Universidad Nacional de Colombia

Análisis de Modelos de Difusión por Imágenes de Resonancia Magnética Nuclear con Machine Learning

Leonar Steven Prieto González

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Física
Bogotá D.C., Colombia
Mayo, 2022

Análisis de Modelos de Difusión por Imágenes de Resonancia Magnética Nuclear con Machine Learning

Leonar Steven Prieto González

Trabajo de grado presentado como requisito parcial para optar al título de:
Magister en Física Médica

Director:
Luis Agulles Pedrós, Dr. rer. nat

Línea de Investigación:
Radiología e Imágenes Diagnósticas
Grupo de Investigación:
Grupo de Física Médica

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Física
Bogotá D.C., Colombia
Mayo, 2022

A mis padres y mi familia por su constante e infinito apoyo durante la realización de este trabajo y por brindarme la fortaleza necesaria para cumplir mis sueños y mis proyectos.

Resumen

Análisis de Modelos de Difusión por Imágenes de Resonancia Magnética Nuclear con Machine Learning

En este trabajo se presenta la caracterización de las curvas de atenuación por difusión vóxel a vóxel de 4 conjuntos de imágenes (uno de próstata, dos de cerebro humano y uno de cerebro *ex vivo* de una neoplasia benigna). Esta caracterización incluye la determinación de los valores de difusión (D), pseudo-difusión (D^*), perfusión (f) y curtosis (K) usando los métodos de ADC (mono-exponencial) e IVIM (bi-exponencial) con y sin curtosis. Estos valores son utilizados como referencia para entrenar y probar la validez de varios algoritmos de machine learning (ML) que permiten disminuir el tiempo en la caracterización de la atenuación. Para decidir si en un vóxel existe difusión se implementan algoritmos de clasificación (Extra-Tree Classifier (ETC), Regresión logística (LR), C-Support vector (SVC), Extra-Gradient Boost (XGB) y perceptrón multicapa (MLP)), evaluados mediante la precisión y el test AUC. Mientras que para estimar los parámetros característicos se implementan métodos de regresión (Regresión lineal (LinR), regr. polinómica (Poly), XGB, Ridge, Lasso, Random Forest (RF), ElasticNet y support-vector machine (SVM)) que son evaluados mediante diferentes métricas de regresión, particularmente, la raíz del error cuadrático medio de la validación cruzada ($\mathrm{RMSE_{CV}}$). El objetivo de este trabajo es aplicar estas herramientas de ML para el análisis de difusión por imágenes de resonancia magnética. Se obtuvieron como mejores clasificadores el ETC y el MLP con una precisión del 94.1 % y 91.7 % respectivamente. Para la estimación de parámetros el mejor algoritmo fue RF; D posee un $\mathrm{RMSE_{CV}}$ del 8.39 %, D^* del 3.57 %, f con 4.52 % y K con 3.53 %. Aunque estos resultados pueden ser considerados satisfactorios, es posible que otros algoritmos que no se tuvieron en cuenta en este trabajo puedan reportar un mejor desempeño. El tiempo promedio que los algoritmos de ML tardan en caracterizar 100.000 vóxeles es $18,998 \pm 0,135$ s mientras que mediante métodos convencionales es de 4408 ± 351 s.

Palabras clave: Imágenes por resonancia magnética, IRM, difusión tisular, ADC, DWI, movimiento incoherente intravóxel, IVIM, aprendizaje automático..

Abstract

Analysis of nuclear magnetic resonance imaging diffusion models using machine learning

In this work we present the characterization of the voxel-by-voxel diffusion attenuation curves of 4 sets of images (one from the prostate, two from the human brain and one from the ex vivo brain of a mini pig). This characterization includes the determination of the values of diffusion (D), pseudo-diffusion (D^*), perfusion (f), and kurtosis (K); using ADC (mono-exponential) and IVIM (bi-exponential) considering kurtosis when convenient. These values are used as a reference to train and test the validity of several machine learning (ML) algorithms that allow to reduce the CPU time to characterize the attenuation curves. To decide if there is diffusion in a voxel, classification algorithms are implemented; (Extra-Tree Classifier (ETC), Logistic Regression (LR), C-Support Vector (SVC), Extra-Gradient Boost (XGB) and Multilayer Perceptron (MLP)), were evaluated by precision and the AUC tests. On the other hand, regression methods; (Linear Regression (LinR), Polynomial Regr. (Poly), XGB, Ridge, Lasso, Random Forest (RF), Elastic-Net and Support-Vector Machines (SVM)) are implemented to estimate the characteristic parameters and are evaluated using different regression metrics, particularly, root mean square error of cross-validation ($\mathrm{RMSE_{CV}}$). The main objective of this work is to apply these ML tools for diffusion analysis in magnetic resonance images. The ETC and the MLP showed the best classifiers with accuracies of 94.1 % and 91.7 %, respectively. For parameters estimation, the best algorithm was RF; D has an $\mathrm{RMSE_{CV}}$ of 8.39 %, D^* of 3.57 %, f of 4.52 % and K of 3.53 %. Although these results can be considered satisfactory, it is possible that other algorithms that were not taken into account in this work may report better performance.The average time that ML algorithms take to characterize 100.000 voxels is $18{,}998 \pm 0{,}135$ s while using conventional methods it is 4408 ± 351 s.

Keywords: Magnetic Resonance Imaging, MRI, tissue diffusion, ADC, DWI, Intravoxel Incoherent Motion, IVIM, Machine Learning.

Contenido

Resumen vii

1. Introducción 1

2. Aspectos teóricos 5

2.1. Imágen por Resonancia Magnética (IRM) . . . 5

2.1.1. Resonancia Magnética Nuclear (RMN) . . . 5

2.1.2. Magnetización macroscópica . . . 7

2.1.3. Tiempos de relajación (T_1 y T_2) . . . 9

2.1.4. Ecuaciones de Bloch . . . 11

2.2. Difusión molecular . . . 12

2.2.1. Imágenes pesadas por difusión (DWI) . . . 16

2.2.2. Modelo mono-exponencial . . . 16

2.2.3. Coeficiente Aparente de Difusión (ADC) . . . 17

2.2.4. Movimiento Incoherente Intravoxel (IVIM) . . . 17

2.2.5. Modelo bi-exponencial . . . 18

2.2.6. Efecto de la curtosis . . . 20

2.3. Técnicas de Machine Learning (ML) . . . 21

2.3.1. Aprendizaje supervisado . . . 21

2.3.2. Evaluación de la predicción . . . 24

2.3.3. Métricas de regresión . . . 30

3. Metodología 33

3.1. Determinación de parámetros . . . 33

3.1.1. Set MR701 . . . 36

3.1.2. Set IVIM . . . 39

3.1.3. Set CFIN . . . 41

3.1.4. Set PigBrain . . . 42

3.2. Preparación de las imágenes . . . 44

3.2.1. Mapas de r^2 y r^2 modificado . . . 44

3.2.2. Discretización de datos . . . 50

3.3. Implementación de los algoritmos de clasificación 52
3.3.1. Transformación de los datos 53
3.3.2. Construcción de los modelos para r^2 55
3.3.3. Construcción de los modelos para D, D* f y K 55

4. Resultados y análisis **57**
4.1. Clasificación de r^2 57
4.2. Predicción de D, D* f y K 59
4.3. Tiempo de procesamiento 65

5. Conclusiones y recomendaciones **66**
5.1. Conclusiones 66
5.2. Recomendaciones 68

A. Anexo: Disponibilidad de los códigos **69**

Bibliografía **70**

1. Introducción

La imagen de resonancia magnética (IRM) pesada por difusión (DWI) es una forma de imagen que hace uso de secuencias especificas de pulsos de radiofrecuencia (RF)y gradiente de campo magnético para medir el movimiento browniano que sucede durante la difusión molecular del agua (principalmente) en un tejido para generar contraste de forma no invasiva [1]. La difusión molecular en los tejidos no es libre y muestra las interacciones con obstáculos, como macromoléculas, fibras, membranas, capilares y demás. De este modo, los patrones de difusión pueden revelar detalles microscópicos particulares de tejidos normales o afectados. [2] El uso generalizado de DWI en diagnostico ha dado lugar a una gran cantidad de estudios y datos prometedores hacia la introducción de esta secuencia en los protocolos de IRM de rutina principalmente en tumores [2–4].

El uso de la tecnología de formación de IRM convencional por si solo tiene ciertas limitaciones, solo puede obtener información morfológica y no puede reflejar la microestructura interna de los tejidos tumorales, lo que dificulta la evaluación del riesgo de estos [3]. La DWI, por su parte, proporciona información funcional con respecto a la difusión libre del agua en espacios intra e intercelulares que en los tumores dependen principalmente de la celularidad [2].

La técnica de imagen de movimiento incoherente intravoxel (IVIM) se ha aplicado con éxito para diagnóstico de enfermedades tumorales en hígado, riñón, páncreas, etc. La sensibilidad a la difusión esta relacionada con un parámetro especifico, el *valor b* (o *b-value*), que varía según la amplitud, duración y tiempo entre dos gradientes de campo magnético [5]. La restricción de la difusión del agua puede luego analizarse cuantitativamente con el calculo del coeficiente de difusión aparente (ADC), que describe la atenuación de la señal del tejido con valores b crecientes [6].

Actualmente, la DWI se utiliza de forma rutinaria en varios escenarios clínicos, y sigue siendo un tema activo de investigación. Se ha probado en varios tipos de cánceres para diferenciar lesiones malignas de las benignas, distinguir diferentes grados de evolución, predecir y evaluar la respuesta a tratamientos e identificar tumores residuales o recurrentes en exámenes de seguimiento [3,7].

Con IVIM se pueden estudiar los movimientos de fluidos dentro de un vóxel durante un tiempo de medición y analizar la distribución de las velocidades y su amplitud. Con esta técnica se puede estimar la perfusión en tejidos sin medios de contraste, ya que el flujo sanguíneo en capilares orientados aleatoriamente imita un proceso de pseudo-difusión [2].

La adquisición de un conjunto de imágenes DWI es de las más veloces dentro de las imágenes de RM. Sin embargo, el diagnóstico de una enfermedad como la isquemia cerebral puede tomar varias horas o incluso días [8]. Por esta razón se pretende aplicar algoritmos de Machine Learning (ML) a este tipo de imágenes con el fin de buscar una reducción en el tiempo de diagnóstico mediante el pronóstico de ocurrencia de la enfermedad. Además, puede ayudar a ser capaces de distinguir si un tumor es maligno o no, basándose en la distribución del flujo sanguíneo en su interior [1,7,9,10].

La disminución en el tiempo de diagnóstico implica que se pueda tratar mucho más rápido al paciente y aumentar su probabilidad de recuperación. También conlleva a la reducción del tiempo de uso de la maquina de resonancia magnética y, por lo tanto, una mayor cantidad de exámenes realizados con esta en el mismo periodo de tiempo. Esto se traduce una reducción en costos tanto para los pacientes como para los centros médicos. Adicionalmente, este trabajo puede extenderse al estudio de IRM en difusión de órganos como pulmones, en donde se no se presenta difusión de liquido sino de gases [11].

El inicio de la DWI con IRM se da con Stejskal y Tanner en 1965 [12], proponiendo el efecto de un gradiente de campo magnético de espín-eco (SE), para el estudio de la difusión. Posteriormente, en 1986, D. Le Bihan *et al.* [13] modifican esta técnica para obtener imágenes de movimiento incoherente intravoxel (IVIM) mediante el uso de pulsos de gradiente adecuados con campos magnéticos de 0,5 T. Analizaron imágenes obtenidas del cerebro de pacientes, en términos de un coeficiente de difusión aparente (ADC) que incorpora el efecto de todos los IVIM y encontraron diferencias entre varios tejidos normales y patológicos.

En lo que concierne a la aplicación de la DWI a oncología, sus aplicaciones extienden principalmente a imágenes tumorales de cerebro, pero también a otros órganos como la próstata, los pulmones, la mama, el hígado, el páncreas, el sistema digestivo, urinario, linfático, y reproductor, así como huesos o tejido suave [7,14].

Respecto al análisis de cerebro, la DWI se incluye en el protocolo de RM de tumores cerebrales de rutina, que generalmente incluye dos valores b (entre 0 y 1000 s/mm^2) [7]. Si

bien las secuencias convencionales son útiles para la detección de tumores, la DWI ayuda a lograr el diagnostico, brindando información sobre el grado y tipo de tumor y así monitorear su respuesta al tratamiento. La DWI puede diferenciar los abscesos, que presentan una difusión reducida y un ADC bajo (entre 0.28 y 0.73×10^{-3} mm^2/s), de tumores necróticos, que muestran un ADC mas alto. Los valores b altos y la IVIM, que permiten un estudio detallado de la difusion del agua, pueden ayudar a diferenciar el tumor del edema peritumoral [15].

En tórax, el problema principal para la DWI es que esta secuencia es muy sensible a la pulsación cardiaca y la respiración, lo que puede generar artefactos y errores de calculo del ADC. Además, la baja densidad de protones de los pulmones crea una señal baja en cualquier secuencia. Sin embargo, Swerkersson *et al.* [16] demostraron la alta repetibilidad de las mediciones de ADC de adenocarcinomas de pulmón entre la respiración libre y la DWI provocada por la respiración, lo que confirma la posible aplicación de esta herramienta en los tumores pulmonares [7].

El cáncer gástrico es muy celular y muestra un ADC significativamente mas bajo ($1.115 \pm 0.156\times10^{-3}$ mm^2/s) que la pared gástrica normal ($1.621 \pm 0.292\times10^{-3}$ mm^2/s) o enfermedades benignas ($1.790 \pm 0.359\times10^{-3}$ mm^2/s) [17].

Por su parte, la resonancia magnética multiparamétrica (mpMRI) se usa ampliamente para detectar, localizar y clasificar el cáncer de próstata (CaP). Para aumentar la precisión de la DWI, se deben evaluar tanto el mapa ADC como las imágenes DWI de alto valor b, (generalmente aplicando tres valores b), con el mas bajo en 0 s/mm^2 y el mas alto en 1400/1500 s/mm^2. Sin embargo, no existe un valor b alto óptimo ampliamente aceptado en la literatura. Las imágenes con valores b mas altos suprimen el tejido de fondo sano, destacando los tejidos de CaP con difusión restringida y ADC mas bajo [18,19].

En la actualidad, se cuenta con gradientes ultra altos y secuencias aceleradas. Estas tecnologías han llevado a una mejora drástica en la resolución espacial, angular y de difusión que es factible *in vivo*. Sin embargo, todavía no alcanzan la escala en la que las propiedades microestructurales de las células [2].

Con relación al diagnóstico mediante imágenes de próstata usando técnicas de machine learning, se ha evaluado la repetibilidad de la radiómica prostática derivada de DWI y los métodos de aprendizaje automático para la caracterización del CaP usando dos exámenes por 112 pacientes diagnosticados concluyendo que la relación superficie-volumen para DWI de próstata tiene una buena clasificación de los datos ocultos desempeñándose de

manera similar en ambos casos (61 % y 81 %) en contraste con múltiples características radiómicas convencionales [20].

En este trabajo se planea aplicar herramientas de ML y aprendizaje automático para el análisis de IRM pesadas por difusión (DWI). Para conseguirlo, el primer paso es identificar y caracterizar por métodos convencionales una serie de imágenes usando diferentes modelos de difusión molecular. Posteriormente, clasificar la prevalencia de cada uno de estos métodos en cada vóxel bajo estudio para continuar con la identificación y adecuación de modelos de ML para diferenciar las DWI. Finalmente, aplicar y analizar los modelos de ML anteriormente estudiados, a la diferenciación de tejidos en diferentes sets de imágenes DWI.

Se aplicarán los modelos ADC e IVIM sobre un conjunto de IRM previamente seleccionadas y procesadas, las cuales son utilizadas para entrenar la IA. Estas imágenes son tomadas de bases de datos publicas [21–25] (disponibles en DYPI y Hvidovre Hospital). Gracias a la información contenida en los metadatos dicom, se puede extraer el valor b de cada imagen en cada corte. Posteriormente, se puede graficar el valor b vs. la intensidad de la señal en cada uno de los vóxeles bajo estudio. Una vez obtenidas estas gráficas en escala semi-log, se dividen en secciones representativas de la difusión celular ($b < 200$ s/mm^2) y de la difusión vascular ($b > 200$ s/mm^2). Estos valores pueden variar en función de la calidad requerida para el ajuste de las curvas. Cada sección puede encajar en diferentes modelos (mono-exponencial o bi-exponencial con y sin curtosis), siendo necesario decidir para cada vóxel si es suficiente con encontrar el valor de difusión D o si es necesario encontrar parámetros adicionales (fracción de perfusión f, pseudo-difusión D^* o curtosis K).

Algunos vóxeles serán fácilmente diferenciables en cuanto a la existencia o ausencia de difusión y su valor aproximado, sin embargo, otros no arrojan un resultado concluyente. Esto dará lugar a una segmentación. Los dos métodos para atacar este problema consisten en: la visión de vecinos mas cercanos (diferente a las técnicas de ML *"nearest neighbors"*) y la visión de predicción probabilística. La vision de vecinos mas cercanos consiste en verificar si existe difusión de los vóxeles alrededor de cada vóxel cuyo resultado haya sido no concluyente. Su valor dependerá de la ponderación de la difusión en función de la cercanía de cada uno de los vóxeles vecinos. El enfoque de la predicción probabilística consiste en aplicar varios algoritmos de ML para definir con cierto grado de precisión el valor de los parámetros asociados a la difusión en cada vóxel.

2. Aspectos teóricos

2.1. Imágen por Resonancia Magnética (IRM)

La IRM es una técnica no invasiva adquisición de imágenes imágenes anatómicas (y a veces funcionales) tridimensionales, haciendo uso de radiación electromagnética en el rango energético de la radiofrecuencia y por lo tanto, sin los riesgos inherentes de la radiación ionizante [26]. Se usa principalmente para el diagnóstico de enfermedades y como técnica complementaria para la ejecución de tratamientos. Funciona aprovechando el fenómeno de la resonancia magnética nuclear (RMN) en la cual se alinean los núcleos atómicos mediante un fuerte campo magnético para posteriormente perturbar este alineamiento con el uso de otro campo magnético alterno perpendicular de la misma frecuencia. Cuando los núcleos se recuperan de dicha perturbación, liberan energía cuya frecuencia depende de las características físicas del medio en el que se encuentren [5,27].

2.1.1. Resonancia Magnética Nuclear (RMN)

De acuerdo con la mecánica cuántica, las partículas elementales poseen un momento angular $\mathbf{L}$ cuantizado, es decir, se encuentra definido por valores discretos y su magnitud está dada por [28,29]:

$$|\mathbf{L}| = \hbar\sqrt{s(s+1)} \tag{2.1}$$

siendo s el número cuántico de espín, que solo puede ser entero o semientero y $\hbar$ representa la constante de Planck reducida. Por su parte, la dirección del momento angular se determina mediante un segundo número cuántico m, que solamente puede tener ciertas orientaciones discretas con respecto a una dirección dada. La componente del momento angular en la dirección z, viene dada por:

$$L_z = \hbar m_s \tag{2.2}$$

donde m_s es el número cuántico secundario de momento angular total y puede tomar $2s+1$ valores de acuerdo con la ecuación (2.3):

$$m_s = -s,\ -s+1,\ -s+2,\ ...,\ s-2,\ s-1,\ s \tag{2.3}$$

Para partículas como electrones, protones y neutrones, el número cuántico de espín s es igual a 1/2. Sin embargo, para los núcleos atómicos, s no puede calcularse simplemente mediante la suma de sus componentes individuales, pero usando tanto el número másico A como el número atómico Z (siguiendo la notación ${}^{A}_{Z}\mathrm{X}$), el numero de espín s puede deducirse por medio de las siguientes reglas [30]:

- Para núcleos con un número A impar, s es semientero (1/2, 3/2, 5/2, ..., por ejemplo, ${}^{1}_{1}\mathrm{H}$, ${}^{13}_{6}\mathrm{C}$, ${}^{15}_{7}\mathrm{N}$, ${}^{23}_{11}\mathrm{Na}$, ${}^{31}_{15}\mathrm{P}$).
- Para núcleos con un número A par y un número Z también par, s es cero (por ejemplo, ${}^{12}_{6}\mathrm{C}$, ${}^{16}_{8}\mathrm{O}$, ${}^{32}_{16}\mathrm{S}$).
- Para núcleos con número A par y numero Z impar, el número s es un número entero (1, 2, 3, ..., por ejemplo, ${}^{2}_{1}\mathrm{H}$, ${}^{14}_{7}\mathrm{N}$).

El momento magnético $\boldsymbol{\mu}$ de las partículas elementales se relacione con el momento angular $\mathbf{L}$ mediante:

$$\boldsymbol{\mu} = \gamma \mathbf{L} \tag{2.4}$$

donde γ es la constante giromagnética equivalente a

$$\gamma = \frac{e}{2m_p} g_n, \quad \text{para núcleos. Con unidades de} \quad [\gamma] = \frac{\text{rad}}{\text{s} \cdot \text{T}} \tag{2.5}$$

Dado que el momento angular está cuantizado, el momento magnético también lo estará. La componente μ_z del momento magnético a lo largo del eje longitudinal corresponde a:

$$\mu_z = \gamma \hbar m_s \tag{2.6}$$

Por un lado, en un campo magnético externo $\mathbf{B}_0$, el torque $\boldsymbol{\tau}$ estará dado por:

$$\boldsymbol{\tau} = \left(\frac{d\mathbf{L}}{dt}\right) = \boldsymbol{\mu} \times \mathbf{B}_0 \qquad \text{o} \qquad \frac{d\boldsymbol{\mu}}{dt} = \gamma \boldsymbol{\mu} \times \mathbf{B}_0 \tag{2.7}$$

lo que significa que $\boldsymbol{\mu}$ cambia su orientación (precesa) en relación con $\mathbf{B}_0$ (Figura **2.1**a).

Por otro lado, la partícula adquiere una energía magnética en la componente longitudinal del momento magnético dada por:

$$E = -\boldsymbol{\mu} \cdot \mathbf{B}_0 = -\mu_z B_0 = -\gamma \hbar m_s B_0 \tag{2.8}$$

Ya que m_s es un número cuántico entero o semientero, los niveles de energía se encuentran cuantizados. Entonces, para una partícula de espín $s = 1/2$, solo hay dos niveles de energía ($m_s = -1/2$ y $m_s = +1/2$) y la diferencia de energía será:

$$\Delta E = \gamma \hbar B_0 \tag{2.9}$$

El fenómeno de resonancia en RMN se logra aplicando un campo magnético oscilante ortogonal a μ_z con frecuencia ν_0, tal que la energía es igual a la energía magnética dada por la ecuación (2.9), y como la energía de la onda electromagnética está dada por

$$\Delta E = h\nu_0 \tag{2.10}$$

al igualar las ecuaciones (2.9) y (2.10) resulta en la *ecuación de Larmor* [31]:

$$\nu_0 = \left(\frac{\gamma}{2\pi}\right) B_0 \quad \longrightarrow \quad \omega_0 = \gamma B_0 \tag{2.11}$$

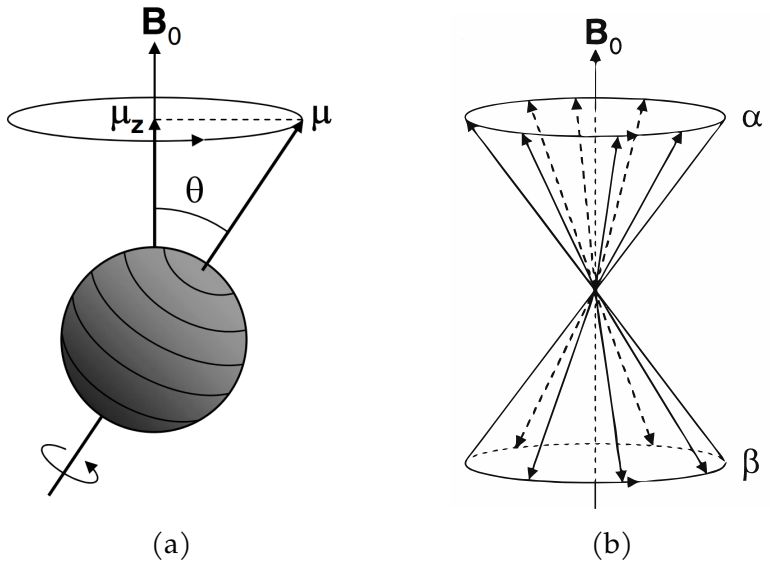


Figura 2.1.: (a) Un espín nuclear precesando en un campo magnético externo $\mathbf{B}_0$. El momento magnético de espín $\boldsymbol{\mu}$ tiene una precesión alrededor de $\mathbf{B}_0$, en el que se cuantizan la orientación θ y la amplitud μ_z. (b) En un conjunto macroscópico de espín nuclear $\pm 1/2$, los espines se distribuyen entre dos orientaciones posibles de acuerdo con la ecuación de Boltzmann. Tomado de [32]

.

2.1.2. Magnetización macroscópica

Para un núcleo de espín $s = 1/2$ y $m = \pm 1/2$ resulta un ángulo $\theta = 54,74^o$ relativo al eje $+z$ o $-z$, respectivamente (Figura 2.1a). Por lo tanto, los núcleos de espín $s = 1/2$ están distribuidos en la superficie de dos conos invertidos uno respecto al otro girando alrededor del vector de campo magnético $\mathbf{B}_0$ a la frecuencia de Larmor (Figura 2.1b). Para un núcleo de espín 1/2, los dos estados de espín $m = +1/2$ (μ paralelo a $\mathbf{B}_0$) y $m = -1/2$ (μ

antiparalelo a $\mathbf{B}_0$) se denominan estados de espín α y β, respectivamente.

Si no se considera solamente el comportamiento individual de los espines nucleares sino que se tiene en cuenta que una muestra macroscópica contiene muchos espines, que se distribuyen aleatoriamente en los conos entonces habrá una pequeña diferencia en la población de estos estados de espín debida a pequeñas diferencias de energía entre ellos. La diferencia entre las poblaciones de espines se puede calcular por medio de la ecuación de Boltzmann [33]. Para el caso mostrado en la figura **2.1b**, la diferencia de energía $\Delta E = h\nu$ resulta en una distribución poblacional dada por:

$$\left(\frac{n_\alpha}{n_\beta}\right) = e^{\Delta E/kT} = e^{h\nu/kT} \tag{2.12}$$

donde n_α es el número de espines en el estado de baja energía (α), n_β es el número de espines en el estado de alta energía (β), k es la constante de Boltzmann y T es la temperatura. Puesto que a temperatura normal, h es mucho menor que la energía térmica kT, la función exponencial se puede simplificar mediante los primeros términos de su expansión en series de Taylor:

$$\left(\frac{n_\alpha}{n_\beta}\right) = 1 + \left(\frac{h\nu}{kT}\right) + \cdots \tag{2.13}$$

Para una muestra macroscópica que contiene 10^6 espines nucleares a una temperatura T de 37 °C (≈ 310 K) y en un campo magnético de 9,4 T, correspondiente a ν = 400 MHz, la diferencia de poblaciones entre los estados de espín α y β es de solamente 30 espines [32]. Dado que la señal cruda recibida es proporcional a la diferencia entre estas poblaciones, se aprecia que la RMN es una técnica poco sensible en comparación con otras formas de espectroscopia, donde la diferencia de energía es mucho mayor. La magnetización total $\mathbf{M}$, de una muestra macroscópica es la suma de todos los momentos magnéticos individuales $\boldsymbol{\mu}_i$. Dado que los momentos magnéticos se distribuyen aleatoriamente sobre los conos, no existirá una componente neta de $\mathbf{M}$ en el plano transversal xy. Sin embargo, debido a la diferencia de población habrá un componente neto de $\mathbf{M}$ paralelo a $\mathbf{B}_0$ a lo largo del eje $+z$. En equilibrio térmico la magnitud de la magnetización longitudinal es:

$$|\mathbf{M}_0| = \sum_{i=1}^{n} \mu_i = n_\alpha \mu_z + n_\beta \mu_z = \left(\frac{h}{4\pi}\right) \gamma \left(n_\alpha - n_\beta\right) \tag{2.14}$$

Si en la ecuación (2.13) se tiene $(h\nu/kT) \ll 1$ y $n = n_\alpha + n_\beta$ donde n es el número total de espines nucleares en la muestra macroscópica, la diferencia de población será:

$$(n_\alpha - n_\beta) \approx \frac{nh\nu}{2kT} \tag{2.15}$$

Entonces, la amplitud del vector de magnetización $\mathbf{M}_0$ en equilibrio térmico será:

$$M_0 = \left(\frac{\gamma h}{2\pi}\right)^2 \left(\frac{nB_0}{4kT}\right) \tag{2.16}$$

De la ecuación (2.16) se pueden deducir varias características importantes relacionadas con la sensibilidad de los experimentos de RMN. La dependencia cuadrática de $\mathbf{M}_0$ con γ implica que los núcleos que resuenan a alta frecuencia también generan señales de RMN relativamente intensas. El hidrógeno tiene el γ más alto de los núcleos comúnmente encontrados [34] y, por lo tanto, tiene la intensidad más alta. La dependencia lineal de $\mathbf{M}_0$ de la intensidad del campo magnético $\mathbf{B}_0$ significa que la sensibilidad mejora al mismo tiempo que los campos magnéticos son más intensos.

De hecho, este argumento ha provocado un impulso constante hacia una mayor intensidad del campo magnético que actualmente oscila entre 1,5 T y 17,5 T o mayores [30]. Finalmente, la relación inversamente proporcional de $\mathbf{M}_0$ con la temperatura T indica que la sensibilidad se puede mejorar a temperaturas de muestra más bajas [32]. No obstante, la última opción no es realista para aplicaciones *in vivo*.

La sensibilidad experimental real está determinada por diversos factores, como el volumen de la muestra, la abundancia natural del núcleo estudiado, la relación giromagnética, el ruido de la muestra, los parámetros de relajación y la intensidad del campo magnético [32].

2.1.3. Tiempos de relajación (T_1 y T_2)

El decaimiento de la polarización de espín de RMN inducida por radiofrecuencia (RF) se caracteriza en términos de dos procesos separados, cada uno con sus propias constantes de tiempo. Un proceso, llamado T_1, es responsable de la pérdida de intensidad de resonancia después de la excitación del pulso. El otro proceso, llamado T_2, caracteriza el ancho o la amplitud de las resonancias. En la RMN convencional, T_1 limita la tasa de repetición de pulsos y afecta el tiempo total que se puede adquirir un espectro de RMN. Los valores de T_1 oscilan entre milisegundos y varios segundos, según el tamaño de la molécula, la viscosidad de la muestra, su temperatura y la posible presencia de especies paramagnéticas como, por ejemplo, O_2 o iones metálicos [35,36].

Relajación espín-red (T_1)

Una tasa de crecimiento de interacción constante a partir de las interacciones de protones con la red, implica que la tasa de cambio de la magnetización longitudinal, $dM_z(t)/dt$,

es proporcional a la diferencia $M_0 - M_z$. La constante de proporcionalidad se determina experimentalmente y representa la inversa de la escala de tiempo de la tasa de crecimiento [30,32].

$$\frac{dM_z}{dt} = \frac{1}{T_1}(M_0 - M_z), \quad \text{para} \quad \mathbf{B}_0 \parallel \hat{\mathbf{z}} \tag{2.17}$$

donde T_1 es el tiempo de relajación de la red de espín experimental. El parámetro de relajación T_1 varía entre $\sim 10^1$ y $\sim 10^3$ milisegundos para protones en tejido humano sobre las intensidades de campo $\mathbf{B}_0$ de interés (mayores a 0,01 T). Los valores típicos para varios tejidos se muestran en la tabla **2.1**.

Tejido	T_1 (ms)	T_2 (ms)
Grasa	250	60
Músculo	900	50
Materia blanca	600	80
Materia gris	950	100
Sangre	1200	150
Líquido cefalorraquídeo	4500	2200

Tabla 2.1.: Valores (aproximados) representativos para los tiempos de relajación T_1 (espín-espín) y T_2 (espín-red), para componentes de hidrógeno de diferentes tejidos en el cuerpo humano a 37 °C con B_0 = 1,5 T.

La solución de la ecuación (2.17) muestra que después de la aplicación de un pulso de RF, la magnetización longitudinal adopta una forma exponencial que muestra la evolución desde el valor inicial, $M_z(0)$, hasta el valor de equilibrio, M_0

$$M_z(t) = M_z(0)e^{-t/T_1} + M_0(1 - e^{t/T_1}) \tag{2.18}$$

Esta solución corresponde al caso en que $\mathbf{B}_0 = B_0 \cdot \hat{\mathbf{z}}$ y M_0 es el valor de equilibrio. Si $\mathbf{M}$ es inclinado hacia el plano xy entonces $M_z(0) = 0$ y la recuperación es:

$$M_z(t) = M_0(1 - e^{t/T_1}) \tag{2.19}$$

Este comportamiento se presenta en la figura **2.2**.

Relajación espín-espín (T_2)

La caracterización de la tasa de reducción en la magnetización transversal genera otro parámetro empírico, el tiempo de relajación "espín-espín". La ecuación (2.7) se modifica

mediante la adición de un término de decaimiento [30,32]:

$$\frac{d\mathbf{M}_\perp}{dt} = \gamma \mathbf{M}_\perp \times \mathbf{B}_0 - \frac{1}{T_2}\mathbf{M}_\perp \tag{2.20}$$

El término adicional conduce al decaimiento exponencial de cualquier valor inicial de $\mathbf{M}_\perp$. En el marco de referencia giratorio, esta ecuación diferencial tiene la siguiente forma:

$$\frac{d\mathbf{M}'_\perp}{dt} = -\frac{1}{T_2}\mathbf{M}_\perp \tag{2.21}$$

con la solución:

$$\mathbf{M}_\perp(t) = \mathbf{M}_\perp(0)e^{-t/T_2} \tag{2.22}$$

Esta ecuación describe el decaimiento exponencial de la magnetización transversal en cualquier marco de referencia. Este comportamiento se presenta en la figura **2.2**.

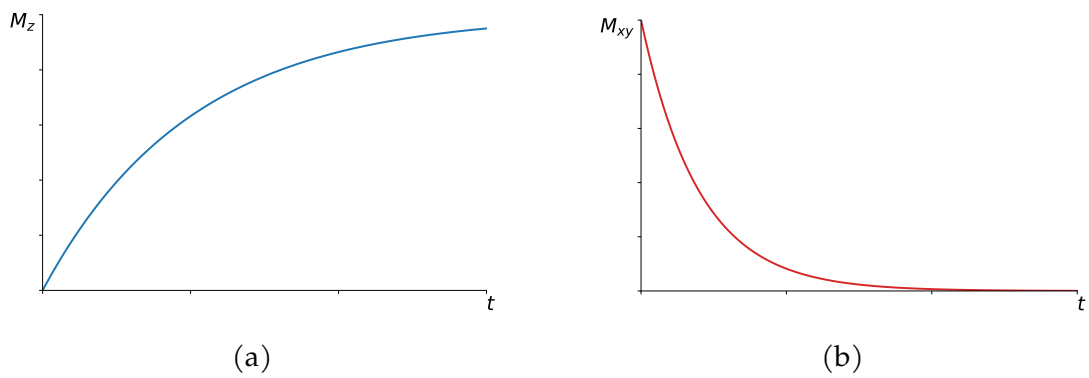


Figura 2.2.: Tiempos de relajación T_1 (*azul*) y T_2 (*rojo*). Las curvas representan las soluciones mostradas por las ecuaciones 2.17 y 2.21.

El tiempo de relajación T_2 es del orden de $\sim 10^1$ milisegundos para los protones en la mayoría de tejidos humanos (tabla **2.1**). Este parámetro es aproximadamente constante en los rangos usuales de B_0 para un tejido específico. Los valores de T_2 suelen ser mucho más cortos en sólidos y mucho más largos en líquidos.

2.1.4. Ecuaciones de Bloch

Las ecuaciones de Bloch son un conjunto de ecuaciones macroscópicas que se utilizan para calcular la magnetización nuclear en función del tiempo $\mathbf{M}(t)$ cuando están presentes

los tiempos de relajación T_1 y T_2. La expresión de movimiento para un solo momento magnético se puede generalizar para la magnetización total, resultando en:

$$\frac{d\mathbf{M}(t)}{dt} = \gamma \mathbf{M}(t) \times \mathbf{B}(t) \tag{2.23}$$

donde $\mathbf{B}(t)$ puede incluir componentes variables en el tiempo además del campo magnético estático $\mathbf{B}_0$. En el equilibrio térmico, en ausencia de campos magnéticos adicionales, la ecuación (2.23) expresa el hecho de que la componente z de la magnetización $\mathbf{M}$ es constante, es decir, $M_z'(t) = 0$.

El componente de la magnetización de un pulso de radiofrecuencias que está polarizado linealmente a lo largo del eje x en el marco de referencia del laboratorio se puede escribir como:

$$\mathbf{B}_1(t) = |B_1| \cos(\omega t)\hat{\mathbf{x}} \tag{2.24}$$

En presencia de dos campos magnéticos $\mathbf{B}_0$ y $\mathbf{B}_1$, la ecuación (2.23) se puede expandir para mostrar claramente las ecuaciones de Bloch en el marco de referencia del laboratorio junto con sus términos de relajación [37]:

$$\frac{dM_x(t)}{dt} = \gamma \left[M_y(t)B_0 - M_z(t)B_{1y}\right] - \frac{1}{T_2}M_x(t) \tag{2.25}$$

$$\frac{dM_y(t)}{dt} = \gamma \left[M_z(t)B_{1x} - M_x(t)B_0\right] - \frac{1}{T_2}M_y(t) \tag{2.26}$$

$$\frac{dM_z(t)}{dt} = \gamma \left[M_x(t)B_{1y} - M_y(t)B_{1x}\right] - \frac{1}{T_1}(M_z(t) - M_0) \tag{2.27}$$

O

$$\frac{\partial \mathbf{M}(\mathbf{r},t)}{\partial t} = \gamma \mathbf{M}(t) \times \mathbf{B}(\mathbf{r},t) - \frac{M_x(t)\hat{\mathbf{x}} + M_y(t)\hat{\mathbf{y}}}{T_2} - \frac{(M_z(t) - M_0)\hat{\mathbf{z}}}{T_1} \tag{2.28}$$

2.2. Difusión molecular

La difusión es el movimiento de traslación aleatorio (browniano) de moléculas o iones que es impulsado por energía térmica interna de un sistema. Aunque los tiempos T_1 y T_2, son afectados por el campo magnético, la difusión es independiente de este. La RMN es la única técnica que permite la visualización y el cálculo de la difusión molecular *in vivo* directamente a partir de los movimientos moleculares de traslación libre del agua [32,38].

Ecuaciones de Fick

En 1855, Adolf Fick [39] ideó dos ecuaciones diferenciales que describían cuantitativamente la difusión de moléculas a través de membranas ultrafinas. De acuerdo con la primera ley de difusión de Fick, el flujo de partículas $\mathbf{J}(\mathbf{r}, t)$, en la posición espacial $\mathbf{r}$ y el tiempo t, es directamente proporcional al gradiente de concentración o densidad de partículas $\rho(\mathbf{r}, t)$ de acuerdo con [40]:

$$\mathbf{J}(\mathbf{r}, t) = -D\boldsymbol{\nabla}\rho(\mathbf{r}, t) \tag{2.29}$$

donde D es el coeficiente de difusión. La conservación de la masa se expresa mediante la ecuación de continuidad:

$$\frac{\partial}{\partial t}\rho(\mathbf{r}, t) = -\boldsymbol{\nabla} \cdot \mathbf{J}(\mathbf{r}, t) \tag{2.30}$$

La combinación de estas ecuaciones conduce a la segunda ley de difusión de Fick:

$$\frac{\partial}{\partial t}\rho(\mathbf{r}, t) = D\boldsymbol{\nabla}^2\rho(\mathbf{r}, t) \tag{2.31}$$

Cuya solución para la difusión isotrópica con la condición inicial $\rho(\mathbf{r}, 0) = \delta(\mathbf{r}_0 - \mathbf{r})$ y condiciones de frontera $\rho(\mathbf{r} \to \infty, t) = 0$ es:

$$\rho(\mathbf{r}, t) = \left(\frac{1}{\sqrt{4\pi Dt}}\right)^3 e^{-\mathbf{r}'^2/4Dt} \tag{2.32}$$

donde, de hecho, ρ es análogo a la probabilidad de que una partícula se haya movido de la posición $\mathbf{r}_0$ a la posición $\mathbf{r}$ en el tiempo t y donde $\mathbf{r'} = \mathbf{r} - \mathbf{r}_0$.

De la ecuación (2.32) se puede apreciar que las unidades en el SI del coeficiente de difusión D son:

$$[D] = \frac{\mathrm{m}^2}{\mathrm{s}} \tag{2.33}$$

En coeficiente de difusión para líquidos es del orden de 10^{-5} cm^2/s, para gases del orden de 10^{-1} cm^2/s y para sólidos de 10^{-9} cm^2/s [41].

Estadísticamente el radio promedio de la distribución de partículas se calcula considerando una distribución gaussiana cuya mediana seria nula ya que la probabilidad de desplazarse en cualquier dirección es la misma pero la varianza de la distancia recorrida es proporcional tanto al tiempo como al cuadrado del radio promedio de la distribución de

acuerdo con la ecuación de Einstein–Smoluchowski:

$$\left\langle \mathbf{r}'^2 \right\rangle = 2Dt \qquad \text{(en 1 dimensión)} \tag{2.34}$$

$$\left\langle \mathbf{r}'^2 \right\rangle = 4Dt \qquad \text{(en 2 dimensiones)} \tag{2.35}$$

$$\left\langle \mathbf{r}'^2 \right\rangle = 6Dt \qquad \text{(en 3 dimensiones)} \tag{2.36}$$

donde $\left\langle \mathbf{r}'^2 \right\rangle$ es el desplazamiento cuadrado medio de las moléculas [42].

Ecuación de Bloch-Torrey

La ecuación de Bloch-Torrey es una generalización de las ecuaciones de Bloch, que incluye términos añadidos debido a la transferencia de magnetización por difusión.

Para describir cuantitativamente la difusión en presencia de gradientes de campo magnético variables en el tiempo, lo más conveniente es seguir el formalismo descrito por Torrey [115], que posteriormente fue utilizado por Stejskal y Tanner [114]. Las ecuaciones de Bloch ampliadas para incluir los efectos de la difusión están dadas a partir de las ecuaciones de Bloch (2.28) por [43]:

$$\frac{\partial \mathbf{M}}{\partial t} = \gamma \mathbf{M} \times \mathbf{B} - \frac{M_x \hat{\mathbf{x}} + M_y \hat{\mathbf{y}}}{T_2} - \frac{(M_z - M_0)\hat{\mathbf{z}}}{T_1} + D\boldsymbol{\nabla}^2 \mathbf{M} \tag{2.37}$$

Esta ecuación diferencial parcial posee soluciones exactas en el continuo dependiendo de las condiciones iniciales del problema y diferentes posibles situaciones como que la activación del gradiente de campo magnético sea dependiente del tiempo, que el gradiente sea sinusoidal o que la densidad inicial de espines sea gaussiana en vez de ser uniforme. Otros tipos de soluciones incluyen las soluciones en el espacio discreto asumiendo gradientes constantes o de dos pulsos [44]. Sin embargo una solución representativa en el campo médico es aquella que no asume sistemas infinitamente extendidos sino los que están confinadas en un volumen por paredes como capilares, venas, arterias o tejido. Esta solución tiene la forma:

$$\mathbf{M}_+(\mathbf{r}, t) = A e^{-i\gamma \mathbf{r} \cdot \int_0^t \mathbf{G}(t')\, dt'} \tag{2.38}$$

donde $\mathbf{M}_+ = \mathbf{M}_x + i\mathbf{M}_y$ en la que se han ignorado los términos de relajación, $\mathbf{G} = \mathbf{G}(t)$, que es independiente de la posición, viene de la sustitución $\mathbf{B} = \mathbf{G} \cdot \mathbf{r}\,\hat{\mathbf{z}}$ y $A = A(t)$ puede obtenerse a partir de [30]:

$$\ln\left(\frac{A(t)}{A(0)}\right) = -D\gamma^2 \int_0^t dt'' \int_0^{t''} \mathbf{G}^2(t')dt' \tag{2.39}$$

El valor de $\mathbf{G}$ se corresponde con el gradiente de campo magnético de una secuencia de obtención de una IRM, mientras que A resulta ser la atenuación de la intensidad de la señal en un vóxel de dicha imagen que posteriormente se denotará como S/S_0.

Técnica de Stejskal y Tanner

La RMN es capaz de medir con precisión la difusión de moléculas en líquidos isotrópicos. En su artículo sobre ecos de espín, Hahn [45] describió la medición de la difusión en presencia de un gradiente constante, que posteriormente fue modificado y desarrollado por Carr y Purcell [46]. Sin embargo, no fue hasta el desarrollo de un método de eco de espín de gradiente pulsado por Stejskal y Tanner [12] que las mediciones de difusión de RMN se convirtieron en una técnica de rutina para medir los coeficientes de difusión molecular. La figura **2.3** muestra el métodos de gradiente pulsado ideado por Stejskal y Tanner comúnmente utilizados con gradientes constantes, para la medición de la difusión incorporada en una secuencia de eco de espín de Hahn.

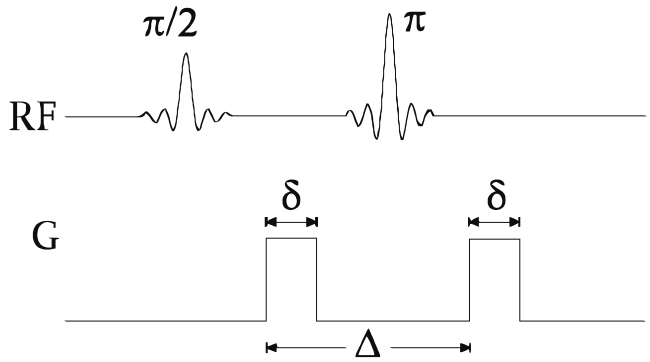


Figura 2.3.: Diagrama de la secuencia convencional de Stejskal-Tanner para el cálculo de la difusión.

La secuencia de eco de espín de Hahn se basa en la aplicación de dos pulsos de RF: un pulso de $\pi/2$ seguido de un pulso de π. El método de eco de espín es un procedimiento para recuperar la pérdida de señal debida a falta de homogeneidad en el campo estático externo $\mathbf{B}_0$. De este modo lo efectos de T_1 se reducen, aunque existe otro efecto que surge de las mismas faltas de homogeneidad que pueden causar pérdida de señal: la "autodifusión" de moléculas a través de campos externos no homogéneos. Este efecto puede llegar a dominar el tiempo de relajación T_2, especialmente en líquidos y gases [32].

La integral del gradiente en el tiempo para la secuencia de la figura **2.3** está descrita por:

$$\int_0^{t''} G(t')dt' = \begin{cases} Gt'' & \text{para} \quad t'' < \delta \\ G(t'' - (\Delta + \delta)) & \text{para} \quad \Delta < t'' < \Delta + \delta \end{cases} \tag{2.40}$$

Así, la ecuación (2.39) se convierte en:

$$\ln A = -D\gamma^2 \int_0^{\delta} G^2 t''^2 dt'' - D\gamma^2 \int_{\Delta}^{\Delta+\delta} G^2 (t'' - (\Delta - \delta))^2 dt'' \tag{2.41}$$

finalmente, resolviendo y simplificando se obtiene que:

$$\ln A = -D\gamma^2 G^2 \delta^2 \underbrace{\left(\Delta - \frac{1}{3}\delta\right)}_{t_{\text{diff}}} \tag{2.42}$$

La potenciación en difusión es función de la duración del gradiente δ, de su amplitud G y del intervalo entre los gradientes de difusión Δ . La duración del gradiente y el intervalo entre gradientes determina el tiempo durante el que se mide la difusión. Este se denomina tiempo de difusión t_{diff} [30,32].

2.2.1. Imágenes pesadas por difusión (DWI)

Las imágenes ponderadas por difusión (DWI) son imágenes crudas donde la señal se ha sensibilizado a la difusión mediante el uso de pulsos de gradiente. El grado de sensibilización de difusión se define por el valor b. El contenido DWI se ve afectado por muchos parámetros distintos de la difusión, principalmente los tiempos de relajación T_1 y T_2. Debido a que estos parámetros pueden no tener el mismo comportamiento que la difusión para diferentes condiciones fisiológicas o patológicas, las variaciones en la intensidad de la imagen pueden ser difíciles de interpretar [42]. Por lo tanto, siempre que sea posible, se deben preferir las imágenes de difusión cuantitativas.

2.2.2. Modelo mono-exponencial

Si se aplica sobre un vóxel la secuencia clásica espín-eco se obtiene una señal S_0. Si se repite la secuencia activando un gradiente en una dirección determinada, los núcleos atómicos que se mueven en dicha dirección presentarían un mayor desfase y por tanto la señal S detectada será menor. Estas dos señales están relacionadas mediante la ecuación:

$$\frac{S}{S_0} = e^{-bD} \tag{2.43}$$

donde el valor de b (*b-value*) corresponde a:

$$b = \gamma^2 G^2 \delta^2 \, t_{\text{diff}} \qquad \text{con} \quad [b] = \frac{\text{s}}{\text{mm}^2} \tag{2.44}$$

Aunque este valor b se da a menudo en la literatura, nunca puede representar el valor real, ya que implica un tiempo de subida de gradiente infinitamente corto. Las formas de onda

de gradiente más realistas serían los gradientes semisinusoidales o de forma trapezoidal.

El confinamiento por paredes biológicas naturales también juega un papel importante en la medida de la difusión. Para difusión libre y sin restricciones, el desplazamiento cuadrático medio aumenta linealmente con el tiempo, lo que de acuerdo con las ecuaciones (2.34)-(2.36) conduce a un coeficiente de difusión constante independiente del tiempo de difusión.

En presencia de límites físicos, la difusión puede parecer libre y sin restricciones durante tiempos de difusión breves. Sin embargo, para tiempos de difusión más largos, un número creciente de partículas encuentra la barrera física, lo que conduce a un desplazamiento cuadrático medio reducido, que para un compartimento esférico se aproxima a las dimensiones del compartimento. El desplazamiento cuadrático medio restringido conduce a un coeficiente de difusión cada vez más pequeño con tiempos de difusión crecientes [47].

Para que la atenuación por difusión sea medible, el producto bD tiene que ser del orden de la unidad. El valor típico de D en el tejido cerebral es de $\sim 10^{-5}$ cm^2/s (0,001 mm^2/s) [42].

2.2.3. Coeficiente Aparente de Difusión (ADC)

Es posible obtener mapas puros de difusión de agua al adquirir dos imágenes con diferentes valores b, a partir de la ecuación (2.43) mediante:

$$\mathrm{ADC} = \frac{\ln(S_0/S_1)}{b_1 - b_0} \tag{2.45}$$

donde b_0 y b_1 son los *b-values* correspondientes a las señales S_0 y S_1 de dos imágenes de difusión. El cálculo se puede realizar en regiones de interés o en una base vóxel por vóxel, lo que proporciona imágenes calculadas ADC. Esta ecuación proporciona el verdadero coeficiente de difusión D solo si la difusión de agua se comporta libremente, es decir, la distribución de desplazamientos impulsados por difusión sigue una distribución gaussiana. En esta ecuación se reemplaza el coeficiente de difusión exacto, D, con un ADC para indicar que se esperaba la desviación de la difusión verdadera. Además, el ADC también puede incluir un componente que se origina en el flujo sanguíneo en los recipientes aleatorios a través del efecto de movimiento incoherente de intravoxel [6, 48].

2.2.4. Movimiento Incoherente Intravoxel (IVIM)

Las imágenes IVIM son un concepto y un método introducido y desarrollado inicialmente por D. Le Bihan *et al.* en 1986 [13] para evaluar cuantitativamente todos los movimien-

tos traslacionales microscópicos que podrían contribuir a la señal adquirida con DWI. En este modelo, el tejido biológico contiene dos ambientes distintos: la difusión molecular del agua en el tejido y la microcirculación de la sangre en la red capilar (perfusión). El concepto es que el agua que fluye en los capilares (a nivel de vóxel) se mueve aleatoriamente (pseudo-difusión), siempre que no haya un flujo coherente neto en ninguna dirección [48].

Pseudo-difusión y perfusión

La microcirculación de la sangre en la red capilar puede contemplarse como un movimiento incoherente a escala del vóxel y aparece entonces la microcirculación como un movimiento al azar que puede interpretarse como una *pseudo-difusión*. El factor D^* (factor de pseudo-difusión) experimentalmente se comprueba que es mayor que D (en un orden de magnitud). Según la teoría IVIM se puede asumir que el vóxel está formado por dos compartimentos: el formado por el flujo capilar y el estático en donde solo se manifiesta la difusión. El factor con que el compartimento dependiendo de la microcirculación contribuye al volumen del vóxel es muy pequeño (menos de un 5 % en el cerebro) [2]. Teniendo en cuenta estos factores se ha demostrado que el efecto de la contribución de la microcirculación a la atenuación de la señal tan sólo se manifiesta a valores de b bajos (menores a $\sim 180\,\mathrm{s/mm^2}$). Pero a partir de ciertos valores de b (mayores a $\sim 200\,\mathrm{s/mm^2}$) solo se manifiesta la difusión. Por tanto el ADC es un parámetro que mezcla difusión y perfusión pero con factores b altos refleja únicamente la influencia de la difusión [49].

La perfusión sanguínea describe el aporte de sangre a un determinado elemento de tejido. Junto con la sangre, la perfusión asegura el aporte de oxígeno y nutrientes a las células. Para ello, sangre arterial oxigenada entra en la red capilar y tanto el oxígeno como los nutrientes son transportados a través de las paredes de los capilares para ser utilizados por las células. Por otro lado, los desechos resultantes del metabolismo tisular son eliminados junto con la sangre venosa desoxigenada a la red venosa capilar [2].

La fracción de perfusión f en IRM se contempla desde el punto de vista hemodinámico sin tener en cuenta el transporte fuera del espacio vascular. De esta forma se puede definir como la cantidad de sangre aportada al lecho capilar de un determinado volumen de tejido en un determinado intervalo de tiempo [42].

2.2.5. Modelo bi-exponencial

Como ya se ha mencionado antes, la componente de microcirculación sanguínea decae aproximadamente 10 veces más rápido que la componente de difusión tisular (o celular),

lo que permite describir los dos fenómenos independientemente. La atenuación total de la señal de resonancia magnética S/So, es simplemente la suma de los componentes tisulares y sanguíneos (suponiendo que el intercambio de agua entre la sangre y los tejidos sea insignificante durante el tiempo de codificación), tomando la forma de un decaimiento bi-exponencial:

$$\frac{S}{S_0} = F_{\text{Pseu}} + F_{\text{Diff}} \tag{2.46}$$

$$\frac{S}{S_0} = fe^{-bD^*} + (1-f)e^{-bD} \tag{2.47}$$

Sin embargo, dada la pequeña fracción de la sangre que fluye en comparación con el contenido total de agua en el tejido, el efecto IVIM generado por perfusión aparece más como una desviación a valores b pequeños en el decaimiento de la señal mono-exponencial impulsada por la difusión tisular (figura **2.4**) [2].

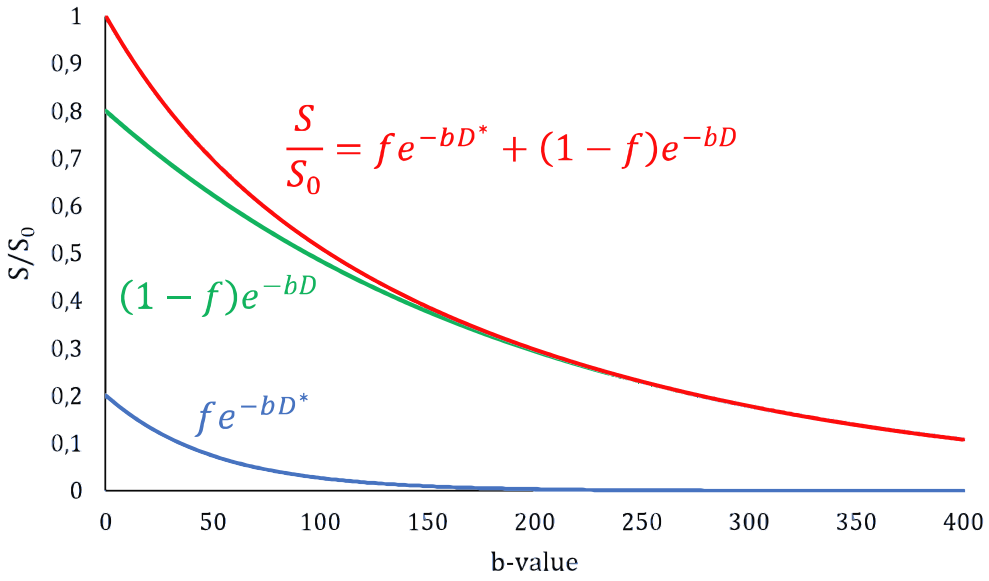


Figura 2.4.: Efectos de la difusión y la pseudo-difusión en la señal de resonancia magnética. Atenuación por difusión (*verde*), atenuación por pseudo-difusión (*azul*), señal obtenida (*rojo*). La difusión tisular y la componente de flujo sanguíneo contribuyen por separado a la señal, lo que resulta en una forma bi-exponencial. Sin embargo, como $D^* \gg D$ el efecto IVIM aparece como una desviación de la decaimiento de la señal de difusión tisular a valores b bajos.

2.2.6. Efecto de la curtosis

No tener en cuenta los efectos de difusión no gaussianos podría dar lugar a estimaciones erroneas de los parámetros f y D^*, en un modelo tisular sin perfusión. La difusión no gaussiana que se hace visible a valores altos de b se origina a partir de la interacción de las moléculas difusoras de agua con obstáculos, como las membranas celulares, y da como resultado una curvatura de la atenuación de la señal. El modelado del componente de difusión tisular con un modelo mono-exponencial (como el ADC) subestima la "difusión genuina" (ADC $< D$) cuando se incluye la señal a valores b altos, lo que, a su vez, crea un efecto IVIM artificioso. La difusión no gaussiana debe manejarse adecuadamente utilizando modelos adecuados, como el modelo de curtosis para evitar tales efectos erróneos en el modelo IVIM [2,49,50].

En todos los casos, una estimación precisa de los parámetros IVIM requiere primero un manejo adecuado del componente de difusión que generalmente representan más del 90 % de la señal total a valores b bajos. Aunque el concepto ADC ha sido extremadamente útil, se basa en el supuesto de difusión libre gaussiana. Se ha establecido que el proceso de difusión en los tejidos no es gaussiano, ya que la gráfica semi-log de la señal de decaimiento (figura **2.5**), en tejidos biológicos exhibe una curvatura no predicha por la ecuación (2.43) [51].

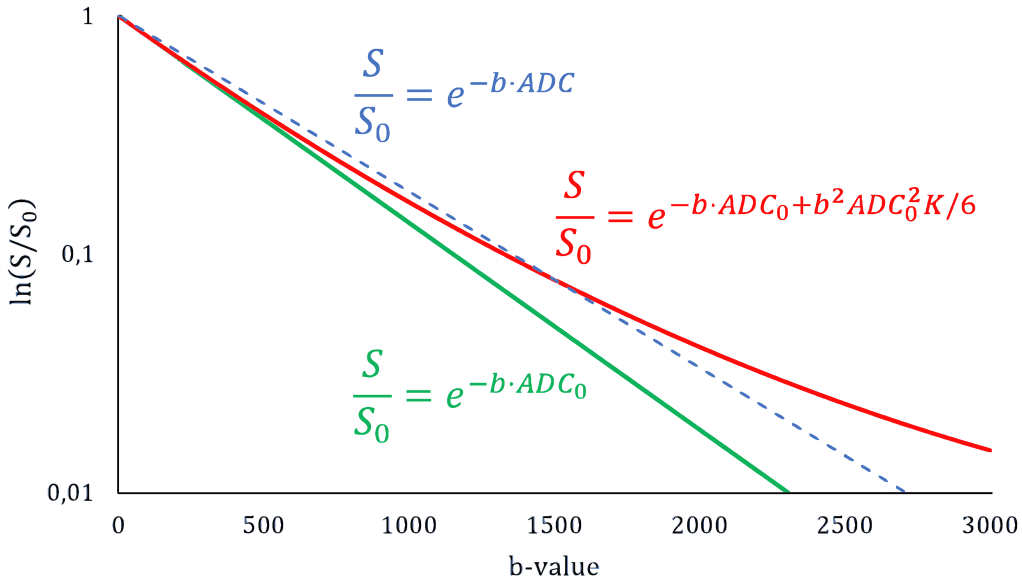


Figura 2.5.: Efecto de la curtosis en la difusión no gaussiana. Atenuación por difusión (*verde*), difusión mono-exponencial (*azul*) y señal obtenida (*rojo*).

Se han sugerido varios modelos para manejar empíricamente este comportamiento no gaussiano, como el modelo polinómico o de curtosis [52, 53], el modelo bi-exponencial [54], el modelo estadístico [55], el modelo exponencial "estirado" [56], el modelo triexponencial [57], el modelo de difusión anómala [58], entre otros, con éxito, dependiendo del contexto y los tejidos bajo examen. Con el modelo de curtosis se tiene:

$$F_{\text{diff}} \approx e^{-b \cdot ADC_0 + b^2 \cdot ADC_0^2 \cdot K/6} \tag{2.48}$$

donde ADC_0 es el ADC virtual que se obtendría cuando b se acerca a 0. El coeficiente adimensional K (curtosis) caracteriza el grado de desviación del comportamiento de la señal de un decaimiento mono-exponencial ($K = 0$ cuando los desplazamientos moleculares impulsados por la difusión siguen una distribución gaussiana), un indicador de la heterogeneidad del entorno de difusión. Por lo tanto, la ecuación (2.47) ahora puede reescribirse como [49]:

$$\frac{S}{S_0} = f e^{-bD^*} + (1-f) e^{-bD + b^2 D^2 K/6} \tag{2.49}$$

La curtosis por tanto, debe ser un valor adimensional.

2.3. Técnicas de Machine Learning (ML)

El aprendizaje automático o Machine Learning (ML) es un campo de investigación que se encarga de formular algoritmos que pueden aprender a través de la experiencia y mediante el uso de información previamente procesada. Los algoritmos de ML toman como base algunos datos de muestra, conocidos como datos de entrenamiento (*training data*), para realizar predicciones o tomar decisiones sin estar explícitamente programados para hacerlo. Los algoritmos de ML se utilizan ampliamente en medicina y tratamiento de imágenes donde es difícil o inviable desarrollar algoritmos convencionales para realizar ciertas tareas. En este caso, se busca reducir el tiempo de estimación de los parámetros característicos de los modelos ADC e IVIM en imágenes de difusión por RMN.

2.3.1. Aprendizaje supervisado

El aprendizaje supervisado es el medio por el cual una máquina aprende con base en pares de datos de "entrada-salida" de ejemplo previamente etiquetados. Cada ejemplo consiste en un vector de entrada y un valor de salida deseado con los que produce una función de inferencia, que se puede utilizar para predecir el resultado de salida en nuevos vectores de entrada. Un escenario óptimo permitirá que el algoritmo determine correctamente las

etiquetas de clase para instancias desconocidas. Esto requiere que el algoritmo de aprendizaje generalice a partir de los datos de entrenamiento a situaciones desconocidas de una manera razonable.

Algunos factores a considerar para elegir y aplicar un algoritmo de aprendizaje incluyen los siguientes [59]:

Heterogeneidad de los datos: Para los vectores de características que incluyen varios tipos de datos, algunos algoritmos son más fáciles de aplicar que otros. Para la predicción de parámetros como la difusión que requieren que las características de entrada sean numéricas, algoritmos como la regresión lineal, la regresión logística, support vector machines (SVM), las redes neuronales (RNA) y los métodos de vecinos más cercanos (KNN) suelen tener mejores resultados. Para que otros parámetros como la pseudo-difusión se puedan considerar dentro del mismo grupo de entrenamiento es necesario que estos se escalen (un orden de magnitud por encima) a rangos similares. Una ventaja de algoritmos como los árboles de decisión es que manejan fácilmente datos heterogéneos.

Redundancia en los datos: Si las características de entrada contienen información redundante (por ejemplo, características altamente correlacionadas como la existencia de valores de perfusión cuando se considera el cálculo de la pseudo-difusión), algunos algoritmos de aprendizaje como la regresión lineal, regresión logística y métodos basados en distancia podrían no funcionar adecuadamente sin imponer alguna forma de regularización.

Presencia de interacciones y no linealidades: Si cada una de las características hace una contribución independiente a la salida, entonces los algoritmos basados en funciones lineales como la regresión lineal o logística, SVM y funciones de distancia como KN generalmente funcionan bien. Sin embargo, si hay interacciones complejas entre las características, entonces los algoritmos como RNA funcionan mejor, porque están diseñados específicamente para descubrir estas interacciones. La estimación de parámetros en modelos ADC e IVIM son una mezcla de ambos tipos de funciones; ADC en el espacio semi-log es generalmente lineal mientras que IVIM puede serlo o no dependiendo de los valores b a estudiar. Para determinar el algoritmo con el mejor desempeño, se suelen comparar varios de ellos para determinar experimentalmente cuál funciona mejor para el problema en cuestión.

Manejo de datos faltantes

Dentro del preprocesamiento de los datos se debe considerar como se manejará la información que falta en los conjuntos de ejemplo. Si algunos de estos datos faltan, entonces el algoritmo escogido puede generalizar erróneamente y arrojar resultados equivocados.

Existen principalmente dos formas de manejar los datos faltantes [60]: eliminando la fila en particular y completando la información mediante la media entre los datos conocidos (*data imputing*). Eliminar las filas con valores faltantes no es viable ni eficiente ya que conduce a una pérdida enorme de información. Calcular la media de los datos conocidos es una estrategia más eficaz para asignar a los parámetros de difusión, pseudo-difusión, fracción de perfusión o curtosis, ya que se pueden estimar tomando como referencia otros datos conocidos y, además permite evaluar la dispersión de esta media para evaluar la precisión de la asignación.

Estandarización y escalado de características

Cuando los predictores son numéricos, la escala en la que se miden, así como la magnitud de su varianza pueden influir en los resultados de los modelos. A veces aquellas variables independientes que se midan en una escala mayor o que tengan más varianza dominarán el modelo, aunque no sean los que más relación tienen con la variable respuesta. Por ello es necesario ubicar las variables en el mismo rango y en la misma escala para que una variable no domine sobre ninguna otra. Los métodos principales para estandarizar y normalizar datos son [61]:

Normalización: se trata de transformar los datos de forma que todos los predictores estén aproximadamente en la misma escala. Los valores de S/S_0 ya se encuentran normalizados a 1 para cada vóxel, pero no se encuentran normalizados respecto a los demás vóxeles de la imagen y mucho menos respecto a distintas imágenes. Esta transformación se obtiene mediante la ecuación (2.50).

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{2.50}$$

Centrado: consiste en restar a cada valor x la media μ del predictor al que pertenece (ecuación (2.51)). Como resultado de esta transformación, todos los predictores pasan a tener una media de cero (o cercana a cero) y se centrarán en torno al origen.

$$x_0 = x - \mu \tag{2.51}$$

Estandarización (z-score): en esta técnica se divide cada predictor entre su desviación estándar después de haber sido centrado, de esta forma, los datos pasan a tener una distribución normal conforme a la ecuación (2.52).

$$z = \frac{x_0}{\sigma} = \frac{x - \mu}{\sigma} \tag{2.52}$$

División del conjunto de datos (training y test)

Como parte del preprocesamiento se suele dividir el conjunto de datos en un subconjunto de entrenamiento (*training*) y un subconjunto de pruebas (*test*). Este seccionamiento de datos permite mejorar el rendimiento y precisión de un algoritmo.

Si se entrena un modelo de ML con el conjunto global de datos y posteriormente se prueba con un conjunto de datos perteneciente al mismo conjunto, su precisión será alta (o perfecta) ya que previamente conocía la salida que debería entregar para la entrada dada. Así se puede pensar que el modelo es bueno cuando en realidad solo está duplicando información. Pero si el modelo se entrena solamente con una fracción de los datos y luego se prueba con datos de su complemento (datos desconocidos hasta ese momento), entonces se podrá verificar su rendimiento comparando el resultado obtenido con el resultado esperado.

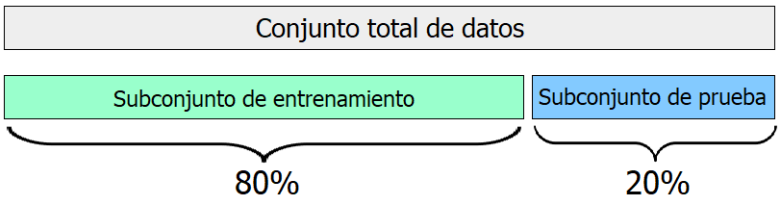


Figura 2.6.: División típica de un conjunto de datos en datos de entrenamiento y datos de prueba.

Para dividir el conjunto de datos en el subconjunto de entrenamiento y el subconjunto de prueba, se suele escoger aleatoriamente el 80 % de los pares "entrada-salida" para entrenar y el 20 % restante para pruebas como se muestra en la figura 2.6, no obstante, estas proporciones pueden cambiar.

2.3.2. Evaluación de la predicción

Una vez entrenado un algoritmo de ML, es necesario verificar su precisión y evaluar su rendimiento, de este modo determinar si es adecuado para ejecutar la función para la que se ha desarrollado. Existen diferentes métodos para probar su eficiencia, algunos de estos mejor diseñados para algoritmos clasificadores, y otros para algoritmos de predicción o de regresión.

Cross-validation

La validación cruzada o *cross-validation* es una técnica para validar la eficiencia del modelo entrenado en el subconjunto de datos de entrenamiento de entrada sobre un subconjunto

nunca antes visto. En otras palabras, es una técnica para comprobar cómo un modelo estadístico generaliza a un conjunto de datos independiente.

En ML, siempre existe la necesidad de probar la estabilidad del modelo. Para este propósito, se reserva una muestra particular del conjunto de datos, que ya formaba parte del conjunto de entrenamiento. Después de eso, se prueba el modelo en esa muestra antes de la implementación, y este proceso completo se somete a validación cruzada. Esto es algo diferente de la división general entrenamiento-prueba.

Existen varios métodos utilizados para realizar una validación cruzada, entre ellos: la aproximación al conjunto de validación, la Leave-P-out cross-validation, Leave-one-out, K-fold, K-fold estratificada o Holdout entre otros. De todas estas, la validación cruzada k-fold estratificada es una de las más comunes. En esta se divide el conjunto de datos de entrada en k grupos de muestra de tamaños iguales, cada uno de los cuales tiene aproximadamente la misma respuesta o salida en promedio. Esto último con el fin de evitar sesgos debido a la alta varianza entre muestras. Para cada conjunto de aprendizaje, la función de predicción utiliza $k - 1$ muestras, y el subconjunto sobrante se utiliza como conjunto de prueba (figura 2.7) [62].

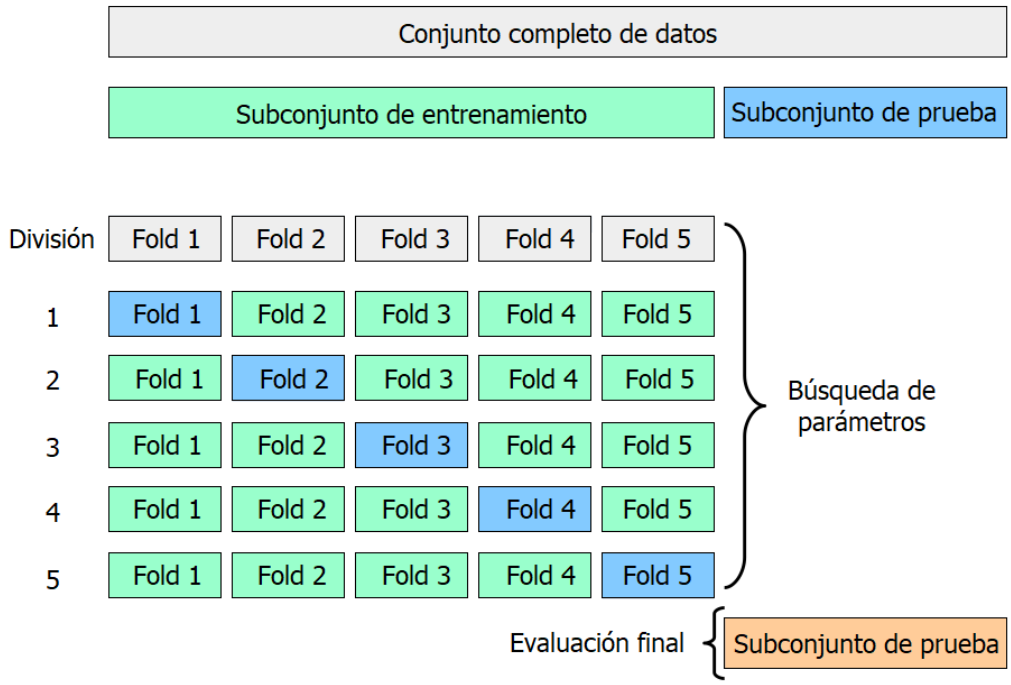


Figura 2.7.: Ejemplo de k-fold cross-validation para $k = 5$. Adaptado de [62].

Los pasos para la validación cruzada k-fold estratificada son:

- Reorganizar aleatoriamente los datos de entrada de forma que el valor promedio de su salida sea más o menos el mismo.
- Dividir el conjunto de datos en k grupos de igual tamaño.
- Reservar el primer subconjunto para ejecutar la evaluación del rendimiento del modelo.
- Repetir el procedimiento, reservando el segundo grupo como conjunto de validación.
- Continuar el procedimiento hasta que todos los k subconjuntos se hayan utilizado como prueba.

La medida de rendimiento reportada por k-fold cross-validation es entonces la contribución de los valores calculados en todo el proceso. Este enfoque puede ser computacionalmente costoso, pero no desperdicia demasiados datos como sucede con otros modelos [63].

Matrices de confusión

Una matriz de confusión o *confusion matrix* es una tabla que permite la visualización del rendimiento de un algoritmo. Cada fila de la matriz representa las instancias en una clase real, mientras que cada columna representa las instancias en una clase predicha. El nombre proviene del hecho de que permite observar fácilmente si el sistema está confundiendo dos clases, es decir, etiquetando erróneamente una como otra [64].

Posee dos dimensiones: la del valor real y la del valor predicho; y conjuntos idénticos de clases en ambas dimensiones (por lo general 2 clases). Muestra el número de verdaderos positivos, falsos negativos, falsos positivos y verdaderos negativos (tabla **2.2**) lo cual permite un análisis más detallado que simplemente observar la proporción de clasificaciones correctas. La precisión arrojará resultados engañosos si el conjunto de datos está desequilibrado; es decir, cuando el número de observaciones en diferentes clases varía mucho.

Una matriz de confusión no se limita a la clasificación binaria sino que también se puede usar en clasificadores de varias clases [65], no obstante, una matriz de confusión con demasiadas clases puede resultar contraproducente debido a la saturación visual de la información a clasificar.

Valor real	Negativo (N)	Verdadero Negativo (TN)	Falso Positivo (FP)
	Positivo (P)	Falso Negativo (FN)	Verdadero Positivo (TP)
	Total (P+N)	Negativo (n)	Positivo (p)
		Predicción	

Tabla 2.2.: Matriz de confusión 2x2 para un clasificador binario.

La tabla 2.2 muestra una distribución típica de los valores de una predicción respecto de los valores reales en una matriz de confusión donde:

TP: Verdaderos positivos o éxitos
TN: Verdaderos negativos o rechazos correctos
FP: Falsos positivos o falsas alarmas o error tipo I
FN: Falsos negativos o error de tipo II

Además de los anteriores valores, se pueden obtener otras cantidades que dan información sobre el rendimiento del algoritmo (entre muchos otros):

TPR: Sensibilidad, recall, tasa de aciertos o tasa de verdaderos positivos

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{P}} = \frac{\mathrm{TP}}{(\mathrm{TP} + \mathrm{FN})} \tag{2.53}$$

FPR: Fall-out o tasa de falsos positivos

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{N}} = \frac{\mathrm{FP}}{(\mathrm{FP} + \mathrm{TN})} \tag{2.54}$$

ACC: Precisión (accuracy)

$$\mathrm{ACC} = \frac{(\mathrm{TP} + \mathrm{TN})}{(\mathrm{P} + \mathrm{N})} \tag{2.55}$$

TNR: Especificidad, selectividad o tasa de verdaderos negativos

$$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{N}} = \frac{\mathrm{TN}}{(\mathrm{FP} + \mathrm{TN})} = 1 - \mathrm{FPR} \tag{2.56}$$

Curva ROC

Una curva ROC (característica operativa del receptor), es una gráfica que ilustra la capacidad de diagnóstico de un sistema clasificador binario a medida que varía su umbral de discriminación y se genera trazando la tasa de verdaderos positivos (TPR) contra la tasa de falsos positivos (FPR) en varios ajustes de umbral [66].

La curva ROC poblacional representa 1−Especifidad frente a la Sensibilidad para cada posible valor umbral o punto de corte en la escala de resultados de la prueba en estudio. Es decir, $y = f(x)$, donde:

$$\text{ROC} = \begin{cases} y = \text{Sensibilidad} = \text{TPR} \\ x = 1 - \text{Especifidad} = \text{FPR} \end{cases} \tag{2.57}$$

Sin embargo, ante la dificultad de obtener datos poblacionales, se puede aproximar por la curva ROC muestral, que representa la fracción de falsos positivos en abscisas frente a la fracción de verdaderos positivos en ordenadas. Puesto que en ambos ejes se tienen probabilidades, la curva ROC, tanto muestral como poblacional, estará contenida en el cuadrado $[0, 1] \times [0, 1]$. Además, por convenio, se considera que los valores de y son, en general, mayores que x. Por tanto, la curva estará contenida en el triángulo: $\{(x, y) \mid 0 \leq x \leq y \leq 1\}$. Si por la naturaleza de la prueba los resultados estuviesen invertidos (valores de y menores que x) habría que invertir su salida [66].

El mejor método de predicción posible arrojaría un punto en la esquina superior izquierda o coordenada (0,1) del espacio ROC, que representa una sensibilidad del 100 % (sin falsos negativos) y una especificidad del 100 % (sin falsos positivos). Un clasificador aleatorio tiende hacia la línea diagonal.

AUC o estadístico c

El AUC o área bajo la curva ROC se usa a menudo para resumir en un solo número la capacidad diagnóstica de un clasificador. El AUC se define simplemente como el área del espacio ROC que se encuentra debajo de la curva ROC [67]:

$$\text{AUC} = \int_{x=0}^{1} \text{ROC}(x)\, dx \tag{2.58}$$

Sin embargo, en el espacio ROC hay regiones donde los valores de FPR o TPR son inaceptables o no viables en la práctica. Por ejemplo, la región donde la FPR es superior a 0,8 implica que más del 80 % de los valores negativos se clasifican incorrectamente como positivos (un resultado pobre para la mayoria de aplicaciones). Como consecuencia, el AUC

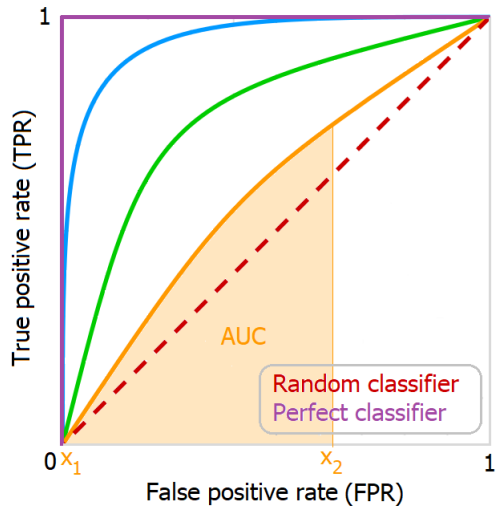


Figura 2.8.: Ejemplos de curvas ROC. La linea roja punteada representa un clasificador aleatorio. A medida que el clasificador mejora su desempeño, las curvas se desplazan hacia la curva morada (clasificador perfecto). En amarillo se muestra el área bajo la curva (AUC) amarilla desde $x_1 = 0$ hasta x_2. Adaptado de https://www.themachinelearners.com/curva-roc-vs-prec-recall/

calculado en todo el espacio ROC (es decir, con FPR y TPR que van de 0 a 1) puede proporcionar indicaciones engañosas. Para superar esta limitación del AUC, se suele calcular el área bajo la curva ROC que corresponde a valores de aceptables de FPR y TPR [66].

$$\text{AUC} = \int_{x_1}^{x_2} \text{ROC}(x)\, dx \tag{2.59}$$

El AUC es ampliamente utilizado, especialmente para comparar los rendimientos de dos (o más) clasificadores binarios: el clasificador que logra el AUC más alto se considera mejor. Por otra parte, el valor AUC es deseable porque es invariante de escala, es decir, mide qué tan bien se clasifican las predicciones, en lugar de sus valores absolutos y es invariante ante el umbral de clasificación, es decir, mide la calidad de las predicciones del modelo independientemente del umbral elegido [68, 69].

El AUC varía entre 0 y 1, donde un modelo cuyas predicciones son 100 % erróneas tiene un AUC de 0.0 mientras que uno cuyas predicciones son 100 % correctas tiene un AUC de 1.0.

2.3.3. Métricas de regresión

La regresión se refiere a problemas de modelado predictivo que implican predecir un valor numérico. Es diferente de la clasificación que implica predecir una etiqueta de clase. A diferencia de la clasificación, no puede utilizar la precisión de la clasificación para evaluar las predicciones realizadas por un modelo de regresión. En su lugar, debe utilizar métricas de error diseñadas específicamente para evaluar las predicciones realizadas sobre problemas de regresión [62,70].

La habilidad o el desempeño de un modelo de regresión se reporta como un error en sus predicciones. Si está prediciendo un valor numérico como la distancia que se mueve una partícula por unidad de tiempo, no se espera que el modelo prediga el valor exactamente; en cambio, sería más útil saber qué tan cerca estaban las predicciones de los valores esperados.

MAE

El error absoluto medio (MAE) es una métrica de riesgo correspondiente al valor esperado de la pérdida de error absoluto. Si $\hat{y}_i$ es el valor predicho de la i-ésima muestra, y y_i es el valor verdadero correspondiente, entonces el error absoluto medio estimado sobre n muestras se define como el promedio aritmético de los errores absolutos $|\varepsilon_i| = |y_i - \hat{y}_i|$, así:

$$\text{MAE}(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} |y_i - \hat{y}_i| \tag{2.60}$$

El error absoluto medio utiliza la misma escala que los datos que se están midiendo. Esto se conoce como una medida de precisión dependiente de la escala y, por lo tanto, no se puede usar para hacer comparaciones entre series utilizando diferentes escalas [62,71].

MSE

El error cuadrático medio (MSE), una métrica de riesgo correspondiente al valor esperado del error o pérdida al cuadrado. Mide el promedio de los errores al cuadrado, es decir, la diferencia entre el estimador y lo que se estima. Si $\hat{y}_i$ es el valor predicho de la i-ésima muestra, y y_i es el valor verdadero correspondiente, entonces el error cuadrático medio estimado sobre n muestras se define como:

$$\text{MSE}(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2 \tag{2.61}$$

RMSE

En una analogía con la desviación estándar, tomando la raíz cuadrada del MSE produce el error de la raíz cuadrada de la media o la desviación de la raíz cuadrada media RMSE, que tiene las mismas unidades que la cantidad que se estima; para un estimador no sesgado, el RMSE es la raíz cuadrada de la varianza, conocida como la desviación estándar [62,71].

$$\mathrm{RMSE}(y, \hat{y}) = \sqrt{\frac{1}{n}\sum_{i=0}^{n-1}(y_i - \hat{y}_i)^2} \tag{2.62}$$

La RMSE es una medida de precisión, para comparar errores de predicción de diferentes modelos para un conjunto de datos en particular y no entre conjuntos de datos, ya que depende de la escala.

R^2 score

El coeficiente de determinación o R^2 representa la proporción de varianza (de y) que ha sido explicada por las variables independientes en el modelo. Proporciona una indicación de la bondad de ajuste y, por lo tanto, una medida de qué tan bien es probable que el modelo prediga muestras no vistas, a través de la proporción de varianza explicada.
Como dicha varianza depende del conjunto de datos, R^2 puede no ser significativamente comparable entre diferentes conjuntos de datos. La mejor puntuación posible es 1.0 y puede ser negativa (porque el modelo puede ser arbitrariamente peor). Un modelo constante que siempre predice el valor esperado de y, sin tener en cuenta las características de entrada, obtendría una puntuación R^2 de 0. Si $\hat{y}_i$ es el valor predicho de la i-ésima muestra y y_i es el valor verdadero correspondiente para las muestras totales, el R^2 estimado se define como [62,72]:

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{2.63}$$

donde

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad \text{y} \qquad \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}\varepsilon_i^2 \tag{2.64}$$

Se debe tener en cuenta que R^2 no está ajustado y se calcula sin corregir el sesgo en la varianza de la muestra de y.

RMSE (cross-validation)

Es la contribución de todos los errores cuadráticos medios calculados a partir de la validación cruzada de k muestras (*k-fold cross-validation*). Es análogo a la incertidumbre combinada de k mediciones donde el resultado es la raíz cuadrada de la suma de los cuadrados de cada una de las contribuciones individuales y también tiene las mismas unidades que la cantidad que se estima [62,71,72]:

$$\mathrm{RMSE}_{\mathrm{CV}} = \sqrt{\frac{1}{k}\sum_{j=1}^{k}(\mathrm{RMSE}_j)^2} = \sqrt{\frac{1}{k}\sum_{j=1}^{k}\left(\frac{1}{n_j}\sum_{i=0}^{n_j-1}(y_{ij}-\hat{y}_{ij})^2\right)} \tag{2.65}$$

La $\mathrm{RMSE}_{\mathrm{CV}}$ es una medida de precisión, para comparar errores de predicción de diferentes modelos para un conjunto de datos en particular y entre conjuntos de datos de la misma cantidad.

3. Metodología

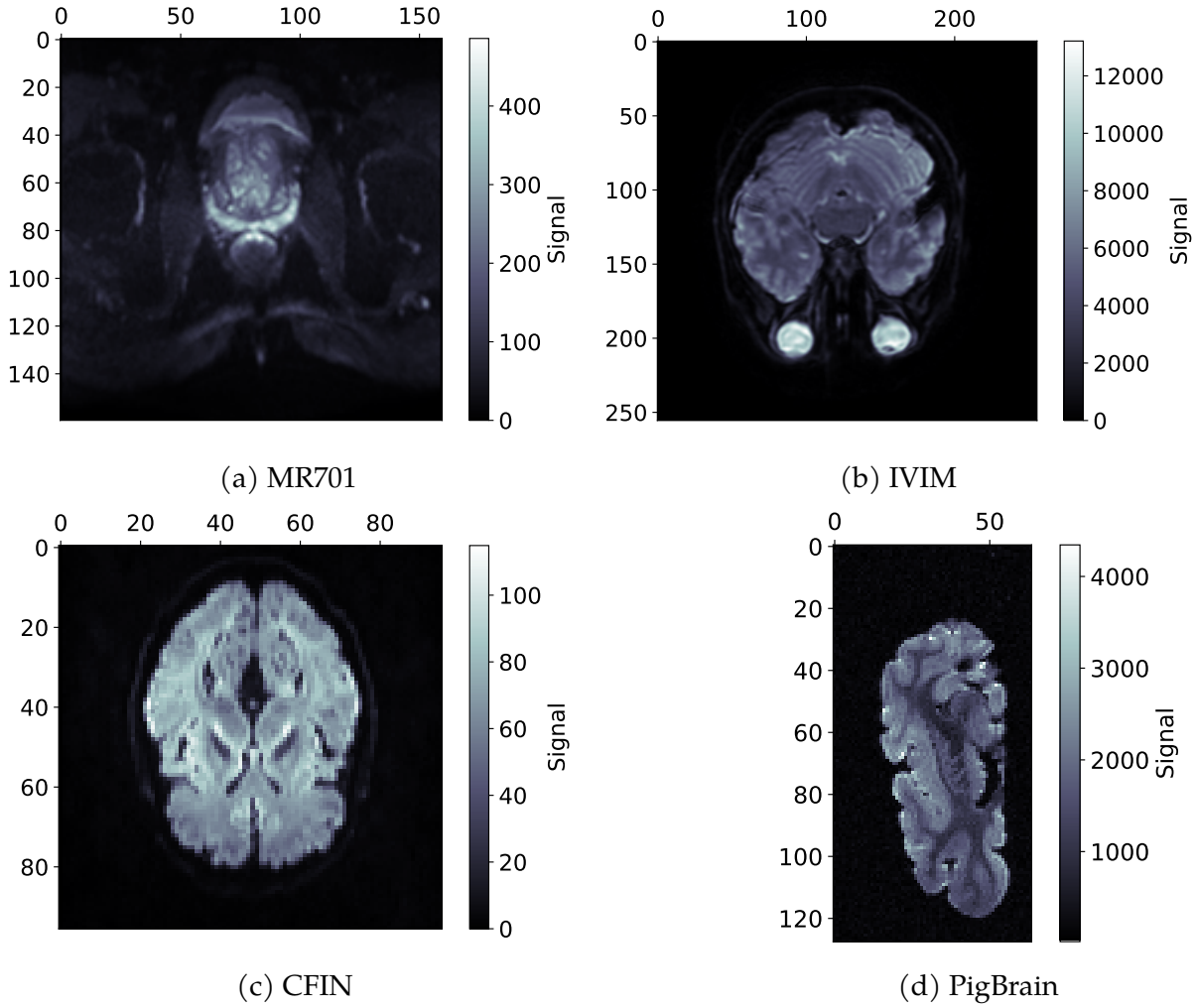
Se utilizaron 4 sets de imágenes con características especiales y diferentes entre sí [21, 22, 25]. El primer set, llamado `MR701`, es un conjunto de imágenes de la próstata de un adulto, el set `IVIM` [24] contiene imágenes de cabeza y cuello humanas, el set `CFIN` [23] posee imágenes de cráneo de humano y, finalmente, el set `PigBrain` [25] tiene imágenes *ex vivo* de una neoplasia benigna en la mitad del cerebro de un minicerdo de Göttingen. Todos los sets de imágenes son originalmente cortes axiales. En la figura **3.1** se muestran algunas imágenes representativas de cada set y en la tabla **3.1** se resumen las características más relevantes de cada uno de ellos. Los sets `IVIM`, `CFIN` y `PigBrain` son de descarga libre, disponibles en DYPI y Hvidovre Hospital, se encuentran en formato NIFTI (`.nii`) y DICOM (`.dcm`), fácilmente manejables en la versiones 3.x de Python mediante los paquetes `NiBabel` y `pydicom`.

El primer paso es obtener la curva o función que mejor caracterice a cada vóxel basándose en los valores de b y su respectiva señal en cada instante, para posteriormente implementar algoritmos de predicción y clasificación con los resultados obtenidos.

Estos conjuntos de imágenes no sufrirán modificaciones tales como reescalamiento, rotación o ninguna de las demás técnicas de *data augmentation* típicamente usados en machine learning para incrementar la cantidad de datos de entrenamiento ya que se cuenta con suficiente información para los fines requeridos. La estrategia por la que se optará es la de aplicar las técnicas de clasificación a vóxeles individuales y la única modificación que se realizará a las imágenes es la normalización de la señal a la unidad, lo que obligará a que las curvas características crucen el eje y cerca a 1.

3.1. Determinación de parámetros

La determinación de los valores de difusión, pseudo-difusión, fracción de perfusión y curtosis, así como el nivel de confianza con el que se calculan estos parámetros, son la base del desarrollo de este trabajo, ya que estos serán los datos numéricos que se usarán como referencia para la comprobación cuantitativa de la eficacia de los modelos de predicción.



(a) MR701

(b) IVIM

(c) CFIN

(d) PigBrain

Figura 3.1.: Cortes de muestra de cada uno de los 4 sets de imágenes utilizados.

Set de imágenes	MR701	IVIM	CFIN	PigBrain
Cantidad de cortes	10	54	19	4
Dimensiones por corte	x: 160 y: 160	x: 256 y: 256	x: 96 y: 96	x: 64 y: 128
Valores de b (s/mm²)	[0, 20, 40, 60, 80, 100, 120, 140, 180, 200, 300, 400, 500]	[0, 10, 20, 30, 40, 60, 80, 100, 120, 140, 180, 200, 300, 400, 500, 600, 700, 800, 900, 1000]	[200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400, 2600, 2800, 3000]	[1075, 2475, 3069, 4000, 5911, 8181, 11295]
Parámetros a determinar	D, D^*, f, r^2	D, D^*, f, r^2	D, K, r^2	D, K, r^2

Tabla 3.1.: Resumen de las principales características de los sets de imágenes utilizados para el desarrollo de este trabajo.

En la caracterización de todos los vóxeles se han implementado algoritmos de regresión lineales y cuadráticos y el cálculo de su respectivo coeficiente de determinación r^2. La regresión de segundo grado es un enfoque reservado para el cálculo de la contribución por efecto de la curtosis, que de acuerdo a lo expuesto en el capítulo 2, solo comienza a ser notable a valores de b superiores a 3000 s/mm^2 aproximadamente.

Muchas veces la relación S/S_0 se comporta de forma consistente en todo el rango de valores b, es decir, se obtienen valores de r^2 cercanos a 1 sin ningún problema. Sin embargo, en muchos vóxeles (principalmente de bordes anatómicos), pueden presentarse valores atípicos de S/S_0 que hace que r^2 se aleje considerablemente de 1. Sin embargo, no se puede simplemente descartar la existencia de difusión, pseudo-difusión, perfusión o curtosis debido al bajo nivel de confianza en la regresión de estos datos. Para estos casos, se han desarrollado funciones dentro de los algoritmos de regresión que permiten descartar como máximo uno de los puntos en caso de que la regresión mejore sin dicho dato.

También se han establecido restricciones para las curvas de ajuste, partiendo de la forma más completa de la ecuación característica del modelo IVIM con curtosis (3.1)

$$\frac{S}{S_0} = A\exp(-bD^*) + B\exp(-bD + b^2D^2K) \tag{3.1}$$

donde:

$$A + B = 1 \quad , \quad K \geq 0 \quad , \quad 0 \leq D < \infty \quad , \quad 0 \leq D^* < \infty$$

Y otras consideraciones como que en ningún momento la difusión en un tejido puede ser mayor que la difusión libre del agua ($D \leq D_{agua}$), la fracción de perfusión debe representar un porcentaje ($0 \leq f < 1$) y que la relación entre pseudo-difusión y difusión en el mismo momento y ubicación debe cumplir ($D^*/D < 1$).

Ya que la determinación de parámetros es un proceso muy tardado y computacionalmente costoso, puesto que debe ser ejecutado para todos los vóxeles y todos los cortes de cada set de imágenes, se ha implementado una condición adicional dentro del código que consiste en descartar los vóxeles en los cuales la señal es menor a una fracción del promedio de la señal de todo el corte:

$$\forall\, S_{i,j,k} \geq (\bar{S}_k/N)\, \exists\, \{D_{i,j} \cap D^*_{i,j} \cap f_{i,j} \cap K_{i,j}\} \cup r^2_{i,j} \tag{3.2}$$

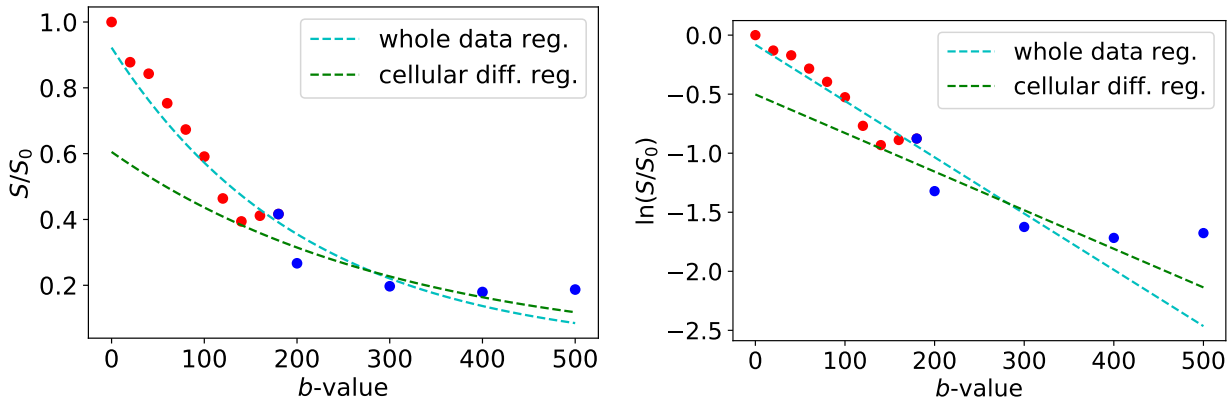
donde $S_{i,j,k}$ es la señal de un vóxel en la posición (i, j) del corte k, $\bar{S}_k$ es la señal promedio de dicho corte, N es un valor real que permite variar la exigencia de la condición y los demás son los parámetros a determinar.

3.1.1. Set MR701

El algoritmo define el límite entre difusión celular y difusión vascular alrededor de $b = 180$ s/mm^2 con base en la mejor estimación de r^2. En las figuras 3.2a y 3.2b se muestra la señal en la región de difusión vascular en rojo y en la región de difusión celular en azul. Luego se obtiene el ajuste lineal del conjunto total de datos (semi-log) (figura 3.2b) para un vóxel, si el coeficiente de determinación de dicho ajuste es lo suficientemente bueno (*i.e.* $r^2 \geq 0{,}95$), basta con el modelo mono-exponencial para caracterizar a este vóxel.

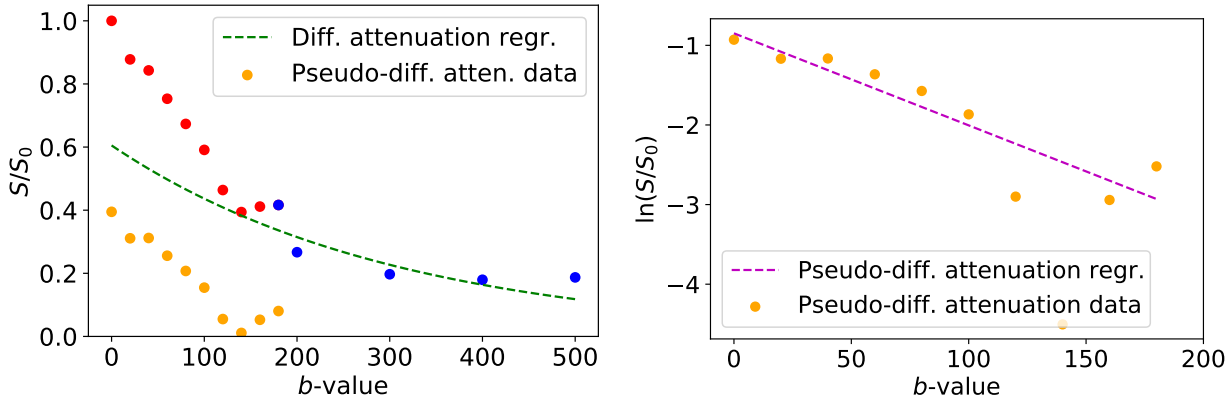
Sin embargo, para este conjunto de datos, esto no se cumple la condición. Para este set sería errado incrementar el orden del ajuste (como se mencionó anteriormente, es un enfoque reservado para el cálculo de la curtosis), así que, si el valor de r^2 para la regresión lineal es menor a 0.95, el modelo mono-exponencial no describe correctamente el comportamiento de la difusión en el vóxel, por lo cual, ahora se debe determinar si existe efecto IVIM y si el modelo bi-exponencial caracteriza mejor la curva de difusión.

Para esto, se calcula el ajuste lineal de los datos (semi-log) correspondientes solamente a la región de difusión celular ($b \geq 180$ s/mm^2) (Figura 3.2b), la recta obtenida se extrapola hasta $b = 0$ s/mm^2 y esta se resta de cada valor de señal en la región de difusión vascular (Figura 3.3a), obteniendo de forma más clara la atenuación por pseudo-difusión, la cual se ajusta a una nueva recta como se muestra en la figura 3.3b.
Finalmente se suman las curvas de difusión y pseudo-difusión obtenidas, para conseguir una descripción bi-exponencial de los datos (Figura 3.4). Para este ajuste bi-exponencial



(a) lin-lin (b) semi-log

Figura 3.2.: Datos de la señal normalizada en función del valor de b para el vóxel en la posición $(63, 61)$ del corte 5. Se muestra el ajuste lineal del conjunto de datos en todo el intervalo de b (linea cian discontinua con $r^2 = 0{,}93$) y ajuste lineal de los datos en la región de difusión celular (linea verde discontinua con $r^2 = 0{,}76$). De acuerdo con el algoritmo, en ambos casos se descarta el peor dato para obtener un mejor ajuste. Set `MR071`.



(a) lin-lin (b) semi-log

Figura 3.3.: Ajuste lineal de los datos en la región de difusión celular (linea verde punteada con $r^2 = 0{,}76$). A partir de esta regresión se obtienen los datos de pseudo-difusión (naranja), los cuales se ajustan a una nueva función lineal (linea rosa punteada con $r^2 = 0{,}84$). Set `MR071`.

se calcula el coeficiente de determinación r^2 mediante la ecuación (3.3):

$$r^2 = 1 - \frac{S_{res}}{S_{tot}} \tag{3.3}$$

donde S_{res} es la suma de los cuadrados de los residuos

$$S_{res} = \sum_i (y_i - f_i)^2 \tag{3.4}$$

y S_{tot} es la suma de los cuadrados de los datos respecto a su media

$$S_{tot} = \sum_i (y_i - \bar{y})^2 \tag{3.5}$$

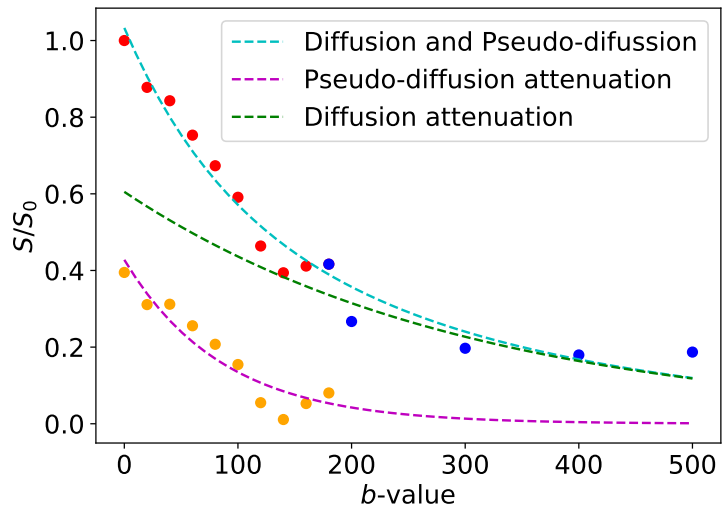


Figura 3.4.: Ajuste final al modelo bi-exponencial (linea cian punteada con $r^2 = 0{,}97$) construida a partir de las contribuciones de atenuación por difusión y pseudo-difusión. Set `MR071`.

Si el valor de r^2 calculado para el modelo bi-exponencial es menor que el calculado para modelo mono-exponencial, se preserva el ajuste de este último.

Los valores obtenidos para esté vóxel de este corte en particular resultan ser:

$$D = 0{,}00326842(34)\ \mathrm{mm}^2/\mathrm{s} \tag{3.6}$$

$$D^* = 0{,}011562(81)\ \mathrm{mm}^2/\mathrm{s} \tag{3.7}$$

$$f = 0{,}395 \tag{3.8}$$

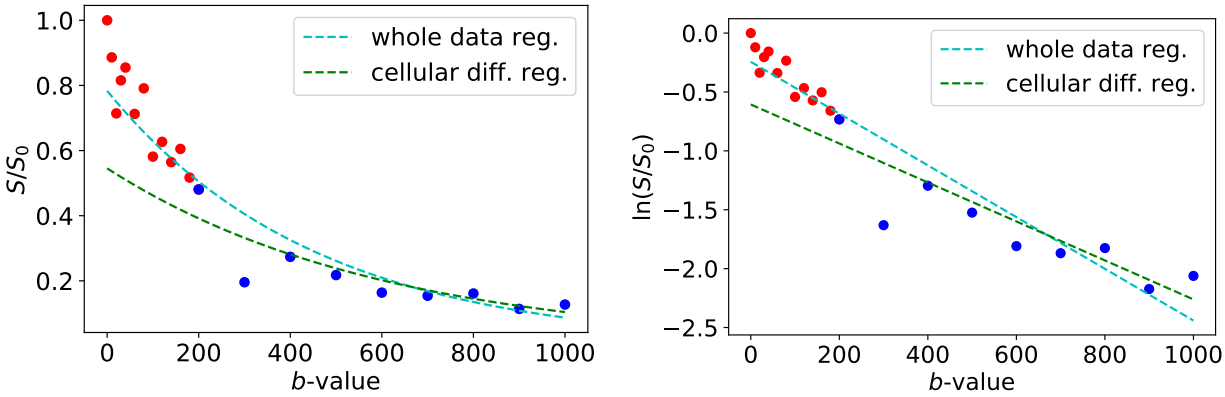
$$K = 0{,}0 \tag{3.9}$$

$$r^2 = 0{,}968 \tag{3.10}$$

El mismo procedimiento se repite para todos los vóxeles en todos los cortes del set de imágenes `MR701`.

3.1.2. Set IVIM

Al igual que con el set `MR701`, el algoritmo define el límite entre difusión celular y difusión vascular alrededor de $b = 200$ s/mm^2 (Figuras 3.5a y 3.5b) con base en el mejor valor de r^2 para el ajuste de todos los datos. Posteriormente, se obtiene el ajuste lineal del conjunto total de datos (semi-log) (Figura 3.5b), si el coeficiente de determinación de dicho ajuste es lo suficientemente bueno (*i.e.* $r^2 \geq 0{,}95$), basta con el modelo mono-exponencial para caracterizar a este vóxel. Aunque este conjunto de datos posee información de señal para valores de b hasta 1000 s/mm^2 seguiría siendo incorrecto incrementar el orden del ajuste ya que no se espera que exista efecto significativo de la curtosis.



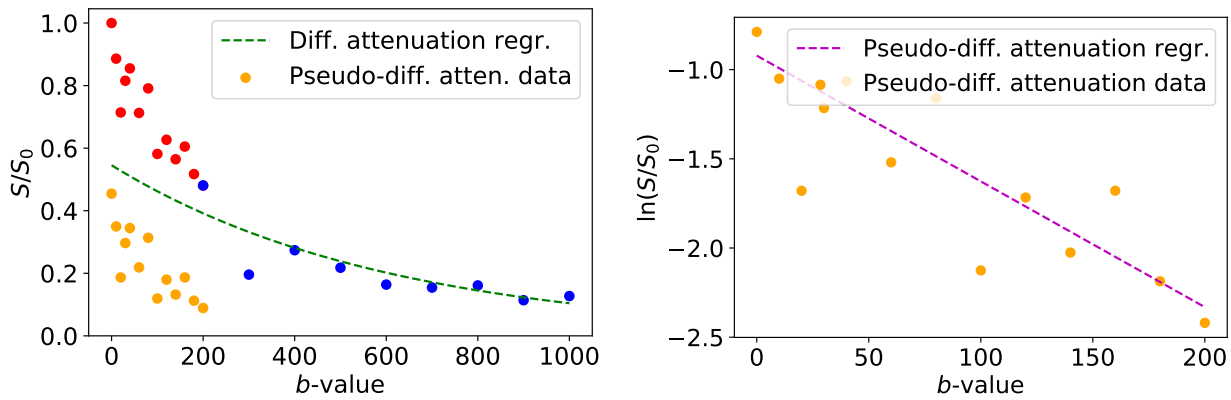
(a) lin-lin (b) semi-log

Figura 3.5.: Datos de la señal normalizada en función del valor de b para el vóxel en la posición (150, 100) del corte 17. Se muestra el ajuste lineal del conjunto de datos en todo el intervalo de b (linea cian punteada con $r^2 = 0{,}91$) y ajuste lineal de los datos en la región de difusión celular (linea verde punteada con $r^2 = 0{,}90$). De acuerdo con el algoritmo, en ambos casos se descarta el peor dato para obtener un mejor ajuste. Set `IVIM`.

Este set se caracteriza por poseer una gran cantidad de datos para valores de b entre 0 y 200 s/mm^2, entonces, ahora se debe determinar el efecto del movimiento incoherente intra-vóxel (IVIM) mediante la caracterización de los datos usando el modelo bi-exponencial.

Para esto, se calcula el ajuste lineal de los datos (semi-log) correspondientes solamente a

la región de difusión celular ($b \geq 200$ s/mm^2), la recta obtenida se extrapola hasta $b = 0$ s/mm^2 y esta se resta de cada valor de señal en la región de difusión vascular como se muestra en la figura 3.6a, y obteniendo de forma clara la atenuación por pseudo-difusión, la cual se ajusta a una nueva recta (figura 3.6b).



(a) lin-lin (b) semi-log

Figura 3.6.: Ajuste lineal de los datos en la región de difusión celular (linea verde punteada con $r^2 = 0{,}90$). A partir de esta regresión se obtienen los datos de pseudo-difusión (naranja), los cuales se ajustan a una nueva función lineal (linea rosa punteada con $r^2 = 0{,}91$). Set `IVIM`.

Finalmente se suman las curvas de difusión y pseudo-difusión obtenidas para conseguir una descripción bi-exponencial de los datos y su correspondiente valor de r^2 mediante la ecuación (3.3). Al igual que con el set `MR701`, si el valor de r^2 calculado para el modelo bi-exponencial es menor que el calculado para modelo mono-exponencial, se preserva el ajuste de este último.

Los valores obtenidos para esté vóxel de este corte en particular resultan ser:

$$D = 0{,}00180033(10)\ \text{mm}^2/\text{s} \tag{3.11}$$
$$D^* = 0{,}007870(56)\ \text{mm}^2/\text{s} \tag{3.12}$$
$$f = 0{,}390 \tag{3.13}$$
$$K = 0{,}0 \tag{3.14}$$
$$r^2 = 0{,}946 \tag{3.15}$$

Dada la cantidad de cortes del set de imágenes `IVIM`, el mismo procedimiento se repite para todos los vóxeles de los cortes múltiplos de 5.

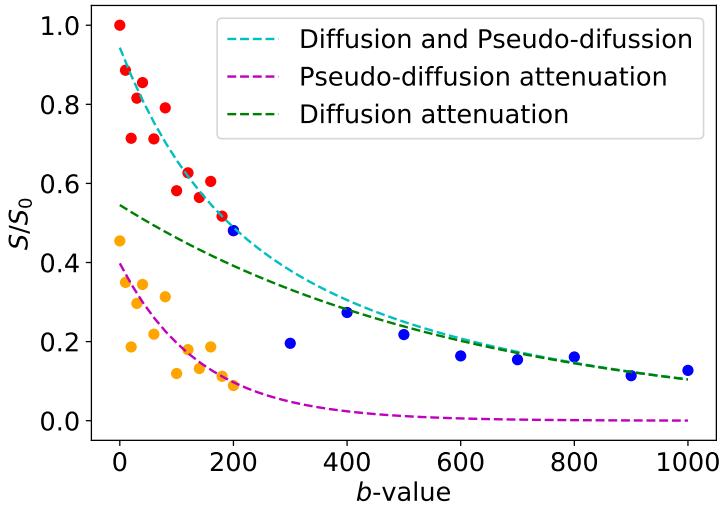
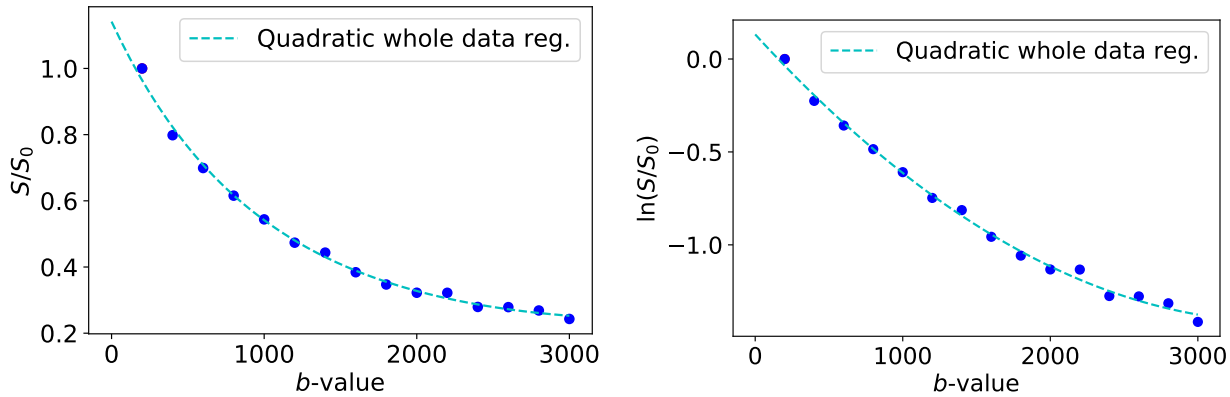


Figura 3.7.: Ajuste final al modelo bi-exponencial (linea cian punteada con $r^2 = 0{,}95$) construida a partir de las contribuciones de atenuación por difusión y pseudo-difusión. Set `IVIM`.

3.1.3. Set CFIN

Este conjunto de imágenes, varía de los dos anteriores en que no posee datos por contribución de la pseudo-difusión, y el efecto por movimiento incoherente intra-vóxel no es detectable, por lo tanto, en este set, no se define límite entre difusión celular y difusión vascular, además sería suficiente el modelo mono-exponencial para caracterizar las curvas de difusión de estas imágenes. Sin embargo, dado que existen datos para valores de b hasta 3000 s/mm^2, el efecto de la curtosis podría llegar a ser detectado.

Primero se obtiene el ajuste cuadrático del conjunto total de datos (semi-log) (Figura 3.8b), en este caso, si $r^2 \geq 0{,}9$ se acepta la regresión y el modelo mono-exponencial (eventualmente con curtosis) para caracterizar a este vóxel. Por otro lado, si $r^2 < 0{,}9$, se rechaza la regresión y se asume que no existe difusión en el vóxel, en caso de existir, se verificará durante la preparación de las imágenes de la siguiente sección.



(a) lin-lin (b) semi-log

Figura 3.8.: Datos de la señal normalizada en función del valor de b para el vóxel en la posición (50, 45) del corte 8. Ajuste cuadrático del conjunto de datos en todo el intervalo de b (linea cian punteada con $r^2 = 0{,}91$). De acuerdo con el algoritmo, se descarta el peor dato para obtener un mejor ajuste. Set `CFIN`.

Los valores obtenidos para esté vóxel de este corte en particular resultan ser:

$$D = 0{,}00087239(66)\ \mathrm{mm}^2/\mathrm{s} \tag{3.16}$$

$$D^* = 0{,}0\ \mathrm{mm}^2/\mathrm{s} \tag{3.17}$$

$$f = 0{,}0 \tag{3.18}$$

$$K = 0{,}753(45) \tag{3.19}$$

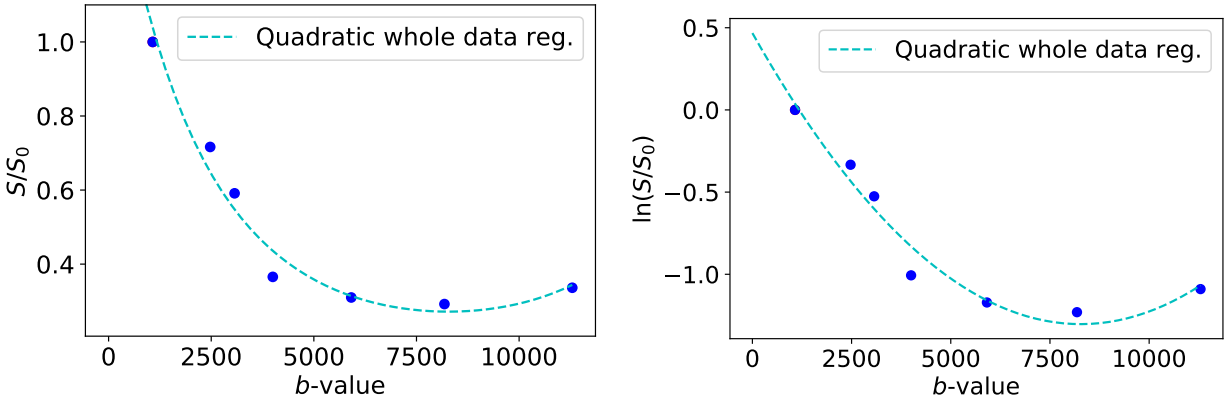
$$r^2 = 0{,}995 \tag{3.20}$$

Dada la cantidad de cortes del set de imágenes `CFIN`, el mismo procedimiento se repite para todos los vóxeles de los cortes pares.

3.1.4. Set PigBrain

Este conjunto de imágenes es similar al set `CFIN` en cuanto a que no posee datos para estimar la contribución de la pseudo-difusión, no se define límite entre difusión celular y difusión vascular y es suficiente con el modelo mono-exponencial para caracterizar las curvas de difusión de este set. Pero a diferencia del anterior conjunto, existen datos para valores de b bastante grandes (hasta aproximadamente 11300 s/mm^2), por lo cuál, el efecto de la curtosis es mucho más notorio que en el set `CFIN`.

Se obtiene el ajuste cuadrático del conjunto total de datos (semi-log) (Figura 3.9b), en este caso, si $r^2 \geq 0{,}9$ se acepta la regresión y el modelo mono-exponencial con curtosis para ca-



(a) lin-lin (b) semi-log

Figura 3.9.: Datos de la señal normalizada en función del valor de b para el vóxel en la posición $(30, 44)$ del corte 2. Ajuste cuadrático del conjunto de datos en todo el intervalo de b (linea cian punteada con $r^2 = 0{,}96$). En este caso, no se descarta ningún dato para obtener un mejor ajuste ya que se cuenta con pocos datos. Set `PigBrain`.

racterizar a este vóxel. Por otro lado, si $r^2 < 0{,}9$, se rechaza la regresión y se asume que no existe difusión en el vóxel. Nuevamente, en caso de existir difusión, se verificará durante la preparación de las imágenes de la siguiente sección.

Los valores obtenidos para esté vóxel de este corte en particular resultan ser:

$$D = 0{,}00042791(66)\ \mathrm{mm}^2/\mathrm{s} \tag{3.21}$$

$$D^* = 0{,}0\ \mathrm{mm}^2/\mathrm{s} \tag{3.22}$$

$$f = 0{,}0 \tag{3.23}$$

$$K = 0{,}847(50) \tag{3.24}$$

$$r^2 = 0{,}961 \tag{3.25}$$

El mismo procedimiento se repite para todos los vóxeles en todos los cortes del set de imágenes `PigBrain`.

3.2. Preparación de las imágenes

Para poder utilizar los resultados obtenidos mediante la metodología planteada en la sección 3.1, es necesario preparar y organizar apropiadamente la información obtenida a partir de las imágenes DWI y datos crudos de todos los sets previamente presentados. Para preparar las imágenes se la utilizado la misma estrategia y procedimiento en todos los sets de imágenes, no obstante, en concordancia con lo expuesto en la tabla **3.1**, en cada set se calculan solo algunos parámetros. En todos los casos se ha calculado el valor de difusión D o de la pseudo-difusión D^* (o ambos) junto con su correspondiente coeficiente de determinación r^2 y solo en algunos casos se calcula fracción de perfusión f o curtosis K.

Al repetir el procedimiento para todos los vóxeles de cada corte, se obtienen valores numéricos de estos parámetros, que pueden ser ubicados en mapas de difusión, pseudo-difusión, perfusión, curtosis y r^2 respectivamente, como se muestra en las siguientes figuras (**3.10**, **3.11**, **3.12** y **3.13**)

3.2.1. Mapas de r^2 y r^2 modificado

Los mapas de r^2 representan la precisión los ajustes en la ubicación de cada vóxel. Para definir los datos con los que se realizará el entrenamiento de los algoritmos de clasificación, se aplica a estos un filtro que consiste en un algoritmo que lleva a cero a todos los vóxeles con valores de r^2 menores a 0.9 ya que muestran ajustes poco confiables, mientras que los valores iguales o mayores a 0.9 se llevan a 1, obteniendo un nuevo mapa de r^2 con la ubicación de los vóxeles con ajustes mejores que el 90 %. El mapa de r^2 de la figura **3.13** adaptado con esta consideración se muestra en la figura **3.15**a.

Al observar la imagen modificada de r^2 puede ser intuitivo pensar que un vóxel con el valor 0 rodeado de vóxeles en 1 debería corresponder a un valor de 1, o que un vóxel en 1 rodeado de vóxeles en 0 debería corresponder a un valor de 0, es decir, si existe difusión en una región anatómica muy pequeña y esta se encuentra inmersa en una región con ausencia de difusión, no debería existir difusión en este espacio y viceversa.

Para esto, la imagen modificada de r^2 se pasa por una serie de filtros basados en la siguiente formulación. Para una imagen 2D, dado un vóxel R ubicado en la posición (i, j) en el mapa r^2 modificado, su nuevo valor después de aplicar el filtro de suavizado dependerá de los valores de los vóxeles vecinos más próximos y su distancia relativa a ellos; los vóxeles arriba, abajo, a la izquierda y a la derecha son los más cercanos y su distancia relativa será 1, mientras que los vóxeles de las diagonales tendrán una distancia relativa de $\sqrt{2}$. Al final, si la suma ponderada de todos los vóxeles vecinos supera la unidad, significa que

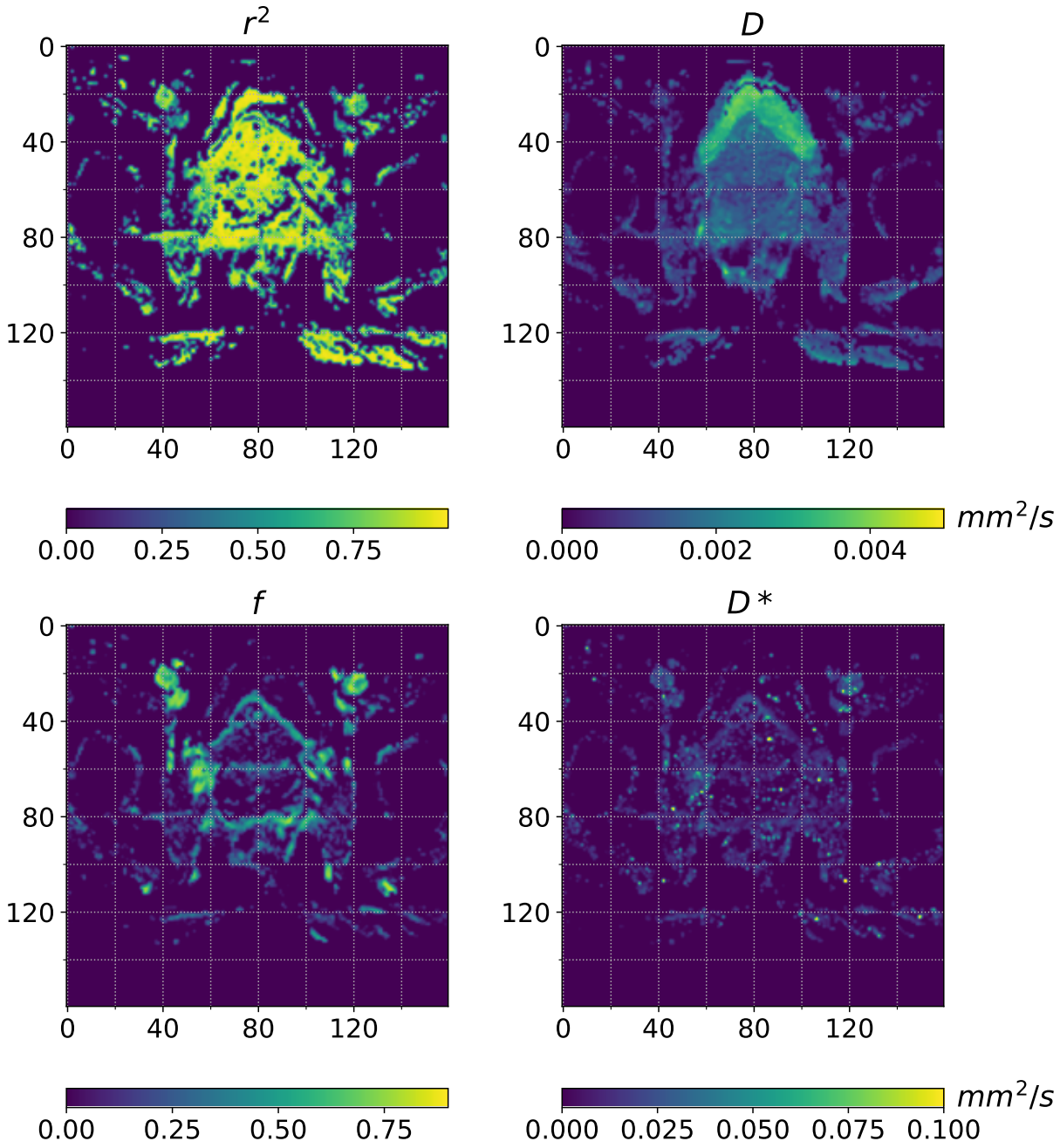


Figura 3.10.: Visualización de mapas de r^2 (*izquierda-arriba*), difusión D (*derecha-arriba*), fracción de perfusión f (*izquierda-abajo*) y pseudo-difusión D^* (*derecha abajo*) del corte 8 del set `MR071`.

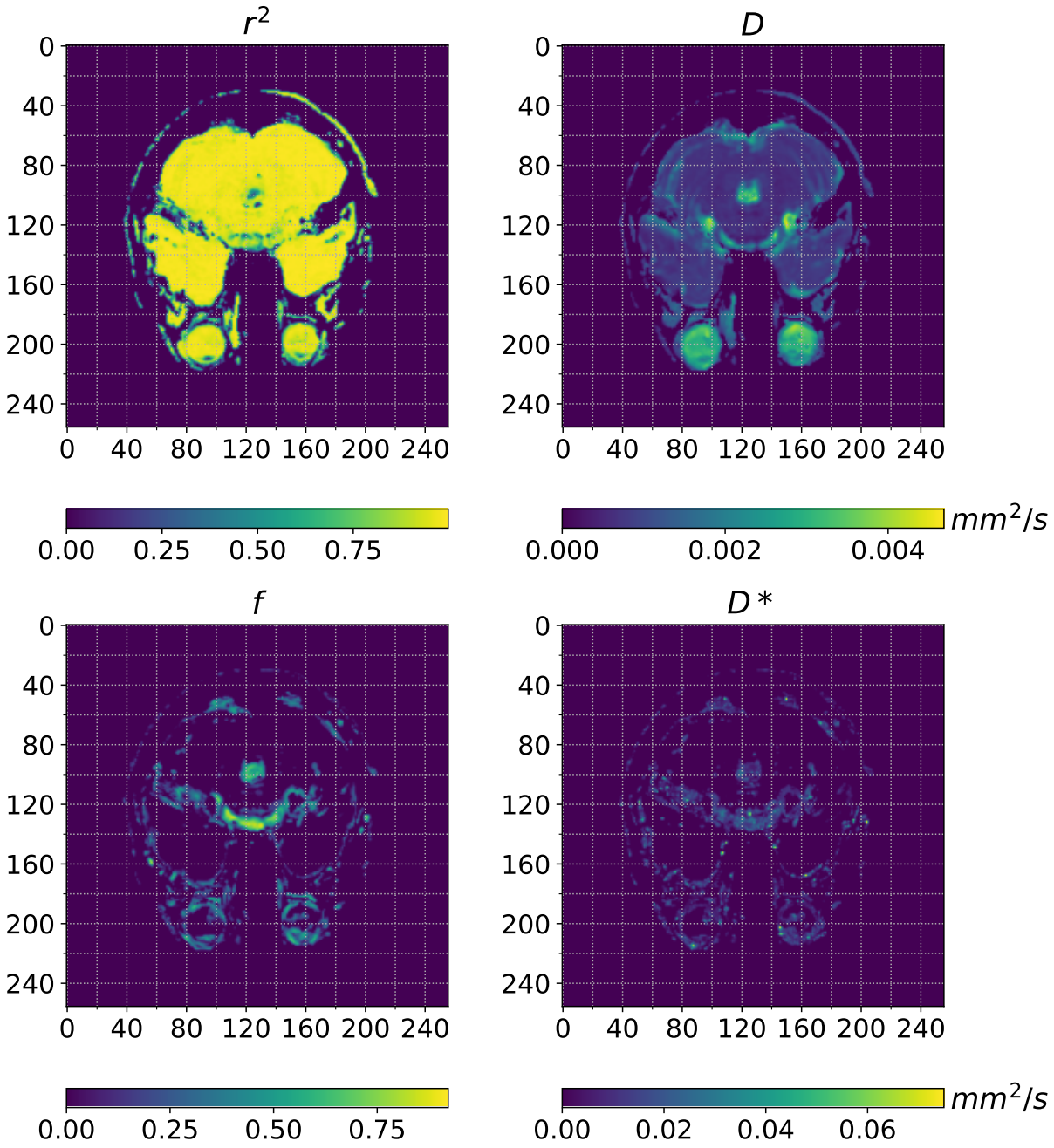


Figura 3.11.: Visualización de mapas de r^2 (*izquierda-arriba*), difusión D (*derecha-arriba*), fracción de perfusión f (*izquierda-abajo*) y pseudo-difusión D^* (*derecha abajo*) del corte 15 del set `IVIM`.

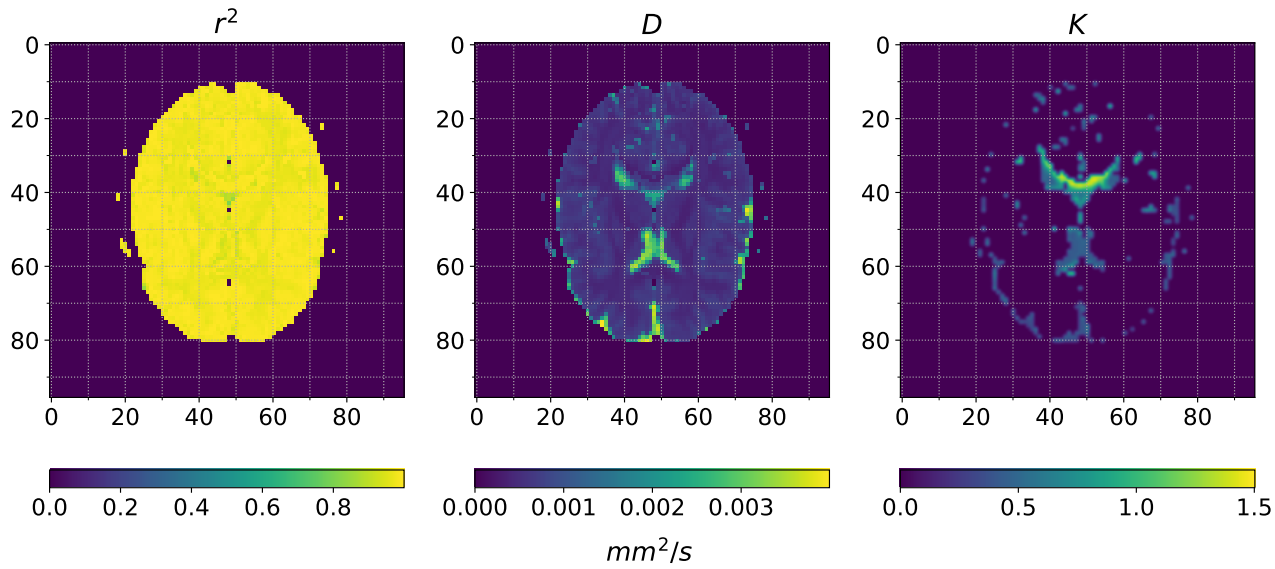


Figura 3.12.: Visualización de mapas de r^2 (*izquierda*), difusión D (*centro*) y curtosis K (*derecha*) del corte 11 del set `CFIN`.

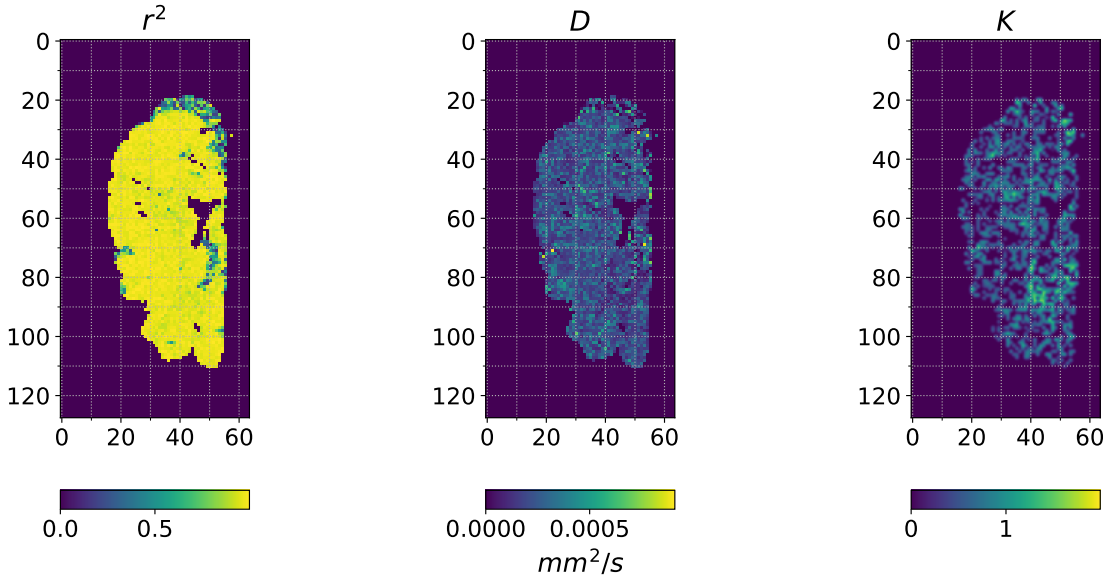


Figura 3.13.: Visualización de mapas de r^2 (*izquierda*), difusión D (*centro*) y curtosis K (*derecha*) del corte 2 del set `PigBrain`.

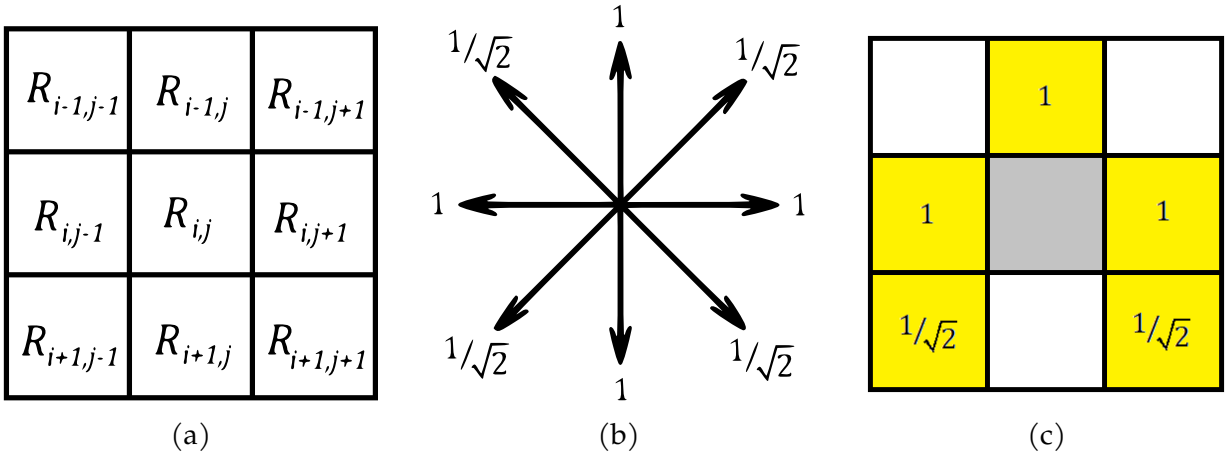
la mayoría de vóxeles se encuentran inicialmente marcados con 1 y por lo tanto, al vóxel $R_{i,j}$ se le asignará también 1. Por otro lado, si la suma ponderada no supera la unidad, significa que la mayoría de los vóxeles alrededor de este contienen inicialmente un cero, y por tanto se le asignará 0 a este vóxel. Esto se aprecia claramente en las ecuaciones (3.26) y (3.27).

$$W_A = \frac{1}{4} \cdot [R_{i-1,j} + R_{i,j-1} + R_{i+1,j} + R_{i,j+1}] \tag{3.26}$$

$$W_B = \frac{1}{4\sqrt{2}} \cdot [R_{i-1,j-1} + R_{i+1,j-1} + R_{i-1,j+1} + R_{i+1,j+1}] \tag{3.27}$$

donde W_A representa la suma ponderada de los vecinos más cercanos y W_B la suma ponderada de los vecinos de las diagonales.

$$\text{Si } W_A + W_B \geq 1 \quad \Rightarrow \quad R_{i,j} = 1$$



(a) (b) (c)

Figura 3.14.: Ejemplo de la aplicación del filtro de suavizado a un vóxel particular (i, j).

La figura **3.14** muestra un ejemplo de la implementación del filtro de suavizado. En la matriz de la figura **3.14a** se observa la nomenclatura de las ubicaciones de los vóxeles alrededor del vóxel de interés $R_{i,j}$. Los pesos asignados de acuerdo a la distancia relativa se muestran en la figura **3.14b**. Si se tuviera una distribución de vóxeles como los de la figura **3.14c**, en la cual se desea conocer el nuevo valor del vóxel central (en gris) en función de sus vecinos (los vóxeles amarillos poseen 1 y los blancos 0), y aplicando las ecuaciones (3.26) y (3.27) se obtendría que:

$$W_A = \frac{1}{4} \cdot [1 + 1 + 0 + 1] = \frac{3}{4} \tag{3.28}$$

$$W_B = \frac{1}{4\sqrt{2}} \cdot [0 + 1 + 0 + 1] = \frac{1}{2\sqrt{2}} \tag{3.29}$$

Así, $W_A + W_B \approx 1{,}1 \geq 1$, entonces el nuevo valor de $R_{i,j}$ será 1.

Para el filtro 3D se sigue el mismo fundamento, en el cual se ponderan los valores de los vecinos de acuerdo a su cercanía con el vóxel de interés. En estos casos se tendrá un factor hasta de $\sqrt{3}$ por cercanía y se debe tener en cuenta hasta 26 vóxeles vecinos. Las ecuaciones (3.30), (3.31) y (3.32) describen de forma análoga al caso 2D, la suma ponderada de todos los vóxeles alrededor de un vóxel en la posición (i, j).

$$W_A = \frac{1}{6} \cdot [R_{i-1,j,k} + R_{i,j-1,k} + R_{i+1,j,k} + R_{i,j+1,k} + R_{i,j,k-1} + R_{i,j,k+1}] \tag{3.30}$$

$$\begin{aligned} W_B = \frac{1}{12\sqrt{2}} \cdot [&R_{i-1,j,k-1} + R_{i+1,j,k-1} + R_{i-1,j,k+1} + R_{i+1,j,k+1} \\ &+R_{i-1,j-1,k} + R_{i-1,j+1,k} + R_{i+1,j-1,k} + R_{i+1,j+1,k} \\ &+R_{i,j-1,k-1} + R_{i,j+1,k-1} + R_{i,j-1,k+1} + R_{i,j+1,k+1}] \end{aligned} \tag{3.31}$$

$$\begin{aligned} W_C = \frac{1}{8\sqrt{3}} \cdot [&R_{i-1,j-1,k+1} + R_{i-1,j+1,k+1} + R_{i+1,j-1,k+1} + R_{i+1,j+1,k+1} \\ &+R_{i-1,j-1,k-1} + R_{i-1,j+1,k-1} + R_{i+1,j-1,k-1}] \end{aligned} \tag{3.32}$$

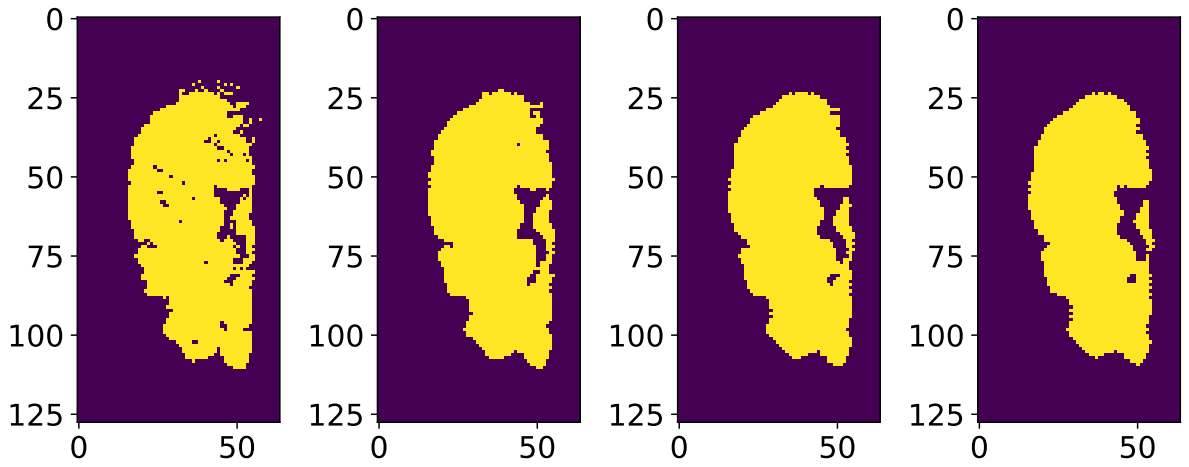
donde W_A, W_B y W_C representan las sumas para los vecinos más cercanos de acuerdo a su distancia relativa.

$$\text{Si } W_A + W_B + W_C \geq 1 \quad \Rightarrow \quad R_{i,j} = 1$$

La figura **3.15** muestra el proceso de filtrado de una imagen de r^2 modificada que ha pasado 3 veces por el filtro 2D desarrollado. La imagen se pasa por el filtro las veces que se considere necesario para reducir el efecto de aislamiento de vóxeles individuales, pero, ya que el filtro tiende a eliminar vóxeles conforme se repite su aplicación, se debe tener cuidado de no modificar demasiado la imagen r^2 original. Para esto se define un criterio que indica en donde no se debe continuar con el filtrado de imagen, este es:

$$\frac{A_n}{A_0} < 0{,}9 \tag{3.33}$$

donde A_0 es el área cubierta por los vóxeles con el valor de 1 (amarillo) del mapa de r^2 modificado sin pasar por el filtro y A_n es el área cubierta la cantidad por los vóxeles con valores de 1 del mapa de r^2 modificado que se ha pasado n veces por el filtro de suavizado. Esto significa que si la reducción del área amarilla de la imagen original es de al menos el 10 %, el proceso de filtrado se detiene.



(a) Sin filtro. (b) Con filtro (1 vez). (c) Con filtro (2 veces). (d) Con filtro (3 veces).

Figura 3.15.: Mapa de r^2 modificado sin filtro de suavizado y luego de aplicar el filtro de suavizado hasta 3 veces.

El proceso de modificación de los mapas de r^2 ayuda a incluir o descartar vóxeles en los cuales no se consiguió una buena regresión lineal o cuadrática y permitir a los algoritmos de predicción generalizar a partir de conjuntos de datos mas diversos y menos homogéneos.

Se debe tener en cuenta que el tratamiento de cada set de imágenes desde el inicio de la metodología hasta este punto puede llegar a tardar varios días, dependiendo de la velocidad de procesamiento del equipo utilizado, cantidad de núcleos, uso de GPU, entre muchos otros factores, por lo que se recomienda guardar una copia de cada mapa una vez obtenido para reducir el tiempo de ejecución y avanzar en el progreso para posteriores pasos.

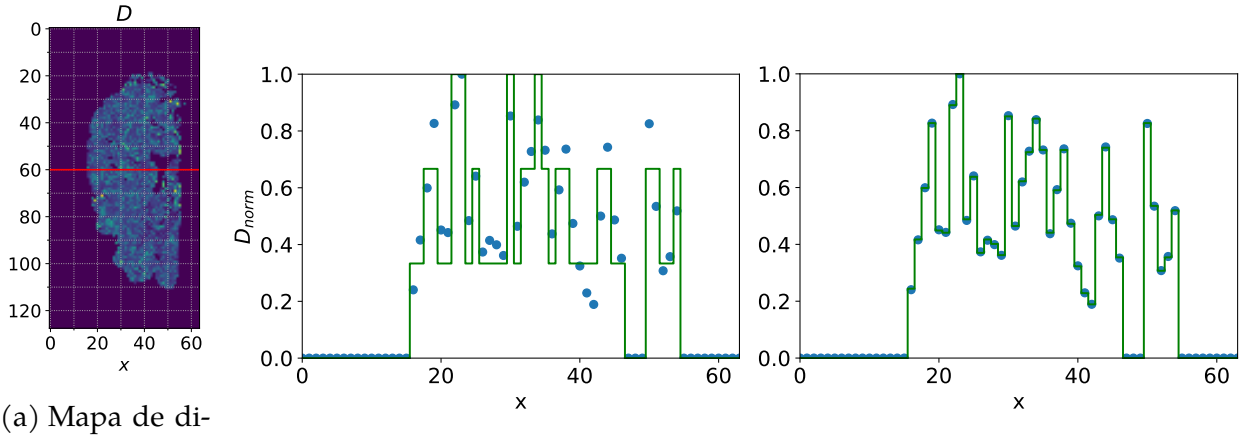
3.2.2. Discretización de datos

Los demás mapas (D, D^*, f, K) deben ser modificados de diferente forma ya que contienen valores continuos de difusión, pseudo-difusión, fracción de perfusión y curtosis. Estos mapas son sometidos a una discretización lineal, es decir que sus intensidades se restringen a valores concretos a lo largo de su dominio, similar al proceso de digitalización de una imagen analógica. Para esto, se establece un mínimo y un máximo (0 y 1 cuando los datos están normalizados), este intervalo se divide en 2^n secciones, donde n es el análogo al número de bits en una digitalización común. En estas imágenes, mientras mayor sea n mayor será la resolución de contraste de la imagen resultante, y así mismo, la cantidad de

recursos al implementar el tratamiento de dicha imagen.

Por ejemplo, para un mínimo de 0, un máximo de 1 y $n = 1$, todos los valores menores a 0.5 se convierten en 0 y todos los valores mayores a 0.5 se convierten en 1. Si $n = 2$, los valores menores a 0.125 se convierten en 0, valores entre 0.125 y 0.375 se convierten en 0.25, y así sucesivamente. Un ejemplo de discretización a 2 bits de una linea horizontal en un corte de difusión se muestra en la figura 3.16b.

En el caso de estos sets de imágenes, se define $n = 7$, lo cual permite un total de 128 valores como resultado como se muestra en la figura 3.16c como ejemplo de discretización de la misma línea de vóxeles del ejemplo para 2 bits. Esta distretización se repite para todos los mapas obtenidos de D, f, D^* y K. La discretización de los valores a 7 bits, cuando la señal está normalizada, permite 128 valores intermedios, cada uno con una amplitud de aproximadamente 1/128, es decir, que la máxima desviación sería de 0.0078 unidades.



(a) Mapa de difusión D.

(b) Discretización a 2 bits.

(c) Discretización a 7 bits.

Figura 3.16.: Ejemplo de discretización de los datos aplicado a la fila $x = 60$ (linea horizontal roja) del conjunto de datos del mapa de difusión normalizado (D_{norm}) del corte 2 usando 2 y 7 bits respectivamente. Set `PigBrain`.

El máximo valor encontrado para la difusión D es $4{,}88 \times 10^{-3}$ mm^2/s, para la pseudo-difusión D^* es $9{,}89 \times 10^{-2}$ mm^2/s, para la fracción de perfusión f es $0{,}91$ y para la curtosis K es $1{,}95$. Estos valores servirán para normalizar los datos mediante la ecuación (2.50).

A diferencia del ejemplo mostrado en la figura 3.16, los datos que originalmente eran iguales a cero debido a ausencia de señal, se marcan con -1, mientras que en otros vóxeles donde puede existir un valor de dichos parámetros pequeño pero no desconocido, se etiquetan con cero.

3.3. Implementación de los algoritmos de clasificación

Una vez preparadas las imágenes se genera un vector de X_b de 37 posiciones correspondientes a todos los valores conocidos de b organizados como sigue:

$$X_b = [0, 10, 20, 30, 40, 60, 80, 100, 120, 140, 180, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1075, 1200, 1400, 1600, 1800, 2000, 2200, 2400, 2475, 2600, 2800, 3000, 3069, 4000, 5911, 8181, 11295]$$

Posteriormente se genera un nuevo vector $X_{i,j}$ de 37 posiciones, correspondientes a los valores de S/S_0 de cada vóxel en una posición (i, j) de cada corte.

Puesto que no todos los sets de imágenes poseen todos los valores de b del vector X_b, los valores desconocidos se marcan con el valor -1 para posteriormente ser tratados como datos faltantes. Por ejemplo para el vóxel (30, 44) del corte 2 del set `PigBrain` (mismo de la figura 3.9), se obtendría el siguiente vector $X_{i,j}$:

$$X_{30,44} = [-1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, 1.0, -1, -1, -1, -1, -1, -1, -1, 0.72, -1, -1, -1, 0.59, 0.37, 0.31, 0.29, 0.34]$$

Para tener en cuenta el efecto de los vóxeles alrededor del que está ubicado en la posición (i, j), se puede incluir el equivalente del vector $X_{i,j}$ de dichos vóxeles. Finalmente, la unión de todos los vectores obtenidos, en forma ordenada, será la entrada X para los algoritmos.

$$X = [X_{i-1,j-1},\ X_{i-1,j},\ X_{i-1,j+1},\ X_{i,j-1},\ X_{i,j},\ X_{i,j+1},\ X_{i+1,j-1},\ X_{i+1,j},\ X_{i+1,j+1}] \tag{3.34}$$

El vector X tiene un tamaño de 37×9 valores, sin embargo, esta formulación se puede complementar incluyendo los valores del vector $X_{i,j,k}$, donde $k = -1, 0, 1$ es el corte anterior, actual y posterior relativo a cada vóxel, de este modo el vector de entrada X resultante tendría un tamaño de $37 \times 9 \times 3$ valores. Para el entrenamiento inicial de los algoritmos, se trabajará únicamente con los datos del corte particular, es decir $k = 0$, por lo cual el tamaño de X será de tan solo 333 posiciones (o 33 si solamente se toma el vóxel de interés).

La salida Y por su parte, se compone de los datos de r^2 modificado luego de la aplicación de los filtros de suavizado, y de los valores "discretizados" de D, D^* y K. Como los algoritmos utilizados clasifican los datos en categorías de acuerdo a su valor, se evaluará cada salida independientemente de las demás. Es decir, la salida correspondiente a r^2 tendrá 2 posibles valores: 0 o 1, mientras que la salida correspondiente a D tendrá 2^7 posibles valores, al igual que las salidas de D^*, f y K.

$$Y_{i,j} = [r^2_{i,j},\ D_{i,j},\ D^*_{i,j},\ f_{i,j},\ K_{i,j}] \tag{3.35}$$

La salida de Y para el vóxel (30, 44), corte 2 del set `PigBrain` sería

$$Y_{30,44} = [1,\ 0.00042366,\ 0,\ 0,\ 0.84297] \tag{3.36}$$

Del mismo modo que con los datos faltantes de entrada, a las salidas desconocidas se marcan como -1; para el ejemplo anterior:

$$Y_{30,44} = [1,\ 0.00042366,\ -1,\ -1,\ 0.84297] \tag{3.37}$$

Finalmente se generan 5 archivos `.csv` con todos los vectores X y r^2, D, D^*, f y K de todos vóxeles, de todas las imágenes y de todos los sets. Esto, con el fin de definir los datos de entrenamiento (*train*) y prueba (*test*) para los algoritmos usados. Cada archivo tendrá entonces un total de 1'101.824 vectores X como la población de datos para entrenar y probar los algoritmos pero se trabajará con varias muestras aleatorias de $\sim 10\,\%$ del tamaño de la población de datos.

3.3.1. Transformación de los datos

La mayoría de los datos obtenidos son los que corresponden a cero difusión, y son obtenidos de lugares por fuera del espacio anatómico, por lo tanto, se debe considerar que el conjunto global de datos se encuentra desbalanceado o sesgado. Esto primero se debe comprobar, y en caso de ser cierto, se debe deben estandarizar los datos para que estos se distribuyan de de una forma más o menos normal o gaussiana.

La estrategia para verificar la distribución es mediante la utilización del método estadístico de *QQ plots* o gráficos quantil-quantil que permiten evaluar rápidamente el sesgo de una distribución.

Para la **3.17** se ha tomado una sección de 10000 datos aleatorios de todos los sets, en la cual se observa que la distribución de los datos de $X_{i,j}$ para $b = 180$ s/mm^2 se encuentra sesgada hacia la derecha y, además, esto se puede verificar comparando la pendiente de los datos (azul) en el gráfico Q-Q con la de una distribución normal (rojo).

Usando la librería `PowerTransformer` de *Sklearn*, es posible realizar una transformación mediante un reescalado de los datos para obtener una distribución más gaussiana o normal alrededor de cero (ec. 2.52).

En la figura **3.18** se muestra el resultado de dicha transformación mediante el histograma y el gráfico Q-Q de los mismos datos mostrados en la figura **3.17**.

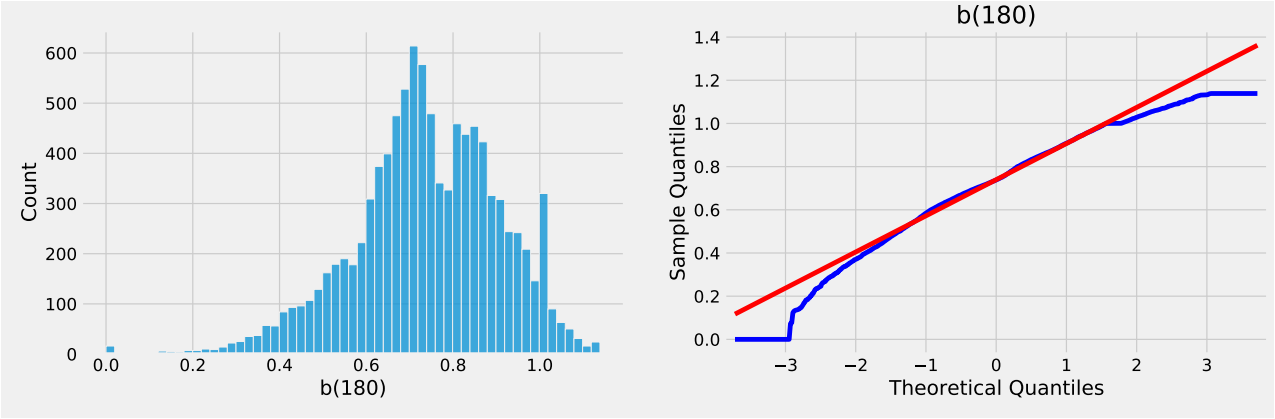


Figura 3.17.: Histograma y gráfico Q-Q del valor $b = 180$ s/mm^2 de aproximadamente 10.000 vóxeles tomados aleatoriamente. Se aprecia un ligero sesgo en la distribución tanto en el histograma como en la pendiente de la gráfica Q-Q.

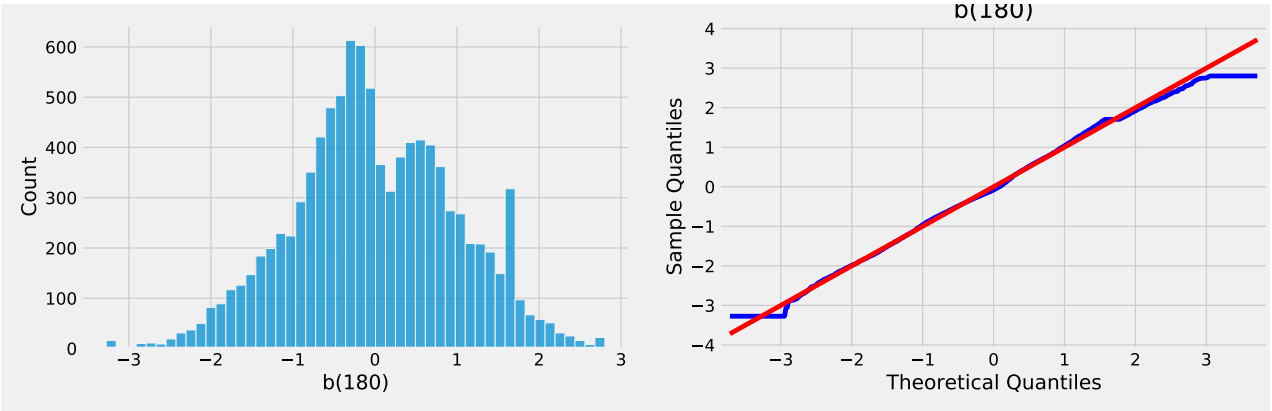


Figura 3.18.: Histograma y gráfico Q-Q de del valor $b = 180$ s/mm^2 de aproximadamente 10.000 vóxeles tomados aleatoriamente. El sesgo en la distribución se ha reducido y se ubica alrededor del cero. El resultado se aprecia en el histograma y en la pendiente de la curva Q-Q.

El mismo proceso se repite en todos los vectores $X_{i,j}$ cuyos datos que presenten algún desbalance o sesgo.

Tanto para los algoritmos de clasificación como para los algoritmos de predicción, el conjunto de datos se divide aleatoriamente en dos subconjuntos cuyo tamaño es del 80 % para la etapa de entrenamiento y del 20 % para la etapa de pruebas. Finalmente, para la implementación de los algoritmos de predicción se toman 2 caminos: uno para los datos de r^2 y otro para los demás grupos de datos D, D^*, f, K.

3.3.2. Construcción de los modelos para r^2

Los algoritmos de predicción para el valor de r^2 se deben basar en modelos de clasificación de clase simple, esto debido a que los valores en la salida Y para r^2 solamente son 0 ó 1. Algunos de los modelos más usados para clasificación de clase simple son los C-Support Vector SVC), Logistic Regression (LR), Random Forest (RF), Extra-Trees Classifier (ETC), XGBoost Classifier (xgb) y Multi-layer Perceptron (mlp), entre muchos otros. Estos algoritmos ya se encuentran validados por el proyecto scikit-learn y son los que se implementan durante este desarrollo.

Cada uno de estos modelos se entrena con los datos preparados y se ordenan de acuerdo al resultado de precisión y test AUC de cada uno.Por cada uno de estos modelos se generan las matrices de confusión y las curvas ROC para la evaluación de la clasificación. Esta metodología se ha tomado parcialmente de la plataforma Kaggle, específicamente del problema "Diabetes Tableau Dashboard & Voting Clf (5 models)"[1].

3.3.3. Construcción de los modelos para D, D* f y K

Los algoritmos de predicción para los valores de D, D^*, f y K se deben basar en modelos de clasificación multi-clase, dado que los valores en la salida Y de estos parámetros tienen 2^7 posibilidades en vez de solamente 0 ó 1 como en el caso anterior. Algunos de los modelos más usados para clasificación multiclase son la regresión Kernel Ridge Regression (Ridge), Least Absolute Shrinkage and Selection Operator (Lasso), ElasticNet (EN), Support Vector Machines (SVR), RandomForest, XGBoost y Polinómica (de grado d) (Poly) y son los que se implementan durante este desarrollo.

Para un modelo de regresión como estos no es conveniente utilizar matrices de confusión o curvas ROC, y por lo tanto no se se evalúa la precisión de las predicciones de la mis-

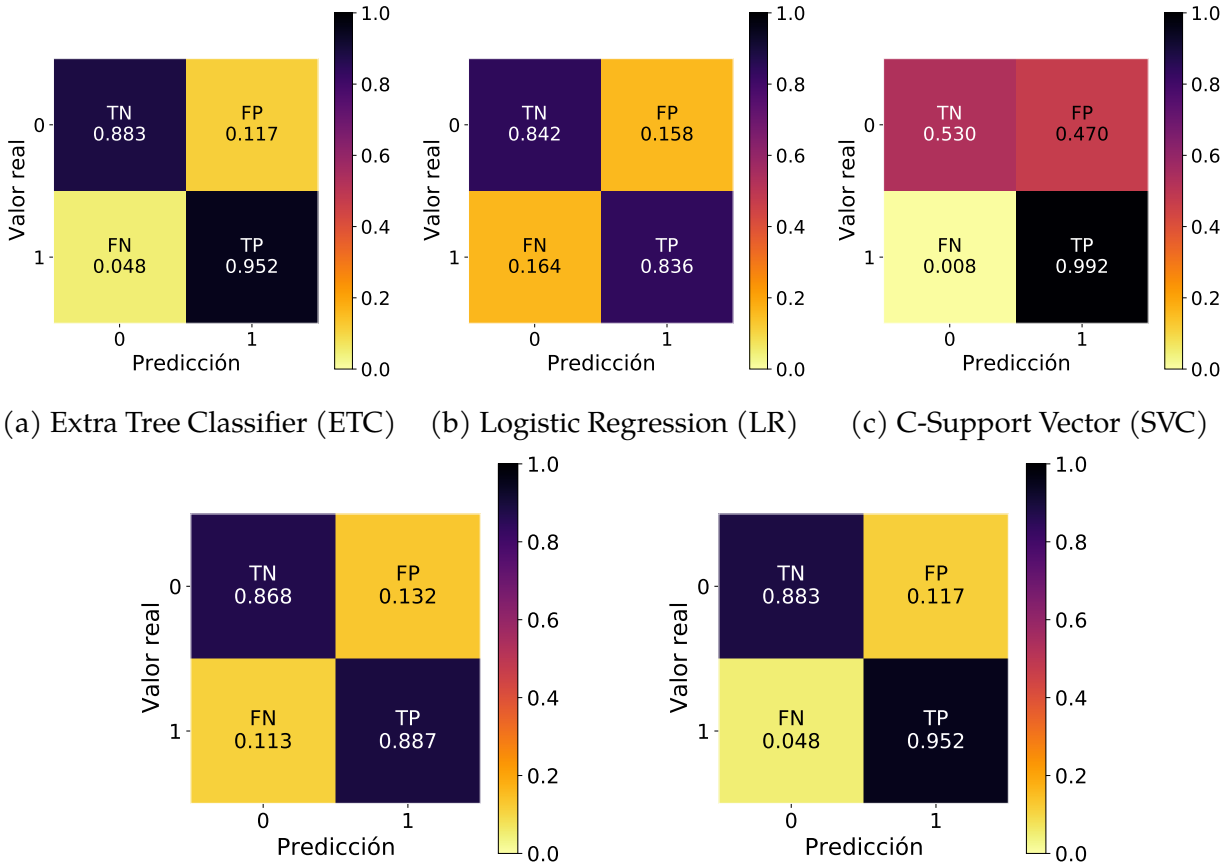
[1] https://www.kaggle.com/code/snikhil17/diabetes-tableau-dashboard-voting-clf-5-models/notebook

ma manera que con los clasificadores de clase simple. En la predicción multiclase se han implementado funciones de cálculo del error absoluto medio (MAE), coeficiente de determinación (R^2), error cuadrático medio (MSE), su raiz (RMSE) y su puntaje por validación cruzada (RMSE cross-valiation); todos estos mediante la librería `sklearn`. Cada uno de los modelos se entrena con los datos preparados y su resultado se ordena de acuerdo al puntaje obtenido por cada uno. Esta metodología se ha tomado parcialmente de la plataforma Kaggle, específicamente del problema "House Price Prediction"[2].

[2] https://www.kaggle.com/code/emrearslan123/house-price-prediction/notebook

4. Resultados y análisis

4.1. Clasificación de r^2



(a) Extra Tree Classifier (ETC) (b) Logistic Regression (LR) (c) C-Support Vector (SVC)

(d) XGBoost (XGB) (e) MultiLayer Perceptron (MLP)

Figura 4.1.: Matrices de confusión normalizadas de los algoritmos implementados para la clasificación de r^2.

En la figura 4.1 se muestran las matrices de confusión normalizadas para los algoritmos de clasificación de r^2 modificado implementados, donde la proximidad a 1 de los valores de la diagonal principal indica el mejor desempeño del algoritmo mediante la precisión (ACC)

de sus predicciones según la ecuación (2.55). La precisión de cada uno de los algoritmos se indica en la tabla **4.1** donde la mejor precisión se obtiene del algoritmo Multi-Layer Perceptron (**4.1**e) y Extra Tree Classifier (**4.1**a).

No obstante, para una clasificación binaria, como es el caso de r^2 modificado, se suele usar preferiblemente el área bajo la curva (AUC) calculada mediante la ecuación (2.58), de las curvas ROC mostradas en la figura **4.2** ya que permite tener en cuenta datos sesgados (si existen) y evitar el *overfitting* o sobreajuste a una única clase cuando solo se busca la mejor relación entre falsos positivos y verdaderos positivos en un modelo.

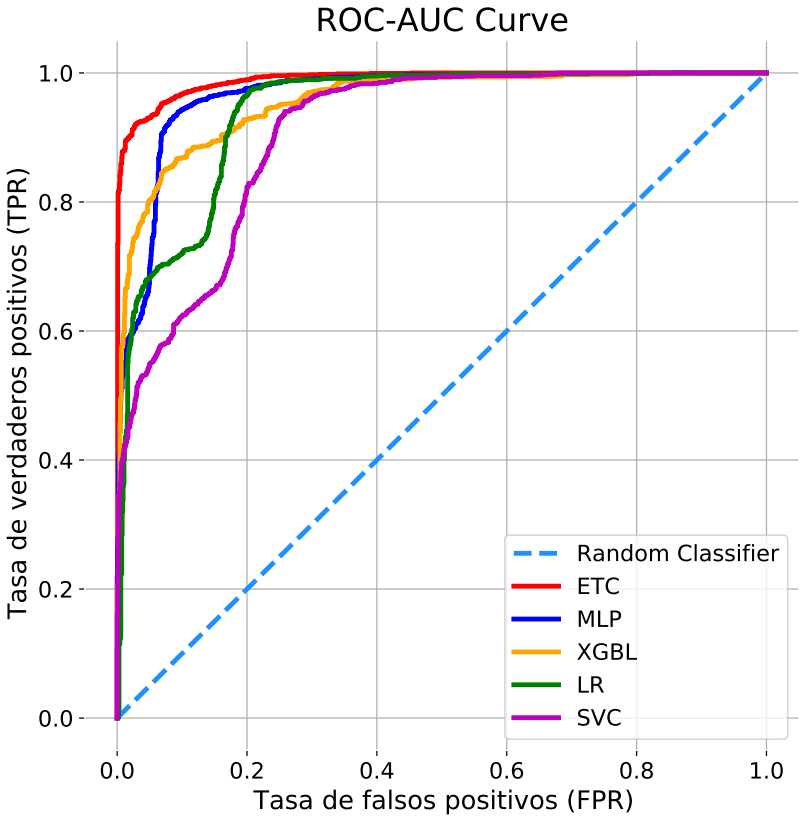


Figura 4.2.: Curvas ROC de los algoritmos implementados para la clasificación de r^2 respecto a un clasificador aleatorio.

La figura **4.2** muestra el gráfico ROC comparativo de los diferentes modelos implementados. Para el conjunto de algoritmos mostrados, se resume la precisión (ACC) y área bajo la curva del gráfico ROC (AUC) en la tabla **4.1**. En esta se muestra ordenadamente los algoritmos de acuerdo a su valor de AUC. Todos los modelos, poseen valores de AUC de al menos el 90 %, sin embargo, los mejores resultados se obtienen del Extra Tree Classifier

(ETC) con el 98.4 % y el Multi-Layer Perceptron (MLP) con el 96.3 %. Por otra parte, la precisión de los algoritmos resulta ser más baja, principalmente para el C-Support Vector (SVC), estando por debajo del 80 %. El Extra-Gradient Boost (XGB), aunque posee un valor de AUC de 93.9 %, su precisión resulta ser menor que la de la regresión logística (LR) con el 83.8 % y 87.7 % respectivamente.

Algoritmo	ACC	AUC
ETC	0.941	0.984
MLP	0.917	0.963
XGB	0.838	0.939
LR	0.877	0.921
SVC	0.761	0.900

Tabla 4.1.: Resultados de precisión (ACC) y área bajo la curva (AUC) de los algoritmos implementados ordenados respecto al valor de AUC. Los valores han sido coloreados de mayor a menor en sus respectivas columnas ACC y AUC empezando con el azul oscuro y terminando en el rojo oscuro.

4.2. Predicción de D, D* f y K

Como se mencionó en el capítulo anterior, para la predicción mediante clasificación multiclase (128 clases) no es conveniente usar matrices de confusión o curvas ROC ya que cada clase adicional incrementaría en uno la cantidad de dimensiones del problema. Para estos casos, se opta por el cálculo de los errores en las predicciones y su ponderación por medio del error absoluto medio (MAE) o el error cuadrático medio (MSE) y su raíz (RMSE) y el valor RMSE de las validaciones cruzadas (cross-validated RMSE o RMSE_{CV}).

Las figuras 4.3, 4.4, 4.5 y 4.6 muestran la raíz de los errores cuadráticos medios obtenidos mediante validación cruzada (RMSE_{CV}) para los algoritmos implementados. Las barras de color más claro indican un menor RMSE_{CV} y, por lo tanto, un algoritmo con un mejor desempeño; un color más oscuro, para los algoritmos menos eficientes, representa barras con un mayor valor de RMSE_{CV}. La escala de colores aplica para los datos dentro de un mismo grupo y no tiene relación entre figuras diferentes. Mientras que en las tablas 4.2, 4.3, 4.4 y 4.5 se muestran los resultados numéricos de los errores calculados ordenados de acuerdo al RMSE_{CV}. Como todos los valores se encuentran normalizados a la unidad, las métricas de regresión también se encuentran calculadas respecto a 1. Este es el valor del error que se esperaría para las predicciones cuando no se han reescalado los valores de difusión, pseudo-difusión, fracción de perfusión y curtosis. Por lo cual, los valores

de MAE, MSE, RMSE y RMSE_{CV} deben ser multiplicados por el factor de normalización de los parámetros D, D^*, f y K respectivamente. Para D (por $4{,}88 \times 10^{-3}$ mm^2/s), para D^* (por $9{,}89 \times 10^{-2}$ mm^2/s), para K (por $1{,}95$) y para la fracción de perfusión f (por $0{,}91$).

Los resultados para la predicción de la difusión D son más homogéneos entre sí ya que es la cantidad que menos valores faltantes posee en todos los sets de imágenes. Como en todos los sets se ha calculado la difusión, pero sólo en algunos se calcula fracción de perfusión f, pseudo-difusión D^* o curtosis K, algunos algoritmos reportan mejores resultados con menor cantidad de datos que otros.

En general, los algoritmos de Random Forest y XGBoost presentan RMSE_{CV} menores que los demás, es decir, son los mejores algoritmos utilizados para la predicción de difusión, pseudo-difusión, perfusión y curtosis. Por su parte los algoritmos de ElasticNet y Lasso presentan los errores más grandes, siendo los menos apropiados para predicciones de valores de pseudo-difusión, perfusión y curtosis.

Los resultados de las métricas de regresión MAE corresponden al error o la diferencia que se obtendría luego de predecir un valor entre 0 y 1, mientras que el MSE tomaría el papel de la desviación de dicha predicción. La RMSE corresponde a los residuos de la estimación y es del mismo orden del MAE pero un poco mayor en todos los casos ya que, al ser más sensible a valores atípicos, tiene en cuenta posibles sesgos. Por su parte, el valor de la RMSE_{CM} se espera que sea similar al valor de la RMSE ya que promedia dichos errores sobre 5 resultados del mismo conjunto de entrenamiento con diferentes muestras. Finalmente, el valor de R^2 muestra la bondad del ajuste de los métodos que se han probado, sin embargo, aunque los demás errores sean pequeños, este valor no será un buen indicador si los datos de entrenamiento se encuentran dispersos respecto a la función de predicción.

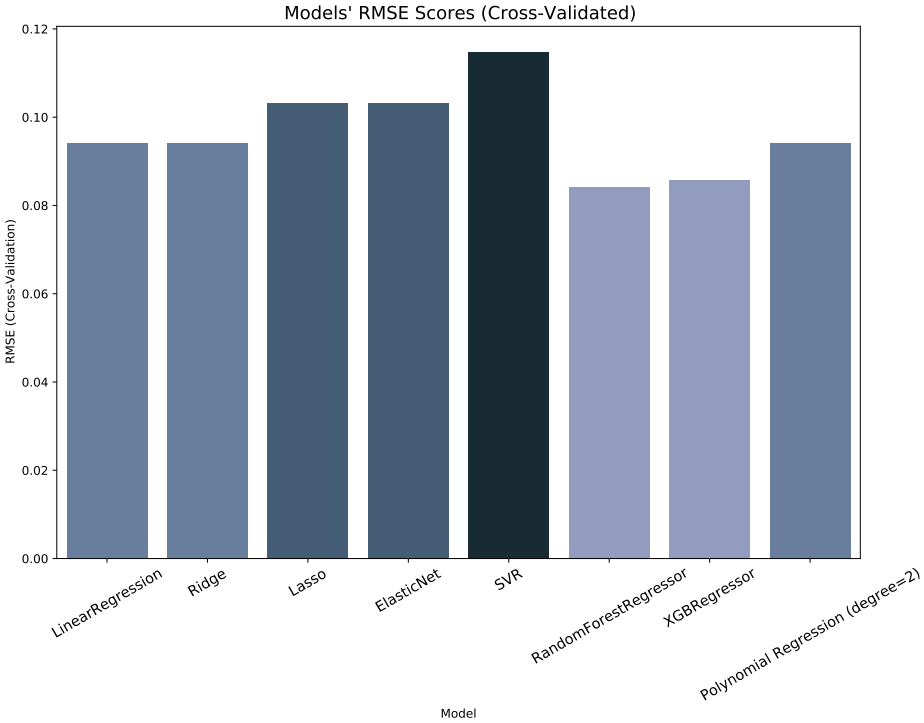


Figura 4.3.: Error cuadrático medio (RMSE) con validación cruzada de cada uno de los algoritmos implementados para la predicción de la difusión D.

Modelo	MAE	MSE	RMSE	R^2	$RMSE_{CV}$
RF	0.0462	0.0071	0.0845	0.292	0.0843
XGB	0.0465	0.0073	0.0856	0.274	0.0858
Ridge	0.0542	0.0084	0.0917	0.166	0.0942
LinR	0.0543	0.0084	0.0918	0.166	0.0942
Poly. (d=2)	0.0577	0.0160	0.1266	0.587	0.0942
Lasso	0.0512	0.0101	0.1004	0.00003	0.1032
ElasticNet	0.0512	0.0100	0.1005	0.00003	0.1032
SVR	0.0743	0.0125	0.1118	0.239	0.1148

Tabla 4.2.: Error absoluto medio (MAE), error cuadrático medio (MSE), raíz del error cuadrático medio (RMSE), coeficiente de determinación de la regresión (R^2) y raíz del error cuadrático medio mediante validación cruzada (cross-validated) de cada uno de los algoritmos implementados para los datos de difusión.

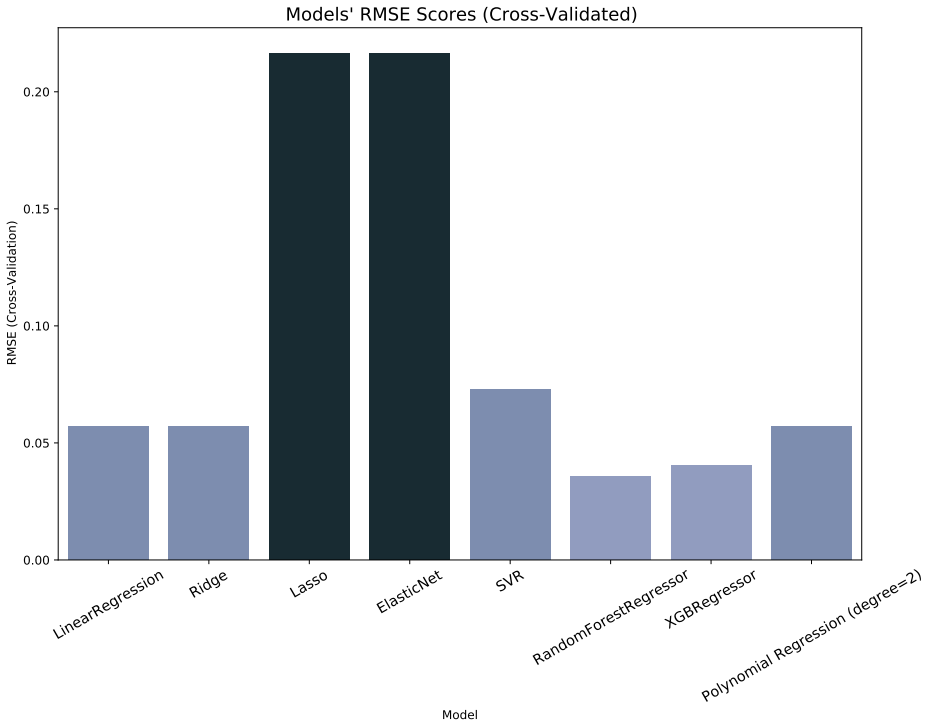


Figura 4.4.: Error cuadrático medio (RMSE) con validación cruzada de cada uno de los algoritmos implementados para la predicción de la pseudo-difusión D^*.

Modelo	MAE	MSE	RMSE	R^2	$RMSE_{CV}$
RF	0.0159	0.00117	0.0342	0.973	0.0357
XGB	0.0213	0.00149	0.0385	0.965	0.0402
LinR	0.0382	0.00314	0.0560	0.926	0.0571
Poly. (d=2)	0.0166	0.00112	0.0335	0.974	0.0571
Ridge	0.0385	0.00313	0.0560	0.926	0.0572
SVR	0.0458	0.00580	0.0761	0.864	0.0727
Lasso	0.118	0.0426	0.2064	0.000006	0.2165
ElasticNet	0.117	0.0426	0.2064	0.000005	0.2165

Tabla 4.3.: Error absoluto medio (MAE), error cuadrático medio (MSE), raíz del error cuadrático medio (RMSE), coeficiente de determinación de la regresión (R^2) y raíz del error cuadrático medio mediante validación cruzada (cross-validated) de cada uno de los algoritmos implementados para los datos de pseudo-difusión.

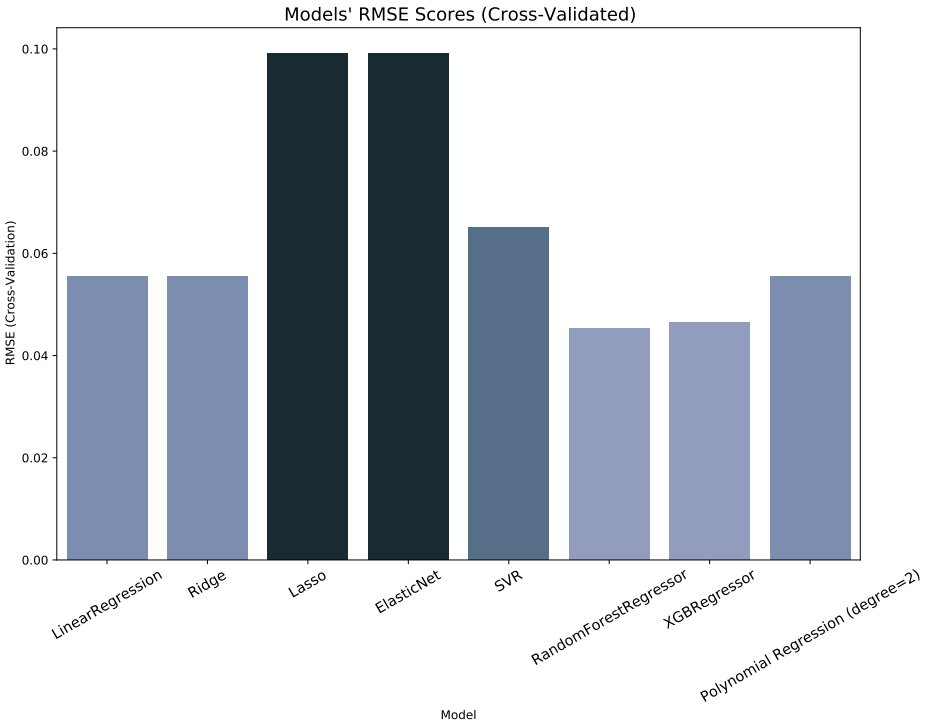


Figura 4.5.: Error cuadrático medio (RMSE) con validación cruzada de cada uno de los algoritmos implementados para la predicción de la fracción de perfusión f.

Modelo	MAE	MSE	RMSE	R^2	$RMSE_{CV}$
RF	0.0182	0.0023	0.0477	0.766	0.0454
XGB	0.0024	0.0024	0.0490	0.753	0.0465
Ridge	0.0337	0.0034	0.0585	0.648	0.0555
LinR	0.0337	0.0034	0.0585	0.647	0.0555
Poly. (d=2)	0.0242	0.0024	0.0491	0.751	0.0555
SVR	0.0477	0.0042	0.0645	0.571	0.0651
Lasso	0.0542	0.0097	0.0986	0.0004	0.0991
ElasticNet	0.0542	0.0097	0.0987	0.0003	0.0992

Tabla 4.4.: Error absoluto medio (MAE), error cuadrático medio (MSE), raíz del error cuadrático medio (RMSE), coeficiente de determinación de la regresión (R^2) y raíz del error cuadrático medio mediante validación cruzada (cross-validated) de cada uno de los algoritmos implementados para los datos de fracción de perfusión.

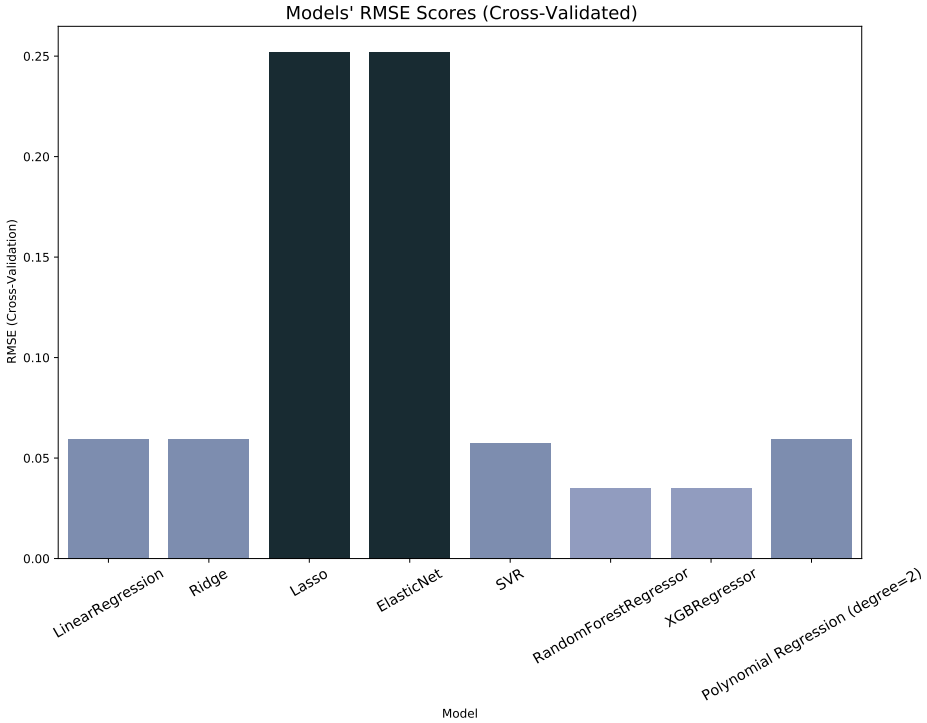


Figura 4.6.: Error cuadrático medio (RMSE) con validación cruzada de cada uno de los algoritmos implementados para la predicción de la curtosis K.

Modelo	MAE	MSE	RMSE	R^2	$RMSE_{CV}$
RF	0.0186	0.0013	0.0355	0.980	0.0352
XGB	0.0197	0.0012	0.0351	0.980	0.0353
SVR	0.0364	0.0031	0.0555	0.951	0.0577
Ridge	0.0368	0.0034	0.0581	0.946	0.0595
LinR	0.0368	0.0034	0.0581	0.946	0.0595
Poly. (d=2)	0.0174	0.0013	0.0366	0.979	0.0595
Lasso	0.161	0.063	0.0251	0.0001	0.2522
ElasticNet	0.161	0.062	0.0251	0.0001	0.2522

Tabla 4.5.: Error absoluto medio (MAE), error cuadrático medio (MSE), raíz del error cuadrático medio (RMSE), coeficiente de determinación de la regresión (R^2) y raíz del error cuadrático medio mediante validación cruzada (cross-validated) de cada uno de los algoritmos implementados para los datos de curtosis.

4.3. Tiempo de procesamiento

Para evaluar el tiempo de procesamiento, se calculó el tiempo que los algoritmos convencionales implementados en este trabajo tardan en encontrar los parámetros D, D^*, f y K de 100.000 vóxeles para cada uno de los grupos de imágenes (tabla 4.6). La tabla 4.7 muestra el tiempo que tardan los métodos de ML presentados anteriormente luego de haber sido entrenados para encontrar individualmente cada uno de estos parámetros en la misma cantidad de vóxeles. Esto último, debido a que los métodos de ML se especializan en encontrar un parámetro a la vez a partir de un vóxel de cualquier corte o cualquier set de imágenes. Los valores encontrados fueron calculados a partir de 10 conjuntos diferentes de 100.000 vóxeles cada uno y se presentan junto a su respectiva desviación estándar.

El tiempo de promedio de cálculo de los 4 parámetros principales para 100.000 vóxeles mediante métodos convencionales teniendo en cuenta todos los sets de imágenes es de 4408 ± 351 segundos o $73{,}5 \pm 5{,}9$ minutos. Por su parte, el tiempo de procesamiento de 100.000 vóxeles usando los algoritmos de ML propuestos para encontrar los 4 parámetros es de $18{,}998 \pm 0{,}135$ segundos.

Set de imágenes	Tiempo promedio de procesamiento de 100.000 vóxeles (s)
MR701	5984(394)
IVIM	4512(428)
CFIN	3416(301)
PigBrain	3723(279)

Tabla 4.6.: Tiempos promedio de cálculo de difusión, pseudo-difusión, fracción de perfusión y curtosis de 100.000 vóxeles en cada set de imágenes mediante métodos convencionales.

Parámetro	Tiempo promedio de procesamiento de 100.000 vóxeles (s)
Difusión (D)	4.835(28)
Pseudo-difusión (D^*)	5.632(53)
Fracción de perfusión (f)	4.545(42)
Curtosis (K)	3.986(12)

Tabla 4.7.: Tiempos promedio de cálculo de difusión, pseudo-difusión, fracción de perfusión y curtosis de 100.000 vóxeles usando métodos de ML.

5. Conclusiones y recomendaciones

5.1. Conclusiones

La caracterización experimental del decaimiento de la señal en vóxeles individuales es el primer paso para la determinación de los parámetros que describen físicamente una imagen de resonancia magnética nuclear pesada por difusión, pues permite cuantificar y comprobar la existencia del fenómeno físico de la difusión a partir de las componentes elementales de la imagen. Para caracterizar estas curvas de decaimiento se han tomado como referencia algunos modelos de difusión tisular propuestos previamente por otros autores. Entre estos se incluyen el modelo ADC (mono-exponencial) y el modelo IVIM (bi-exponencial) simple y con curtosis. Cada uno de estos modelos ha permitido determinar los parámetros de interés (D, D^*, f y K) mediante regresiones lineales o polinómicas vóxel por vóxel para 4 conjuntos de imágenes distintos.

Entre los modelos implementados en este trabajo, los mapas de ADC son los que presentan valores de difusión más consistentes en cuanto a la poca dispersión de los datos para magnitud D. Además, en todas las imágenes prevalece la existencia de difusión sin otros efectos adicionales. Por su parte, de acuerdo con la teoría, se comprueba que la pseudo-difusión D^*, presenta valores de un orden de magnitud por encima de D, sin embargo, es un parámetro con una dispersión mucho más grande ya que algunos vóxeles se estimaron valores atípicos bastante mayores producto de los métodos prácticos de adquisición de las imágenes. La fracción de perfusión f, conforme a lo esperado y a su definición, arroja valores en el intervalo entre 0 y 1, aunque con una distribución asimétrica sesgada a la izquierda y una moda alrededor de 0.22. Finalmente, la curtosis K ha sido estimada solamente para los sets de imágenes con valores b suficientemente grandes para que el efecto sea detectable ($b \geq 3000$ s/mm^2). La distribución de estos valores varía entre imágenes, siendo más o menos gaussiana con media alrededor de 1 para el set `PigBrain`. Mientras que para el set `CFIN` la media también se encuentra alrededor de 1 pero los datos presentan un leve sesgo hacia la derecha.

La implementación de los filtros de suavizado las imágenes de r^2 modificado también es un paso importante en el preprocesamiento de las imágenes, ya que aquí, se introducen in-

homogeneidades en la información y los datos de entrenamiento (al considerar datos que se habrían descartado inicialmente por falta de precisión en la regresión) pero además, se corrigen posibles errores ocurridos durante la caracterización de las curvas de atenuación de los vóxeles. La discretización de las imágenes a 7 bits no es un punto esencial dado que la evaluación del error en algoritmos de predicción también aplica para cantidades continuas. No obstante, se muestra que, mediante la discretización se limita el error a múltiplos de 0.0078 para predicciones realizadas en datos normalizados (entre 0 y 1).

Además se han implementado algoritmos de clasificación para determinar, a partir de los mapas de r^2 modificado, en cuales lugares de una imagen existe difusión, pseudo-difusión, fracción de perfusión o curtosis con un nivel de confianza igual o mayor al 90 %. De los algoritmos de clasificación puestos a prueba, los métodos ETC (Extra Trees Classifier) y MLP (Perceptrón multicapa) reportaron los mejores resultados de precisión (94.1 % y 91.7 % respectivamente) y test de área bajo la curva ROC (98.4 % y 96.3 % respectivamente) (tabla **4.1**). Respecto a los algoritmos de regresión para D, D^*, f y K, RF (Random Forest) y XGB (Extra-Gradient Boost) han sido, en todos los casos, los métodos con menores errores respecto a la raíz del error cuadrático medio de una validación cruzada (RMSE_{CV}) para una división *k-fold* de 5 muestras. Los mejores resultados (menores errores) para la predicción de difusión, pseudo-difusión, fracción de perfusión y curtosis de acuerdo con la métrica de RMSE_{CV}, se obtuvieron de los algoritmos RF y XGB. En contraparte, el error más grande para predicción de D fue para el algoritmo SVR mientras que para los demás parámetros (D^*, f y K), los algoritmos Lasso y ElasticNet fueron los de peor desempeño de acuerdo con su puntuación RMSE_{CV}. No obstante, es posible que otros algoritmos que no estén considerados dentro en esta investigación presenten mejores resultados en clasificación o regresión.

Finalmente, se ha calculado el tiempo que los algoritmos de ML tardan en encontrar los valores de los parámetros D, D^*, f y K frente al tiempo que toma encontrar los mismos parámetros mediante métodos convencionales, ambos sobre un total de 100.000 vóxeles. Por medio de los métodos convencionales este tiempo es de 4408 ± 351 segundos mientras que mediante ML es de $18,998 \pm 0,135$ segundos, siendo este aproximadamente 232 veces más rápido que el primero (tablas **4.6** y **4.7**). Se debe tener en cuenta que el tiempo de cálculo usando los algoritmos de ML se ha calculado luego de que estos ya se encontraran entrenados. El tiempo de entrenamiento no se ha medido puesto que este varía ampliamente entre un método y otro y en función de la cantidad de datos usados para el entrenamiento.

5.2. Recomendaciones

En este trabajo solamente se toman en cuenta algunos de los modelos más utilizados en investigación para la caracterización de la difusión, sin embargo, existen otros que permiten detectar y diferenciar otros efectos. Por ejemplo, podría considerarse la caracterización de atenuación mediante el modelo tri-exponencial [57] o mediante "signal spoiling" [73]. Además, para mejorar la precisión en el cálculo de los coeficientes de determinación por métodos convencionales, podría considerarse el efecto del ruido Riciano para IRM [74].

Por su parte, los algoritmos pueden mejorarse sustancialmente para reducir el tiempo de ejecución (mediante procesamiento en paralelo) o disminuir los recursos computacionales mediante la implementación de Cython, una herramienta que permite ejecutar código C y C++ en Python. La ventaja de C/C++ sobre Python está en que normalmente se ejecuta más rápido y tiene una mejor eficiencia en bucles como `for` o `while`.

Otra proyección (un poco optimista) que se podría esperar a partir de esta investigación es, no solamente utilizar las ventajas del ML para predicción de parámetros sobre vóxeles individuales, sino sobre imágenes completas, mediante técnicas de súper-resolución o rellenado de imágenes incompletas y con la opción de permitir el diagnóstico, tratamiento o predicción de padecimientos con base en una cantidad reducida de imágenes crudas DWI.

A. Anexo: Disponibilidad de los códigos

Los códigos en Python, tanto de la preparación y tratamiento de las imágenes como del preprocesamiento y entrenamiento de los algoritmos de Machine Learning, se encuentran disponibles para su consulta en linea a través de la herramienta Google Colaboratory (Colab).

Cálculo de parámetros D, D^*, f y K, preparación de las imágenes y filtrado 2D:

- Set `MR701`
- Set `IVIM`
- Set `CFIN`
- Set `PigBrain`

Preparación de los vectores de entrenamiento y prueba, algoritmos de clasificación de r^2, matrices de confusión y curvas ROC:

- Clasificación de mapas r^2 modificado

Preparación de los vectores de entrenamiento y prueba, algoritmos de predicción y métricas de regresión:

- Difusión (D)
- Pseudo-difusión (D^*)
- Fracción de perfusión (f)
- Curtosis (K)

Bibliografía

[1] Valerij G. Kiselev. Fundamentals of diffusion mri physics. *NMR in biomedicine*, 30, 3 2017.

[2] Denis Le Bihan. What can we see with ivim mri? *NeuroImage*, 187:56–67, 2 2019.

[3] Mara Cercignani and Mark A. Horsfield. The physical basis of diffusion-weighted mri. *Journal of the neurological sciences*, 186 Suppl 1, 5 2001.

[4] Radiopaedia.org. Diffusion-weighted imaging — radiology reference article —, 2021.

[5] Laura Andrea Pastor Luque and Luis Agulles Pedrós. Modelación de difusión en irm, 2020.

[6] Denis Le Bihan. Apparent diffusion coefficient and beyond: What diffusion mr imaging can tell us about tissue structure. *https://doi.org/10.1148/radiol.13130420*, 268:318–322, 8 2013.

[7] Carmelo Messina, Rodolfo Bignone, Alberto Bruno, Antonio Bruno, Federico Bruno, Marco Calandri, Damiano Caruso, Pietro Coppolino, Riccardo De Robertis, Francesco Gentili, Irene Grazzini, Raffaele Natella, Paola Scalise, Antonio Barile, Roberto Grassi, and Domenico Albano. Diffusion-weighted imaging in oncology: An update. *Cancers*, 12:1–28, 6 2020.

[8] Saef Izzy, Daniel B. Rubin, Firas S. Ahmed, Feras Akbik, Simone Renault, Katelyn W. Sylvester, Henrikas Vaitkevicius, Jennifer A. Smallwood, Michael M. Givertz, and Steven K. Feske. Cerebrovascular accidents during mechanical circulatory support: New predictors of ischemic and hemorrhagic strokes and outcome. *Stroke*, 49:1197–1203, 2018.

[9] Ashok Srinivasan, R. Dvorak, K. Perni, S. Rohrer, and S. K. Mukherji. Differentiation of benign and malignant pathology in the head and neck using 3t apparent diffusion coefficient values: Early experience. *American Journal of Neuroradiology*, 29:40–44, 1 2008.

[10] Young Jin Choi, In Sook Lee, You Seon Song, Jeung Il Kim, Kyung Un Choi, and Jong Woon Song. Diagnostic performance of diffusion-weighted (dwi) and dynamic contrast-enhanced (dce) mri for the differentiation of benign from malignant soft-tissue tumors. *Journal of magnetic resonance imaging : JMRI*, 50:798–809, 9 2019.

[11] Lluís Agulles Pedrós, R. Acosta, Peter Blümler, and Hans Spiess. Diffusion influence on gas magnetic resonance imaging. *Universitas Scientiarum*, 13:281–289, 09 2008.

[12] E. O. Stejskal and J. E. Tanner. Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient. *The Journal of Chemical Physics*, 42:288, 7 1965.

[13] D. Le Bihan, E. Breton, D. Lallemand, M. L. Aubin, J. Vignaud, and M. Laval-Jeantet. Separation of diffusion and perfusion in intravoxel incoherent motion mr imaging. *https://doi.org/10.1148/radiology.168.2.3393671*, 168:497–505, 8 1986.

[14] Olaf Dietrich, Andreas Biffar, Andrea Baur-Melnyk, and Maximilian F. Reiser. Technical aspects of mr diffusion imaging of the body. *European Journal of Radiology*, 76:314–322, 2010.

[15] Fumiyuki Yamasaki, Kaoru Kurisu, Kenichi Satoh, Kazunori Arita, Kazuhiko Sugiyama, Megu Ohtaki, Junko Takaba, Atushi Tominaga, Ryosuke Hanaya, Hiroyuki Yoshioka, Seiji Hama, Yoko Ito, Yoshinori Kajiwara, Kaita Yahara, Taiichi Saito, and Muhamad Thohar Arifin. Apparent diffusion coefficient of human brain tumors at mr imaging. *Radiology*, 235:985–991, 6 2005.

[16] Signe Swerkersson, Oscar Grundberg, Karl Kölbeck, Andreas Carlberg, Sven Nyrén, and Mikael Skorpil. Optimizing diffusion-weighted magnetic resonance imaging for evaluation of lung tumors: A comparison of respiratory triggered and free breathing techniques. *European Journal of Radiology Open*, 5:189, 1 2018.

[17] Francesco Giganti, Elena Orsenigo, Antonio Esposito, Damiano Chiari, Annalaura Salerno, Alessandro Ambrosi, Luca Albarello, Elena Mazza, Carlo Staudacher, Alessandro Del Maschio, and Francesco De Cobelli. Prognostic role of diffusion-weighted mr imaging for resectable gastric cancer. *Radiology*, 276:444–452, 8 2015.

[18] Sungmin Woo, Chong Hyun Suh, Sang Youn Kim, Jeong Yeon Cho, Seung Hyup Kim, and Min Hoan Moon. Head-to-head comparison between biparametric and multiparametric mri for the diagnosis of prostate cancer: A systematic review and meta-analysis. *AJR. American journal of roentgenology*, 211:W226–W241, 11 2018.

[19] Weigen Yao, Jie Liu, Jiaju Zheng, Pengcong Lu, Shufang Zou, and Ying Xu. Study on diagnostic value of quantitative parameters of intravoxel incoherent motion diffusion-weighted imaging (ivim-dwi) in prostate cancer. *American Journal of Translational Research*, 13:3696, 2021.

[20] Harri Merisaari, Pekka Taimen, Rakesh Shiradkar, Otto Ettala, Marko Pesola, Jani Saunavaara, Peter J. Boström, Anant Madabhushi, Hannu J. Aronen, and Ivan Jambor. Repeatability of radiomics and machine learning for dwi: Short-term repeatability study of 112 patients with prostate cancer. *Magnetic resonance in medicine*, 83:2293–2309, 6 2020.

[21] Garyfallidis E, Brett M, Amirbekian B, Rokem A, van der Walt S, Descoteaux M, Nimmo-Smith I, and Dipy Contributors. Dipy, a library for the analysis of diffusion mri data. frontiers in neuroinformatics. 8, 2014.

[22] Rafael Neto Henriques, Ariel Rokem, Eleftherios Garyfallidis, Samuel St-Jean, Eric Thomas Peterson, and Marta Morgado Correia. [re] optimization of a free water elimination two-compartment model for diffusion tensor imaging. *bioRxiv*, page 108795, 2 2017.

[23] Brian Hansen and Sune Nørhøj Jespersen. Data for evaluation of fast kurtosis strategies, b-value optimization and exploration of diffusion mri contrast. *Scientific Data 2016 3:1*, 3:1–5, 8 2016.

[24] Eric Peterson. Ivim dataset. figshare. dataset., 2016.

[25] Tim B. Dyrby, William F.C. Baaré, Daniel C. Alexander, Jacob Jelsing, Ellen Garde, and Lise V. Søgaard. An ex vivo imaging pipeline for producing high-quality and high-resolution diffusion-weighted imaging datasets. *Human brain mapping*, 32:544–563, 4 2011.

[26] Andrés Felipe Giraldo Román and Luis Agulles Pedrós. Análisis en imágenes de resonancia magnética pesadas por difusión. Master's thesis, Universidad Nacional de Colombia, 2014.

[27] National Institute of Biomedical Imaging and Bioengineering (NIBIB). Imagen por resonancia magnética (irm).

[28] J. J. (Jun John) Sakurai. *Advanced quantum mechanics*. Addison-Wesley Pub. Co, 1 edition, 1967.

[29] David J. Griffiths and Darrell F. Schroeter. *Introduction to Quantum Mechanics*. Cambridge University Press, 3 edition, 8 2018.

[30] E. Mark Haacke, Robert W. Brown, Michael R. Thompson, and Ramesh Venkatesan. *Magnetic Resonance Imaging - Physical Principles and Sequence Design*. John Wiley & Sons, 1 edition, 1999.

[31] Joseph Larmor. Ix. a dynamical theory of the electric and luminiferous medium.—part iii. relations with material media. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 190:205–300, 12 1897.

[32] Robin A. De Graaf. *In vivo NMR spectroscopy: Principles and techniques*. John Wiley & Sons, 2 edition, 2007.

[33] F Reif. *Fundamentals of statistical and thermal physics*. McGraw-Hill, Inc, 1 edition, 1965.

[34] Evert J Blink. Basic mri physics, 2010.

[35] Malcolm H Levitt. *Spin Dynamics Basics of Nuclear Magnetic Resonance Second edition*. Ltd John Wiley & Sons, second edition edition, 2008.

[36] Ray Freeman. Spin choreography : basic steps in high resolution nmr. page 391, 1999.

[37] F. Bloch. Nuclear induction. *Physical Review*, 70:460, 10 1946.

[38] Jean Philibert. One and a half century of diffusion: Fick, einstein, before and beyond. 2005.

[39] Adolf Fick. Ueber diffusion. *Annalen der Physik*, 170:59–86, 1855.

[40] J Crank. *The mathematics of diffusion*. Clerendon Press, 2 edition, 1975.

[41] Ángel Franco García. Difusión. ley de fick, 2010.

[42] Jaume Gili and Julio Alonso. IntroducciÓn biofÍsica a la resonancia magnÉtica en neuroimagen, 2004.

[43] H. C. Torrey. Bloch equations with diffusion terms. *Physical Review*, 104:563, 11 1956.

[44] V M Kenkre, Eiichi Fukushima, and D Sheltraw. Simple solutions of the torrey-bloch equations in the nmr study of molecular diffusion. *JOURNAL OF MAGNETIC RESONANCE*, 128:62–69, 1997.

[45] E. L. Hahn. Spin echoes. *Physical Review*, 80:580, 11 1950.

[46] H. Y. Carr and E. M. Purcell. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Physical Review*, 94:630, 5 1954.

[47] Jonathan H. Burdette, Allen D. Elster, and Peter E. Ricci. Acute cerebral infarction: quantification of spin-density and t2 shine-through phenomena on diffusion-weighted mr images. *Radiology*, 212:333–339, 1999.

[48] D. Le Bihan, E. Breton, D. Lallemand, M. L. Aubin, J. Vignaud, and M. Laval-Jeantet. Separation of diffusion and perfusion in intravoxel incoherent motion mr imaging. *Radiology*, 168:497–505, 1988.

[49] D. (Denis) Le Bihan, Mami Lima, Christian Federau, and Eric E. Sigmund. *Intravoxel incoherent motion (IVIM) MRI : principles and applications*. Jenny Stanford Publishing, 1 edition, 2018.

[50] Andrew J. Steven, Jiachen Zhuo, and Elias R. Melhem. Diffusion kurtosis imaging: An emerging technique for evaluating the microstructural environment of the brain. *http://dx.doi.org/10.2214/AJR.13.11365*, 202, 12 2013.

[51] Mami Iima and Denis Le Bihan. Clinical intravoxel incoherent motion and diffusion mr imaging: Past, present, and future. *Radiology*, 278:13–32, 1 2016.

[52] Steren Chabert and Paola Scifo. Diffusion signal in magnetic resonance imaging: Origin and interpretation in neurosciences. *Biological Research*, 40:385–400, 2007.

[53] Jens H. Jensen, Joseph A. Helpern, Anita Ramani, Hanzhang Lu, and Kyle Kaczynski. Diffusional kurtosis imaging: the quantification of non-gaussian water diffusion by means of magnetic resonance imaging. *Magnetic resonance in medicine*, 53:1432–1440, 2005.

[54] Robert V. Mulkern, Steven J. Haker, and Stephan E. Maier. On high b diffusion imaging in the human brain: ruminations and experimental insights. *Magnetic resonance imaging*, 27:1151, 10 2009.

[55] Dmitriy A. Yablonskiy, G. Larry Bretthorst, and Joseph J.H. Ackerman. Statistical model for diffusion attenuated mr signal. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 50:664, 10 2003.

[56] Kevin M. Bennett, Kathleen M. Schmainda, Raoqiong Bennett, Daniel B. Rowe, Hanbing Lu, and James S. Hyde. Characterization of continuously distributed cortical water diffusion rates with a stretched-exponential model. *Magnetic resonance in medicine*, 50:727–734, 10 2003.

[57] Qiang Zeng, Feina Shi, Jianmin Zhang, Chenhan Ling, Fei Dong, and Biao Jiang. A modified tri-exponential model for multi-b-value diffusion-weighted imaging: A method to detect the strictly diffusion-limited compartment in brain. *Frontiers in Neuroscience*, 12:102, 2 2018.

[58] Matt G. Hall and Thomas R. Barrick. From diffusion-weighted mri to anomalous diffusion imaging. *Magnetic resonance in medicine*, 59:447–455, 2008.

[59] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of Machine Learning*. The MIT Press, 1 edition, 2012.

[60] C. E. Brodley and M. A. Friedl. Identifying mislabeled training data. *Journal Of Artificial Intelligence Research*, 11:131–167, 6 2011.

[61] Joaquín Amat Rodrigo. Machine learning w python & scikit-learn.

[62] Scikit learn 1.0.2 documentation. Cross-validation: evaluating estimator performance, 2022.

[63] Markus Ojala@tkk Fi and Gemma C Garriga. Permutation tests for studying classifier performance markus ojala. *Journal of Machine Learning Research*, 11:1833–1863, 2010.

[64] David Powers and Ailab. Evaluation: From precision, recall and f-measure to roc, informedness, markedness & correlation. *J. Mach. Learn. Technol*, 2:2229–3981, 01 2011.

[65] S. Madeh Piryonesi and Tamer E. El-Diraby. Data analytics in asset management: Cost-effective prediction of the pavement condition index. *Journal of Infrastructure Systems*, 26:04019036, 12 2019.

[66] Ana Rocío Rocío Del Valle Benavides and Juan Manuel Muñoz Pichardo. Curvas roc (receiver-operating-characteristic) y sus aplicaciones, 2017.

[67] Kelly H. Zou, A. James O'Malley, and Laura Mauri. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation*, 115:654–657, 2 2007.

[68] Google Developers. Machine learning crash course.

[69] Christopher D Brown and Herbert T Davis. Receiver operating characteristics curves and related decision measures: A tutorial. 2005.

[70] Morris H. DeGroot and Mark J. Schervish. *Probability and Statistics*, volume 1. Fourth edition edition, 2011.

[71] A. Mood, F. Graybill, and D Boes. *Introduction to the Theory of Statistics*. McGraw-Hill, 3rd edition, 1974.

[72] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[73] Nadège Corbin and Martina F Callaghan. Imperfect spoiling in variable flip angle t 1 mapping at 7t: Quantifying and minimizing impact. *Magn Reson Med*, 00:1–16, 2021.

[74] Saurav Basu, Thomas Fletcher, and Ross Whitaker. Rician noise removal in diffusion tensor mri. *Lecture Notes in Computer Science* (*including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*), 4190 LNCS - I:117–125, 2006.