de generar las observaciones reales a partir de los parámetros estimados $\hat{\beta}$. Su expresión matemática es:

$$\log L(\hat{\beta}) = \sum_{i=1}^n \left[y_i \log(\hat{p}_i) + (1 - y_i) \log(1 - \hat{p}_i) \right],$$

donde y_i es la clase observada (0/1) y \hat{p}_i la probabilidad predicha por el modelo. En términos prácticos, valores más altos de log-verosimilitud indican que el modelo describe mejor los datos, mientras que valores más bajos reflejan un peor ajuste.

Importancia relativa de predictores Tonidandel y LeBreton, 2011 La importancia relativa de los predictores permite conocer qué peso tiene cada variable dentro del modelo. En la regresión logística, se puede estimar observando cuánto disminuye la log-verosimilitud al eliminar un predictor, y expresando esa reducción como un porcentaje de la ganancia total conseguida por el modelo completo. De esta forma, se obtiene una visión clara y comparativa de qué variables aportan más a la capacidad explicativa global del modelo.

Anexo D Detalles de implementación

Con el fin de garantizar la reproducibilidad de los experimentos, en esta sección se describen los recursos de *hardware* y *software* empleados, así como las librerías y configuraciones clave del entorno de ejecución.

Los experimentos se han ejecutado en servidores de cálculo de altas prestaciones con soporte para GPU, necesarios para el entrenamiento eficiente de modelos de aprendizaje profundo sobre volúmenes de imagen médica. El entorno contaba con múltiples procesadores multinúcleo y varias GPU NVIDIA, lo que ha permitido cubrir tanto las fases de entrenamiento intensivo como las tareas posteriores de inferencia y análisis.

En particular, los entrenamientos de los modelos basados en DL (segmentación de hemorragia con nnUNet) se han llevado a cabo sobre una GPU NVIDIA A100 (40 GB VRAM) en combinación con procesadores Intel Xeon Gold 6338 (2.0 GHz, 64 hilos) y 256 GB de memoria RAM del sistema. Los modelos entrenados con ML (clasificación en GBM usando HSI o modelos basados en radiómica), así como las operaciones de preprocesamiento, postprocesamiento, evaluación y segmentación automática sobre nuevas cohortes se ejecutan eficientemente en CPU, sin necesidad de recurrir a aceleración por GPU.

De este modo, el uso de GPU se limita a la fase de entrenamiento, mientras que para las fases de inferencia y análisis estadístico la CPU resulta suficiente, lo que subraya la viabilidad de su despliegue en entornos clínicos con recursos computacionales más limitados.

El desarrollo se ha realizado en Python, gestionando las dependencias mediante contenedores Docker, lo que permite un control estricto de versiones y portabilidad del código a diferentes entornos de ejecución. Dentro de los contenedores se incluyen las siguientes bibliotecas principales:

- **Bibliotecas de aprendizaje profundo:** PyTorch Paszke et al., 2019 como marco principal para la implementación y entrenamiento de redes neuronales, junto a nnU-Net Isensee et al., 2021 para tareas específicas de imágenes médicas.
- **Bibliotecas de *machine learning*:** scikit-learn para algoritmos clásicos (SVM, RF, regresión logística) y funciones de validación cruzada; XGBoost T. Chen y Guestrin, 2016 en tareas complementarias de clasificación.
- **Procesamiento de imagen:** SimpleITK y OpenCV para manipulación de volúmenes médicos y preprocesado espectral.
- **Cálculo científico y análisis:** NumPy y SciPy para operaciones matriciales; Pandas para gestión tabular de resultados.
- **Visualización y análisis:** Matplotlib y Seaborn para la generación de figuras; SHAP para el análisis de interpretabilidad de los modelos predictivos Lundberg y Lee, 2017.

Para optimizar la reproducibilidad y el rendimiento de los modelos se siguieron las siguientes prácticas:

- **Paralelización:** se habilitó entrenamiento multi-GPU mediante las utilidades de PyTorch (*DataParallel/DistributedDataParallel*), lo que permite ampliar el tamaño efectivo de lote y acelerar el tiempo de entrenamiento.
- **Gestión de datos:** los volúmenes de TC se convirtieron a formato NIfTI y los cubos hiperespectrales a arreglos `.npy`, lo que facilitó la estandarización de *pipelines* y redujo tiempos de entrada/salida.