



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



ESCUELA TÉCNICA
SUPERIOR INGENIERÍA
INDUSTRIAL VALENCIA

TRABAJO FIN DE MASTER EN INGENIERÍA BIOMÉDICA

**VALIDACIÓN CLÍNICA DE ALGORITMOS
DE PROCESADO EN IMAGEN MÉDICA.
APLICACIÓN EN RESONANCIA
MAGNÉTICA MULTIPARAMÉTRICA DE
PRÓSTATA.**

AUTOR: MARÍA MANCEBO GONZÁLEZ

TUTOR: IGNACIO BOSCH ROIG

Curso Académico: 2019-20

AGRADECIMIENTOS

“Me gustaría aprovechar la ocasión para agradecer a mis tutores la confianza que han depositado en mí y por darme la oportunidad de trabajar en este proyecto. También agradecer al equipo de QUIBIM la formación, la ayuda y la paciencia que han tenido conmigo, pero sobre todo por recibirme con los brazos abiertos.

A mi familia y amigos, a los de siempre y a los que acaban de llegar, gracias por apoyarme siempre en las decisiones que he tomado, por confiar en mí y por enseñarme a hacerlo. Soy todo lo que soy, gracias a ellos.”

MARCO COLABORATIVO

El presente trabajo se ha llevado a cabo durante las prácticas extracurriculares en la empresa QUIBIM SL.

El objetivo que persigue QUIBIM SL es mejorar la salud humana mediante la aplicación de técnicas avanzadas e innovadoras de procesamiento de imágenes radiológicas basadas en modelos matemáticos e inteligencia artificial, midiendo de esta forma cuantitativamente procesos que se producen en el cuerpo humano a través de la extracción de biomarcadores de imagen. Esta tecnología está dirigida para la práctica clínica, proyectos de investigación y ensayos clínicos. A través de su plataforma QUIBIM Precision® (www.precision.quibim.com) ofrece la posibilidad de que los usuarios puedan aplicar estos modelos de alta computación a sus imágenes sin necesidad de instalar ningún programa en su ordenador.

Resumen

Hoy en día, el cáncer de próstata es considerado uno de los carcinomas malignos más prevalentes entre la población masculina, siendo en España, el segundo más frecuente y el tercero en mortalidad. El tratamiento consta principalmente de procedimientos quirúrgicos, lo que puede producir en el paciente, efectos secundarios y secuelas. Por ello, surge la necesidad del diagnóstico precoz, de establecer el riesgo de progresión de los tumores y valorar a cada paciente antes de tomar cualquier decisión terapéutica.

Con el propósito de obtener una mayor cantidad de información sobre la patología, una de las técnicas más utilizadas, es la Imagen por Resonancia Magnética (RM), con la cual, de manera no invasiva, se puede clasificar de forma local y regional los estados en pacientes con cáncer. Los avances en la tecnología han derivado en el desarrollo de RM multiparamétrica, en el que se combina secuencias T2W anatómicas con evaluaciones funcionales y fisiológicas, incluyendo imágenes de difusión e imágenes de perfusión.

Al existir una excesiva variación en la interpretación de las imágenes de cáncer de próstata, para estandarizar el proceso de evaluación RM multiparamétrica y su aplicación, se creó el Sistema de Información y Datos en Imagen de Próstata (PI-RADS), formado por directrices básicas para la adquisición de las imágenes y la evaluación de cada lesión analizando la información proveniente de las distintas secuencias.

El presente trabajo consiste en la validación de los algoritmos de preprocesado involucrados en el análisis de las distintas secuencias, siendo estos: selección de la función de entrada arterial en imágenes de perfusión DCE y registros intersecuencia. Se lleva a cabo un proceso de validación completo e independiente para cada algoritmo de preprocesado, obteniéndose métricas y medidas estadísticas para evaluar el buen funcionamiento de los distintos algoritmos en una base de datos de pacientes representativos de la práctica clínica.

Palabras clave: “cáncer de próstata; resonancia magnética; validación; algoritmo”

Resum

Hui en dia, el càncer de pròstata és considerat un dels carcinomes malignes més prevalents entre la població masculina, sent a Espanya, el segon més freqüent i el tercer en mortalitat. El tractament consta principalment de procediments quirúrgics, la qual cosa pot produir en el pacient, efectes secundaris i seqüeles. Per això, sorgix la necessitat del diagnòstic precoç, d'establir el risc de progressió dels tumors i valorar cada pacient abans de prendre qualsevol decisió terapèutica.

Amb el propòsit d'obtindre una major quantitat d'informació sobre la patologia, una de les tècniques més utilitzades, és la Imatge per Ressonància Magnètica (RM), amb la qual, de manera no invasiva, es pot classificar de forma local i regional els estats en pacients amb càncer. Els avanços en la tecnologia han derivat en el desenrotllament de RM multiparamètrica, en el que es combina seqüències T2W anatòmiques amb avaluacions funcionals i fisiològiques, incloent imatges de difusió i imatges de perfusió.

A l'existir una excessiva variació en la interpretació de les imatges de càncer de pròstata, per a estandarditzar el procés d'avaluació RM multiparamètrica i la seua aplicació, es va crear el Sistema d'Informació i Dades en Imatge de Pròstata (PI- RADS), format per directrius bàsiques per a l'adquisició de les imatges i l'avaluació de cada lesió analitzant la informació provinent de les distintes seqüències.

El present treball consistix en la validació dels algorismes de preprocessat involucrats en l'anàlisi de les distintes seqüències, sent estos: selecció de la funció d'entrada arterial en imatges de perfusió DCE i registres interseqüència. Es du a terme un procés de validació complet i independent per a cada algorisme de preprocessat, obtenint-se mètriques i mesures estadístiques per a avaluar el bon funcionament dels distints algorismes en una base de dades de pacients representatius de la pràctica clínica.

Paraules clau: “càncer de pròstata; ressonància magnètica; validació; algoritmo”

Abstract

One of the most prevalent malignant carcinomas among the male population in these days is prostate cancer, also it is the second most frequent in Spain and the third in mortality. The treatment consists mainly of surgical procedures, which can produce side effects and sequels. Therefore, the need for early diagnosis arises, to establish the risk of tumor progression and to evaluate each patient before taking any therapeutic decision.

In order to obtain a greater amount of information about this pathology, one of the most used techniques is Magnetic Resonance Imaging (MRI), which with the use of contrast agents, states in cancer patients can be locally and regionally classified. Advances in technology have led to the development of Multiparametric MRI (mpMRI), which combines anatomical T2W sequences with functional and physiological assessments, including diffusion and perfusion imaging.

As there is excessive variation in the interpretation of prostate cancer images, in order to standardize the mpMRI evaluation process and its application, the Prostate Image-Reporting and Data System (PI-RADS) was created, consisting of basic guidelines for the evaluation of each lesion with this type of analysis.

The present work is about the validation of a series of pre-processing algorithms for the evaluation of prostate cancer using mpMRI and the recommendations of PI-RADS. These algorithms are the selection of the arterial input function in DCE perfusion images and the registration of images. A complete and independent validation process is carried out for each type of analysis, obtaining metrics and statistical measures that allow differentiating and demonstrating the validity of the application under study.

Keywords: “prostate cancer; magnetic resonance; validation, algorithm”

Índice General

PARTE I. MEMORIA

Capítulo 1. Introducción	1
1.1 Patología	1
1.1.1 Anatomía de la próstata	2
1.1.2 Cáncer de próstata	3
1.1.3 Diagnóstico en el cáncer de próstata	4
1.2 Imagen por Resonancia Magnética	4
1.2.1 Imágenes T2W	6
1.2.2 Imágenes de perfusión DCE	7
1.2.3 Imágenes de difusión DWI	8
1.3 Antecedentes	9
1.3.1 Selección de la función de entrada arterial	10
1.3.2 Registro inter secuencia de imágenes	12
1.4 Hipótesis y Objetivos	13
Capítulo 2. Materiales	15
2.1 Base de Datos Global	15
2.1.1 Base de datos para la validación de la AIF	15
2.1.1 Base de datos de la validación del registro de imágenes T2W-DCE	17
2.2 Software y formatos de imagen	19
Capítulo 3. Metodología	20
3.1 Validación de la selección de AIF	20
3.1.1 Validación cualitativa AIF	22
3.1.2 Validación cuantitativa AIF	24
3.2 Validación del Registro T2W-DCE	27

3.2.1	<i>Validación cualitativa del registro</i>	27
Capítulo 4.	Resultados	36
4.1	Validación de la Selección AIF	36
4.1.1	<i>Validación cualitativa AIF</i>	36
4.1.2	<i>Validación cuantitativa AIF</i>	38
4.1.3	<i>Evaluación interna del algoritmo AIF</i>	40
4.2	Validación del Registro T2W-DCE.....	42
4.2.1	<i>Primeros resultados obtenidos</i>	42
4.2.2	<i>Resultados validación clínica cualitativa</i>	44
4.2.3	<i>Evaluación interna del algoritmo de registro</i>	45
Capítulo 5.	Discusión	48
5.1	Validación de la Selección de AIF	48
5.1.1	<i>Discusión de los resultados cualitativos</i>	48
5.1.2	<i>Discusión de los resultados cuantitativos</i>	49
5.2	Validación del Registro T2W-DCE.....	51
Capítulo 6.	Conclusiones	55
Bibliografía		57
 PARTE II. PRESUPUESTO		
	Introducción y objetivo.....	64
	Presupuesto desglosado	64
	<i>Costes de Personal</i>	64
	<i>Costes de Software</i>	64
	<i>Costes de Hardware</i>	65
	Presupuesto total.....	65

Índice de Figuras

Figura 1. a) Tumores más frecuentemente diagnosticados en el mundo (estimación para 2018, ambos sexos), b) Estimación de la prevalencia de tumores en España para el 2018 (prevalencia a los 5 años) [SEOM, 2020].....	1
Figura 2. Plano sagital de la anatomía de la próstata [Harvard Health Blog, 2011]	2
Figura 3. AFT: tejido fibromuscular anterior; CZ: zona central; ED: conducto eyaculador; NVB: haz neurovascular; PUT: tejido periuretral; PZ: zona periférica; U: uretra; TZ: zona de transición; B: base; M: mediana; A: ápice (Copyrigh (Choi et al., 2007)).	2
Figura 4. Ejemplo de imágenes RM T2W	6
Figura 5. Ejemplo de imágenes RM DCE	7
Figura 6. Ejemplo de curva de concentración en imágenes de perfusión. En rojo se muestra la curva resultante con los datos en bruto y en blanco la curva ajustada tras aplicar modelos farmacocinéticos	7
Figura 7. Ejemplo de imagen RM DWI para diferentes valores de b.	8
Figura 8. Gráficos de distribución de la base de datos utilizada según el país de origen y el proveedor del equipo utilizado para la toma de imágenes DCE.....	17
Figura 9. Gráficos de distribución de la base de datos utilizada según el país de origen y el proveedor del equipo utilizado para la toma de imágenes DCE y T2W	19
Figura 10. Ejemplo de la AIF media obtenida con el algoritmo de selección automática	20
Figura 11. ROI dibujada manualmente en las arteria ilíacas.....	21
Figura 12. Ejemplo de AIF obtenida de forma manual. Se muestran los resultados de ambas arterias ilíacas y la media de ambas.	22

Figura 13. Comparación de tres casos con clasificación cualitativa: En c) se puede ver como el algoritmo ha seleccionado correctamente ambas arterias ilíacas, mientras que en b), selecciona las arterias y otras zonas, y en a) todo lo seleccionado está fuera de los vasos.	23
Figura 14. Curvas de la función de entrada arterial obtenidas con los métodos manual y automático. (a) Comparación AIF método manual y automático. (b) Comparación AIF método manual (global e individual) y automático.	25
Figura 15. Parámetros cuantitativos extraídos de la curva AIF. H = intensidad de pico, t_0 = tiempo de llegada del bolo, ttp = tiempo de pico, WHH = ancho a media altura del pico.	26
Figura 16. Ejemplo de gráficas de correlación (a) y Bland-Altman (b) con la representación de límites de concordancia (línea de puntos) [Bland et al, 1999]	26
Figura 17. Ejemplo de imágenes utilizadas en la clasificación: secuencia anatómica T2W y secuencias antes y después de registrar (alineadas y registradas)	29
Figura 18. Comparación de la aplicación del registro utilizando la segmentación de la próstata en dos casos.	30
Figura 19. Ejemplo de tablero de ajedrez entre secuencias T2W y secuencias T2W-DCE	31
Figura 20. Ejemplo de tableros de ajedrez para cada corte de las secuencias para un caso	33
Figura 21. Ejemplo de obtención de bounding box de la próstata en imágenes T2W ..	34
Figura 22. Ejemplo de un caso en el que el algoritmo actúa de forma correcta cuando no hay movimiento inicial	34
Figura 23. Ejemplo de un caso en el que el algoritmo actúa de forma incorrecta al introducir una deformación en la imagen.	34
Figura 24. Ejemplo de un caso en el que el algoritmo actúa correctamente cuando hay movimiento inicial.....	35

Figura 25. Gráficas de correlación y Bland-Altman para evaluar la similitud de la intensidad de pico entre los métodos automático y manual de selección de la AIF.....	38
Figura 26. Gráficas de correlación y Bland-Altman para evaluar la similitud del tiempo de pico entre los métodos automático y manual de selección de la AIF.....	39
Figura 27. Diferencias entre tres casos clasificados de forma diferente según la validación cualitativa.....	41
Figura 28. Diagrama de cajas y bigotes para los ratios	41
Figura 29. Ejemplo de introducción de una deformación en la imagen al aplicar el algoritmo.....	42
Figura 30. Representación de los parámetros cuantitativos para la evaluación interna según la clasificación cualitativa	47
Figura 31. Clasificación de los métodos de validación del registro	53

Índice de Tablas

Tabla 1. Especificaciones técnicas recomendadas por PI-RADS para secuencias DCE .16
Tabla 2. Especificaciones técnicas recomendadas por PI-RADS para secuencias T2W .18
Tabla 3. Esquema general del sistema de clasificación para la validación del registro ...28
Tabla 4. Resumen de resultados de la validación cualitativa37
Tabla 5. Coeficientes de similitud de Jaccard entre las dos clasificaciones37
Tabla 6. Resumen de resultados de la validación cualitativa revisada37
Tabla 7. Comparativa de los resultados en un conjunto de datos reducido para ambos métodos de registro43
Tabla 8. Resultados de la clasificación en la base de datos total44
Tabla 9. Clasificación final de los casos correctos e incorrectos del registro45
Tabla 10. Interpretación de la fuerza de relación basada en el valor r50
Tabla 11. Cuadro de precios Personal64
Tabla 12. Cuadro de precios Software65
Tabla 13. Cuadro de precios Hardware65
Tabla 14. Cálculo del presupuesto total del trabajo66

Lista de Acrónimos

AEMPS Agencia Española de Medicamentos y Productos Sanitarios

AIF *Arterial Input Function*

CaP Cáncer de Próstata

CV *Coefficient of Variation*

CZ *Central Zone*

DCE *Dynamic Contrast Enhanced*

DICOM *Digital Imaging and Communications in Medicine*

DWI *Diffusion-Weighted Imaging*

FDA *Food and Drug Administration*

IA Inteligencia Artificial

PCA *Principal Component Analysis*

PI-RADS *Prostate Imaging-Reporting and Data System*

PSA *Prostatic Specific Antigen*

PZ *Peripheral Zone*

RM Resonancia Magnética

RMmp Resonancia Magnética Multiparamétrica

ROI *Region of Interest*

RPC *Reproducibility Coefficient*

SD *Standard Deviation*

T2W *T2-Weighted images*

TZ *Transitional Zone*

PARTE I
MEMORIA

Capítulo 1. Introducción

1.1 Patología

El término cáncer hace referencia a un conjunto de enfermedades relacionadas entre sí y con características independientes, como las causas de aparición, su evolución o los tratamientos que requieren. Sin embargo, todas ellas presentan la capacidad de multiplicación y diseminación sin control de las células por el organismo [aecc, 2018].

Todas las células del organismo tienen una función determinada, que una vez esta es cumplida, o cuando las células han sido dañadas, son reemplazadas por otras nuevas, lo que permite el mantenimiento de la integridad y el correcto funcionamiento del organismo. Cuando se produce la alteración de los mecanismos de reproducción celular, se inicia una división descontrolada de las células, dando lugar, con el tiempo, a un tumor o nódulo. Además, si adquieren la facultad de invadir tejidos y de proliferar en otras partes del organismo, se denomina tumor maligno o cáncer [aecc, 2018].

Hoy en día, el cáncer constituye una de las principales causas de morbilidad y mortalidad del mundo, siendo 18,1 millones la cifra de casos nuevos en el mundo en el año 2018. Además, se estima que el número de casos aumentará en las dos próximas décadas hasta 29,5 millones al año [SEOM, 2020].

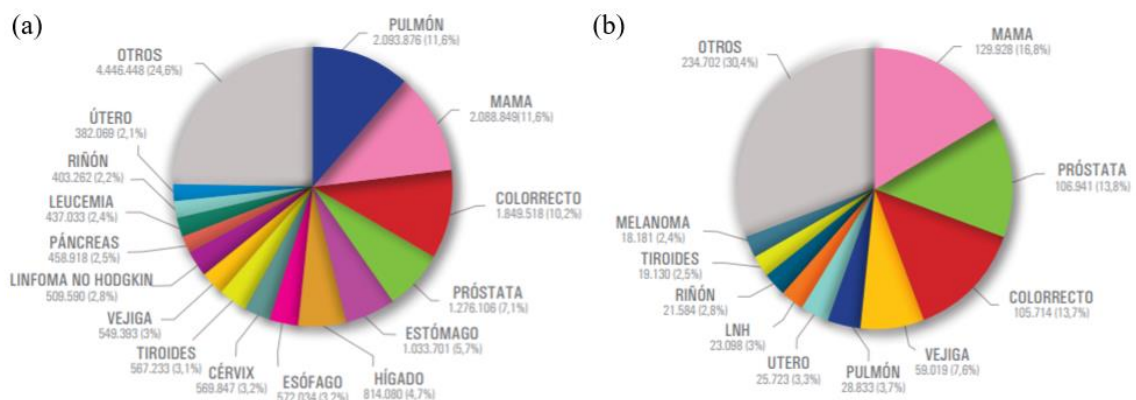


Figura 1. a) Tumores más frecuentemente diagnosticados en el mundo (estimación para 2018, ambos sexos), b) Estimación de la prevalencia de tumores en España para el 2018 (prevalencia a los 5 años) [SEOM, 2020]

Los tumores más frecuentemente diagnosticados en el año 2018 (Figura 1a) fueron los de pulmón, colonrectal, estómago, mama y próstata, siendo estos dos últimos, a su vez, los de mayor prevalencia, como se puede ver en la Figura 1b [SEOM, 2020]. La supervivencia de los pacientes está relacionada con la prevalencia, siendo esta mayor, a menor mortalidad. Sin embargo, también puede suponer una disminución de la calidad de vida de las personas afectadas, incrementando así la necesidad y mejora del diagnóstico precoz y de precisión.

1.1.1 Anatomía de la próstata

La próstata es una glándula exocrina del sistema reproductor masculino. Se encuentra situada debajo de la vejiga y alrededor de la uretra como se muestra en la Figura 2. Respecto a sus dimensiones, aproximadamente mide 3 cm de largo y 2.5 cm de ancho, y su peso es desde 7 g hasta 16 g para un adulto [Lemaitre, 2016]. El tamaño de la próstata incrementa durante su desarrollo físico en dos momentos, inicialmente durante la pubertad hasta alcanzar su tamaño normal, y posteriormente de nuevo, después de los 60 años, crecimiento conocido como hiperplasia prostática benigna [Parfait, 2010].

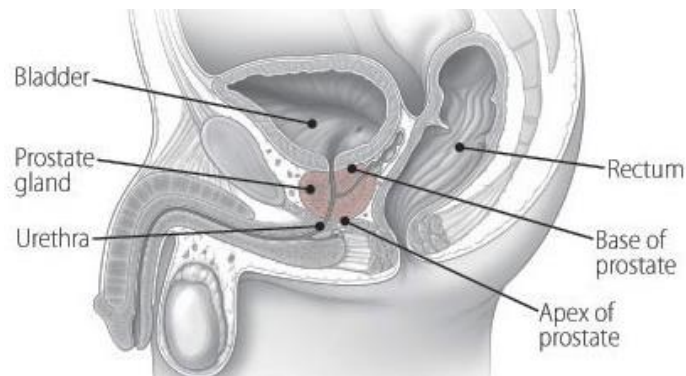


Figura 2. Plano sagital de la anatomía de la próstata [Harvard Health Blog, 2011]

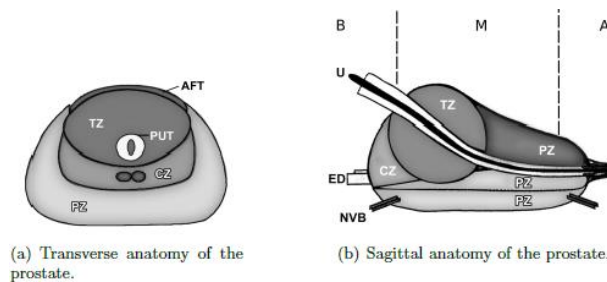


Figura 3. AFT: tejido fibromuscular anterior; CZ: zona central; ED: conducto eyaculador; NVB: haz neurovascular; PUT: tejido periuretral; PZ: zona periférica; U: uretra; TZ: zona de transición; B: base; M: mediana; A: ápice (Copyright (Choi et al., 2007)).

De forma general, se divide la próstata en tres zonas: periférica (PZ) la cual representa el 70% de la glándula, central (CZ) con el 20-25% y de transición (TZ), siendo esta última la más pequeña, refiriéndose al 5% de la próstata [Lemaitre, 2016]. Además, como se puede ver en la Figura 3b, existe una segunda clasificación por la que la glándula se divide en tres porciones longitudinales.

1.1.2 Cáncer de próstata

Hoy en día, el cáncer de próstata (CaP) es uno de los carcinomas malignos más prevalentes entre la población masculina, siendo en España, el segundo más frecuente y el tercero en mortalidad [aecc, 2019]. El número de casos nuevos en España supera los 35.000 casos, estimándose que, 1 de cada 9 hombres será diagnosticado de cáncer de próstata a lo largo de su vida [ACS, 2019]. Por tanto, al caracterizarse esta enfermedad por tener, tanto una incidencia como prevalencia elevada, esto afecta directamente a la necesidad de establecer un diagnóstico eficaz y precoz

El cáncer de próstata se caracteriza por evolucionar de dos formas, por un lado, están los tumores de crecimiento lento, constituyendo un 85% del total, y tumores de desarrollo rápido con tendencia de metástasis a otros órganos, principalmente los huesos. En el caso de darse esta última situación, al tratarse de una enfermedad incurable, afecta gravemente a la morbilidad y mortalidad del paciente [Lemaitre, 2016]. No solo se incrementa la necesidad del diagnóstico precoz, ya que distinguir entre estos dos tipos de tumores es fundamental para establecer el tratamiento a seguir, sino también, de establecer el riesgo de progresión de los tumores y valorar a cada paciente antes de tomar cualquier decisión terapéutica.

El CaP tiene mayor tendencia a aparecer en unas regiones que en otras de la próstata. Alrededor del 70-80% se origina en PZ, mientras que del 10-20% en TZ. Únicamente el 5% de los tumores comienzan en CZ, sin embargo, son estos los que aparecen de forma más agresiva y suelen derivar en metástasis [Lemaitre, 2016].

1.1.3 Diagnóstico en el cáncer de próstata

El diagnóstico precoz permite un tratamiento exitoso dado que el cáncer sigue aislado en la próstata, sin embargo, si el diagnóstico tiene lugar en etapas tardías, puede provocar el aumento de metástasis y un aumento de la morbilidad y la mortalidad [Turkbey y Choyke, 2012]. Existen diferentes pruebas para detectar el cáncer de próstata a tiempo, siendo las dos más importantes, el tacto rectal y la prueba del antígeno prostático específico (PSA). Al caracterizarse la evolución del CaP por ser bastante lenta, es importante evitar el sobrediagnóstico, ya que se trataría innecesariamente a muchos varones que nunca tendrían problemas reales con este tipo de cáncer [aecc, 2019].

El PSA es una proteína cuya determinación sirve como marcador en la prueba diagnóstica para establecer la sospecha de CaP. Niveles elevados de PSA no es determinante de cáncer, ya que puede deberse además a infecciones de próstata, irritaciones, exámenes rectales recientes, entre otros [Parfait, 2010]. Tras la realización de esta prueba y al alcanzarse los valores de PSA establecidos, se pasa a la realización de una biopsia prostática para llegar a un diagnóstico seguro que excluya o confirme la presencia de CaP.

Las biopsias son evaluadas a través de la escala de Gleason [Gleason, 1977], con la cual se puntúa la agresividad del tumor, clasificándose en poco agresivo (Puntuación de Gleason < 7) o muy agresivo (Puntuación de Gleason ≥ 7). Sin embargo, estos resultados no son los suficientemente precisos, ya que no detectan focos cancerígenos de menor tamaño o pueden confundir la agresividad [Lemaitre, 2016].

1.2 Imagen por Resonancia Magnética

Hoy en día, la imagen médica es considerada un recurso fundamental para el diagnóstico, tratamiento y pronóstico de numerosas patologías, sirviendo como apoyo en la toma de decisiones de los sanitarios, aportando información tanto cualitativa como cuantitativa, gracias al análisis y procesado de estas con herramientas de inteligencia artificial y técnicas de *Machine Learning*.

Desde las primeras técnicas de imagen diagnóstica debidas a la utilización de los Rayos X, el campo de la imagen médica, catalogado por un gran número de clínicos como el avance técnico con mayor impacto en la práctica clínica [Howell, 2016], ha sufrido una

enorme evolución desde estas primeras radiografías, apareciendo así nuevas técnicas con imágenes obtenidas por resonancia magnética y medicina nuclear, entre otras. De esta forma, se permite obtener información del cuerpo humano mediante métodos no invasivos.

Los avances técnicos en Imagen de Resonancia Magnética (IRM o RM) en los últimos años han provocado que ésta pase a considerarse la técnica principal de elección en situaciones de sospecha o diagnóstico de cáncer de próstata [Vilanova et al., 2010]. La RM es una tecnología de imágenes no invasiva con la que se obtienen imágenes tridimensionales detalladas, sin el uso de ningún tipo de radiación [NIH, 2013], lo que evita posibles futuros daños biológicos.

El equipo de RM está formado principalmente por un imán que genera un campo magnético de gran intensidad (0.5-7 T) [Laader et al., 2017] y obliga a los protones del cuerpo a alinearse con ese campo. Al aplicar un pulso de radiofrecuencia perpendicular a través del paciente, estos protones son estimulados y empiezan a girar fuera del equilibrio. Al cesar este pulso, los protones tardan un tiempo en volver a realinearse con el campo magnético inicial, siendo entonces cuando los sensores de RM son capaces de detectar la energía liberada. El tiempo que tarda en restablecerse la magnetización longitudinal hasta un 63% del valor inicial se denomina T1 y el tiempo correspondiente en la pérdida de la magnetización transversal se conoce como T2. Tanto el tiempo como la energía cambian dependiendo del entorno y la naturaleza química de las moléculas. Esta técnica ofrece imágenes cuya resolución y contraste dependen de la modalidad empleada [NIH, 2013].

Las imágenes de resonancia magnética se dividen en dos grupos, anatómicas y funcionales, obteniendo así, información de las diferentes estructuras anatómicas y de la fisiología del organismo, respectivamente. A continuación, se muestran las características de diferentes técnicas de ambos grupos, además de su utilidad y aplicación en el campo del cáncer de próstata, a través de su combinación resultante en la resonancia magnética multiparamétrica (RMmp) [Vilanova et al., 2010].

La RMmp es el estudio de RM que combina imágenes anatómicas potenciadas en T2 con imágenes funcionales y fisiológicas, como imágenes potenciadas en difusión (DWI) o imágenes de perfusión (DCE), y en ocasiones otro tipo de técnicas como la espectroscopía protónica in vivo [Weinreb, 2016]. Este tipo de tecnología ha incrementado la capacidad de detección del cáncer clínicamente significativo, lo que supone la reducción de la

mortalidad y el aumento de la confianza en las enfermedades latentes con menor probabilidad de causar morbilidad en la vida de un hombre [Weinreb, 2016], en el orden de reducir las biopsias y tratamientos innecesarios.

1.2.1 Imágenes T2W

Dentro del grupo de RM anatómica se encuentran las imágenes potenciadas en T2 (T2W, del inglés *T2-Weighted*), ampliamente utilizadas para el diagnóstico, detección, localización y estadificación del CaP [Lemaitre, 2016].

Este tipo de secuencias están determinadas por las variaciones de contraste debidas a las propiedades del tejido, es decir, dependiendo de la cantidad de agua libre existente en los tejidos, en la imagen aparecerán como regiones hiperintensas o hipotensas. Debido a que existen una gran cantidad de procesos patológicos en los que se producen incrementos de la cantidad de agua libre, este tipo de imagen es ampliamente utilizada en su detección, por ejemplo, los cánceres de próstata presentan en la zona periférica se ven como una lesión hipointensa, rodeada de tejido normal hiperintenso [Fernández, 2020].

Este tipo de imágenes dotan de contexto anatómico y estructural a las secuencias funcionales de RMmp, ya que es posible apreciar con detalle tanto la próstata, como las estructuras que la rodean, como las vesículas seminales o la vejiga. De esta forma, llegan a servir de mapa anatómico en el análisis de las secuencias de difusión y perfusión, en las cuales es más complicado distinguir las estructuras que aparecen.

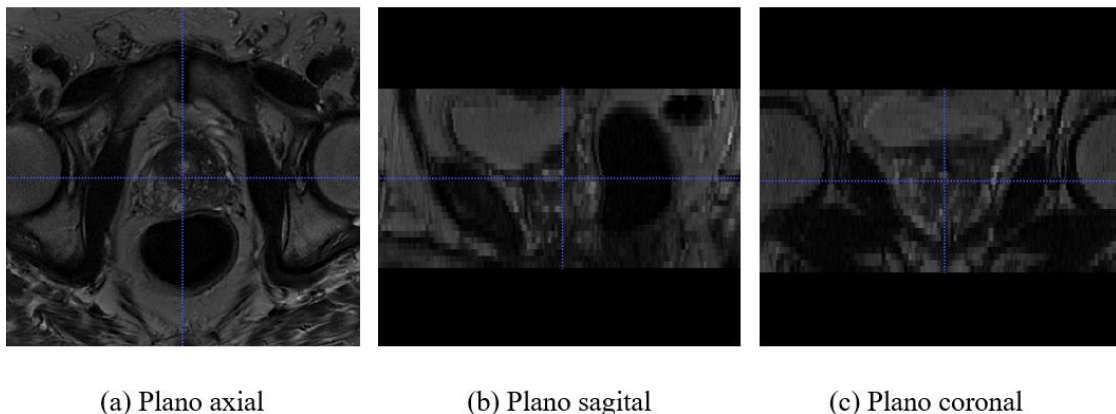


Figura 4. Ejemplo de imágenes RM T2W

1.2.2 Imágenes de perfusión DCE

Imagen de perfusión de contraste por resonancia magnética (DCE, del inglés, *Dynamic Contrast Enhanced*) es una técnica de imagen RM basada en el uso de un medio de contraste, generalmente derivados del gadolinio, para la evaluación y obtención de métricas cuantitativas de la fisiología vascular en tejidos con diferentes situaciones biológicas, centrándose en la caracterización de tumores y respuestas a tratamientos [Sanz Requena et al., 2015]. Para ello, se adquieren un conjunto de secuencias de T1 en diferentes instantes de tiempo, denominados como volúmenes o tiempos dinámicos, antes, durante y después de la inyección del agente de contraste, para observar la evolución del contraste en un área específica. En la Figura 5 se muestran diferentes dinámicos de DCE en un mismo paciente, y como se puede ver, existe un aumento de la intensidad en las arterias ilíacas con el paso del contraste, la cual va disminuyendo con el paso del tiempo.

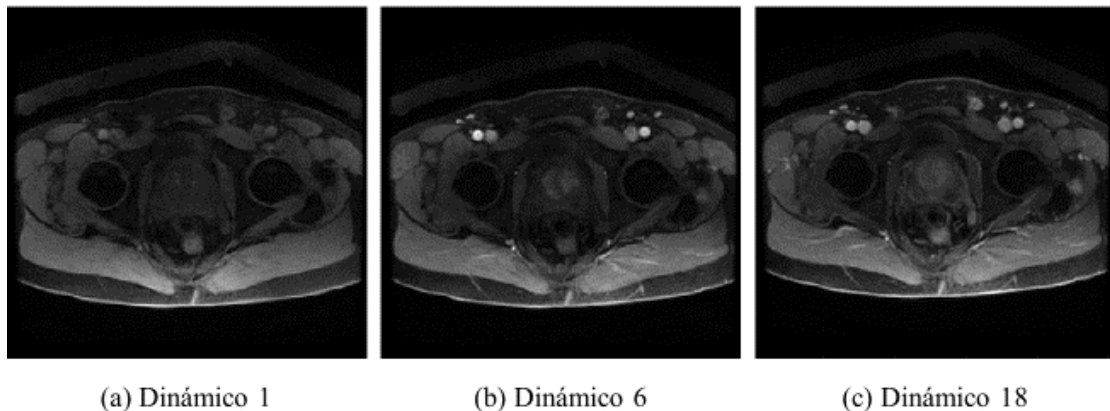


Figura 5. Ejemplo de imágenes RM DCE

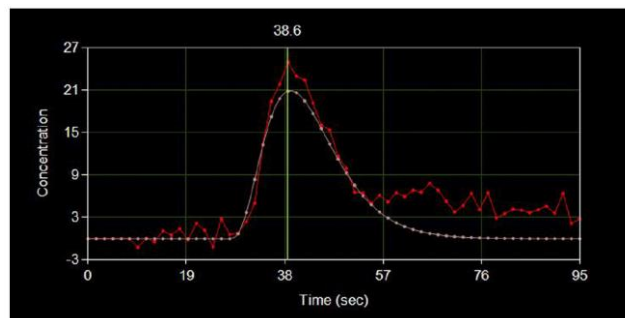


Figura 6. Ejemplo de curva de concentración en imágenes de perfusión. En rojo se muestra la curva resultante con los datos en bruto y en blanco la curva ajustada tras aplicar modelos farmacocinéticos

El análisis de las secuencias DCE suele realizarse de forma cualitativa a partir de la evaluación de la curva de realce de contraste de las imágenes, la cual representa la concentración del agente de contraste frente al tiempo. La curva se caracteriza por tener una rápida subida, que disminuye poco a poco a medida que el agente va desapareciendo (Figura 6). Sin embargo, con el objetivo de obtener un método estandarizado del análisis de perfusión se recurre a la obtención de métricas cuantitativas y representativas.

El análisis cuantitativo de la secuencia DCE permite obtener un conjunto de parámetros que miden las características de la microvascularización de un órgano o tejido, lo que lo convierte en una herramienta de gran interés para la detección de estados de angiogénesis y su modificación con la enfermedad y el tratamiento [Tofts et al., 1999]. Para ello se relaciona la concentración de contraste en el tumor y en un vaso relevante cercano a éste mediante modelado farmacocinético.

1.2.3 Imágenes de difusión DWI

Un segundo ejemplo de RM funcional son las imágenes potenciadas en difusión (DWI, del inglés *Diffusion-Weighted Imaging*), siendo en este caso el contraste debido a la variación del movimiento de las moléculas de agua en diferentes tejidos [Koh and Collins, 2007].

Basándose en el hecho de que la señal en las imágenes DWI está inversamente correlacionada con el grado de movimiento aleatorio del agua, se utilizan los gradientes para codificar la ubicación espacial de los núcleos [Huisman, 2003]. Se aplica un gradiente en una dirección y un segundo gradiente en la dirección contraria, cancelando el primero. De esta forma, las moléculas de agua inmóviles dependerán del gradiente inicial volviendo a estar en fase, mientras que las moléculas en movimiento permanecerán desfasadas, resultando en una señal de menor intensidad.

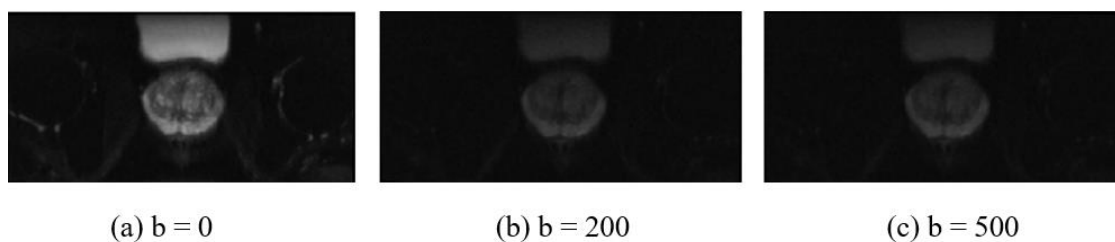


Figura 7. Ejemplo de imagen RM DWI para diferentes valores de b .

Un parámetro característico de este tipo de adquisiciones es el parámetro *b-value*, del cual dependen la sensibilidad de difusión, la magnitud de los pulsos, su separación, así como el intervalo entre un gradiente y otro [Huisman, 2003]. La señal de RM decae a medida que este parámetro incrementa, correspondiendo la imagen adquirida a la secuencia T2 cuando $b = 0$. Como se puede apreciar en la Figura 7, a medida que aumenta el *b-value*, va disminuyendo la intensidad de la señal.

En tejidos sanos, las moléculas de agua se caracterizan por moverse de forma aleatoria con cierta libertad, sin embargo, esto cambia en tejidos patológicos debido a procesos como la angiogénesis, hinchazón celular, fibrosis o necrosis, lo que limita este movimiento y permite que se visualice en las imágenes de difusión al aumentar el *b-value* y caer la señal de intensidad en estas regiones.

1.3 Antecedentes

Como se ha comentado anteriormente, actualmente el diagnóstico y la detección del CaP incluye tanto el tacto rectal como la medición del antígeno PSA, sin embargo, estos métodos aumentan considerablemente las ratios de incidencia en la población masculina y del sobretreatmento de la enfermedad. Por ello, se comenzaron a utilizar otro tipo de técnicas relacionadas con la imagenología, con el objetivo de poder conseguir una mayor precisión en esta tarea.

La RM lleva utilizándose aproximadamente desde los años 80 en la evaluación de forma no invasiva de la próstata, basándose inicialmente en el uso de las secuencias de pulsos T1 y T2 para clasificar de forma local y regional los estados del tumor en los pacientes [Weinreb, 2019]. El desarrollo de la resonancia magnética multiparamétrica ha permitido introducir nuevos avances en el diagnóstico, aumentando la capacidad de distinguir entre tejido benigno patológico y cáncer clínicamente significativo.

El principal problema surge con la excesiva variación existente en la interpretación y en los informes de los exámenes RM de próstata, por lo que, con el objetivo de estandarizar este proceso se publicó en 2012 la primera versión del “Sistema de información y datos en imagen de próstata” (PI-RADS, del inglés “*The Prostate Imaging – Reporting and Data System*”), actualizado posteriormente a la versión 2 y versión 2.1 en el año 2019. En este documento se recogen las directrices básicas y necesarias para realizar una

evaluación completa de cáncer de próstata a través del análisis de las imágenes RMmp. Por tanto, el principal objetivo de PI-RADS es estandarizar y disminuir las posibles variaciones en el análisis, ofreciendo mejoras en la detección y caracterización de este tipo de lesiones en la próstata. Además, evalúa la probabilidad de que el paciente presente CaP en una escala del 1 al 5, siendo 1 una baja probabilidad de que el paciente presente CaP y 5 muy alta [Steiger and Thoeny, 2016].

De esta forma, el protocolo PI-RADS facilita el uso de algoritmos de procesamiento de imagen de forma universal en secuencias T2W, DCE y DWI, con el propósito principal de obtener métricas y resultados cuantitativos que sirvan de apoyo en la toma de decisiones médicas. Existen diferentes técnicas de procesado dependiendo del tipo de imagen y de la información que se quiere obtener, por ejemplo, la aplicación de filtros para suavizar, corregir o resaltar alguna zona concreta o la segmentación de estructuras y lesiones anatómicas. Algunos de los procedimientos básicos en la RMmp, y en los que se centra este trabajo son la selección de la función de entrada arterial (AIF) en secuencias DCE y el registro de imágenes.

Una vez se han desarrollado los algoritmos de preprocesado, es necesario realizar una evaluación de su funcionamiento y de los resultados obtenidos, comprobando así la bondad del método. Este trabajo se centra en realizar una validación clínica de los algoritmos de procesado comentados anteriormente, basándose en la bibliografía existente sobre el tema y siguiendo las directrices básicas de validación de productos sanitarios establecido por los organismos y agencias nacionales e internaciones, como la Agencia Española de Medicamentos y Productos Sanitarios (AEMPS) o la Administración de Medicamentos y Alimentos (FDA, del inglés, *Food and Drug Administration*).

1.3.1 Selección de la función de entrada arterial

Como se ha introducido en el apartado *1.2.2 Imágenes de perfusión DCE*, la función de entrada arterial (AIF, del inglés, *Arterial Input Function*) hace referencia a la concentración del agente de contraste en los vasos sanguíneos en función del tiempo, suponiendo su medida y aplicación al modelo farmacocinético de Tofts [Tofts et al.,

1999], la descripción de términos fisiológicos como permeabilidad vascular, volumen extracelular y flujo sanguíneo [Rijpkema et al., 2001].

Existen numerosos modelos farmacocinéticos adaptados a distintos tejidos, siendo el más utilizado para el CaP el modelo de Tofts [Tofts et al., 1999], por ser el más aceptado de forma habitual en el análisis del tejido de próstata. En este modelo se simplifica el contenido de cada vóxel mediante una representación muy elemental del tejido basada en compartimentos. Una de las suposiciones fundamentales del mismo implica que en cada vóxel de la imagen, el tejido de interés es alimentado por una arteriola que en el modelo representa la función de entrada arterial (AIF, del inglés, *Arterial Input Function*), necesaria para la cuantificación de los parámetros del modelo.

Actualmente los modelos farmacocinéticos están en vías de investigación y validación clínica, demostrándose en diferentes estudios que, a pesar de ser un método de análisis reproducible e independiente del paciente, presenta también una alta variabilidad [García et al., 2008]. El hecho de que no exista un método de referencia para la obtención de la función arterial supone uno de los factores más limitantes en la reproducibilidad del modelo, ya que es posible cometer errores en su cálculo, afectando de forma directa a los parámetros cuantitativos resultantes.

Existen diferentes técnicas para medir la AIF, diferenciándose en métodos de selección manual y métodos automáticos [Rijpkema et al., 2001], estos últimos basados en la aplicación de algoritmos de detección de funciones individualizadas para cada paciente. En algunas ocasiones, se utiliza una AIF de referencia para una población [Sanz Requena et al., 2015], ya que esta función no varía con patologías como el cáncer. Sin embargo, esta generalización no tiene en cuenta la hemodinámica individual, además de que pierde la capacidad de caracterizar individualmente a los pacientes.

El método más recomendado para la extracción de la AIF es el automático, existiendo diferentes formas para su cálculo, sin embargo, la selección manual del paso del contraste al medio sanguíneo es recomendada en aquellos casos en los que no es posible obtener la AIF debido a la anatomía, movimientos del paciente y otros artefactos propios de la imagen [Sanz Requena et al., 2015].

Con el objetivo de extraer de forma automática esta función sin variabilidad, el equipo de QUIBIM propuso un método basado en el análisis de componentes principales (PCA)

para extraer con la mínima variabilidad aquellos vóxeles con un comportamiento arterial puro de una imagen volumétrica de DCE de un grupo de pacientes. PCA es una técnica estadística basada en aprendizaje no supervisado aplicada para el análisis exploratorio de un conjunto con gran cantidad de datos, siendo los objetivos claves, reducir la dimensionalidad de las variables o información redundante y visualizar los datos [Gil, 2018]. Este método consiste en la extracción automática de la AIF en cada caso concreto aplicando PCA al volumen DCE y encontrando la máxima correlación entre las componentes principales con una AIF de referencia poblacional [Sanz Requena et al., 2015].

1.3.2 Registro inter secuencia de imágenes

Una etapa fundamental en el análisis de imágenes médicas es el registro de imágenes, con la que se pretende encontrar una relación espacial entre dos o más imágenes, alineándolas de modo que la posición real o anatómica de una coincida con la de la otra. Para ello, una de las imágenes es considerada la móvil, y la otra, es la fija, siendo el registro la transformación espacial de la imagen móvil sobre la imagen fija.

Según las secuencias utilizadas en el protocolo PI-RADS [Weinreb et al., 2016], para el análisis de imágenes RM de próstata, únicamente es necesario dos tipos de registro, entre imágenes de la misma modalidad adquiridos en distintos instantes de tiempo (registro intra secuencia) y entre imágenes de diferentes modalidades (registro inter secuencia). Ya que en este trabajo nos hemos centrado en el segundo tipo de registros comentado, se continuará con la explicación de este.

El registro inter secuencia es el llevado a cabo entre secuencias anatómicas, T2W, y secuencias funcionales, DCE y DWI; y consiste en adaptar características dimensionales y corregir movimientos o deformaciones entre secuencias.

A pesar de los numerables avances en las distintas técnicas de imagen médica, no existe un procedimiento unificado o una teoría general para el registro de dichas imágenes. Actualmente se emplean distintas transformaciones, clasificadas en lineales y no lineales. En las primeras, se encuentran la transformación rígida, compuesta por rotaciones y translaciones; y la afín, añadiéndole cizallamiento y reescalado a la anterior. En el

segundo grupo de transformaciones se encuentra la elástica, caracterizada por la aplicación de un campo de deformación [Pascau et al., 2002].

En esta ocasión, el algoritmo analizado es el registro entre secuencias T2W y secuencias DCE, para lo cual, el equipo de QUIBIM ha desarrollado uno propio basado en un registro lineal. Según estudios anteriores [Vos et., 2010] [Patel et al., 2011] [Chapellow et al., 2011], la transformada óptima entre estas dos secuencias en la región de la próstata es la transformada afín, utilizando además como métrica de similitud la información mutua. De esta forma, el algoritmo que se va a validar en este trabajo corresponde a una transformación afín, la cual, una vez los dinámicos de la DCE están registrados entre sí, es suficiente registrar uno con la T2W y aplicar la misma transformación al resto [Fernández, 2019].

1.4 Hipótesis y Objetivos

El objetivo principal de este trabajo es llevar a cabo un proceso de validación clínica para dos algoritmos de preprocesado de imágenes de resonancia magnética multiparamétrica aplicado a cáncer de próstata.

Los algoritmos en cuestión se refieren a la selección de la función de entrada arterial en secuencias DCE y el registro inter secuencia entre imágenes anatómicas T2W y funcionales DCE, estando los dos integrados dentro del flujo de trabajo de qp-Prostate®, solución propuesta por QUIBIM que consiste en un software de procesado, análisis y visualización de imágenes de próstata con uso por parte de usuarios capacitados o profesionales de la salud. Ambos algoritmos son una etapa fundamental en el procesado de imágenes, teniendo una implicación directa en el diagnóstico final.

Puesto que se trata de un producto sanitario que saldrá al mercado y cuya ejecución se realiza de forma automática, es necesario proveer junto al producto una serie de informes de validación en los que se demuestre de forma empírica sobre un conjunto de datos representativo de la práctica clínica, el correcto funcionamiento de todos los procesos que tienen lugar. Además, también resulta interesante establecer métricas de control de calidad del propio algoritmo, que, sin llegar a considerarse de validación clínica propiamente dicha, permitan controlar aquellos casos en los que la selección de la AIF o el registro no sean correctos, sirviendo de aviso para su corrección antes de llegar a un

diagnóstico final incorrecto. Este segundo objetivo permite complementar el proceso de validación, y de la misma forma, el objetivo principal de este trabajo.

Como se ha comentado anteriormente, según el protocolo PI-RADS [Weinreb, 2019], no existe una metodología unificada para el análisis de imágenes RMmp, por lo tanto, tampoco existe ningún procedimiento de validación clínica de referencia. Con este trabajo no solo se pretende validar una parte del procesado llevado a cabo por qp-Prostate, sino llevar a cabo una revisión bibliográfica de diferentes técnicas llevadas a cabo en estudios anteriores similares, que permita encontrar el mejor método de evaluación, y quizás en proyectos futuros constituir una metodología de referencia.

Para cada algoritmo de análisis se ha planteado un proceso de validación independiente, a partir de la revisión de estudios sobre la selección de la función de entrada arterial en imágenes de perfusión en CaP [Rijpkema et al., 2001] [García et al., 2008] [Sanz-Requena et al., 2015]; y de la consecuente revisión del Trabajo Final de Máster de Matías Fernández Patón [Fernández, 2019] en el cuál está planteado el diseño del registro intersecuencia utilizado, junto con diferentes estudios sobre las validaciones existentes [Zhe, 2012] [Razlighi, 2013].

Por tanto, el propósito principal del presente trabajo es el desarrollo de dos metodologías de validación clínica completas e independientes de dos de los algoritmos de preprocesado involucrados en el análisis de imágenes de resonancia magnética multiparamétrica en cáncer de próstata, a través de su evaluación cuantitativa y cualitativa, obteniéndose métricas y medidas estadísticas que evalúen el buen funcionamiento en un conjunto de pacientes.

Capítulo 2. Materiales

2.1 Base de Datos Global

Para la validación clínica global de todos los algoritmos que componen qp-Prostate®, se ha recopilado una base de datos de un total de 333 casos de imágenes de RM en próstata. Estos casos provienen de diferentes instituciones, centros de investigación y hospitales: Hospital Universitario y Politécnico de La Fe (Valencia), DASA (Brasil), Centro Uruguayo de Imagenología Molecular (Uruguay) y VOXEL (Polonia).

Todos estos casos poseen características y especificaciones técnicas diferentes, lo que permite obtener un conjunto de datos variado y adecuado para evaluar los diferentes métodos. Sin embargo, no todas las imágenes son válidas para la validación en este trabajo, pues es necesario que cumplan todos los requisitos establecidos en el PI-RADS según las secuencias utilizadas, por lo que, antes de comenzar el análisis, se ha realizado un estudio de la base de datos, para seleccionar todos aquellos casos útiles y que constituyan una población representativa para la validación clínica. Además, no se recogieron todos los casos a la vez, es decir, no todos los casos estaban disponibles para su uso en cada una de las validaciones, aun así, se han obtenido muestras de casos variadas para cada análisis con los datos disponibles en el momento.

2.1.1 Base de datos para la validación de la AIF

Con el objetivo de llevar una validación clínica del algoritmo lo más adecuada posible, la base de datos utilizada debe ser elegida y preparada cuidadosamente para representar la variabilidad dentro del dominio del escenario clínico, ya que las propiedades de las imágenes de RM se ven afectadas por factores como el origen, los proveedores, las instituciones, las máquinas de adquisición y parámetros técnicos como la resolución temporal, tamaño de píxel y el campo magnético, entre otros, recogido en la Tabla 1.

La base de datos final para este caso está formada por 242 secuencias de DCE de 242 pacientes diferentes con distintas propiedades de adquisición para poder evitar la

limitación comentada. Todos los casos han sido recogidos, por colaboración de QUIBIM con diferentes instituciones de todo el mundo, desde abril del 2017 hasta abril de 2020.

Según el PI-RADS [Weinreb, 2019], existen diversas especificaciones técnicas que las imágenes DCE deben cumplir para poder ser utilizadas en la validación clínica. En la Tabla 1 se muestran las características más relevantes que se han tenido en cuenta para la selección de los casos:

Tabla 1. Especificaciones técnicas recomendadas por PI-RADS para secuencias DCE

PARÁMETROS TÉCNICOS SECUENCIA DCE	
SECUENCIA	Transverse Dynamic Contrast-Enhanced (DCE) MR perfusion sequence
CAMPO MAGNÉTICO	1.5 – 3 T
RATIO OBSERVACIÓN	[2.1 – 19.4] minutes
RESOLUCIÓN TEMPORAL	[3 – 15] seconds
TAMAÑO PÍXEL	[0.5 – 1.56] mm
MATRIZ ADQUISICIÓN	[150 – 320] x [120-224] mm
ESPACIO ENTRE PÍXELES	[1.5 – 4] mm
GAP	[-2 – 0] mm
GROSOR CORTE	[3 – 4] mm
TR	[2.77 – 6.374] ms
TE	[1.04 – 2.864] ms
ÁNGULO DE GIRO	[10 – 40] °

Las instituciones que han colaborado con QUIBIM para la recogida de datos se muestran a continuación, además de la distribución en la Figura 8a. Todos los radiólogos y clínicos colaboradores han distribuido estos casos de forma anonimizada.

- Hospital Universitario y Politécnico La Fe (Valencia, España).
- Centro Uruguayo de Imagenología Molecular CUDIM (Montevideo, Uruguay).
- DASA (Brasil).
- Stony Brook Medicine Hospital (Nueva York, Estados Unidos).

Si se dividen los casos utilizados según los proveedores del equipo de imagen de RM, se obtiene la siguiente distribución (Figura 8b):

- 100 casos de General Electric, con modelos: DISCOVERY MR750w, Signa HDxt, Signa Pioneer y Optima MR360.
- 81 casos de Philips, con el modelo Achieva.

- 61 casos de Siemens Healthcare, con modelos: MAGNETOM Verio y MAGNETOM Prisma_fit.

Según el campo magnético, la distribución de los casos con 1,5T y 3T es de 2% y de 98%, respectivamente. El número de casos de 1,5T es significativamente menor porque, de manera general, la mayoría de los hospitales y clínicas privadas utilizan equipos de 3T. Sin embargo, aunque el porcentaje sea menor, se consideró importante incluir todos los casos posibles y variados en la validación. En la base de datos global sí existe un número mayor de casos con 1,5T, como se podrá ver en el apartado 2.1.1 *Base de datos de la validación del registro de imágenes T2W-DCE*, sin embargo, no pudieron ser utilizados en este caso, ya sea porque no cumplían el resto de los parámetros técnicos definidos, o porque fueron recogidos después de este análisis.

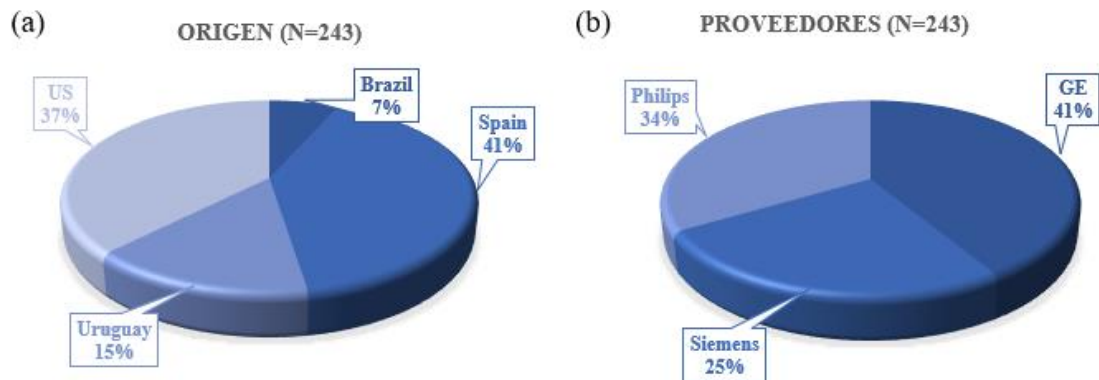


Figura 8. Gráficos de distribución de la base de datos utilizada según el país de origen y el proveedor del equipo utilizado para la toma de imágenes DCE

2.1.1 Base de datos de la validación del registro de imágenes T2W-DCE

Al igual que en el apartado anterior, con el objetivo de elaborar una base de datos específica para la validación del algoritmo de registro inter secuencia entre las imágenes T2W y las imágenes de perfusión DCE, se han revisado las indicaciones técnicas que sugiere el protocolo PI-RADS [Weinreb, 2019] en el conjunto total de casos recogidos por QUIBIM, y se han sido seleccionado 112 de los 333 casos. Los criterios seguidos para ello son los mismos que los planteados en el apartado anterior, se han revisado los parámetros técnicos de las secuencias DCE, mostradas en la Tabla 1, y los parámetros de las secuencias T2W, recogidos en la Tabla 2. Además, con el objetivo de obtener una

muestra variada, también se han tenido en cuenta factores como el origen, los proveedores, las instituciones, las máquinas de adquisición y el campo magnético empleado.

Tabla 2. Especificaciones técnicas recomendadas por PI-RADS para secuencias T2W

PARÁMETROS TÉCNICOS SECUENCIAS T2W	
SECUENCIA	Fast-spin echo (FSE) / Turbo-spin-echo (TSE)
CAMPO MAGNÉTICO	1.5-3 T
TAMAÑO PÍXEL	[0.28-0.94] mm
MATRIZ ADQUISICIÓN	[220-512] x [160x-512] mm *376x54
ESPACIO ENTRE PÍXELES	[2.1 – 6.5] mm
GAP	[0-2] mm
GROSOR CORTE	[2.1-6] mm
TR	[2020-13634] ms
TE	[80-170] ms
ÁNGULO DE GIRO	[90-160] °

El conjunto de imágenes seleccionado proviene de dos instituciones, mostrándose a continuación y su distribución para este caso en la Figura 8a. Además, coincide que todos los casos procedentes de Estados Unidos han sido obtenidos con una intensidad de campo igual a 3T, y las imágenes de Polonia, igual a 1,5T, obteniendo así la misma distribución mostrada en la Figura 9a.

- Stony Brook Medicine Hospital (Nueva York, Estados Unidos).
- VOXEL S.A. (Polonia)

En cuanto a los proveedores del equipo de imagen de RM, ambas instituciones utilizan los siguientes y distribuyéndose de la siguiente manera:

- 71 casos de GE con modelos: MAGNETOM Prisma_fit 3T, Signa HDxt (1.5T and 3T) and Signa Pioneer 3T.
- 41 casos de Siemens Healthcare y con el modelo MAGNETOM Prisma_fit 3T.

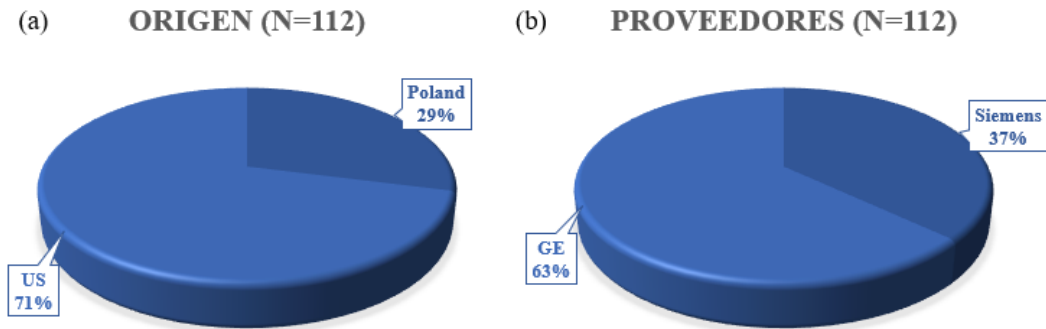


Figura 9 Gráficos de distribución de la base de datos utilizada según el país de origen y el proveedor del equipo utilizado para la toma de imágenes DCE y T2W

2.2 Software y formatos de imagen

Para la evaluación de ambos algoritmos de preprocesado se ha utilizado el *software* Matlab®, con el cuál, a partir del uso de funciones propias de la plataforma y otras especificadas en capítulos posteriores, ha sido posible la visualización y análisis de todas las imágenes utilizadas para este trabajo.

En cuanto a los formatos de imagen, el más utilizado y reconocido para la adquisición y almacenamiento es el formato DICOM (*Digital Imaging and Communication On Medicine*), caracterizado por estar compuesto por una cabecera de metadatos con información de la adquisición, datos del paciente, del dispositivo y del hospital, entre otros; y por otra parte con los datos propios de la imagen en forma de cadena de valores codificados que contienen la información de los píxeles que componen la imagen.

Capítulo 3. Metodología

En este apartado se va a explicar detalladamente todos los pasos que se han ido llevando a cabo hasta el diseño y planteamiento definitivo de una metodología de validación clínica. Como se ha comentado anteriormente, cada algoritmo de análisis requiere una evaluación diferente e independiente, ajustada a las necesidades y requerimientos de cada situación. Por ello, y para evitar posibles confusiones, tanto en este capítulo, como en los capítulos correspondientes a los resultados obtenidos y a su correspondiente discusión, el trabajo se ha dividido en dos partes diferenciadas, según sea la validación clínica del algoritmo de selección de la función de entrada arterial AIF en imágenes de perfusión DCE y la validación clínica del registro inter secuencia entre imágenes anatómicas T2W y funcionales DCE.

3.1 Validación de la selección de AIF

Para comprobar el correcto funcionamiento del algoritmo de selección automática de la función de entrada arterial, se van a realizar dos tipos de validaciones, cuantitativa y cualitativa, en las cuales, se compararán los resultados obtenidos tanto con el método manual como con el automático, obteniéndose métricas de evaluación que permitan conocer la especificidad de los resultados.

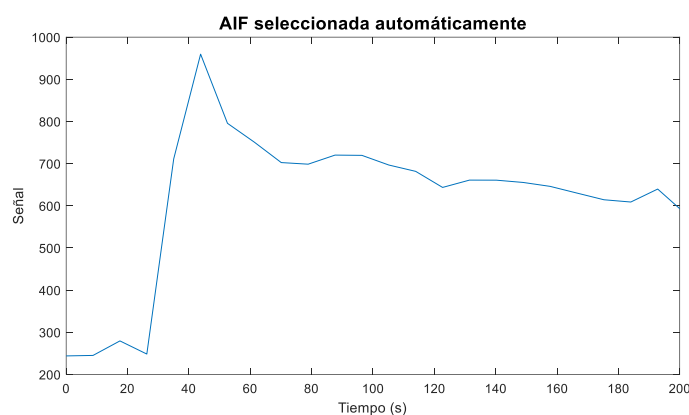


Figura 10. Ejemplo de la AIF media obtenida con el algoritmo de selección automática

Por un lado, los 242 casos fueron analizados con el algoritmo de selección automática en cuestión. Aunque es posible que el algoritmo señale píxeles en ambas arterias ilíacas (derecha e izquierda), además de otros grandes vasos, la AIF media (Figura 10) es obtenida de todos los píxeles seleccionados. Esta AIF media es la que se utiliza para comparar con la AIF manual en cada caso.

Para el método manual de obtención de la AIF, es necesario establecer unos criterios de selección en primer lugar, con el objetivo de obtener un método reproducible que pueda ser llevado a cabo por cualquier técnico especializado en imagen médica o clínico. En estudios anteriores [Sanz Requena et al., 2015], se ha segmentado únicamente la arteria ilíaca externa derecha, sin embargo, en este estudio se ha considerado más apropiado dibujar regiones de interés (ROI) tanto en las arterias ilíacas derecha e izquierda, obteniendo así una AIF media, igual que con el algoritmo. Las ROI han sido dibujadas dentro de cada arteria ilíaca en el corte central de cada secuencia. Es muy importante que la ROI no contuviese los bordes de la arteria, por tanto, era preferible obtener una segmentación de menor tamaño, siempre que ésta solo estuviese en el interior del vaso. En la Figura 11 se muestra un ejemplo de cómo se han dibujado las ROI sobre las arterias ilíacas en un paciente:

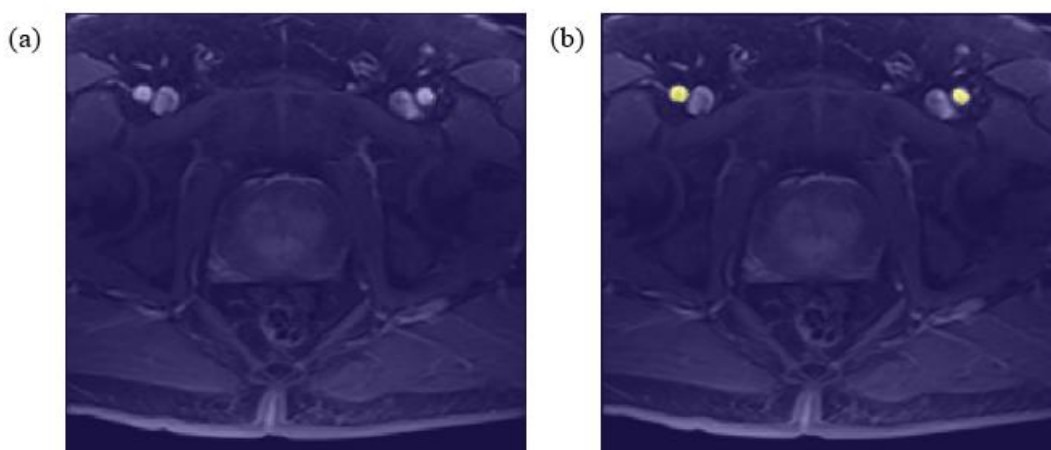


Figura 11. ROI dibujada manualmente en las arteria ilíacas.

Una vez se han segmentado las arterias ilíacas, si se aplica la máscara resultante sobre la imagen original, podemos conocer el valor correspondiente de concentración de señal para cada píxel del área de interés. Por lo que, llevando a cabo esto para cada volumen dinámico, se obtiene para cada tiempo su valor, representado en forma de curva, como se puede ver en la Figura 12.

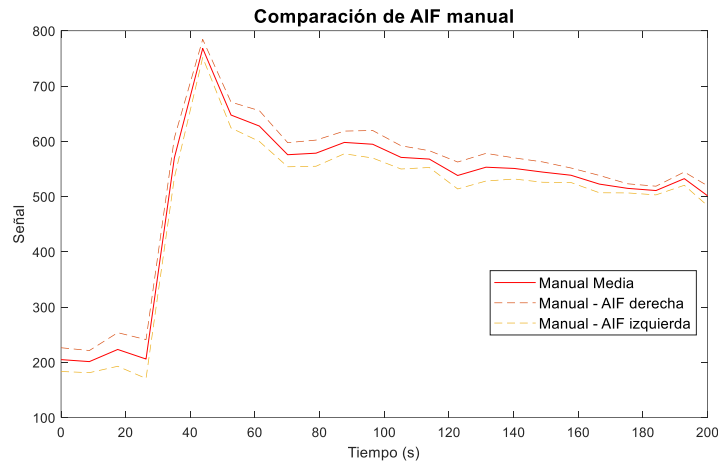


Figura 12. Ejemplo de AIF obtenida de forma manual. Se muestran los resultados de ambas arterias ilíacas y la media de ambas.

La evaluación de la selección automática de la AIF se realizó de forma cualitativa y cuantitativa, obteniendo así una comparación desde dos perspectivas diferentes. En la valoración cualitativa, el objetivo era revisar de forma visual la segmentación automática para comprobar si las arterias ilíacas y los grandes vasos habían sido correctamente seleccionados. Por otro lado, en la validación cuantitativa, se centró en evaluar las curvas AIF obtenidas con ambos métodos.

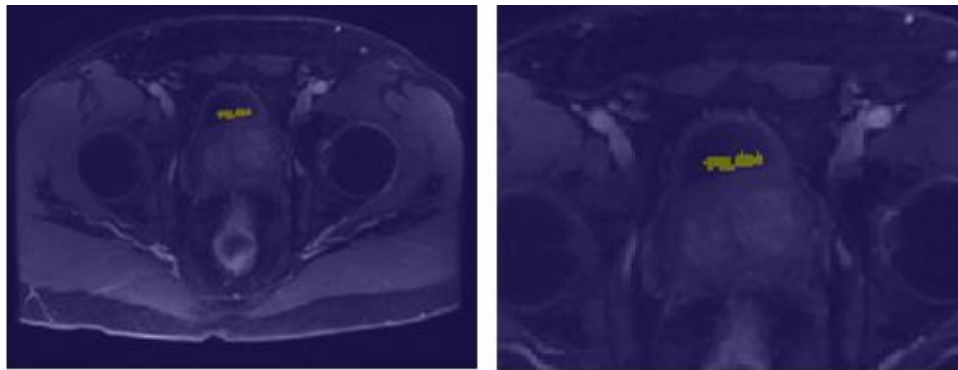
3.1.1 Validación cualitativa AIF

En primer lugar, se ha comprobado la detección automática de la AIF por parte del algoritmo a través de una clasificación visual en forma de escala numérica para determinar si la ROI resultante corresponde a las arterias ilíacas.

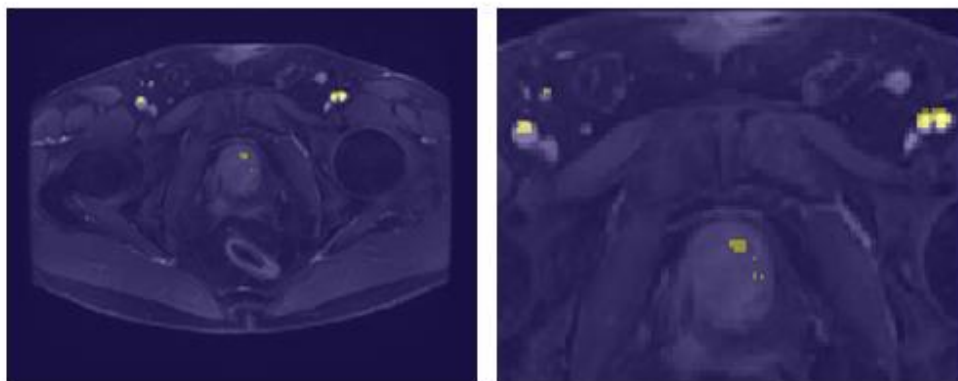
Se ha decidido dividir esta clasificación según el criterio:

- Valor 0: Más de la mitad de los vóxeles están fuera de las arterias ilíacas y de otros vasos grandes.
- Valor 1: Casi todos los vóxeles están dentro de las arterias ilíacas y vasos grandes, siendo no representativo lo que queda fuera.
- Valor 2: Todos los vóxeles están dentro de las arterias ilíacas y vasos grandes. Además, la curva de la función arterial resultante corresponde con la curva de referencia esperada.

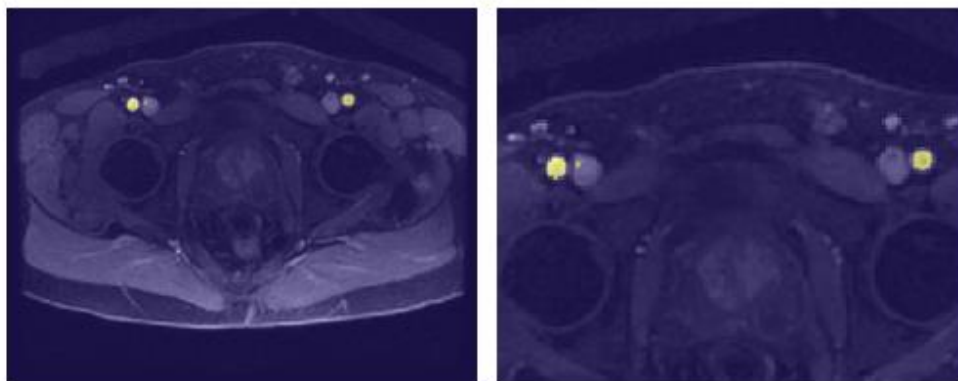
A continuación, se muestra una serie de ejemplos donde se puede ver claramente las diferencias entre los tres grupos, entendiéndose mejor los criterios señalados:



(a) Valor 0



(b) Valor 1



(c) Valor 2

Figura 13. Comparación de tres casos con clasificación cualitativa: En c) se puede ver como el algoritmo ha seleccionado correctamente ambas arterias ilíacas, mientras que en b), selecciona las arterias y otras zonas, y en a) todo lo seleccionado está fuera de los vasos.

Cabe destacar que este tipo de análisis tiene la desventaja que, aun teniendo un criterio de clasificación establecido, no en todos los casos es sencillo diferenciar unas clases de otras. Se han dado casos en los que la máscara tomaba ambas arterias ilíacas, pero luego el algoritmo solo marcaba una de ellas (lo que en primer momento podría clasificarse como 1), pero de forma completamente correcta, por lo que la curva AIF resultante también era correcta, teniendo así una puntuación igual a 2. Tras analizar estos casos de forma más específica, se ha llegado a la conclusión de que, si solo se marca una de las arterias, esto se debe a la aparición de artefactos en uno de los dos lados, la inhomogeneidad en la imagen (gradientes de intensidad a lo largo de ella) o porque sistemáticamente, una de las arterias tiene una llegada de contraste menor y el algoritmo selecciona aquellas regiones con una intensidad máxima o superior a un umbral.

Para evitar la subjetividad y variabilidad del método, la evaluación cualitativa fue realizada una segunda vez, de forma completamente independiente a la primera, por un ingeniero de la empresa QUIBIM.

Por último, se usó el coeficiente de Jaccard, como métrica estadística para medir la similitud entre ambas clasificaciones. Este coeficiente (Ecuación 1) se define como el tamaño de la intersección dividido por el tamaño de la unión de los conjuntos de muestras, y puede ir de 1 a 0. Cuanto más cerca de 1, más similares son los dos conjuntos de muestras [Glen, 2016]. A corresponderá al conjunto de datos de la primera evaluación, y B, a la segunda.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (1)$$

3.1.2 Validación cuantitativa AIF

Con los resultados obtenidos del método automático y del manual, una vez realizada una primera validación cualitativa, el siguiente paso a seguir es la validación de forma cuantitativa, con la cual se pretende obtener parámetros que permitan medir con mayor precisión el correcto funcionamiento de la técnica propuesta. Para ello, se ha trabajado sobre diferentes métricas obtenidas de la comparación entre las curvas AIF, de las cuales se puede ver un ejemplo en la Figura 14.

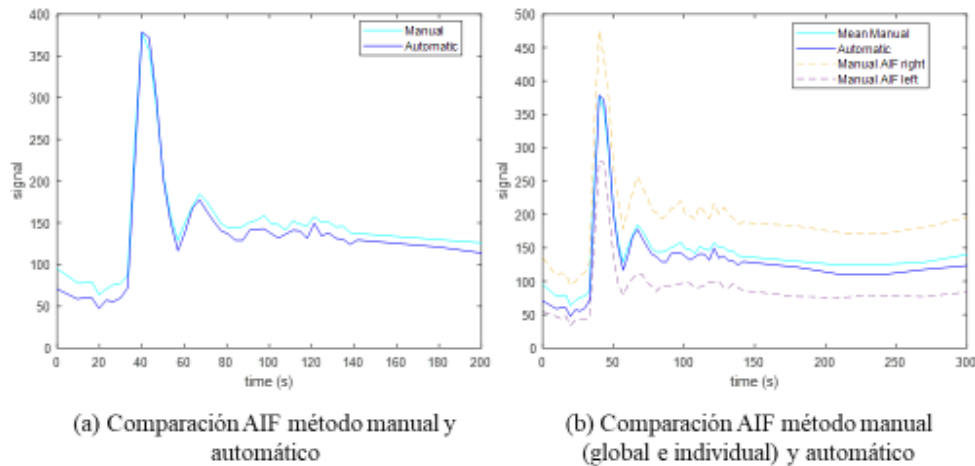


Figura 14. Curvas de la función de entrada arterial obtenidas con los métodos manual y automático. (a) Comparación AIF método manual y automático. (b) Comparación AIF método manual (global e individual) y automático.

Para esta valoración, nos hemos centrado en la comparación de parámetros de la propia curva y no de la máscara AIF, ya que dos de ellas de diferente tamaño en la misma arteria ilíaca pueden dar lugar a diferentes métricas de comparación de segmentación, pero la misma curva AIF.

En primer lugar, se planteó calcular las diferencias entre las curvas medias de ambos métodos a través de la distancia euclídea. De esta manera, se obtuvo para cada tiempo dinámico, la distancia correspondiente entre la curva automática y la curva manual, lo que, a nivel individual en cada caso, nos daba una idea de los resultados obtenidos. Sin embargo, a la hora de comparar todos los casos entre sí, nos dimos cuenta de que el tiempo de adquisición no era una variable constante, lo que introducía un sesgo a la hora de calcular las medias de las distancias euclídeas, por lo que decidimos dejar a un lado esta métrica.

A partir de esto, planteamos las siguientes medidas centrándonos en el primer tercio de los dinámicos (parte correspondiente al pico de la curva), ya que es donde se sitúa el paso del bolo de contraste, y de donde mayor cantidad de información podremos extraer.

Considerando el pico el punto más importante de la curva, se han calculado dos métricas diferentes, la intensidad de pico y el tiempo desde la *baseline* hasta el pico máximo, obteniéndose así, valores independientes para ambas curvas, pudiéndose contrastar en busca de diferencias. En la Figura 15, están representados los principales parámetros cuantitativos que se obtienen de la curva AIF, siendo H y ttp , la intensidad de pico y el tiempo de pico (del inglés, *Height* y *time-to-peak*) estudiadas en este trabajo.

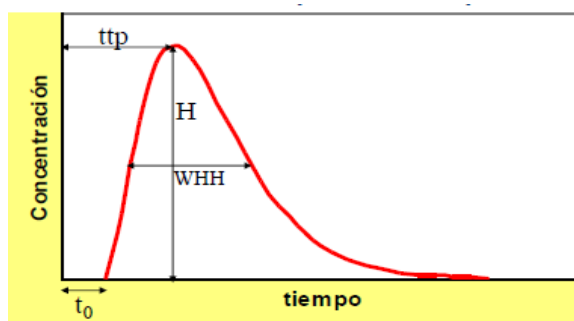


Figura 15. Parámetros cuantitativos extraídos de la curva AIF. H = intensidad de pico, t_0 = tiempo de llegada del bolo, ttp = tiempo de pico, WHH = ancho a media altura del pico.

Para evaluar la similitud entre ambos métodos, se han utilizado gráficas de correlación y Bland-Altman:

- Las gráficas de correlación (Figura 16a) muestran la asociación entre dos métodos cuantitativos, donde cada caso del conjunto de datos se traza como un punto cuyas coordenadas se relacionan con sus valores para las dos métricas.
- Las gráficas de Bland-Altman (Figura 16b) describen la concordancia entre dos medidas cuantitativas construyendo límites de concordancia. Estos límites estadísticos se calculan utilizando la media y la desviación estándar (SD) de las diferencias entre las técnicas, siendo el límite superior $1,96SD$ por encima de la media, y el límite inferior, $1,96$ por debajo de la media [Bland et al, 1999].

La diferencia de las dos medidas se muestra en el eje Y, y la media de ellas en el eje X, siendo recomendable que el 95% de los puntos de datos caigan dentro de los límites de acuerdo.

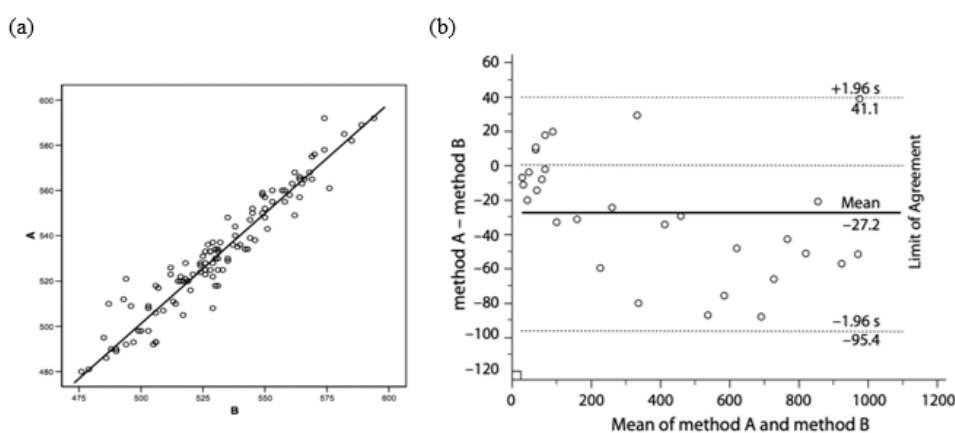


Figura 16. Ejemplo de gráficas de correlación (a) y Bland-Altman (b) con la representación de límites de concordancia (línea de puntos) [Bland et al, 1999]

Como criterio de aceptación, consideraríamos un coeficiente de correlación superior a 0,8 [Evans, 1996] para el caso de la intensidad y el tiempo de pico, y en lo referente a Bland-Altman, se consideraría aceptable un sesgo medio en la intensidad de pico de AIF menor al 10% y un tiempo de pico menor al 5%. El motivo por el que este valor es mayor para la intensidad es porque el método automático solo selecciona los píxeles de mayor intensidad en las arterias ilíacas y otros vasos, mientras que en el método manual se dibuja una ROI en las arterias ilíacas, incluyendo algunos píxeles de menor intensidad que pueden sesgar ligeramente los resultados de la intensidad máxima [Jacobs M. et al., 2016].

3.2 Validación del Registro T2W-DCE

Una vez planteada la metodología para el algoritmo de selección de la AIF, a continuación, se describe cuáles han sido los pasos seguidos en el diseño de la metodología de validación para el registro inter secuencia T2W-DCE.

El objetivo principal de la validación clínica del registro es comprobar si cumple su principal función, en este caso, el registro entre la secuencia anatómica T2W y la funcional DCE, permitiendo llevar la información funcional calculada sobre la DCE al espacio anatómico de la T2W.

Para ello, se plantea realizar una evaluación cualitativa, con la que se pueda comprobar y clasificar la presencia o ausencia de movimiento tras aplicar el algoritmo en cuestión. Además, aun siendo lo más común en esta etapa de preprocesado realizar una validación únicamente cualitativa, se pretende encontrar una serie de parámetros cuantitativos dados por el algoritmo o establecer métricas, con las cuáles, aunque no sean establecidas como validación cuantitativa, permita realizar una evaluación interna del procedimiento.

3.2.1 Validación cualitativa del registro

Apoyándose en la bibliografía [Zhe, 2012] [Razlighi, 2013], las medidas estadísticas más importantes que se utilizan para garantizar la eficacia de los algoritmos de registro en el uso clínico son las métricas de similitud. Sin embargo, no es una tarea fácil, ya que al contrario que para la validación del algoritmo de selección de AIF, no se dispone ni se puede preparar resultados de referencia con los que poder comparar.

El método de validación más utilizado es la inspección visual y cualitativa. La comparación de la combinación de la imagen fija y la imagen móvil, antes y después del registro, pueden reflejar intuitivamente el resultado del registro [Zhe, 2012]. Para ello, es necesario definir un método reproducible, efectivo y de bajo coste temporal con el que poder hacer esta comparación.

Para poder definir la mejor metodología de validación en esta situación es necesario entender cómo funciona el algoritmo en cuestión, qué tipo de resultados se obtienen y qué tipo de imágenes se disponen para ello. De esto último, podemos distinguir entre las secuencias individuales de perfusión DCE y anatómicas T2W, imágenes alineadas e imágenes registradas. Las imágenes alineadas son las imágenes de entrada del algoritmo, y se refieren únicamente a las secuencias DCE y T2W que han sido superpuestas una encima de la otra sin ser corregidas, mientras que las imágenes registradas son el resultado de aplicar el algoritmo de registro sobre las imágenes alineadas, y por tanto el resultado que debe ser validado.

Para comprobar las posibles diferencias en presencia o ausencia de movimiento entre las imágenes alineadas y registradas, se planteó un sistema de clasificación, dividido en dos partes, en el que se evalúan todas las posibilidades existentes al utilizar el algoritmo en las imágenes disponibles. Como se puede ver en la Tabla 3, la primera parte de la clasificación consiste en analizar las imágenes alineadas y comprobar si existe movimiento inicial entre las secuencias DCE y T2W antes de registrarlas, puntuándolas así como 1 y 0. A continuación, a partir de esta agrupación, se evaluó la posibilidad de que el algoritmo corrija el movimiento, en el caso de ser necesario, o, por el contrario, se induzca o se obtenga el mismo resultado que en un principio.

Tabla 3. Esquema general del sistema de clasificación para la validación del registro

Clasificación antes del registro		Clasificación después del registro	
No hay movimiento inicial	0	Igual	0.0
		Se induce	0.1
Hay movimiento inicial	1	Igual	1.0
		Se corrige	1.1
		Se induce más	1.2

Comenzando con la primera parte de la clasificación, la principal problemática que surgió fue la mala visualización de las estructuras anatómicas de las imágenes de perfusión DCE. Aunque las secuencias T2W tengan la principal ventaja de que muestran con detalle la anatomía, al alinearlas con las secuencias DCE se pierde gran parte de esta resolución y visibilidad, como se puede comprobar en la Figura 17, donde se comparan las secuencias T2W con las imágenes alineadas y registradas. Aunque de forma general si se pueden distinguir las estructuras anatómicas, el problema aparece al buscar diferencias entre las imágenes, ya que, en el caso de existir movimiento, ya sea corregido o introducido, puede ser mínimo y ocurrir en los bordes, lo cual no es visible debido a esta pérdida de resolución.

Por lo tanto, la tarea de distinguir un movimiento inicial se dificultó en gran medida, ya que, en la gran mayoría de los casos, aunque se puede intuir las estructuras, no se puede percibir con seguridad los bordes, en este caso de la próstata, siendo en ellos, donde en el caso de que hubiese, se apreciaría el movimiento. Debido a esta limitación, a la metodología de validación se le añade la necesidad de encontrar un buen método de visualización de las estructuras anatómicas, por lo que, antes de comenzar el análisis con toda la base de casos disponibles, se fue probando las diferentes metodologías en algunos, hasta que se estableciese la mejor para la validación.

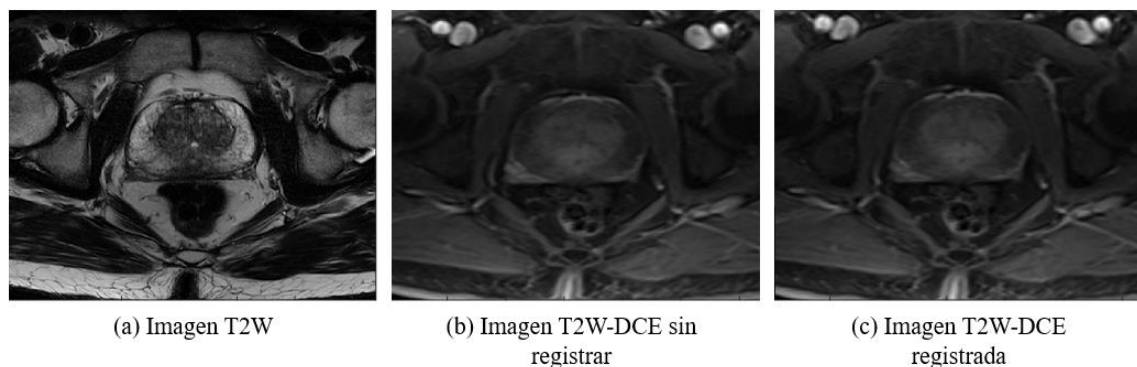


Figura 17. Ejemplo de imágenes utilizadas en la clasificación: secuencia anatómica T2W y secuencias antes y después de registrar (alineadas y registradas)

Ya que la próstata es la única estructura anatómica de la imagen en la que interesa comprobar si actúa de forma correcta el registro, intentando mejorar la visibilidad, la primera idea que se planteó para comprobar la presencia de movimiento inicial fue utilizar la máscara de la próstata, ya validada anteriormente, para que al superponerla con las imágenes alineadas sirviese como guía. El visor de imágenes utilizado fue ITK-Snap®,

el cual ofrece la posibilidad de superponer fácilmente las imágenes con la máscara, además de que permite variar las propiedades de brillo y contraste, la opacidad de la segmentación, y visualizar los volúmenes completos, posibilitando así realizar un análisis individualizado para cada caso en todos los cortes disponibles. Al probar este método con algunos de los casos, aunque la técnica es útil en aquellos casos donde es muy evidente el cambio (Figura 18a), al analizar las imágenes registradas, aunque se aprecia movimiento, no se puede asegurar si es corregido o inducido (Figura 18b), imposibilitando su clasificación.

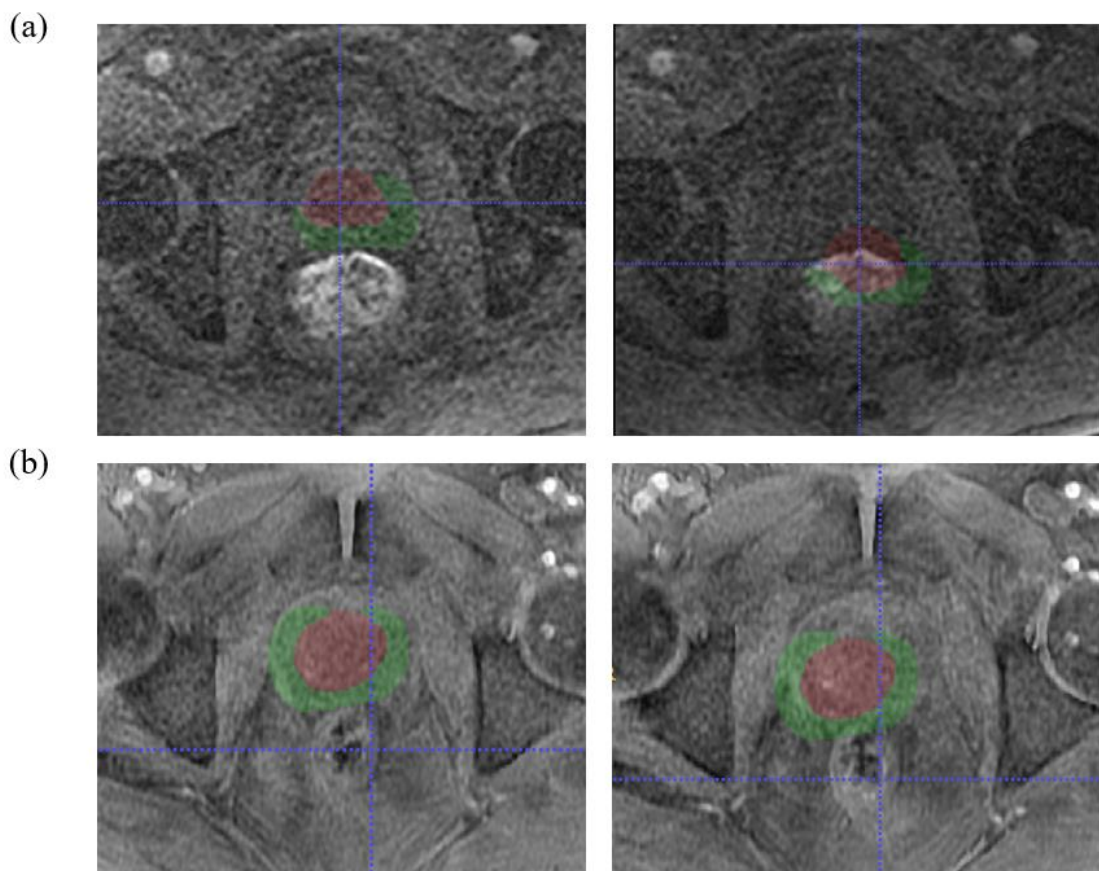


Figura 18. Comparación de la aplicación del registro utilizando la segmentación de la próstata en dos casos.

La siguiente metodología empleada, para mejorar la visualización de las estructuras anatómicas, fue la superposición a las imágenes alineadas y registradas de las secuencias de T2W. Para ello se hizo uso de la posibilidad de modificación de la opacidad en el visor ITK-Snap, lo que permite comprobar la correspondencia de la próstata en ambas imágenes, además de que al mejorar la visualización de las estructuras anatómicas que la rodean, permite aclarar qué tipo de movimiento se está produciendo.

En todos aquellos casos clasificados con movimiento, el siguiente paso del método, fue seleccionar *landmarks* anatómicos (puntos brillantes, zonas características de la próstata en ese paciente, etc.) con los que se pudiese comparar las posibles diferencias antes y después del registro.

Sin embargo, aun obteniéndose resultados satisfactorios en algunos casos en los que se probó esta metodología, se produjo un incremento importante en el tiempo de trabajo, ya que, para poder clasificar los casos según se indica en la Tabla 3, es necesario realizar una primera revisión de todos los casos, tanto de las imágenes alineadas como de la registradas superpuestas con las secuencias T2W, para después seleccionar *landmarks* diferentes para cada paciente.

En cualquier caso, y aunque este tipo de análisis suponga un aumento en el tiempo de trabajo, sigue siendo la mejor opción realizar análisis individualizados de cada caso, siendo además necesario utilizar algún tipo de referencia o guía que permita mejorar la visualización, sin embargo, con el objetivo de diseñar la metodología de validación lo más efectiva posible, tanto el tiempo como en resultados, se dejó de lado este método y se continuó desarrollando otras metodologías que sirvan para todos los casos.

Manteniendo la idea de utilizar las secuencias T2W como guía con las imágenes alineadas y registradas, en vez de compararlas variando la opacidad en el visor de imágenes, siendo una tarea completamente subjetiva por parte de la persona responsable del análisis, se planteó combinar las imágenes DCE-T2W con las secuencias T2W en algo parecido a un tablero de ajedrez, correspondiendo las casillas blancas a fragmentos de una secuencia y las negras a otra, mostrado en la Figura 15. De esta manera, se pretende poder clasificar el registro de forma visual, ya que, en un caso ideal, aun utilizando imágenes diferentes, el resultado en el tablero debería mostrar la imagen completa, sin discontinuidad en la anatomía.

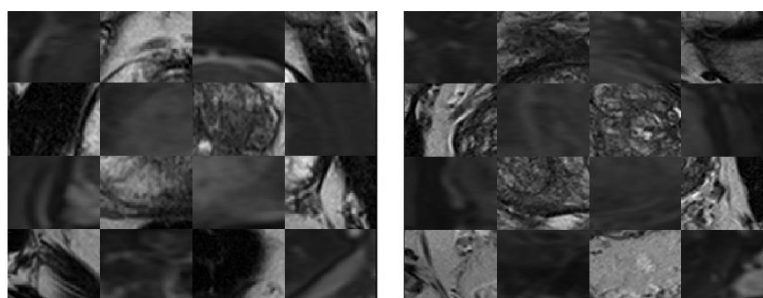


Figura 19. Ejemplo de tablero de ajedrez entre secuencias T2W y secuencias T2W-DCE

Para desarrollar esta metodología, en vez de utilizar un visor de imágenes como en los casos anteriores, se ha desarrollado un código en MatLab®, con el cual no solo se pueden visualizar imágenes de formatos diferentes, sino que existen funciones públicas que permiten elaborar el tablero según se prefiera, entre otras funcionalidades que se comentarán a continuación.

La metodología seguida, aunque en un primer momento pueda resultar de mayor complejidad en comparación con las anteriores, tiene como principal ventaja que, aun analizando los resultados de forma individual para cada paciente, el método está codificado en esta plataforma, siendo el mismo para cada caso, ahorrando así en tiempo, pero manteniendo la personalización.

A continuación, se detallan los pasos seguidos en la metodología implementada:

1. Visualización de las imágenes:

Las secuencias T2W está formada por tres dimensiones, correspondiendo la tercera a los cortes realizados en el plano. Por otro lado, las imágenes alineadas y registradas están formadas por cuatro dimensiones, siendo la tercera los cortes realizados, y la cuarta los tiempos dinámicos característicos de las imágenes de perfusión.

Para poder trabajar con ambas secuencias y combinarlas en una, es necesario que tengan el mismo tamaño de matriz, es decir, en tres dimensiones. Por lo tanto, es necesario escoger uno de los volúmenes dinámicos que componen las imágenes alineadas/registradas. En el algoritmo, se registró un dinámico específico con la T2W, para posteriormente aplicar la misma transformación resultante al resto. Aunque no tendría que suponer ninguna diferencia trabajar con un volumen dinámico y otro, se estableció como criterio en la validación, utilizar el mismo volumen dinámico seleccionado para el registro.

2. Elaboración de los tableros de ajedrez:

Una vez se tienen las tres imágenes (secuencias T2W, alineadas y registradas) con las mismas dimensiones, se procedió a obtener los tableros de ajedrez, por un lado, el correspondiente a la combinación T2W y alineadas, y, por otra parte, T2W y registradas, para cada corte.

Lo ideal sería tener dos únicos tableros, directamente con un corte en el que se apreciase el movimiento, si lo hubiese en ese caso, pero al desconocer ese dato, se aplica este método a todos los cortes de cada caso, se representan todos juntos obteniendo así una visualización general de todos ellos, y se selecciona aquel que permita hacer una mejor comparación. Sin embargo, como se puede ver en la Figura 20, visualizar de esta manera las imágenes dificulta la tarea de encontrar un corte donde se aprecie movimiento, además de aumentar considerablemente el tiempo de evaluación.



Figura 20. Ejemplo de tableros de ajedrez para cada corte de las secuencias para un caso

3. *Bounding Box* de la próstata:

Tras realizar las primeras pruebas del método con un primer número de casos, nos dimos cuenta de que la comparación de las imágenes completas podía introducir errores, por lo que se decidió centrarse únicamente en la zona de la próstata, ya que es donde nos interesa que actúe correctamente el registro, independientemente de la situación inicial de las imágenes.

Por ello, a partir de la segmentación de la próstata se ha construido una *Bounding Box* a su alrededor (Figura 21), ampliando así el área de trabajo y permitiendo una clasificación de mayor precisión. Una vez centradas las imágenes en la zona de interés, en el caso de que el tablero de ajedrez siga sin mostrar unos resultados claros, se ha propuesto de nuevo la utilización de *landmarks* anatómicos sobre las propias imágenes. Además, para obtener una mejor visibilidad de los resultados en el tablero, en el análisis definitivo, se utilizaron mapas de colores diferentes entre las casillas, escala de grises para las secuencias T2W, y escala de verdes para las imágenes alineadas y registradas.

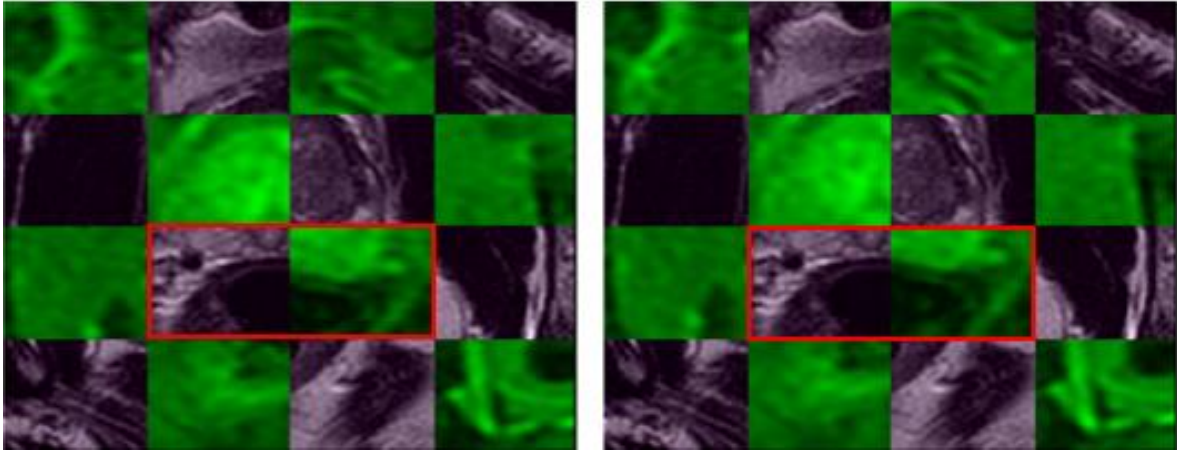


Figura 24. Ejemplo de un caso en el que el algoritmo actúa correctamente cuando hay movimiento inicial

Por último, se ha realizado una evaluación interna del algoritmo con métricas cuantitativas propias extraídas de él. El objetivo de esto es establecer un control de calidad del registro mediante la comparación entre estos valores y los resultados de la validación cualitativa, con el fin de elegir un umbral con el que distinguir los casos correctamente registrados de los incorrectos. Las métricas obtenidas son los tres parámetros de translación correspondientes a los tres ejes (x,y,z) y los valores mínimos y máximos del determinante jacobiano y de la variable de deformación, siendo así en total siete parámetros para comparar. En el apartado 4.2.3 *Evaluación interna del algoritmo de registro*, se definen estas métricas y se incluye su representación gráfica para una mejor comprensión.

Capítulo 4. Resultados

En este apartado se presentan todos los resultados obtenidos tras llevar a cabo las metodologías de validación planteadas en el Capítulo 3. En primer lugar, se muestran los resultados correspondientes a la selección automática de la función de entrada arterial AIF en imágenes DCE, y posteriormente, los resultados relativos al registro intersecuencia entre imágenes T2W y DCE.

4.1 Validación de la Selección AIF

A continuación, se presentan los resultados que se han obtenido siguiendo los criterios y métodos explicados en el apartado *3.1 Validación AIF*. Al igual que anteriormente, los resultados se presentan divididos según el tipo de metodología seguido, cualitativa o cuantitativa.

4.1.1 Validación cualitativa AIF

Debido a la subjetividad de esta clasificación, siendo completamente dependiente de la interpretación de los criterios para cada persona, se ha realizado una segunda evaluación, lo que permite contrastar esta validación. En la Tabla 4 se muestran los resultados, siendo la primera fila la correspondiente a la clasificación inicial, y la segunda, la realizada por un ingeniero experto de QUIBIM.

Las curvas AIF ideales se obtienen de las ROI clasificadas con el valor 2 y si el algoritmo de selección fuese perfecto, todos los casos serían clasificados como tal. Aun así, en ambas clasificaciones, en más del 90% de los pacientes, las ROI han sido seleccionadas correctamente.

Si se comparan ambos resultados, se puede ver que las diferencias entre las dos evaluaciones son mínimas, con relación al número de casos dentro de cada grupo.

Tabla 4. Resumen de resultados de la validación cualitativa

Validación cualitativa	Valor 0	Valor 1	Valor 2
Primera evaluación	9 (3,72%)	8 (3,31%)	225 (92,97%)
Segunda evaluación	8 (3,31%)	6 (2,48%)	228 (94,21%)

Las curvas AIF ideales se obtienen con los casos clasificados con valor 2, y las que seguro que son inutilizables son las puntuadas con valor 0. Se consideró incluir un tercer grupo intermedio para aquellos casos en los que, visualmente no fuese tan sencillo clasificar las ROI, aunque sus curvas AIF si se consideraran válidas. Sin embargo, ya que esta puntuación de 1 es completamente dependiente de la interpretación de los criterios de clasificación expuestos en el apartado 3.1.1 *Validación cualitativa AIF*, se consideró interesante conocer, no solo si coincide la cantidad de casos recogidos en cada grupo, si no, si los casos coinciden dentro de cada grupo.

Ya que ninguna de las evaluaciones puede considerarse de referencia sobre la otra, es decir, ninguna puede ser determinada como *gold standard*, métricas como falsos positivos o negativos, entre otras, no ofrecen información objetiva sobre la coincidencia de casos entre las clasificaciones. Por ello, se ha obtenido el coeficiente de Jaccard total y para cada subgrupo:

Tabla 5. Coeficientes de similitud de Jaccard entre las dos clasificaciones

	Coeficiente Jaccard	Casos coincidentes	Casos totales
Total	0,975	236	242
Valor 0	0,889	8	9
Valor 1	0,400	4	10
Valor 2	0,978	224	229

Los casos totales no coinciden con los datos proporcionados en la Tabla 4 debido a que ninguna de las dos evaluaciones está establecida como referencia, por lo que, para calcular el coeficiente de Jaccard correctamente, se hace un recuento de los casos totales con ambas clasificaciones.

Como se puede ver en la Tabla 5, el coeficiente de Jaccard para el valor 1 es 0,400. Ya que, en un primer momento, se consideró que las curvas AIF para estos casos sí se podrían

utilizar, pero la similitud entre ambas clasificaciones es baja, con el objetivo de evitar cometer errores, todos los casos clasificados como 1 volvieron a ser revisados y puntuados con un valor 0 o 2, separando así los casos inutilizables, de los válidos. La nueva clasificación quedaría de la siguiente manera:

Tabla 6. Resumen de resultados de la validación cualitativa revisada

Validación cualitativa	Valor 0	Valor 2
Primera evaluación	10 (3,72%)	232 (92,97%)
Segunda evaluación	9 (3,31%)	233 (94,21%)

De esta forma, al volverlos a comparar, de los 242 casos analizados, solo en uno de ellos ha habido un desacuerdo entre las dos validaciones, obteniendo así un coeficiente de Jaccard igual a 0,996.

4.1.2 Validación cuantitativa AIF

Una vez establecidas las métricas cuantitativas de validación como la intensidad de pico y el tiempo hasta su alcance, se obtuvieron ambos valores para las curvas de los métodos.

En cuanto al análisis la intensidad de pico mostrado en la Figura 25a, comenzando con la gráfica de correlación, se han obtenido los siguientes resultados cuantitativos:

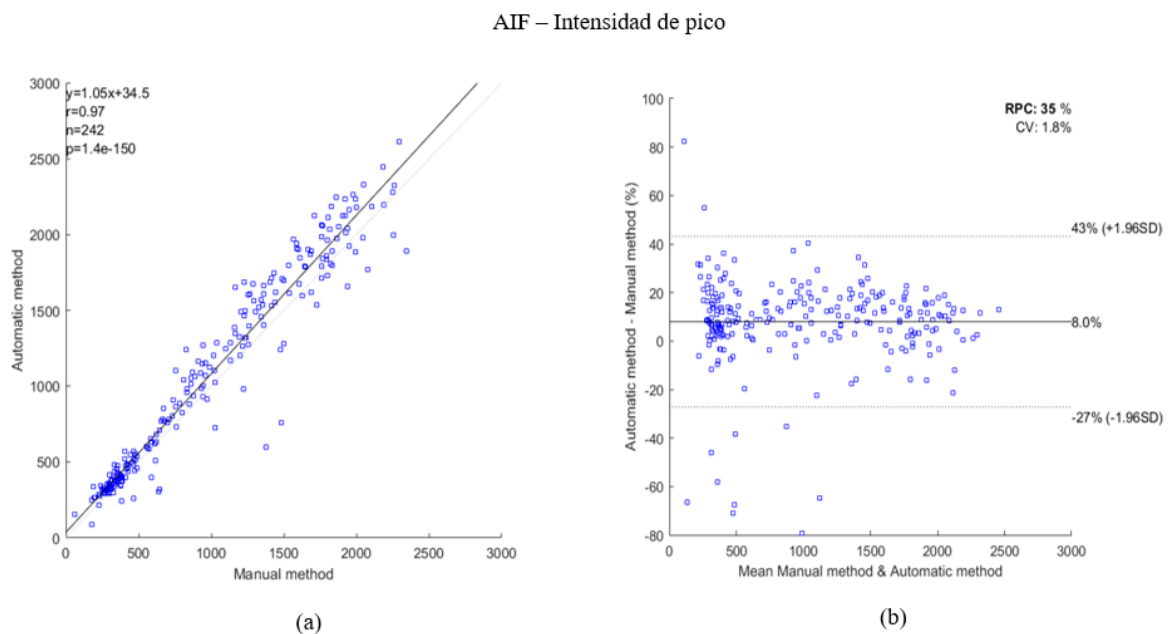


Figura 25. Gráficas de correlación (a) y Bland-Altman (b) para evaluar la intensidad de pico

- $y = 1,05x + 34,5 \rightarrow$ Modela la relación entre dos variables a través del ajuste de los datos con una ecuación lineal.
- $n = 242 \rightarrow$ Representa el número de los casos comparados-
- $r = 0,97 \rightarrow$ El coeficiente de correlación de Pearson (r) es una medida de la relación lineal entre dos variables. Toma valores que van de $+1$ a -1 , correspondiendo a una relación directa e indirecta, respectivamente. Además, cuanto más se acerca a 1 , más lineal es la relación.
- $p = 1,4 \cdot 10^{-150} \rightarrow$ El valor p es la probabilidad de encontrar el resultado actual si el coeficiente de correlación fuera 0 (hipótesis nula). Lo ideal es obtener valores $p < 0,001$, ya que determina que la correlación es significativa.

En lo referente al análisis Bland-Altman representado en la Figura 25b, se obtuvieron los siguientes parámetros:

- $SD = 18\% \rightarrow$ Desviación estándar de las diferencias.
- $RPC = 35\% \rightarrow$ Coeficiente de reproducibilidad ($1,96 \cdot SD$).
- $CV = 1,8\% \rightarrow$ Coeficiente de variación (SD de los valores medios en porcentaje).
- Sesgo medio = 8%
- Límites de concordancia = $-43,27\%$

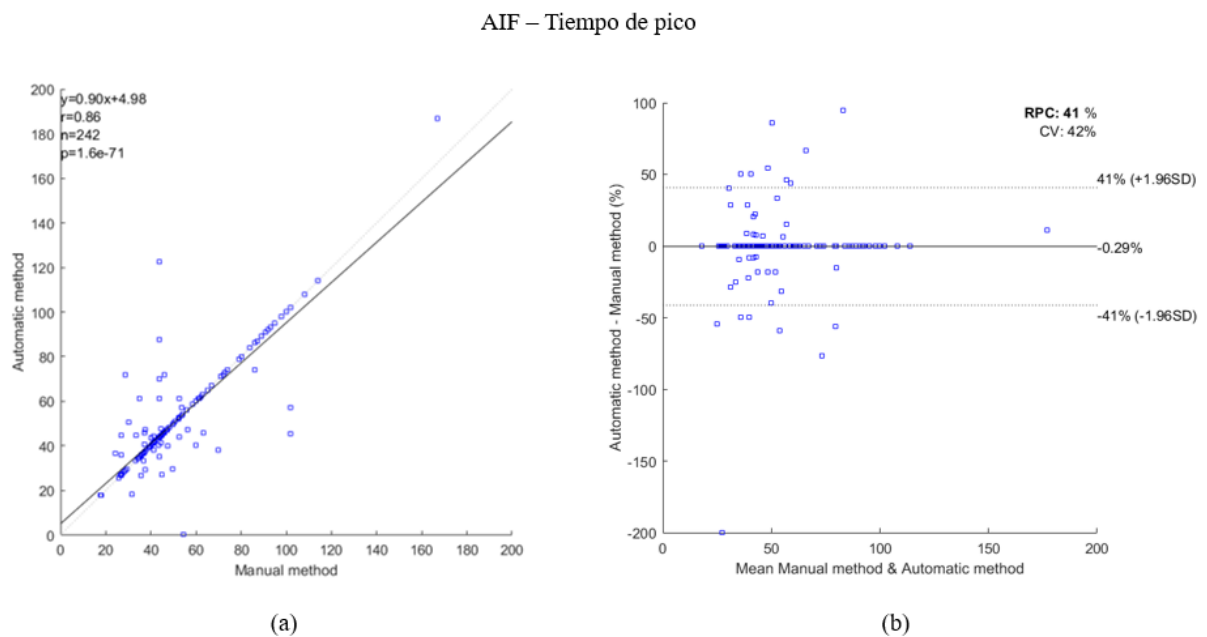


Figura 26. Gráficas de correlación (a) y Bland-Altman (b) para evaluar el tiempo de pico

Al igual que para la intensidad de pico, se muestran en la Figura 26a, los parámetros cuantitativos extraídos de la gráfica de correlación:

- $y = 0,90x + 4,98$
- $n = 242$
- $r = 0,86$
- $p = 1,6 \cdot 10^{-71}$

Por otra parte, las métricas mostradas en la Figura 26b correspondiente a la gráfica Bland-Altman:

- $RPC = 41\%$
- $CV = 42\%$
- $SD = 21\%$
- $\text{Sesgo medio} = -0,29\%$
- $\text{Límites de concordancia} = -41,41\%$

4.1.3 Evaluación interna del algoritmo AIF

Al ir obteniendo y revisando de forma individual las curvas AIF para cada uno de los casos, se fueron encontrando diferencias apreciables en su forma (Figura 27), sobre todo en la pendiente de subida inicial, correspondiente al paso del bolo de contraste. A partir de ello, se planteó la idea de realizar una evaluación interna del algoritmo, en función de la forma de las curvas obtenidas, a partir del cálculo del ratio entre la *baseline* (antes del paso del contraste) y el pico, con el objetivo de buscar e identificar un valor umbral con el que poder clasificar la validez de la AIF seleccionada, antes de ser utilizada en el posterior análisis o modelado farmacocinético.

Para ello, se ha utilizado un diagrama de caja y bigotes en el que se muestran todos los ratios calculados, separados según la clasificación cualitativa realizada en el apartado 4.1.1 *Validación cualitativa AIF*, para comprobar si, dependiendo del valor dado, existen diferencias en el ratio de la pendiente de la curva.

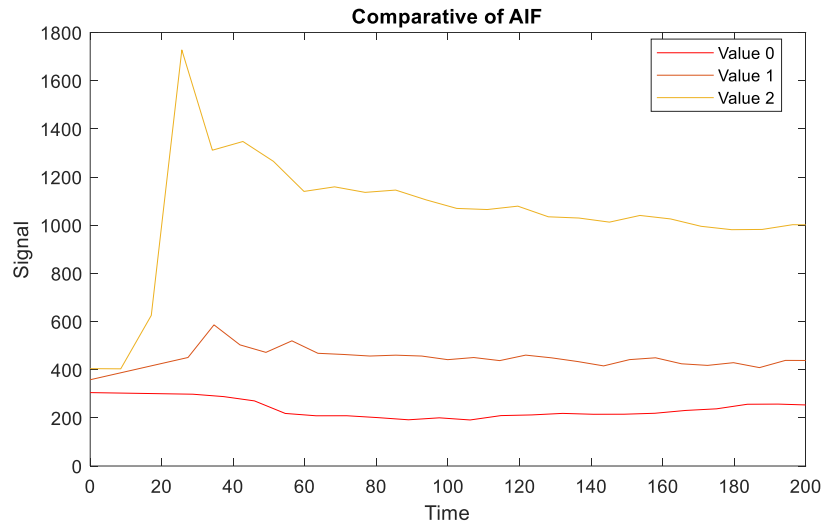


Figura 27. Diferencias entre tres casos clasificados de forma diferente según la validación cualitativa

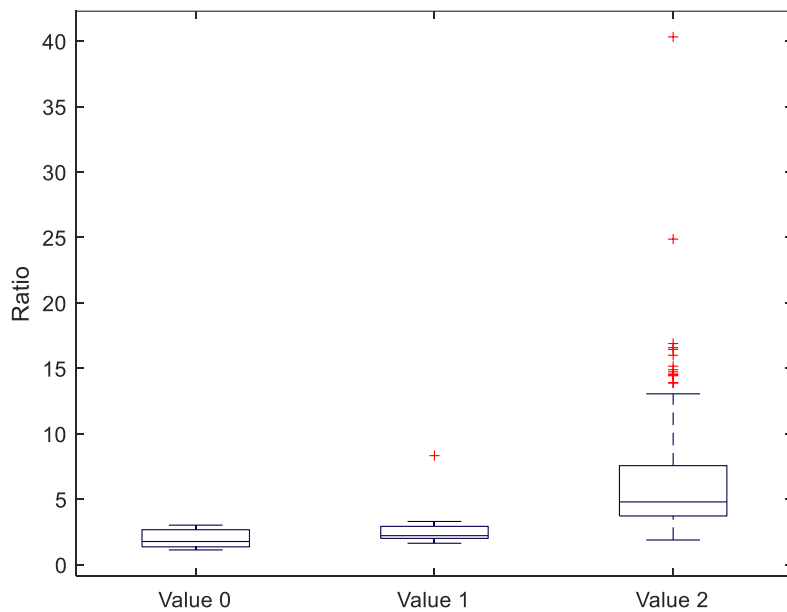


Figura 28. Diagrama de cajas y bigotes para los ratios

Analizando la gráfica de la Figura 28, se puede ver que existe una clara diferencia entre los casos clasificados como 0 y 1, y los clasificados como 2. A partir de esto, se puede establecer que la curva o función AIF es válida para su uso en posteriores modelos en aquellos casos en los que su ratio sea mayor de 3, lo que corresponde al percentil 75 de los valores de 1. Aun así, se ha creído conveniente recomendar que en los casos que se encuentren entre los valores 3 y 4 (el percentil 25 de los valores de 1 es 3,7), se contraste los resultados obtenidos con la predicción.

4.2 Validación del Registro T2W-DCE

A continuación, en este apartado se presentan los resultados obtenidos con la validación clínica del registro inter secuencia DCE y T2W. En primer lugar, se muestran los primeros resultados obtenidos, ya que como se explica más adelante, se han tenido que hacer algunos cambios en el algoritmo en cuestión, para así finalmente poder realizar la validación definitiva del método.

4.2.1 Primeros resultados obtenidos

Como se ha podido ver en el apartado 3.2 sobre la metodología seguida para la validación del registro, se han ido planteando diferentes ideas para llevarla a cabo. Tras descartar las primeras, debido al aumento del coste temporal y a la subjetividad del técnico encargado, antes de realizar la validación definitiva con la base de datos específica para este caso (3.2.1 *Validación cualitativa del registro*), se decidió hacer una primera comprobación de la funcionalidad del método, con un conjunto de imágenes menor.

Se comenzó trabajando con 30 casos escogidos al azar de la base de datos total de 333 casos. De ellos, de la primera clasificación de las imágenes alineadas se observó que únicamente 2 casos presentaban movimiento inicial, y al aplicar el registro se corrigió en ambos. Sin embargo, analizando las imágenes registradas, se detectó que en 8 de los casos se inducía movimiento de manera bastante exagerada, como se puede ver en la Figura 29, lo que suponía un problema ya que este tipo de situaciones es la que más se intenta evitar.

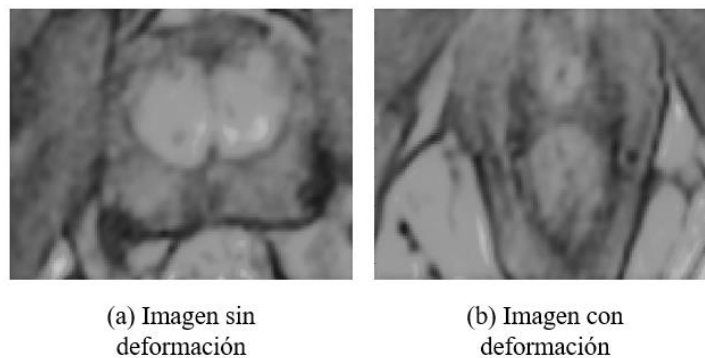


Figura 29. Ejemplo de introducción de una deformación en la imagen al aplicar el algoritmo

Con esto, nos dimos cuenta de que el algoritmo de registro planteado contenía algún tipo de error que provocaba la introducción de movimiento en algunos casos, por lo que se paró el proceso de validación y nos centramos en corregir este error. El departamento

encargado del desarrollo de algoritmos en QUIBIM es el de i+d, y es con el cual se trabajó en conjunto analizando aquellos casos en los que se inducía error. En todos ellos el movimiento introducido era el desplazamiento en el eje vertical, generalmente, lo que indicaba que el problema era debido a la oblicuidad en el plano entre las imágenes DCE y T2W, introducida en la adquisición de las imágenes.

Con las modificaciones llevadas a cabo por los ingenieros de i+d en el algoritmo, relacionadas con una estrategia multirresolución con transformación afín, se consiguió solucionar 7 de los 8 casos, sin penalizar el tiempo de ejecución. A partir de esta solución, surgió otra idea, que consiste en alinear primero las imágenes DICOM y después aplicar el registro afín con Elastix, sin añadir tampoco tiempo de ejecución. De esta forma, se ayudaría al algoritmo dándole las imágenes alineadas y en el mismo espacio, evitando problemas de oblicuidad en el plano de adquisición.

De esta forma, se cuenta con dos algoritmos de registro inter secuencia, diferenciados por el alineamiento de las imágenes DCE y T2W antes de ser registradas. Por lo tanto, es necesario decantarse por uno, ya que solo uno puede ser utilizado en el producto clínico final. Para ello, siguiendo la metodología planteada para la validación, se ha realizado una comparativa entre ambos registros, para poder así elegir uno, y hacer la validación clínica definitiva.

Para este propósito, se tomaron 43 casos al azar, de forma balanceada entre los tipos de adquisición, y se siguió la misma metodología: en primer lugar, se revisó la existencia de movimiento inicial y después el resultado de las imágenes registradas, para ambos registros. En la Tabla 7 se muestran los resultados.

Tabla 7. Comparativa de los resultados en un conjunto de datos reducido para ambos métodos de registro

Imágenes antes del registro		Imágenes registradas			
		Registro 1 (sin alineamiento inicial)		Registro 2 (con alineamiento inicial)	
No hay movimiento inicial	37	Se mantiene igual	31	Se mantiene igual	37
		Se induce	6	Se induce	0
Hay movimiento inicial	6	No se corrige	1	No se corrige	0
		Se corrige	5	Se corrige	6
Total = 43					

Como se puede ver, existe una clara diferencia entre los resultados de ambos registros. Aunque ambos funcionan prácticamente perfectos cuando hay movimiento inicial, aun habiendo arreglado el registro inicial, se sigue induciendo movimiento en una cantidad relevante de casos, mientras que, en el segundo registro, gracias al alineamiento inicial de las imágenes, en ninguno de estos casos, aparece movimiento. Por tanto, se ha seleccionado el segundo registro para su uso definitivo, siendo este del que es necesario hacer la validación clínica.

4.2.2 Resultados validación clínica cualitativa

La base de datos final específicamente seleccionada para este caso, como se explica en el 2.1.1 Base de datos de la validación del registro de imágenes T2W-DCE, consta de 112 casos, y al igual que se ha hecho en los análisis anteriores, se ha revisado si en las imágenes alineadas (las secuencias T2 y DCE superpuestas) existía movimiento inicial o no, y si en las imágenes registradas, en caso de haber, éste es corregido, o por el contrario, si se induce. Las otras dos posibilidades planteadas se refieren a que los resultados se mantengan iguales, es decir, no varían, sea cual sea la situación inicial. En la Tabla 7 se muestra la clasificación y los porcentajes finales:

Tabla 8. Resultados de la clasificación en la base de datos total. Los casos correctos están indicados en verde y los incorrectos en rojo

Resultados		US		Voxel		TOTAL
Sin movimiento inicial	Sin movimiento	58 casos	73,42%	30 casos	90,91%	78,57%
	Deformación	6 casos	7,59%	1 caso	3,03%	6,25%
Movimiento inicial	Movimiento corregido	14 casos	17,72%	1 caso	3,03%	13,39%
	Deformación	1 caso	1,26%	1 caso	3,03%	1,79%
Casos analizados		79 casos		33 casos		112 casos

Al final, el objetivo de esta validación es evaluar y verificar el funcionamiento del algoritmo de registro de las secuencias T2W y DCE independientemente de la situación de las imágenes de entrada, por lo tanto, a partir de estos resultados mostrados en la Tabla 8, se pueden agrupar de acuerdo con la correcta o incorrecta actuación del algoritmo:

- **Registro correcto:** el algoritmo corrige el movimiento en el caso de existir inicialmente. Si desde un primer momento no hay movimiento entre las secuencias, se obtiene el mismo resultado. De la Tabla 8, los registros correctos están señalados en verde, ya que corresponden a los casos sin movimiento o corregidos.
- **Registro incorrecto:** el algoritmo introduce una deformación o induce movimiento en la imagen independientemente de la situación inicial, o no corrige el movimiento en el caso de existir en un principio. Este tipo de resultados están señalados en rojo en la Tabla 8.

La nueva clasificación quedaría de la siguiente manera:

Tabla 9. Clasificación final de los casos correctos e incorrectos del registro

Registro Correcto	Registro Incorrecto
103 casos (91,96%)	9 casos (8,04%)

4.2.3 Evaluación interna del algoritmo de registro

Aun siendo la metodología cualitativa empleada suficiente para la validación del algoritmo de registro en cuestión, al igual que se hizo con el algoritmo de selección de AIF, resulta interesante establecer métricas cuantitativas que permitan la evaluación interna del algoritmo.

El control de calidad interno del registro DCE-T2W planteado sirve para verificar, de forma automática y mediante métricas cuantitativas, la correcta actuación del algoritmo sin revisar las imágenes de forma visual.

Para ello, se han obtenido 7 parámetros cuantitativos específicos del tipo de transformación empleada en el registro de imágenes: tres corresponden a la translación en los ejes $x - y - z$ dada por la transformación afín; dos se refieren a los valores máximos y mínimos del determinante jacobiano; y, por último, los valores mínimos y máximos de la variable de deformación.

A partir de estas métricas, el objetivo es poder comprobar si alguna de ellas sigue alguna tendencia con respecto a la clasificación hecha en el apartado 4.2.2 *Resultados validación*

clínica cualitativa, y poder determinar un valor umbral que determine la bondad del registro. Para ello, se ha vuelto a utilizar los diagramas de cajas y bigotes, ya que es el que mejor puede mostrar, de forma visual, la distribución de los valores para cada parámetro, clasificado según el algoritmo actúa correctamente, ya sea con movimiento inicial o sin él, o introduzca una deformación.

Tanto las variables correspondientes a la translación como el determinante jacobiano han sido calculados en valores absolutos, para poder visualizar mejor la distribución de estos y establecer un umbral de diferenciación positivo. De la Figura 30 se puede concluir lo siguiente:

- De variables de movimiento solo se han obtenido las correspondientes a la traslación, ya que son características de la transformación afín empleada en el registro. Analizando las tres, se ha llegado al consenso que la que mejor define la diferenciación entre un registro correcto (englobando los resultados normales y corregidos) y un registro incorrecto, es la tercera variable de traslación, correspondiente al eje z.
- Las variables de deformación no es necesario obtenerlas en valor absoluto ya que ya se encuentran distribuidas por encima y por debajo del eje 0. Es evidente que hay una clara distinción entre los grupos, secundando así los resultados obtenidos con la validación cualitativa
- Aunque se han obtenido tanto los valores mínimos y máximos del determinante jacobiano, según lo esperado, estos coinciden, por lo que solo es necesario representar uno de ellos. Este parámetro caracteriza la deformación del campo y muestra cuanto es deformada una imagen según el valor se aleje de la media igual a la unidad, es decir, si se obtiene un determinante igual 1.1, este valor hacer referencia a que se ha introducido una variación del 10% en el volumen en comparación con el inicial. En este caso, como solo interesa establecer un umbral, los datos obtenidos se han calculado en valor absoluto y centrado en el 0.

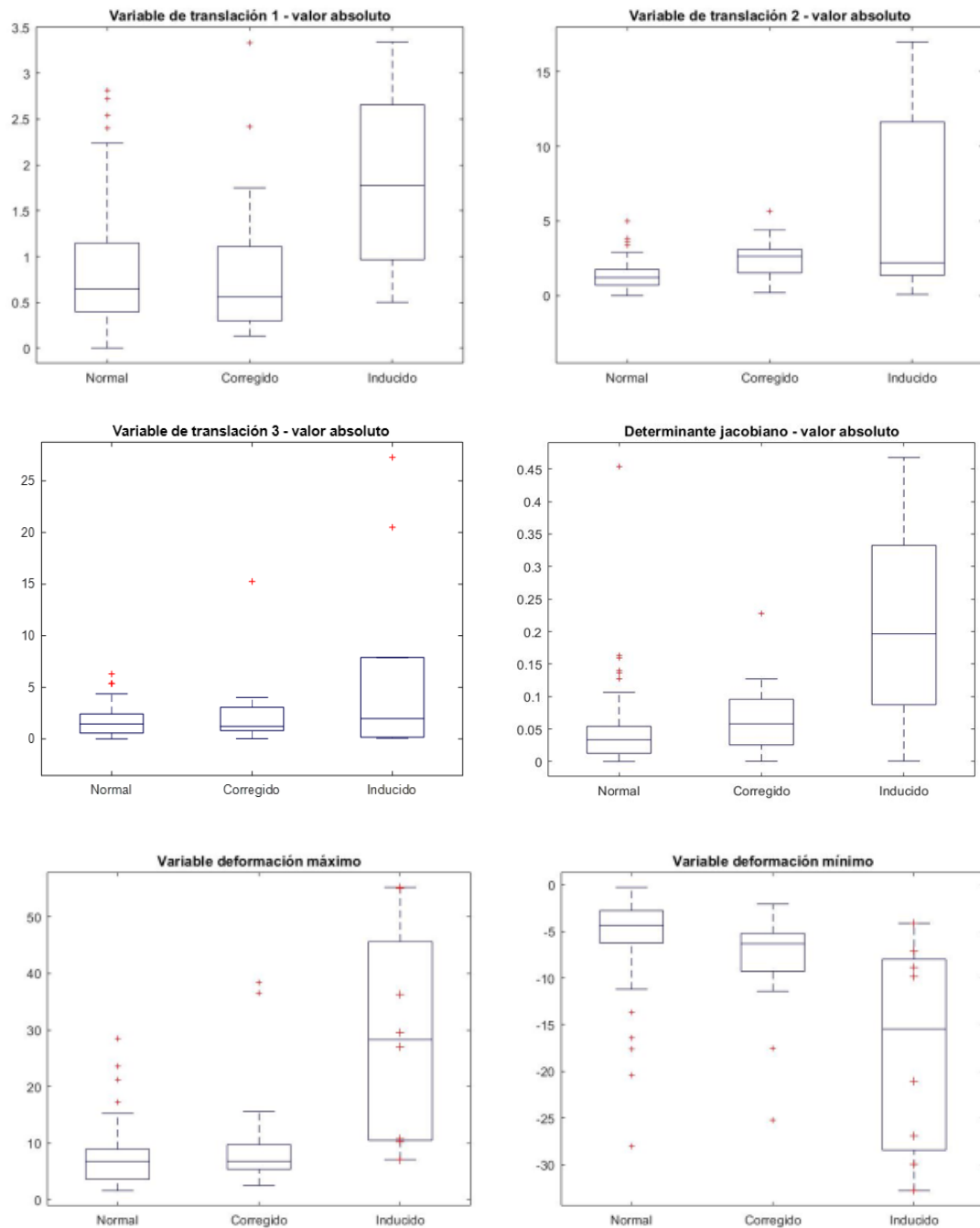


Figura 30. Representación de los parámetros cuantitativos para la evaluación interna según la clasificación cualitativa

A partir de esta información, se ha llegado al consenso de que las dos variables cuantitativas que mejor pueden determinar la bondad de la actuación del algoritmo son el determinante jacobiano y la variable de traslación en el eje y. El umbral escogido para cada una de ellas corresponde al máximo valor del diagrama (no *outlier*) de los casos correctos, siendo así, para la variable de traslación y igual a 4.4, y para el determinante jacobiano, 0,13.

Capítulo 5. Discusión

En este apartado se recoge la discusión correspondiente de la metodología de validación planteada en el Capítulo 3 y los resultados obtenidos presentados en el Capítulo 4. Continuando con la misma dinámica en este trabajo, este apartado se divide en la discusión, por un lado, de la selección automática de la función de entrada arterial AIF en imágenes DCE; y por otro, el registro inter secuencia entre imágenes T2W e imágenes DCE.

5.1 Validación de la Selección de AIF

5.1.1 *Discusión de los resultados cualitativos*

Los resultados de las dos evaluaciones realizadas difieren ligeramente, y esto ocurre en aquellos casos en los que se han producido malentendidos en la interpretación de los criterios establecidos. En cualquier caso, debido a los buenos resultados obtenidos, se puede establecer que el método utilizado para esta validación es correcto.

Aunque se puedan producir malentendidos en la interpretación de los criterios de clasificación establecidos en el apartado 3.1.1 *Validación cualitativa AIF*, para evaluar la selección automática de las ROI en cada caso, los resultados pueden ser claramente diferenciados de forma visual, entre los casos con los que se obtendrá una curva AIF válida y con los que no, es decir, los casos clasificados como 0 pueden ser fácilmente distinguidos de los casos clasificados como 2, existiendo una mayor incertidumbre en la clasificación de los casos con valor 1.

Como se puede ver en la Tabla 5, el coeficiente de Jaccard calculado para toda la base de datos es 0,975, lo que indica la elevada similitud entre las dos evaluaciones. Sin embargo, si se analizan los coeficientes específicos para cada valor, los correspondientes a valor 0 y valor 2 se mantienen también elevados, pero para el valor 1 el coeficiente es 0,400. Este resultado se justifica por la variabilidad de interpretación del criterio de clasificación para este tipo de casos. Por tanto, ya que pueden ser utilizadas las curvas AIF obtenidas de las

segmentaciones clasificadas tanto como 1 como 2, con el fin de evitar introducir errores con los casos con valor 1, se ha reducido a una clasificación binaria, separando los casos inutilizables, de los válidos.

Con esto, se mantiene la clasificación de los casos 0 y 2, y se vuelve a evaluar los casos con valor 1 de forma independiente. Como se puede ver en la Tabla 6, con esta nueva clasificación, solo se difiere en un caso, obteniendo así un coeficiente de Jaccard igual a 0,996.

5.1.2 Discusión de los resultados cuantitativos

En este tipo de comparaciones entre dos métodos, existe un dilema sobre como evaluar su concordancia, confundándose a menudo con correlación. La principal diferencia entre ambos es que:

- La correlación mide la fuerza de la relación lineal entre dos variables. Un cambio en la escala de medición no afecta a la correlación, pero sí a la concordancia. No obstante, dos métodos diseñados para medir la misma variable deben tener una buena correlación.
- Los gráficos de Bland-Altman evalúan visualmente la concordancia entre los dos métodos de medida. La concordancia hace referencia a la conformidad de un método con otro, es decir, es el grado en el que ambas técnicas están de acuerdo sobre el mismo fenómeno observado [Cortés-Reyes et al, 2010].

Por lo tanto, para la comparación entre los métodos de selección automática y manual de la AIF, la correlación y el análisis Bland-Altman son evaluaciones importantes y complementarias.

Al examinar los resultados obtenidos en la evaluación de correlación entre los parámetros obtenidos (intensidad y tiempo de pico), hay una regla empírica para la interpretación de la fuerza de la relación lineal basada en el valor absoluto de r [Evans, 1996], mostrada en la Tabla 10.

Tabla 10. Interpretación de la fuerza de relación basada en el valor r

Valor absoluto de r	Fuerza de relación
0.00-0.19	MUY DÉBIL
0.20-0.39	DÉBIL
0.40-0.59	MODERADA
0.60-0.79	FUERTE
0.80-1.00	MUY FUERTE

Para la intensidad de pico, se ha obtenido de valores $r = 0,97$ y $p = 1,4 \cdot 10^{-150}$. Por lo tanto, la correlación es muy fuerte ($r > 0,8$) y significativa ($p < 0,001$), habiendo así una relación positiva entre los métodos automático y manual. Se obtiene la misma conclusión para el tiempo de pico, ya que los valores obtenidos son $r = 0,86$ y $p = 1,6 \cdot 10^{-71}$, por lo que la relación también es positiva, fuerte y significativa.

Esta evaluación establece una correlación entre los métodos, pero no necesariamente muestra la concordancia. El análisis Bland-Altman proporciona la desviación media de un método con respecto al otro y los límites de concordancia para evaluar si los métodos son comparables o similares.

En el apartado de resultados (4.2.2 Validación cuantitativa AIF), en las Figuras 25 y 26 se proporciona los valores de los ejes, límites y media, como porcentajes. La máxima concordancia entre métodos se obtiene cuando el sesgo medio se aproxima a 0, lo que supondría que con el método de selección automático se obtienen resultados equivalentes al manual, que sirve de referencia. En cuanto a los límites de concordancia, cuanto menor sea el rango, mayor similitud existirá entre los métodos. Si estos límites no exceden la diferencia máxima aceptable entre métodos, se considera que ambos están en concordancia y pueden ser utilizados indistintamente.

Para la intensidad de pico, se obtuvo un sesgo medio del 8%, lo que significa que el método automático sobreestima el valor del pico de la curva sobre el método. Dentro del intervalo de confianza constituido por los límites de concordancia de -43,27%, se encuentran 228 casos de los 242 iniciales, representando así el 94,21%.

El motivo por el que el método automático sobreestima al manual en este caso es porque se obtienen valores de intensidad de pico más altos que con el método manual, debido a que este primero selecciona solo los píxeles que mejor representan la entrada arterial (tanto en las arterias ilíacas como otros vasos grandes). En el caso manual, se dibuja un círculo ROI en las arterias ilíacas, ignorando otros vasos cercanos a la próstata que pueden representar con precisión la función de entrada arterial, además de que pueden incluir píxeles menos intensos que podrían estar presentes en los bordes de la ROI. Por lo tanto, el método automático presenta un valor de intensidad máxima mayor en comparación con el manual, considerándose el porcentaje de 8% de diferencia aceptable, siendo <10% el criterio aceptable.

En cuanto al tiempo de pico, se obtuvo un sesgo de -0,29%, siendo el método automático el que subestima sobre el método manual, aunque este valor determina casi la concordancia perfecta entre métodos debido a la proximidad a la diferencia ideal de 0. Entre los límites de concordancia de -41,41% se encuentran 232 casos, siendo un total de 95,87% de los casos. En este caso Bland-Altman muestra una buena concordancia para el tiempo de llegada al pico, sin ningún sesgo sistemático entre los dos métodos.

Algunos de los puntos que están fuera de los límites de concordancia corresponden a los 9 casos detectados en la evaluación cualitativa (clasificados con valor 0) que seleccionaron incorrectamente los píxeles de las arterias ilíacas u otros grandes vasos y, por lo tanto, dieron lugar a una curva AIF deficiente, con parámetros incorrectos tanto de intensidad como tiempo de pico.

5.2 Validación del Registro T2W-DCE

Para validar la actuación y bondad de la transformación utilizada en el registro intersecuencia entre imágenes T2W y DCE, existen diferentes métodos, dependiendo de la situación y de la información disponible para ello. En este trabajo, se ha centrado en la revisión e inspección visual de los resultados en comparativa con las imágenes iniciales.

Una vez han sido analizados y revisados los resultados del registro intersecuencia en imágenes anatómicas T2W y funcionales DCE, de forma personalizada en los 112 casos de RM escogidos específicamente como conjunto variado y representativo de la práctica clínica, estos han sido clasificados siguiendo un método de evaluación diseñado para la

validación del algoritmo en cuestión. Esta clasificación se muestra en anteriores apartados en la Tabla 7, pudiendo concluir lo siguiente:

- En 88 de los casos (78,57%), no se apreció ninguna diferencia o movimiento inicial en las secuencias alineadas antes de la transformación, y una vez estas imágenes fueron registradas, los resultados fueron los mismos, por lo que el algoritmo actúa correctamente.
- En 15 de los casos (13,39%), si que se pudo ver que existía movimiento en las imágenes alineadas al contrastarlas con las secuencias T2W, pero una vez aplicada la transformación del registro, esta diferencia fue corregida.
- En 7 de los casos (6,25%), al igual que en los primeros comentados, no se vio la existencia de movimiento inicial, sin embargo, al utilizar en ellas el algoritmo de registro, se pudo apreciar claramente la introducción de deformación en la imagen.
- En 2 de los casos (1,79%), al igual que en los segundos mencionados, existía una diferencia en las imágenes alineadas, aun sí, el algoritmo no actúa correctamente corrigiendo este movimiento, sino que, en uno de ellos se obtiene el mismo resultado inicial, y en el otro, se introduce una mayor deformación

En total, únicamente en 9 casos el algoritmo de registro analizado no actúa correctamente, siendo en total en más del 90% de los casos, donde presenta un funcionamiento completamente normal.

Una vez terminada la validación cualitativa del registro, cabe destacar la dificultad de establecer una metodología de evaluación en este tipo de análisis de preprocesado. El registro de imágenes, en comparación con el análisis de los resultados obtenidos de la selección de la función AIF, no devuelve ninguna métrica cuantitativa que vaya a ser utilizada en posteriores análisis, como es la curva AIF, lo que complicaba la evaluación del algoritmo.

En este trabajo se ha intentado diseñar una metodología útil no solo en nuestro caso, sino viable y reproducible en futuras validación de algoritmo de registro con las mismas o diferentes modalidades de imagen. Para ello, se han planteado diferentes soluciones, basadas en estudios anteriores y en la experiencia del equipo de QUIBIM, de las cuales,

a partir de cometer errores y estudios con conjuntos de datos reducidos, se ha llegado hasta el método planteado.

Tal como se señala en revisiones anteriores [Zhe et al., 2012], la forma más fiable de validación es la evaluación de métricas como la exactitud, la solidez, la estabilidad y la facilidad de uso, entre otras. Probablemente, la exactitud es la medida más importante utilizada para garantizar la eficacia de los algoritmos de registro en la práctica clínica.

Sin embargo, la exactitud o la precisión solo puede medirse si se dispone de imágenes registradas de referencia, información que no se suele tener en este tipo de análisis, como ocurre en este trabajo. Aun así, existen diferentes soluciones para realizar validaciones eficaces ante estas situaciones, ya sea teniendo imágenes de referencia o no, gracias a trabajos publicados anteriormente, pueden clasificarse utilizando los criterios planteados a continuación [Zhe et al, 2012]:

Methodology	Ground truth	Data	Metric
Qualitative	Not used	Clinical data	Not used
Quantitative	Synthetic transformation	Physical phantom data	Target registration error
	Anatomical features	Digital synthetic data	Target registration error
		Clinical data	Average distance
Not used	Not used	Clinical data	Volume overlap
			Consistency
			Intensity metric

Figura 31. Clasificación de los métodos de validación del registro

Por lo tanto, según la Figura 31, en aquellas situaciones en las que no se disponga de referencia, las dos soluciones para la validación es la cualitativa, realizada en este caso mediante la inspección visual, y la evaluación cuantitativa mediante el uso de métricas de consistencia y de intensidad, siendo la más utilizada la información mutua [Razlighi et al, 2013]. Sin embargo, debido a que para el algoritmo de registro ha sido utilizado el software Elastix, tras revisar su manual [Klein, 2019], está contraindicado el uso de este tipo de parámetros, ya que, aunque puede dar una idea del funcionamiento del algoritmo, no se recomienda utilizarlo en la validación clínica con este software.

Aun siendo la metodología cualitativa suficiente para la validación clínica del registro T2W-DCE, se planteó evaluar de forma interna el algoritmo utilizando métricas cuantitativas propias. El objetivo de esto es poder verificar de forma automática la actividad del algoritmo de registro durante su aplicación, sin recurrir a la revisión de las imágenes resultantes de forma visual, pudiendo avisar al usuario en tiempo real de la introducción de un error en el análisis global de imágenes de próstata que se ofrece.

De esta forma y como se ha mostrado en la Figura 30 en el apartado *4.2.3 Evaluación interna del algoritmo de registro*, dos de los siete parámetros cuantitativos obtenidos del propio algoritmo, muestran la efectividad del método de registro analizado, ya que muestran una clara diferencia entre los registros considerados correctos e incorrectos. La representación en forma de diagrama de cajas y bigotes permite seleccionar un valor umbral que distinga estos dos grupos. Por un lado, para la variable de traslación y se ha escogido un umbral igual a 4.4, y para el determinante jacobiano, un valor igual a 0,13. Con estos límites establecidos, si se da el caso de superarlos, es posible predecir la inefectividad del registro, ya que supone la introducción de deformación o movimiento tras su aplicación. Gracias a esto, es posible informar al usuario de ello y corregir el error durante el mismo momento de preprocesado, evitando continuar el análisis de las imágenes en caso de producirse un registro incorrecto.

Capítulo 6. Conclusiones

En este trabajo se ha llevado a cabo el desarrollo de dos metodologías de validación clínica completas e independientes de dos de los algoritmos de preprocesado involucrados en el análisis de imágenes de resonancia magnética multiparamétrica en cáncer de próstata. Por un lado, el primer algoritmo analizado corresponde a la selección automática de la función de entrada arterial AIF en imágenes funcionales de perfusión DCE; y por otro, el algoritmo de registro inter secuencia entre imágenes anatómicas y funcionales correspondientes a secuencias T2W y DCE.

El algoritmo de selección automática de la AIF ha sido evaluado desde dos perspectivas diferentes, cualitativa y cuantitativamente. La primera de ellas ha consistido en clasificar de forma visual el funcionamiento del algoritmo a través de la selección de las arterias ilíacas en imágenes DCE, a partir de las cuales se obtiene la curva AIF, necesaria en análisis posteriores. De los 242 casos analizados, en más del 90%, la segmentación automática de los vasos se realizó correctamente, siendo además la similitud entre ambas clasificaciones muy elevada, indicada por el Coeficiente Jaccard igual a 0,975. Respecto a la validación cuantitativa, se ha contrastado con el método manual de referencia, mediante métricas estadísticas extraídas de la curva. La representación gráfica de estos parámetros, intensidad y tiempo de pico, han resultado en una correlación lineal y clínicamente significativa ($r > 0,8$; $p < 0,001$), además de mostrar la concordancia del método automático con el de referencia al obtener sesgos medios menores al 10% en las gráficas Bland-Altman presentadas.

Respecto al algoritmo de registro T2W-DCE, se ha realizado una validación cualitativa, en la cual, revisando individualmente los 112 casos seleccionados, se ha encontrado que en el 91,96% de los pacientes, el registro entre ambas secuencias se ha llevado a cabo correctamente. La dificultad de obtener métricas cuantitativas, que sean representativas y válidas, descartó la posibilidad de desarrollar una metodología de evaluación cuantitativa. Sin embargo, se ha realizado un control de calidad con parámetros propios del algoritmo, con el objetivo de establecer un valor límite con el que poder clasificar el funcionamiento del registro. De las siete métricas evaluadas, se han seleccionado las dos más

representativas, el determinante jacobiano y la traslación en el eje y, con umbrales iguales a 0,13 y 4,4, respectivamente.

Este trabajo presenta una de las partes más importantes del desarrollo de un producto sanitario, la validación de los algoritmos que lo componen. Antes de sacar al mercado cualquier producto, es fundamental realizar una evaluación completa de todas las funcionalidades que ofrece, mediante informes de validación en los que se demuestre de forma empírica el correcto funcionamiento de los procesos que se incluyen.

La tarea de validación de métodos de análisis de imágenes RMmp no es sencilla, sobre todo cuando no existe una metodología clara y unificada para ello. Gracias al protocolo PI-RADS, existen indicaciones a partir de las cuales es posible plantear técnicas que pueden llegar a ser reproducibles, válidas y clínicamente significativas. Aun así, se espera que, en un futuro cercano, se establezcan metodologías de referencia en el análisis de imágenes de cáncer de próstata, lo que supondrá diagnósticos y pronósticos de mayor fiabilidad, mejorando así la calidad de vida de los afectados.

Bibliografía

- [1] Asociación Española Contra el Cáncer [aecc] (2018). *Cáncer de mama*. Recuperado 16 abril 2020, de <https://www.aecc.es/es/todo-sobre-cancer/tipos-cancer/cancermama/que-es-cancer-mama>
- [2] Sociedad Española de Oncología Médica [SEOM] (2020). Las cifras del cáncer en España 2020 Hombres. *Sociedad Española de Oncología Médica*, 36.
- [3] Lemaitre, G. (2016). *Computer-Aided Diagnosis for Prostate Cancer using Multi-Parametric Magnetic Resonance Imaging*. PhD thesis, Universitat de Girona.
- [4] Parfait, S. (2010). *Classification de spectres et recherche de biomarqueurs en spectroscopie par resonance magnetique nucleaire du proton dans les tumeurs prostatiques*. PhD thesis, Universite de Bourgogne.
- [5] Harvard Health Blog (2011). *Prostate Gland - Harvard Health Blog*. Recuperado 13 de mayo de 2020, de <https://www.health.harvard.edu/blog/prostate-gland>
- [6] Choi, Y. J., Kim, J. K., Kim, N., Kim, K. W., Choi, E. K., and Cho, K.-S. (2007). Functional MR imaging of prostate cancer. *RadioGraphics*, 114(1):63-77.
- [7] Asociación Española Contra el Cáncer [aecc] (2019). *Cáncer de próstata*. Recuperado 27 de abril de 2020, de <https://www.aecc.es/es/todo-sobre-cancer/tipos-cancer/cancer-prostata>
- [8] American Cancer Society [ACS] (2019). Cáncer de próstata. Recuperado 27 de abril de 2020, en <https://www.cancer.org/es/cancer/cancer-de-prostata.html>
- [9] Turkbey, B. and Choyke, P. (2012). Multiparametric mri and prostate cancer diagnosis and risk stratification. *Current Opinion in Urology*, 22.
- [10] Gleason, D. (1977). *Urologic pathology: The prostate*, The Veteran's Administration Cooperative Urologic Research Group: histologic grading and clinical staging of prostatic carcinoma, 171198. Lea and Febiger.
- [11] Howell, J. D. (2016). Early clinical use of the x-ray. *Transactions of the American Clinical and Climatological Association*, 127:341-349.

- [12] Vilanova, J. C., Comet, J., Garcia-Figueiras, R., Barceló, J., & Boada, M. (2010). Utilidad de la resonancia magnética en el cáncer de próstata. *Radiologia*, 52(6), 513–524. <https://doi.org/10.1016/j.rx.2010.06.003>
- [13] NIH (2013). Imagen por resonancia magnética (IRM). *Instituto Nacional de Bioingeniería e Imágenes Biomédicas*, 3. www.nibib.nih.gov
- [14] Laader, A., Beiderwellen, K., Kraff, O., Maderwald, S., Wrede, K., Ladd, M. E., Lauenstein, T. C., Forsting, M., Quick, H. H., Nassenstein, K., and Umutlu, L. (2017). 1.5 versus 3 versus 7 tesla in abdominal mri: A comparative study. *PLOS ONE*, 12(11).
- [15] Sanz-Requena, R., Prats-Montalbán, J. M., Martí-Bonmatí, L., Alberich-Bayarri, Á., García-Martí, G., Pérez, R., & Ferrer, A. (2015). Automatic individual arterial input functions calculated from PCA outperform manual and population-averaged approaches for the pharmacokinetic modeling of DCE-MR images. *Journal of Magnetic Resonance Imaging*, 42(2), 477–487. <https://doi.org/10.1002/jmri.24805>
- [16] Weinreb, J. C., Barentsz, J. O., Choyke, P. L., Cornud, F., Haider, M. A., Macura, K. J., ... Verma, S. (2016). *PI-RADS Prostate Imaging - Reporting and Data System: 2015, Version 2*. *European Urology*, 69(1) 16–40. <https://doi.org/10.1016/j.eururo.2015.08.052>
- [17] Fernández, P., (2020). *Biopsia Y RMN Para Detectar Cáncer De Próstata*. Top Doctors. Recuperado 16 julio 2020, de <https://www.topdoctors.es/articulos-medicos/biopsia-y-resonancia-magnetica-en-el-cancer-de-prostata>
- [18] Tofts PS, Brix G, Buckley DL, Evelhoch JL, Henderson E, Knopp MV, et al. (1999) “Estimating kinetic parameters from dynamic contrast-enhanced T1-weighted MRI of a diffusable tracer: standardized quantities and symbols”. *J Magn Reson Imaging*; 10:223-32
- [19] Rijpkema, M., Kaanders, J. H. A. M., Joosten, F. B. M., Van Der Kogel, A. J., & Heerschap, A. (2001). Method for quantitative mapping of dynamic MRI contrast agent uptake in human tumors. *Journal of Magnetic Resonance Imaging*, 14(4), 457–463. <https://doi.org/10.1002/jmri.1207>

- [20] García, J. R., Requena, R. S., & Martínez, A. P. (2008). Influencia de la Selección de la Función de Entrada Arterial en el Modelado Farmacocinético del Síndrome de Hiperestimulación Ovárica a partir de Imágenes de Resonancia Magnética. *Libro de Resúmenes XXVI Congreso Anual de La Sociedad Española de Ingeniería Biomédica, October 2016*. <http://seib.org.es/publicaciones-cientificas-caseib/>
- [21] D. M. Koh and D. J. Collins (2007). Diffusion-weighted MRI in the body: applications and challenges in oncology. *AJR Am J Roentgenol*, 188(6):1622-1635.
- [22] T. A. Huisman (2003). Diffusion-weighted imaging: basic concepts and application in cerebral stroke and head trauma. *Eur Radiol*, 13(10):2283-2297.
- [23] Steiger, P. and Thoeny, H. C. (2016). Prostate MRI based on PI-RADS version 2: how we review and report. *Cancer Imaging*, 16(1).
- [24] Pascau, Javier & Gispert, Juan & Reig, S & Mart'inez, R & Desco, Manuel. (2002). Registro de imágenes en medicina nuclear. *Rev R Acad Cien Exact Fis Nat*. 96. 29-43.
- [25] Vos, P. C., Hambrock, T., Barenstz, J. O., and Huisman, H. J. (2010). Computer-assisted analysis of peripheral zone prostate lesions using T2-weighted and dynamic contrast enhanced T1-weighted MRI. *Physics in Medicine & Biology*, 55(6):1719-1734.
- [26] Patel, P., Chappelow, J., Tomaszewski, J., Feldman, M. D., Rosen, M., Shih, N., and Madabhushi, A. (2011). Spatially weighted mutual information (swmi) for registration of digitally reconstructed ex vivo whole mount histology and in vivo prostate mri. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6269-6272. IEEE.
- [27] Chappelow, J., Bloch, B., Rofsky, N., et al. (2011). Elastic registration of multimodal prostate mri and histology via multiattribute combined mutual information. *Medical Physics*, 38(4):2005-2008.
- [28] Fernández, M. (2019). *Desarrollo y validación de un algoritmo de registro inter e intra secuencia en resonancia magnética multiparamétrica de próstata* (Trabajo Fin de Máster). Universitat de València, Valencia.

- [29] Zhe L, Deng D, Guang-Zhi W. (2012) Accuracy validation for medical image registration algorithms: a review. *Chin Med Sci J.* 2012;27(3):176-181. doi:10.1016/s1001-9294(14)60052-3
- [30] Razlighi, Q. R., Kehtarnavaz, N., & Yousefi, S. (2013). Evaluating similarity measures for brain image registration. *Journal of Visual Communication and Image Representation*, 24(7), 977–987. <https://doi.org/10.1016/j.jvcir.2013.06.010>
- [31] Glen, S., (2016), *Jaccard Index/Similarity Coefficient*. StatisticsHowTo. Recuperado 11 agosto 2020, de <https://www.statisticshowto.com/jaccard-index/>
- [32] Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical methods in medical research*, 8(2), 135–160.
- [33] Evans, JD. et al. (1996). Straightforward statistics for the behavioral sciences. Pacific Grove, CA: Brooks/Cole Publishing.
- [34] Jacobs M. et al. (2016). Evaluation of an automated method for arterial input function detection for first-pass myocardial perfusion cardiovascular magnetic resonance. *Journal of Cardiovascular Magnetic Resonance* volume 18, Article number: 17.
- [35] Cortés-Reyes, Édgar, Rubio-Romero, Jorge Andrés, & Gaitán-Duarte, Hernando. (2010). Métodos estadísticos de evaluación de la concordancia y la reproducibilidad de pruebas diagnósticas. *Revista Colombiana de Obstetricia y Ginecología*, 61(3), 247-255. Recuperado 11 septiembre 2020, de <http://www.scielo.org.co/>
- [36] Klein S., Staring M. (2019). Elastix, the manual. Recuperado 04 septiembre 2020, de <https://elastix.lumc.nl/download/elastix-5.0.0-manual.pdf>

PARTE II
PRESUPUESTO

Introducción y objetivo

El objetivo de esta parte del trabajo es realizar una valoración económica de los recursos empleados en el desarrollo de este trabajo. Para ello, en primer lugar, se van a tratar los costes parciales, entre los que se incluyen el coste de personal y el coste material, para después obtener el presupuesto total del proyecto.

Presupuesto desglosado

Costes de Personal

El coste de mano de obra corresponde a los recursos humanos que han participado en la elaboración de este trabajo. Para realizar este cálculo, se ha tenido en cuenta el salario medio de las tres personas participantes, además del tiempo dedicado al proyecto. En la Tabla 1 se muestra el coste estimado de cada persona y el global del personal:

Tabla 11. Cuadro de precios Personal

Nº	Perfil	Uds.	Cantidad	Precio Unitario (€)	Total
1	Tutor del TFM	h	30	40.00 (€/h)	1200.00 €
2	Cotutor del TFM	h	30	40.00 (€/h)	1200.00 €
3	Autor del TFM	h	600	15.00 (€/h)	9000.00 €
				TOTAL	11400.00 €

Costes de Software

En este apartado se tratan los recursos Software empleados en el trabajo. Este tipo de recursos hace referencia a todas las plataformas utilizadas para la visualización y procesamiento de las imágenes médicas, así como herramientas de programación y aplicaciones ofimáticas de escritorio. En la Tabla 12 se recogen los costes de cada una de las licencias junto con el período y el intervalo de amortización, ya que no todas ellas se han obtenido específicamente para este proyecto:

Tabla 12. Cuadro de precios Software

Nº	Licencia	Uds.	Cantidad	Precio Unitario (€)	Período de amortización (años)	Intervalo amortizado (meses)	Total
1	Office 365	u	1	149.00 €	1 año	12 meses	149.00 €
2	Matlab 2018b	u	1	250.00 €	1 año	12 meses	250.00 €
3	ITK-Snap	u	1	0.00 €	1 año	6 meses	0.00 €
4	Elastix	u	1	0.00 €	1 año	6 meses	0.00 €
						TOTAL	399.00 €

Costes de Hardware

Por último, en la Tabla 13 se incluye el coste del único recurso Hardware utilizado:

Tabla 13. Cuadro de precios Hardware

Nº	Licencia	Uds.	Cantidad	Precio Unitario (€)	Período de amortización (años)	Intervalo amortizado (meses)	Total
1	Ordenador portátil	u	1	799.00 €	5 años	6 meses	79.90 €
						TOTAL	79.90 €

Presupuesto total

Por último, en la Tabla 14 se muestra cuál ha sido el coste total del trabajo, en el que se incluyen los costes parciales desglosado en el apartado anterior, a los que se les ha sumado los porcentajes correspondientes a los gastos generales y el beneficio industrial obtenido. A esta suma, se le añade el impuesto sobre el valor añadido (IVA) igual al 21%.

Tabla 14. Cálculo del presupuesto total del trabajo

DESCRIPCIÓN	IMPORTE
Coste de Personal	11400.00 €
Coste de Software	399.00 €
Coste de Hardware	79.90 €
Total del presupuesto recursos	11878.90 €
13% Gastos Generales	1544.26 €
6% Beneficio Industrial	712.73 €
Suma	14135.89 €
21% IVA	2968.54 €
TOTAL PRESUPUESTO	17104.43 €

El presupuesto total del presente Trabajo de Fin de Máster asciende a la cantidad de **DIECISIETE MIL CIENTO CUATRO EUROS CON CUARENTA Y TRES CÉNTIMOS**.