



Institución Universitaria

Metodología para el apoyo al diagnóstico de cáncer a partir de imágenes de resonancia magnética multiparamétrica, integrando características radiómicas y modelos de aprendizaje profundo

Diana Marcela Marín Castrillón

Instituto Tecnológico Metropolitano
Facultad de Ingenierías
Medellín, Colombia
2019

Metodología para el apoyo al diagnóstico de cáncer a partir de imágenes de resonancia magnética multiparamétrica, integrando características radiómicas y modelos de aprendizaje profundo

Diana Marcela Marín Castrillón

Tesis o trabajo de grado presentada(o) como requisito parcial para optar al título de:
Magíster en Automatización y Control Industrial

Director(a):

Ph.D. Gloria Mercedes Díaz Cabrera

Línea de Investigación:

Máquinas Inteligentes y Reconocimiento de Patrones

Grupo de Investigación:

Automática, Electrónica y Ciencias Computacionales

Instituto Tecnológico Metropolitano

Facultad de Ingenierías

Medellín, Colombia

2019

Resumen

El desarrollo de tecnologías para mejorar el diagnóstico y pronóstico de cáncer es un tema de investigación prioritario a nivel mundial. En particular, el diagnóstico por imagen ha presentado avances importantes, permitiendo la detección de tejidos tumorales en etapas tempranas de desarrollo. Sin embargo, el alto nivel de subjetividad en el diagnóstico genera una alta variabilidad inter e intra-observador, que produce un número importante de diagnósticos errados, en particular para cánceres en tejidos blandos, donde el tumor se confunde con el tejido de soporte.

En los últimos años, una tecnología para el apoyo al diagnóstico y pronóstico de cáncer a partir de imágenes denominada radiómica, ha tomado relevancia debido a sus buenos resultados. Esta tecnología aprovecha las diferentes modalidades de imagen para extraer características en diferentes niveles de representación, que son analizadas por técnicas de minería de datos para la detección de patrones asociados a la enfermedad, aún en estadios no visibles para el experto. Por lo general, el uso de estas tecnologías implica además, la segmentación de la región tumoral, lo que demanda una gran cantidad de tiempo en el proceso de lectura por parte del especialista, o una alta dependencia del método de segmentación, cuando esta se realiza automáticamente. Por otro lado, otra tecnología conocida como aprendizaje profundo, se basa en el uso de extensas arquitecturas de redes neuronales que aprenden directamente a partir de los datos observados, eliminando la necesidad de realizar una extracción o cálculo de características específicas previo al proceso de aprendizaje. Estas técnicas han mostrado resultados prometedores en el análisis y representación de información de imágenes de resonancia magnética, aun cuando la mayoría de aproximaciones han estado orientadas al análisis de información 2D extraída de tumores segmentados, desaprovechando la información 3D que proveen estas imágenes.

Teniendo en cuenta lo anterior, en esta tesis se propuso una metodología para el diagnóstico de cáncer en tejidos blandos, basada en la integración de características radiómicas y profundas, que permitan la descripción de la información 3D contenida en múltiples secuencias de resonancia magnética, obviando procesos de segmentación o extracción de características morfológicas. El desempeño de la metodología se evaluó en la tareas de clasificación de hallazgos positivos y negativos en imágenes de resonancia magnética de seno, obteniendo resultados comparables con los reportados en el estado del arte.

Palabras clave: Radiómica, Características profundas, Resonancia magnética multiparamétrica

Abstract

The development of technologies to improve the diagnosis and prognosis of cancer is a subject of priority research worldwide. In particular, medical imaging has presented significant advances, allowing the detection of tumor tissues in the early stages of development. However, the high level of subjectivity in the diagnosis generates a high inter and intra-observer variability, mainly in the soft tissue cases, due that the tumor is confused with the supporting tissue.

In recent years, radiomic has emerged as a promising technology for developing systems to support the diagnosis and prognosis of cancer based on medical images. This technology takes advantage of different imaging modalities to extract features at different levels of representation, which are analyzed by data mining techniques for pattern recognition associated with this disease, even in the early stages. Generally, the use of these technologies also requires an accurate tumor segmentation, which is time-consuming for the specialist if it is manually performed, or highly dependant on segmentation results if automatic segmentation is implemented. On the other hand, another technology known as deep learning is based on the use of extensive neural network architectures that learn directly from the observed data, eliminating the need to perform the extraction of specific features before the learning process. These techniques have shown promising results in the analysis and representation of information from magnetic resonance images, even though most of the approaches have been oriented to the analysis of 2D information extracted from segmented tumors, missing the 3D information provided by these images.

Considering the above mentioned, in this thesis we proposed a methodology for the diagnosis of soft tissue cancer, based on the integration of radiomics and deep features, which allow the description of 3D information contained in multiple magnetic resonance sequences, obviating processes of segmentation or extraction of morphological features. The performance of the methodology is evaluated in the tares of classification of positive and negative results in magnetic resonance images of the breast, obtaining results comparable with those reported in state of the art.

Contenido

Resumen	v
Lista de figuras	ix
Lista de tablas	1
1 Introducción	2
1.1 El problema del diagnóstico de cáncer basado en MRI	3
1.2 Sistemas de Apoyo al Diagnóstico por Imágenes	4
1.3 Hipótesis	5
1.4 Objetivos	5
1.4.1 Objetivo General	5
1.4.2 Objetivos Específicos	6
1.5 Estructura de la tesis	6
2 Marco Conceptual y Trabajos Previos	7
2.1 Imagen por resonancia magnética	7
2.1.1 Principio físico de la generación de imágenes por resonancia magnética	7
2.1.2 Secuencias de Resonancia Magnética	8
2.2 El flujo de trabajo del diagnóstico por resonancia magnética	11
2.2.1 Descripción de los hallazgos	12
2.2.2 Categorización de los hallazgos	12
2.3 Sistemas de apoyo al diagnóstico	13
2.3.1 Análisis basado en características radiómicas	14
2.3.2 Análisis basado en aprendizaje profundo	16
2.3.3 Transferencia de aprendizaje	21
3 Metodología	22
3.1 Extracción de características de volúmenes de MRI	23
3.1.1 Características radiómicas	24
3.1.2 Obtención de características profundas mediante transferencia de aprendi- zaje	33
3.2 Clasificación	36

3.3	Evaluación	36
4	Resultados	38
4.1	Construcción de base de datos	38
4.1.1	Búsqueda, selección y preparación de estudios	38
4.2	Selección de secuencias	39
4.3	Preparación de los volúmenes de interés	39
4.4	Distribución de la base de datos	40
4.5	Configuración experimental	40
4.6	Selección de filtros para representación basada en características profundas .	41
4.7	Clasificación con características profundas	42
4.8	Clasificación con características Radiómicas	45
4.9	Clasificación con la integración de características profundas y radiómicas . .	45
4.9.1	Significancia estadística de los resultados	47
5	Conclusiones, recomendaciones y trabajo futuro	50
	Bibliografía	53

Lista de Figuras

- 1-1. Imagen de Resonancia Magnética 4
- 1-2. Marco de trabajo de los sistemas CADx 5

- 2-1. Principios físicos de la generación de imágenes por resonancia magnética (IRM) 9
- 2-2. Secuencias de RM utilizadas en la práctica clínica 10
- 2-3. Proceso de entrenamiento de modelo de aprendizaje basado en el análisis de características radiómicas 15
- 2-4. Convolución de filtro con la imagen de entrada 17
- 2-5. Ejemplo de representación de capa de agrupamiento por máximos 18

- 3-1. Metodología propuesta para la fusión de características radiómicas y profundas en sistemas de diagnóstico automático 23
- 3-2. Matriz de co-ocurrencia con dirección 0° y distancia 1 26
- 3-3. Matriz de Longitud de Niveles de Gris en dirección 0° 28
- 3-4. Matriz de tamaño de niveles de gris en una zona 31
- 3-5. Matriz de Dependencia de Niveles de Gris 31
- 3-6. Arquitectura de ResNet50 34
- 3-7. Representación de Filtros ResNet50 34

- 4-1. Comparación entre la selección de filtros con la representación propuesta respecto y la salida del filtro completa 44
- 4-2. Desempeño de la clasificación de hallazgos positivos y negativos en MRI de mama respecto al AUC, evaluando características radiómicas y profundas por separado y la integración de ambas, tanto en la clasificación con SVM (arriba) como en la clasificación con MKL (abajo) 47

Lista de Tablas

- 2-1. Categorización de lesiones estandarizadas por el léxico BIRADS [1] 13

- 3-1. Características radiómicas de primer orden 25
- 3-2. Características radiómicas de segundo orden 27
- 3-3. Características radiómicas de orden superior provenientes de GLRLM 29
- 3-4. Características radiómicas de orden superior provenientes de GLSZM 30
- 3-5. Características radiómicas de orden superior provenientes de GLDM 32

- 4-1. Distribución de estudios de la base de datos empleada para evaluación 40
- 4-2. Desempeño respecto a la media geométrica de la selección de filtros para todas las capas. N indica la cantidad de filtros relevantes con los cuales de obtuvo el desempeño reportado. 42
- 4-3. Desempeño respecto a la media geométrica de la selección de filtros para todas las capas, utilizando el filtro vectorizado. N indica la cantidad de filtros relevantes con los cuales de obtuvo el desempeño reportado. 43
- 4-4. Filtros seleccionados para cada secuencia 43
- 4-5. Desempeño de las estrategias de clasificación evaluadas son SVM 46
- 4-6. Desempeño de las estrategias de clasificación evaluadas son MKL-SVM 46
- 4-7. Desempeños en cada fold para MKL-SVM 48
- 4-8. Resultados del T-test para las tres medidas de desempeño 49

1 Introducción

El cáncer es una patología que afecta tanto a hombres como a mujeres alrededor del mundo, y que puede ser originado en diversos tejidos o glándulas del cuerpo debido a una reproducción celular anormal [2]. En el más reciente estudio llevado a cabo por la Agencia Internacional para la Investigación en Cáncer, se estimó para 2018 una detección de 18.1 millones de nuevos casos y 9.1 millones de muertes relacionadas con estas patologías a nivel mundial [3], con lo cual es evidente la tendencia al aumento, respecto a las cifras de nuevos casos y muertes estimadas para 2012, las cuales fueron de 14.1 millones y 8,2 millones respectivamente [4]. Los tres tipos de cáncer con mayor influencia en las cifras mostradas anteriormente se presentan en tejidos blandos, siendo el cáncer de pulmón el más diagnosticado, seguido del cáncer de seno y de próstata.

Cuando esta patología se presenta en glándulas como la tiroides y próstata o en tejidos blandos como el seno puede ser tratable y tener un buen pronóstico si es diagnosticada e intervenida en etapas tempranas [5]. Por esta razón, en el ámbito clínico se ha aprovechado la evolución de la tecnología para desarrollar diferentes técnicas enfocadas al diagnóstico oportuno del cáncer, dentro de las cuales se encuentran las Imágenes por Resonancia Magnética (MRI, del inglés Magnetic Resonance Imaging), que gracias a su alta relación señal ruido y resolución espacial, ha mostrado superioridad en la detección de pequeñas anomalías en tejidos blandos, respecto a técnicas convencionales como el ultrasonido y los Rayos X [6, 7, 8], brindando además una vista del tejido en tres dimensiones con diferentes tipos de información, estructural y funcional, lo que proporciona mayor información para la toma de decisión diagnóstica por parte de los especialistas.

A pesar de las propiedades brindadas por las imágenes de resonancia magnética para el análisis de tejidos blandos, aún se reporta alta variabilidad inter e intra-sujetos, tanto en el diagnóstico como en el pronóstico del cáncer a partir de esta modalidad de imagen [9, 10] y, a pesar de que se han establecido estándares para el proceso de análisis y descripción de regiones sospechosas identificadas en las imágenes, estudios reportan que existen tasas de desacuerdo entre radiólogos cercanas a 23% en la descripción de patrones de distribución y realce interno de los hallazgos [11], particularmente en los que no son realces tipo masa. Adicionalmente, también se ha reportado un porcentaje de error hasta del 12.8% en la asignación de categoría maligna o benigna a las lesiones durante el diagnóstico. Para abordar

esta problemática, se han propuesto diferentes métodos para el análisis automático de este tipo de imágenes, con el fin de apoyar al especialista en el proceso diagnóstico y pronóstico de la enfermedad. En los últimos años, se han marcado dos tendencias principales en el desarrollo de estos sistemas; por un lado, la tecnología conocida como análisis radiómico, que hace uso de técnicas clásicas de extracción de características de imagen junto con técnicas de aprendizaje computacional, para desarrollar modelos que discriminan los hallazgos e incluso predicen su clasificación histológica [12], para lo cual, una fina segmentación o delineación del tumor es requerida. Por otro lado, los modelos de aprendizaje profundo, los cuales utilizan varias capas de procesamiento no lineal para aprender características directamente de los datos [13, 14]. Sin embargo, el desarrollo de este tipo de sistemas es un tema abierto de investigación, en el que se busca aprovechar mejor la información volumétrica y multiparamétrica que estas imágenes proveen.

1.1. El problema del diagnóstico de cáncer basado en MRI

Como ya se indicó antes, las imágenes de resonancia magnética han mostrado ser útiles para el diagnóstico de cáncer en tejidos blandos, en especial por la capacidad que estas tienen de ofrecer tanto detalles anatómicos, como características funcionales de los tejidos, a partir de modificaciones en los parámetros de adquisición. Esta sin embargo, es igualmente una de las mayores dificultades en la interpretación de estos estudios, pues a diferencia de otras modalidades de imagen, como la mamografía o la ecografía, el especialista debe observar e interpretar varias imágenes a la vez, realizando además una búsqueda o navegación tridimensional, para encontrarse por ejemplo, con hallazgos que se observan en una imagen pero no en otra, u otros con características similares en una, pero que permiten una diferenciación diagnóstica en otra. Así, la decisión sobre una lesión, suele estar asociada a las variaciones de características en conjunto y no sólo a las características aparentes (observables visualmente) en una de las secuencias de imagen, lo que aporta incertidumbre al proceso debido a la cantidad de información diferente que debe visualizarse en el proceso de lectura, contemplando además el cambio de la información a través de los cortes que contienen el hallazgo.

A manera de ilustración, se puede observar en la fila superior de la Figura **1-1** un quiste BIRADS 2, los cuales pueden verse oscuros en las secuencias ponderadas en T1 como la sustracción 2, con un tono más brillante o gris en secuencias como la ADC, y con un tono intenso o brillante en una secuencia T2, considerando que la presencia de sustancias como la sangre, revierten su apariencia, haciéndolas ver brillantes en T1 y oscuras en T2. Por otro lado, en la fila inferior de la figura, se observa una masa probablemente maligna categorizada como BIRADS 4, en la que se presenta un realce en la sustracción 2 indicando el metabolismo aumentado de la masa, una señal hipo-intensa en T2 y en un tono gris en ADC. Por otra parte, a pesar de la calidad de las MRI, algunos hallazgos no tienen un buen contraste con el

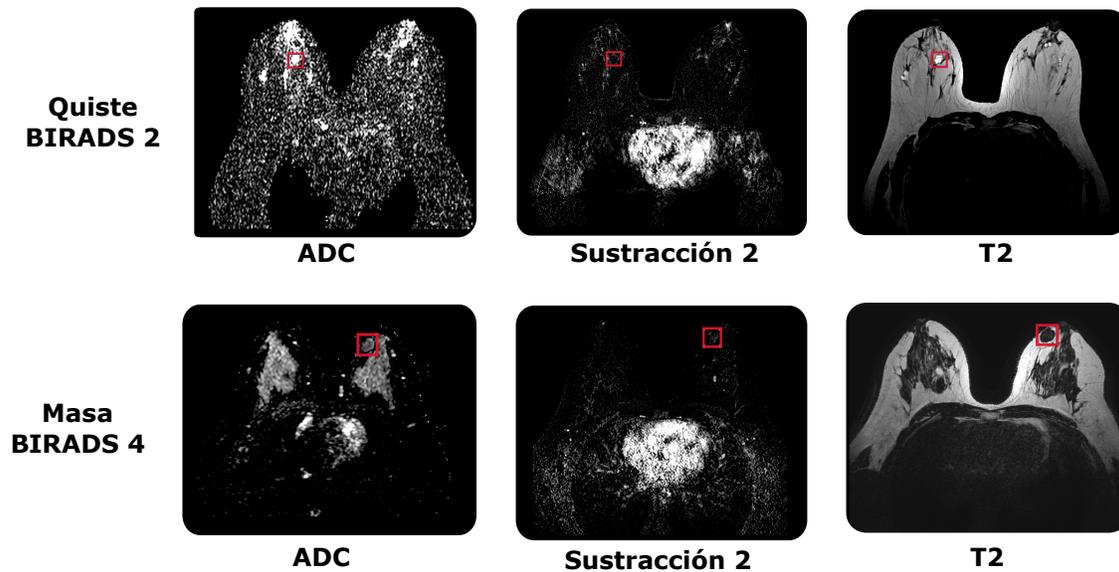


Figura 1-1: Imagen de Resonancia Magnética

tejido base, lo que dificulta la identificación a simple vista de características de la masa que se tienen en cuenta en la práctica clínica para la realización del diagnóstico, y la definición exacta de la región abarcada por el hallazgo.

Claramente, este proceso diagnóstico presenta una alta dependencia del observador y, consume bastante tiempo, aún por parte de especialistas con años de experiencia, por lo cual se reporta una importante variabilidad inter e intra-observador, en especial para los radiólogos noveles o con menor experiencia [15, 16, 17]. Por lo anterior, el desarrollo de sistemas que apoyen la interpretación de imágenes de MRI para el diagnóstico del cáncer es un tema relevante para mejorar el pronóstico de la enfermedad.

1.2. Sistemas de Apoyo al Diagnóstico por Imágenes

Con el objetivo de ayudar a los especialistas en el proceso de diagnóstico y pronóstico de patologías como el cáncer en tejidos blandos, los sistemas de apoyo guiado por computador (CAD, del inglés Computer Assisted Diagnosis), se han convertido en una herramienta potencial, dado que proporcionan información relevante en los diferentes procesos llevados a cabo en la práctica clínica durante el análisis de imágenes médicas. En general, estos sistemas pueden dividirse en dos clases, sistemas de detección asistida por computador (CADE) y sistemas de diagnóstico asistido por computador (CADx). Los CADE son sistemas orientados a la localización de hallazgos en imágenes, con el objetivo de ayudar al especialista con la identificación de Regiones de Interés (ROIs), dado que algunas de ellas son difíciles de per-

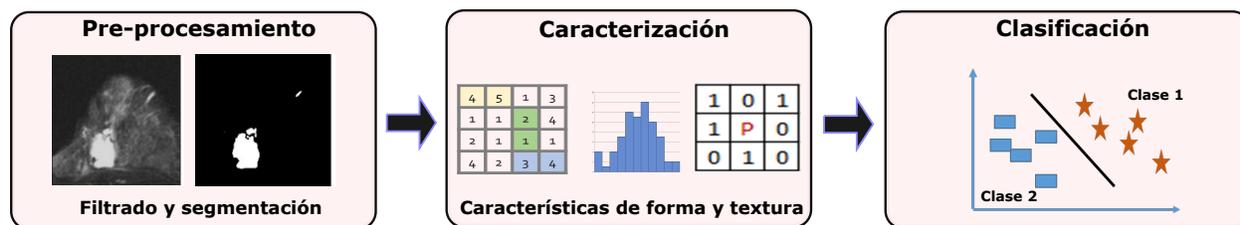


Figura 1-2: Marco de trabajo de los sistemas CADx

cibir, por su tamaño o por los límites ambiguos presentados cuando el hallazgo se contrasta con el tejido base [18]. La detección de regiones por medio de los CADE, está basada principalmente en el uso de algoritmos de segmentación por regiones, que permiten identificar los ROIs que posteriormente se clasifican como tejido normal o sospechoso, siendo estos últimos presentados al especialista como hallazgos sobre los cuales debe hacer el diagnóstico. Por otro lado, los CADx son sistemas que posterior a la identificación de las ROIs por parte de los especialistas, toman los hallazgos y los analizan a través de las etapas mostradas en la Figura 1-2, con el objetivo de presentar información adicional al especialista, que sea usada como segunda opinión para la generación del diagnóstico final, lo que contribuye con el aumento de la exactitud en la categorización de hallazgos [19]. Estos últimos tienen una mayor aceptación entre los especialistas, debido a que no funcionan como caja negra, ofreciendo cierto grado de interpretabilidad para el radiólogo, quien al final toma la decisión diagnóstica [20].

1.3. Hipótesis

La integración de características radiómicas y profundas, que representen la información volumétrica de regiones de interés provenientes de diferentes secuencias de imágenes de resonancia magnética, en un esquema único de decisión basado en aprendizaje automático, permitirá clasificar efectivamente estas regiones en categorías relevantes para el apoyo al diagnóstico de cáncer en tejidos blandos.

1.4. Objetivos

1.4.1. Objetivo General

Proponer una metodología basada en modelos de aprendizaje, para el apoyo al diagnóstico de cáncer en tejidos blandos, a partir de la representación de la información volumétrica de imágenes de resonancia magnética multiparamétrica, que integre esquemas de representación basado en características radiómicas y profundas.

1.4.2. Objetivos Específicos

1. Desarrollar una estrategia de caracterización volumétrica de imágenes de resonancia magnética, que integre características radiómicas y profundas, y que permita describir regiones de interés que puedan contener lesiones asociadas a cáncer en tejidos blandos.
2. Desarrollar una estrategia de aprendizaje para la clasificación de regiones de interés en categorías relevantes para el apoyo al diagnóstico, que integre las características de diferentes secuencias de resonancia magnética.
3. Determinar el desempeño de la arquitectura propuesta en tareas de clasificación asociadas al diagnóstico de cáncer en tejidos blandos.

1.5. Estructura de la tesis

Capítulo 1: En este capítulo se presenta una introducción a la epidemiología del cáncer y la problemática del diagnóstico basado en imágenes. Finalmente se plantea la hipótesis y se presentan los objetivos.

Capítulo 2: En este capítulo se presentan los conceptos fundamentales de la generación de imágenes de resonancia magnética, del diagnóstico de cáncer basado en imágenes y de los sistemas de apoyo al diagnóstico.

Capítulo 3: En este capítulo se presenta la metodología propuesta para el desarrollo de los objetivos planteados, haciendo énfasis en la construcción de la bases de datos, la representación de la información 3D en las diferentes secuencias de resonancia magnética, la clasificación de las regiones de interés y la evaluación de desempeño del método propuesto.

Capítulo 4: En este capítulo se presentan los resultados obtenidos en la evaluación de la metodología propuesta para el apoyo al diagnóstico de cáncer a partir de imágenes tridimensionales de resonancia magnética.

Capítulo 5: En este capítulo es presentada una discusión sobre la metodología propuesta y se formulan las conclusiones considerando los resultados para plantear trabajos futuros.

2 Marco Conceptual y Trabajos Previos

2.1. Imagen por resonancia magnética

La resonancia magnética es una técnica usada para la generación de imágenes de alta resolución sin la utilización de radiación ionizante, que además de proporcionar detalles anatómicos, proporciona información funcional con la cual se puede caracterizar la dinámica de una patología [21]. La resonancia magnética, con respecto a otras técnicas de generación de imágenes como el ultrasonido, los Rayos X y la tomografía por emisión de positrones, tiene una alta resolución, alta relación señal ruido, y presenta una mayor sensibilidad en la detección temprana del cáncer en tejidos blandos [22].

Como se mencionó anteriormente, la resonancia magnética permite obtener múltiples secuencias del tejido analizado, por medio de la modificación de parámetros en el proceso de adquisición, lo que proporciona una alta cantidad de información en la cual se ven reflejadas diferentes características del tejido. Aunque las secuencias utilizadas en el diagnóstico de cáncer, pueden variar según el órgano de interés, a continuación se presenta detalladamente el proceso de generación de imágenes por resonancia magnética, y las secuencias ampliamente utilizadas en el diagnóstico de diferentes tipos de cáncer en tejidos blandos.

2.1.1. Principio físico de la generación de imágenes por resonancia magnética

La generación de IRM, está basada en un fenómeno llamado resonancia magnética nuclear, producido cuando el núcleo de un átomo de hidrógeno interactúa con un campo magnético externo generado por un imán ubicado en el equipo de resonancia [23]. Cuando el paciente es ingresado al resonador, los protones de los átomos de hidrógeno que están organizados en direcciones aleatorias se reorganizan paralelos o antiparalelos al campo magnético. Una cantidad mayor de protones se alinean en la dirección del campo magnético externo (paralelos), con respecto a la cantidad de protones alineados en la dirección contraria al campo (antiparalelos), como lo muestra la Figura 2-1(A), dado que para esta última alineación se necesita más energía.

Los protones alineados de forma antiparalela al campo son anulados por los protones alineados paralelos en la misma dirección y dado que el número de protones paralelos al campo es mayor, después de este proceso, solo quedan protones alineados en la dirección del campo magnético externo, como se ilustra en la Figura 2-1(B).

En el proceso de la generación de las imágenes por resonancia magnética, los protones presentan dos tipos de movimiento, un giro natural al rededor de su eje, al cual se le denomina spin y un movimiento generado cuando se someten al campo magnético, el cual es llamado precesión y consiste en dar vueltas al rededor del eje que apunta en la dirección del campo magnético, es decir, el eje z, lo que da lugar a la generación de una fuerza magnética (F_m) representada por la flecha que atraviesa el protón en la Figura 2-1(C). Esta fuerza magnética tiene dos componentes, uno apuntando en la dirección de Z y otra apuntando en la dirección de X o Y, según la posición del protón (Figura 2-1(D)). Las componentes de estas fuerzas son sumadas en Z y anuladas en X y en Y con las de otros protones. Por lo cual, finalmente quedan solo las componentes de la fuerza en la dirección del campo (Z), a lo que se le llama magnetización longitudinal (M_L) (Figura 2-1(E)).

La magnetización longitudinal no puede ser medida directamente, por lo cual, se perturban los protones que están haciendo precesión en el eje Z. Para la perturbación, es utilizado un pulso de radio frecuencia con la misma frecuencia y velocidad de los protones, Figura 2-1(F). Este pulso transmite energía a los protones, ocasionando que algunos de ellos ganen la energía suficiente para alinearse antiparalelos al campo, y que tanto los paralelos como los antiparalelos se muevan alineados y sincronizados en fase, por lo cual, la componente de la fuerza magnética en X o Y no se anula, y aparece la magnetización transversal, a la vez que se disminuye la magnetización longitudinal, Figura 2-1(H).

Cuando el pulso de radiofrecuencia es desactivado, los protones tratan de volver a su posición inicial, es decir, no alineados Figura 2-1(I), por lo cual se recupera la magnetización longitudinal y se pierde la magnetización transversal, proceso en el que los protones liberan energía que es captada por las antenas de radiofrecuencia (bobinas) del resonador para la generación de las imágenes.

2.1.2. Secuencias de Resonancia Magnética

Como se mencionó anteriormente, en la práctica clínica son usadas diferentes secuencias de MRI para el diagnóstico de cáncer, debido a que cada una brinda información de diferentes características del tejido analizado, con la cual el especialista puede categorizar el hallazgo, Figura 2-2. A continuación, se describen las principales secuencias utilizadas por los especialistas para el diagnóstico y pronóstico de diferentes tipos de cáncer en tejidos blandos.

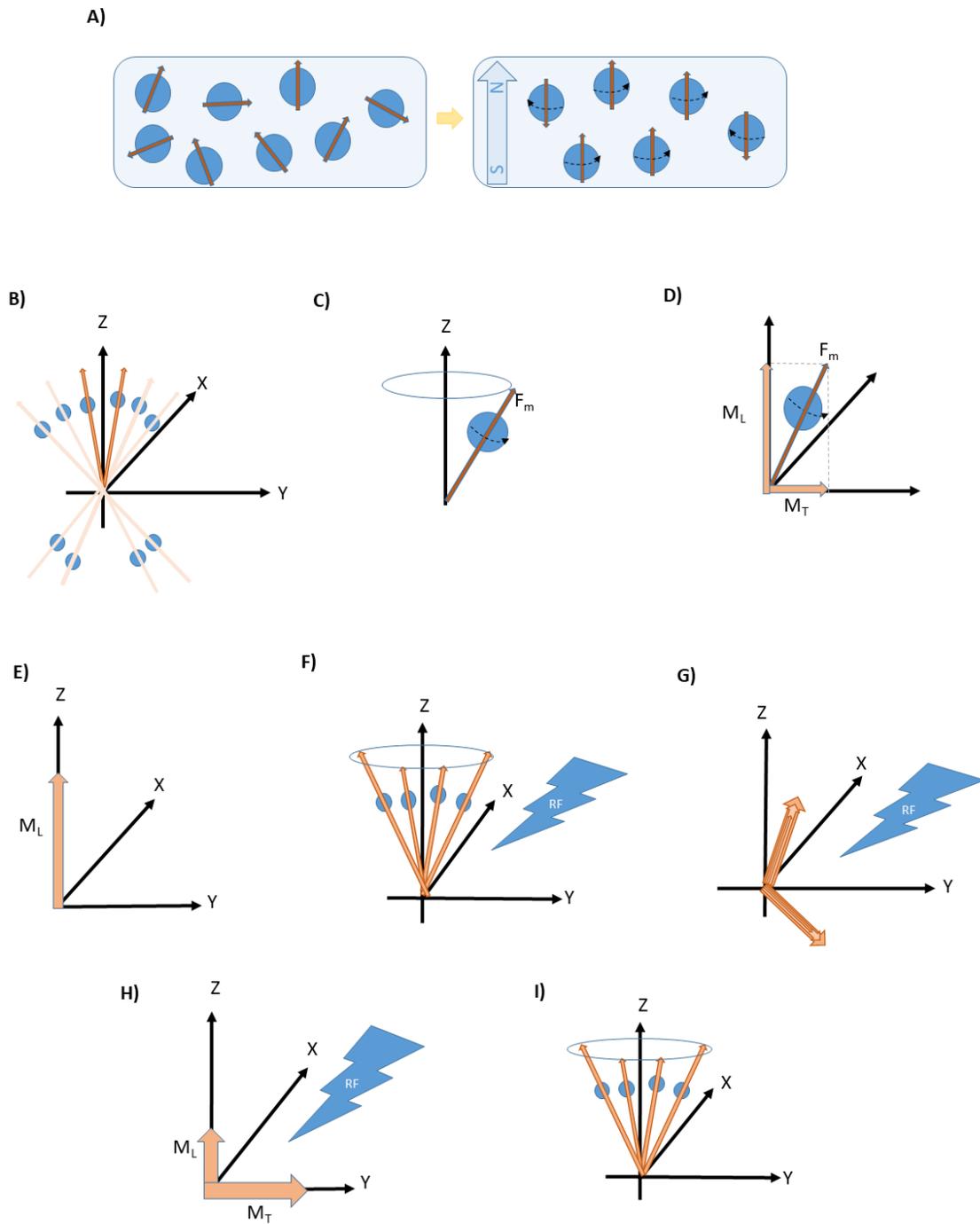


Figura 2-1: Principios físicos de la generación de imágenes por resonancia magnética (IRM)

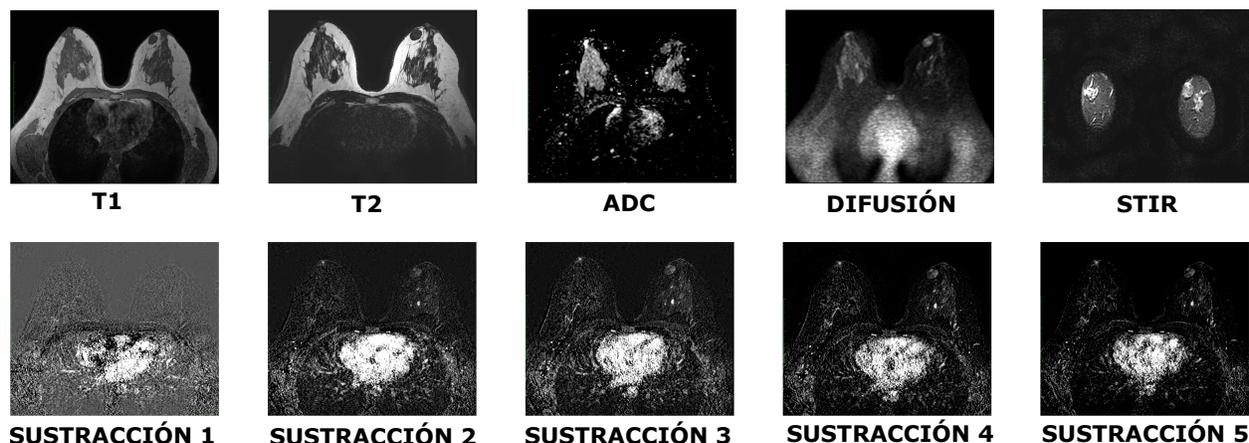


Figura 2-2: Secuencias de RM utilizadas en la práctica clínica

Relajación T1

La relajación T1 describe la recuperación del 63 % de la magnetización longitudinal a lo largo de la dirección del campo magnético al que son sometidos los protones, justo después de retirado el pulso de radiofrecuencia. Esta serie, proporciona buena correspondencia anatómica, pero es poco sensible a cambios patológicos.

Relajación T2

La relajación T2, es el tiempo necesario para que la magnetización disminuya 37% de su intensidad en el plano transversal al campo magnético al que se someten los protones, justo después de retirado el pulso de radiofrecuencia. Esta serie, es muy sensible a cambios patológicos, pero en comparación con T1, no presenta buena correspondencia anatómica.

Imagen de Difusión Ponderada (DWI)

Es una técnica derivada o ponderada en T2, en la cual se visualiza el movimiento aleatorio de las moléculas de agua de los diferentes tejidos. En un medio sin limitación estas moléculas experimentan cambios de velocidad, por lo cual, al llegar a un medio membranoso, como un tumor, puede visualizarse una restricción de dicho movimiento. Para la adquisición de esta serie, se establece un parámetro llamado b , el cual puede tomar valores muy pequeños o cero, y valores altos, por ejemplo 1000. El aumento del valor b , permite tener una mayor sensibilidad a la difusión del agua, mostrando en detalle las atenuaciones altas del movimiento de las moléculas de agua [24].

Mapas de Coeficientes de Difusión Aparente (Mapas-ADC)

Los mapas de difusión aparente son usados para cuantificar la difusión. El ADC es un valor independiente de la magnitud del campo aplicado, y se calcula para cada píxel por medio de un software, utilizando dos DWI adquiridas con diferentes valores de b (0 y 1000 generalmente). Estos valores de ADC calculados son mostrados en una imagen con diferentes intensidades de gris.

Imagen Dinámica Contrastada (DCE)

Son secuencias T1, adquiridas antes (una imagen) y después (cinco imágenes) de la administración de un medio de contraste, con las cuales se pueden visualizar características fisiológicas de tejido, dado que permiten conocer el comportamiento del mismo ante el medio de contraste, como son la cantidad captada y la velocidad con la que sale del tejido. Esta secuencia es ampliamente utilizada a pesar de su naturaleza invasiva, dado que con el análisis del metabolismo de tejido puede ser categorizados los hallazgos, teniendo en cuenta que lesiones malignas por su alta reproducción celular presentan una captación mayor del contraste en los primeros minutos, reflejado en una señal más intensa que la del tejido base o las lesiones benignas.

Sustracción

Son cinco imágenes obtenidas a partir de las imágenes de DCE, dado que son la sustracción de la imagen tomada en el primer tiempo (Pre-contraste) de cada una de las cinco tomadas posterior a la administración del medio de contraste.

STIR

La secuencia STIR (Recuperación corta de la inversión T1), es usada para anular la señal emitida por el tejido graso, lo que permite diferenciar regiones de interés con mayor facilidad. Esta serie es usualmente generada adicionando señales de radiofrecuencia de 180 grados y suprimiendo los tejidos que alcanzan T1 entre 200ms y 300ms, dado que es el tiempo T1 del tejido graso.

2.2. El flujo de trabajo del diagnóstico por resonancia magnética

En la actualidad, el diagnóstico de cáncer por resonancia magnética es realizado por un especialista después del proceso de adquisición de múltiples secuencias de RM, en las cuales identifica las regiones de interés (ROIs) para ser analizadas en la estación de trabajo. Durante

el análisis, el especialista debe realizar diferentes tareas, que incluyen la identificación de regiones sospechosas, descripción de la masa según las características visualizadas en las diferentes secuencias, diagnóstico y categorización de cada hallazgo, utilizando una guía de interpretación diseñada para el correspondiente órgano, entre las que se encuentra el BI-RADS (Breast Imaging Reporting and Data System), el PI-RADS (Prostate Imaging Reporting and Data System) y el LU-RADS (Lung Reporting and Data System), para cáncer de seno, próstata y pulmón respectivamente, ente otros.

2.2.1. Descripción de los hallazgos

Las guías de interpretación diseñadas para diferentes tipo de cáncer, tienen como objetivo principal estandarizar el proceso diagnóstico de forma que se reduzca la variabilidad inter e intra-observador. El proceso de categorización de lesiones con estos léxicos está diseñado para analizar tanto características morfológicas como características relacionadas con la captación del medio de contraste (cinéticas), que permitan determinar la probabilidad de malignidad de cada una de las lesiones tipo masa presentes en el órgano analizado, utilizando descriptores para ambos tipos de características [25].

Descriptores morfológicos

Con el objetivo de analizar la morfología de la masa, se tienen en cuenta dos características relevantes durante el proceso de análisis, la forma y la margen. La primera describe si la masa es redonda, ovalada, lobulada o irregular, y la segunda indica si el borde de la masa es regular, espiculado o irregular, siendo una margen regular indicador de una masa probablemente benigna, y una margen espiculada un indicador de una masa probablemente maligna[26].

Descriptores cinéticos

Las características cinéticas reflejan el comportamiento del tejido frente a la inyección de un medio de contraste. En un tejido fibroglandular normal, la captación de contraste se presenta poco después de la administración del medio de contraste, mostrando un aumento débil y gradual durante el transcurso de la prueba. En las masas, el realce, o aumento de la intensidad de la señal, puede ser reflejado en seis formas distintas: realce heterogéneo, realce homogéneo, realce del borde de la masa, espacios internos oscuros, realce de espacios al interior de la masa o un patrón de realce en el centro de la masa[27].

2.2.2. Categorización de los hallazgos

De acuerdo a las características morfológicas y cinéticas que describen el hallazgo, el radiólogo define una categoría, con base en una escala de probabilidad de 5 puntos; siendo 1,

Tabla 2-1: Categorización de lesiones estandarizadas por el léxico BIRADS [1]

Categoría	Diagnóstico	Prob. de malignidad(%)
0	Incompleto o técnicamente insatisfactorio	N/A
1	Negativo	0
2	Benigno	0
3	Probablemente benigno	< 2
4	Sospechoso de malignidad	3 – 94
5	Altamente sospechoso de malignidad	≥ 95
6	Malignidad comprobada con biopsia	100

un hallazgo normal, 2, un hallazgo con mayor probabilidad de benignidad y 5 un hallazgo con mayor probabilidad de malignidad. Como ejemplo, la Tabla **2-1**, presenta las categorías estandarizadas por el léxico BIRADS.

La interpretación final de un estudio corresponde a la del hallazgo de mayor categoría, aunque el reporte debe describir todos los hallazgos relevantes. Es importante señalar que la categorización de un estudio define la conducta o manejo a seguir. Si la categoría es 0, se debe recomendar estudios adicionales, ya sea de MRI o de otras modalidades de imagen. Si es 1 o 2, el paciente deberá continuar con estudios rutinarios; si la categoría es 3, se deberá realizar seguimiento en un menor lapso de tiempo al rutinario, para determinar si hay cambios significativos que conlleven a una recategorización de las lesiones; y finalmente, si la categoría es 4 o 5, se deberá realizar biopsia para confirmación de malignidad.

2.3. Sistemas de apoyo al diagnóstico

Teniendo en cuenta las variables que influyen en la práctica clínica en el diagnóstico de cáncer basado en imágenes médicas, múltiples investigaciones han sido llevadas a cabo con el objetivo de proponer y desarrollar sistemas CAD que orienten al especialista en la toma de decisiones, dándole a conocer características de los hallazgos que probablemente no son percibidas a simple vista. Para el desarrollo de dichos sistemas, diferentes técnicas de análisis de imágenes e inteligencia computacional han sido implementadas en el estado del arte [28, 29, 30]; las cuales pueden agruparse en dos grandes tendencias, la tecnología radiómica, que se fundamenta en la extracción de un conjunto amplio de características que intentan describir numéricamente las propiedades visuales de las lesiones, y el aprendizaje profundo, que intenta emular el procesamiento realizado por el neocortex, el cual no pre-procesa las señales captadas por los sentidos, sino que las propaga a través de una jerarquía de aprendizaje compleja que aprende a representar la información que permite la discriminación de escenas, con bases en las regularidades que exhiben los objetos [31]. A continuación, se presenta una revisión

sobre estas tecnologías y su aplicación en el desarrollo de sistemas CAD para el análisis de imágenes multiparamétricas.

2.3.1. Análisis basado en características radiómicas

La radiómica es una tecnología derivada del diagnóstico guiado por computador clásico, del cual se diferencia en que es un proceso para diseñar un espacio de características de alta dimensionalidad, en el que se realiza minería de datos, con el objetivo de obtener hallazgos o generar hipótesis [32]. El principal objetivo de la tecnología radiómica es el apoyo a la decisión médica en procesos de diagnóstico, pronóstico y evaluación de respuesta a tratamientos de cáncer, por medio del desarrollo de métodos que permitan la representación de la información visual y logren capturar los aspectos relevantes en cada uno de estos procesos[33].

Como se observa en la Figura 2-3, el análisis basado en tecnología radiómica está basado en varias etapas, como son: la adquisición y generación de la(s) imagen(es), la identificación de las regiones de interés, la representación de la información, que incluye la extracción y selección de características, y la implementación de los modelos de decisión. Con relación a la adquisición y generación de imágenes, generalmente se hace uso de técnicas tradicionales de adquisición como el Ultrasonido (US)[34], la Tomografía Computarizada o por Emisión de Positrones (CT o PET) [35] y la resonancia magnética (RM) [8]. En cuanto al uso de esta última modalidad de imagen para la implementación de la tecnología radiómica, es común la realización de los análisis a partir de resonancia magnética multi-paramétrica para aplicaciones en cáncer de próstata y de resonancia magnética dinámica contrastada para cáncer de seno [36]. Con respecto a la representación de información, esta consiste en procesar las imágenes a fin de obtener la mayor cantidad de información, que será luego sujeto de análisis por parte de los algoritmos de decisión. Finalmente, el último componente se basa en el uso de técnicas de minería de datos que permiten tomar decisiones con relación a la información generada a partir de la información obtenida en la etapa anterior.

Teniendo en cuenta que el componente del análisis radiómico que fue intervenido en este proyecto es la de representación de información, a continuación se describe este con mayor profundidad. En este componente, se suelen encontrar tres etapas: la segmentación de regiones sospechosas, la extracción de características que describen la información visual de las regiones, y la selección de las características relevantes para la tarea [37, 38]. Inicialmente, el objetivo de realizar una segmentación del hallazgo es garantizar que las características se orienten a describir con exactitud la información sobre la región de interés. Con este objetivo, las implementaciones de radiómica en secuencias de MRI 3D han sido llevadas a cabo principalmente en regiones sospechosas de cáncer segmentadas de forma manual o semi-automática [39, 40].

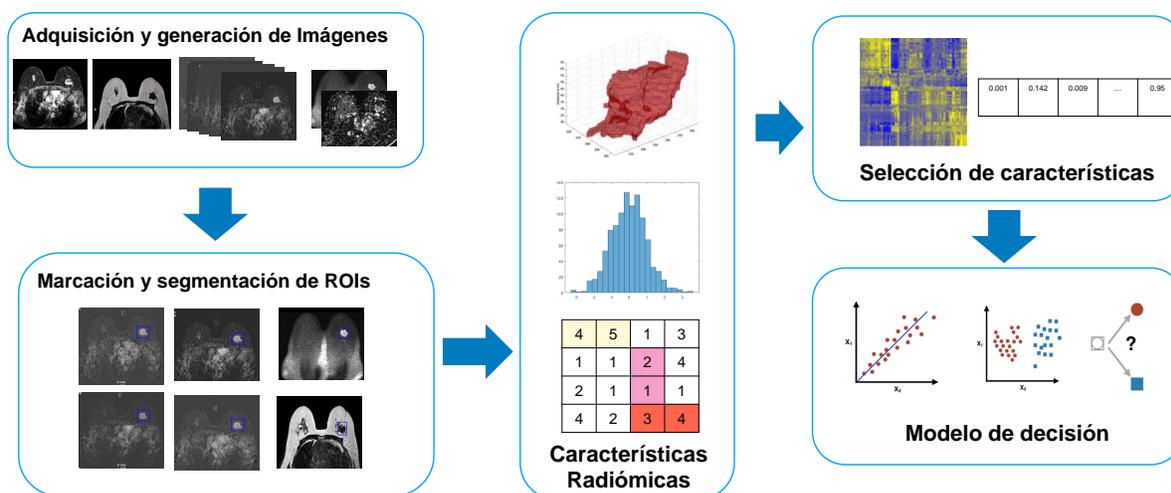


Figura 2-3: Proceso de entrenamiento de modelo de aprendizaje basado en el análisis de características radiómicas

Como su nombre lo indica, las técnicas de segmentación semi-automáticas, incluyen una participación del usuario, ya sea para establecer puntos de inicio para la segmentación, para corregir el resultado después de un proceso automático, o ambas. Entre estas técnicas se encuentran: la segmentación basada en atlas [41], y la segmentación basada en agrupamiento, como el fuzzy C-Mean [42] y la técnica de crecimiento de regiones [36]. A pesar de los esfuerzos por mejorar esta etapa, es claro que, además de que su precisión es altamente dependiente del operador, este tipo de métodos consumen una gran cantidad de tiempo del especialista, lo cual las hace poco eficientes, para procesar grandes volúmenes de información.

Con relación a la etapa de caracterización, esta es quizás la etapa más importante en el proceso de representación de la información y sobre la cual se han realizado la mayor cantidad de publicaciones en el área. La mayoría de trabajos proponen implementar métodos de extracción de características clásicos, reportados en la literatura, y agrupados en dos grandes ramas, aquellos que cuantifican la morfometría del tumor (descriptores geométricos), con base en su contorno y, aquellas que intentan describir la apariencia interna del mismo.

Los descriptores geométricos intentan cuantificar características que son tenidas en cuenta por los especialistas, entre las que se encuentran: área, volumen, radio, circularidad del tumor, entre otras [43, 44]. Estos son usados en casi la totalidad de trabajos que implementan tecnología radiómica [32], y han mostrado ser útiles para el diagnóstico de cáncer de próstata y mama [45, 41] y en la determinación de tipos de cáncer de mama [46]. Por su naturaleza, estos descriptores dependen completamente de una buena segmentación del tumor.

Por otro lado, los descriptores que intentan caracterizar el contenido interno del tumor no dependen de un proceso de segmentación; sin embargo, en la mayoría de los estudios han sido implementados posterior a este proceso. En este grupo de descriptores se encuentran, los basados en histograma, de los cuales se extraen características estadísticas de la frecuencia de los niveles de intensidad en la imagen [47], y los descriptores de textura, que brindan información de la distribución espacial de dichas intensidades. Estos últimos a su vez, están divididos en descriptores locales y globales, dependiendo de si describen la totalidad de la imagen o pequeñas regiones alrededor de un punto de interés. En los modelos de radiómica orientados a aplicaciones en cáncer de próstata, mama y pulmón, los descriptores de textura locales usados comúnmente son: los patrones binarios locales y los basados en histograma de gradiente [48, 49]. En cuanto a la estimación de características de textura a nivel global, son usadas, entre otras, diferentes técnicas de filtrado de frecuencias de la imagen, como la descomposición Wavelet, filtro Gabor, filtro kirsch y transformada discreta de coseno [41]. También es habitual el uso de descriptores de Haralick [50] y la estimación de medidas estadísticas a partir de la matriz de co-ocurrencia y la matriz de longitud de intensidades de gris consecutivas [51]. Adicionalmente, se han incluido en algunos trabajos los descriptores de textura basado en características fractales [52].

En cuanto a los modelos de decisión, la tecnología radiómica se basa en modelos clásicos de aprendizaje de máquina entre los que se destacan las máquinas de vectores de soporte o SVM (del inglés Support Vector Machine), bosques aleatorios (o random forest) y modelos de regresión logística [53, 54].

2.3.2. Análisis basado en aprendizaje profundo

En los últimos años, el aprendizaje profundo basado en redes neuronales ha sido una técnica ampliamente usada en visión artificial para el análisis de diferentes tipos de imágenes, debido a que estos modelos permiten integrar en una sola arquitectura las etapas de caracterización y decisión utilizadas en aprendizaje automático.

Redes Neuronales Profundas

Las redes neuronales profunda son técnicas de aprendizaje de máquina que se adaptan a los datos ingresados a la red por la capa de entrada, obteniendo así representaciones de estos en diferentes niveles de abstracción, a partir de las transformaciones no lineales realizadas por medio de sus capas ocultas. Estos modelos, en una forma similar a las redes neuronales convencionales o perceptrones multicapa implementan una función de activación en las múltiples neuronas agrupadas en las capas, conectando la salida de cada una con las neuronas de otras capas [55].

Entre las arquitecturas de aprendizaje profundo más utilizadas en el reconocimiento y clasificación de imágenes, se encuentran las redes neuronales convolucionales o CNN (del inglés Convolutional Neural Network), dado que estas permite tanto la estimación de características a través de sus capas convolucionales, como la clasificación a partir de estas, por medio de un perceptrón multicapa implementado al final de la arquitectura. Las CNN se componen principalmente de capas convolucionales, de agrupamiento y completamente conectadas para conformar el perceptrón multicapa.

Capa convolucional. Esta capa es el corazón de la CNN dado que permite convertir un volumen de entrada a un volumen de salida a partir de la aplicación de filtros convolucionales, donde cada valor en la matriz de salida corresponde a la respuesta al impulso del grupo de pesos W que conforman el filtro, Figura 2-4.

La salida de una neurona convolucional j está dada por la ecuación 2-1, de donde Y_j es calculada por medio de una combinación lineal de las salidas Y_i de las neuronas de la capa anterior (volumen de entrada), operando cada una de ellas con un núcleo convolucional K_{ij} correspondiente a la conexión. A esta cantidad se le suma además un peso b_j para finalmente pasarla por una función de activación $g(\cdot)$.

$$Y_j = g(b_j + \sum_i K_{ij} \otimes Y_i) \quad (2-1)$$

En este tipo de capas, deben ser considerados diferentes parámetros durante el proceso de convolución, el paso o "stride", que permite determinar el número de salto en píxeles que da el filtro para aplicarse a la imagen completa; el relleno o "padding", con el cual se indican el

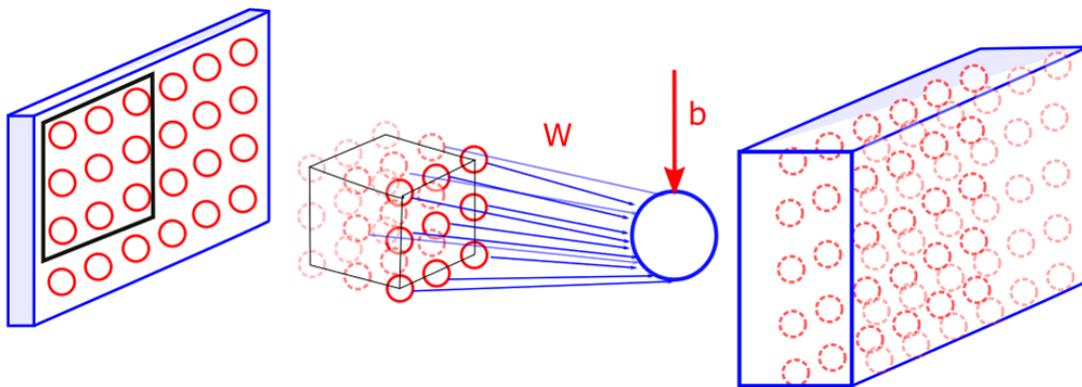


Figura 2-4: Convolución de filtro con la imagen de entrada

número de filas y columnas que se ponen en el borde de la imagen, con el fin de determinar la resolución espacial del volumen de salida; y finalmente la profundidad, con la cual se indica la cantidad de filtros o neuronas de convolución a aplicar sobre la imagen durante la estimación de características en esta capa.

Capa de agrupamiento. Esta capa permite realizar una reducción espacial del volumen de salida, para generar una representación de menor tamaño mientras se avanza por las capas de la arquitectura. Las estrategias de agrupamiento más utilizadas son el agrupamiento por máximo y el agrupamiento por promedio, en los cuales se genera un nuevo valor en la matriz de salida tomando el máximo o el promedio de un grupo de píxeles, como se ilustra en la Figura 2-5

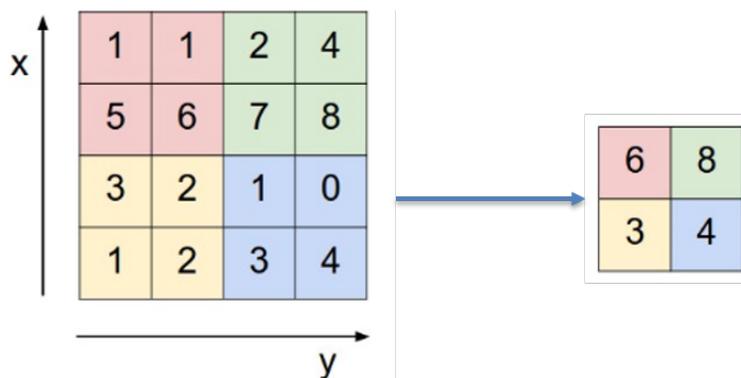


Figura 2-5: Ejemplo de representación de capa de agrupamiento por máximos

Extracción de características profundas

Teniendo en cuenta la capacidad del aprendizaje profundo y específicamente de las arquitecturas CNN para representar información contenida en imágenes, capturando de forma automática características con diferentes niveles de representación, diversos modelos de CNN han sido utilizados en el estado del arte con el objetivo de proponer sistemas de apoyo al diagnóstico que no requieran la implementación de características diseñadas con conocimiento a priori del problema, como las empleadas en el análisis radiómico.

En la práctica clínica la identificación y diagnóstico de lesiones asociadas a cáncer en tejidos blandos es llevado a cabo con la interpretación de imágenes provenientes de modalidades de imagen 3D como las de MRI, por lo cual, algunos autores han propuesto el uso de modelos de aprendizaje profundo que permitan la representación de la información 3D contenida en estas imágenes con el objetivo de contemplar los cambios presentados por el tejido en la región completa abarcada por los hallazgos. Para este fin, diferentes formas de considerar

información 3D en imágenes médicas como entrada a los modelos de aprendizaje profundo han sido presentados en 3 enfoques principales, como se describe a continuación.

Caracterización de cubo 3D que contiene el hallazgo Con el objetivo de aprovechar la capacidad de las redes neuronales profundas de estimar automáticamente características de los datos de entrada para la posterior tarea de clasificación, la mayoría de trabajos propuestos para la representación de información 3D en imágenes médicas han estado orientados al uso de redes neuronales convolucionales profundas, las cuales durante el entrenamiento encuentran los valores óptimos para múltiples filtros 3D que permiten en un proceso de convolución de estos con la imagen, estimar características relevantes para la separación entre clases en las tareas de clasificación. En este enfoque, los datos ingresados a la red, son volúmenes de interés (VOIs) extraídos de la imagen, los cuales contienen el hallazgo completo en las tres dimensiones, x, y, z, es decir, todos los cortes en los que se percibe dicho hallazgo. La mayoría de los trabajos propuestos que emplean este esquema, se orientan a tareas de clasificación relacionadas con diversas patologías cerebrales, como el Alzheimer, Parkinson o esclerosis múltiple [56, 57, 58, 59], teniendo en cuenta que, para realizar el análisis del volumen completo algunos autores segmentan el hallazgo en las tres dimensiones, es decir, lo delimitan manual o automáticamente en cada uno de los cortes en que se observa.

Esta estrategia de representación ha sido empleada en estudios de tejidos blandos para la detección y categorización de cáncer, por ejemplo en seno y pulmón. En cáncer de seno, en [60] se propuso el uso de la imagen de MRI generada con la substracción de T1 pre-contraste con T1 post-contraste pasados 1.20 segundos, segmentando los hallazgos en tres cortes, el central y los dos contiguos con el objetivo de discriminar lesiones malignas y benignas. Respecto al cáncer de pulmón, múltiples trabajos han sido publicados tanto para la detección como para la categorización de lesiones utilizando CT. En la primera aplicación, propuestas tanto para imágenes segmentadas como no segmentadas se han presentado [61, 62], proponiendo en estas últimas el uso de cubos con diferentes resoluciones espaciales para contemplar cantidades diferentes de tejido base, realizando una clasificación por votación [63] entre todas las predicciones de los modelos entrenados con cubos de diferente resolución. Para la categorización de nódulos pulmonares, también se ha realizado la aproximación descrita anteriormente, sin embargo, en [64] las características de las redes entrenadas con diferentes resoluciones fueron concatenadas en un solo vector con el objetivo de realizar la tarea de clasificación con el algoritmo de RF (Del inglés Random Forest), tomando como entrada estas características profundas.

Caracterización cortes en diferentes vistas Con el objetivo de reducir la cantidad de información procesada al analizar los cambios en 3D del tejido de interés, se ha propuesto otra técnica para la utilización y caracterización de hallazgos 3D en sistemas CAD. Esta técnica

plantea el uso de información de dos o tres cortes del hallazgo tomados desde diferentes vistas corporales, considerando como referencia el centro del hallazgo. En [65] se planteó el análisis de una vista axial y una sagital en imágenes de CT, conformando un volumen o imagen de tres canales con estas dos vistas más la imagen de proyección de máxima intensidad en PET para la detección de nódulos pulmonares. Para la clasificación de nódulos pulmonares, en [66] se propuso el uso de tres vistas (axial, coronal y sagital) caracterizadas independientemente con una CNN pre-entrenada con imágenes RGB, las características de cada corte fueron ingresadas en dos clasificadores ampliamente utilizados en sistemas de diagnóstico automático, RF y SVM, definiendo la clase del hallazgo según la mayoría de etiquetas asignadas por el clasificador a cada una de las tres vistas, obteniendo un desempeño ligeramente mayor cuando se utilizó SVM.

Características independientes en cada corte del volumen Con el objetivo de identificar micro hemorragias cerebrales en imágenes de RM, en [67] se propuso caracterizar todos los cortes en los cuales se localizan los posibles hallazgos de forma independiente utilizando una CNN 2D, esto con el fin de evitar el posible sobre ajuste de una CNN 3D ocasionado por la baja cantidad de ejemplos de entrenamiento. Para la clasificación entre verdaderos y falsos hallazgos, las características profundas generadas para cada corte del hallazgo fueron concatenadas e ingresadas a una SVM, obteniendo mejor desempeño tanto en sensibilidad y precisión comparado con RF y una CNN 3D.

Combinación de características radiómicas y profundas

Teniendo en cuenta los avances en el desarrollo de sistemas de apoyo al diagnóstico y pronóstico de cáncer a partir de los métodos de representación presentados anteriormente, algunos trabajos han sido desarrollados con el objetivo de mejorar estos sistemas mediante la integración de características provenientes de ambas tecnologías, con lo cual se han reportado como resultados preliminares mejoras respecto al desempeño de los algoritmos de clasificación. Considerando la cantidad de características profundas que pueden ser extraídas de cada una de las capas de una red convolucional, se han planteado algunos métodos de fusión de estas con características radiómicas de forma y textura, dentro de las cuales se ha explorado la selección de características con métodos como *Relief* o *Symmetric Uncertainty* para obtener vectores de hasta 5 características profundas desde uno arreglo de 4096 [68], y la eliminación de características teniendo en cuenta la desviación media absoluta y el índice C para pasar de 98304 a 1581 características consideradas como factores predictivos, aplicando por último reducción de dimensión para eliminar las características altamente correlacionadas y obtener 150 características profundas [69]. Para el diagnóstico de cáncer de seno, en [70] se propuso la fusión de clasificadores entrenados independientemente con estos dos tipos de características, evaluando esta metodología en tres diferentes modalidades de imagen (ultrasonido, mamografía y MRI). Las características radiómicas estimadas incluyeron tamaño,

forma y textura de los hallazgos y las características profundas usadas fueron tomadas de la salida de las 5 capas de agrupamiento de la arquitectura VGG19 con un tamaño de 64, 128, 256, 512 y 512 características respectivamente. Luego del entrenamiento de cada SVM con cada tipo de característica se elige por votación la etiqueta final del ROI en la modalidad de imagen evaluada, estrategia para la cual se obtuvo el mejor desempeño cuando se usaron imágenes de ultrasonido (AUC=90%).

2.3.3. Transferencia de aprendizaje

La implementación de modelos de aprendizaje profundo para el análisis de imágenes ha ido incrementando en los últimos años, dado que como se mencionó anteriormente, estos permiten realizar en un solo modelo tanto caracterización como clasificación [71]. Por lo anterior, los modelos de aprendizaje profundo son considerados caracterizadores automáticos, porque a diferencia de la tecnología radiómica, en la cual cada característica es calculada a partir de un proceso matemático independiente, las características obtenidas con aprendizaje profundo son generadas posterior a un proceso de convolución de los filtros formados con los pesos entrenados de la red. Sin embargo, el proceso de entrenamiento de algoritmos de clasificación, y en particular, los basados en redes neuronales profundas, requieren de una base de datos extensa para realizar una buena generalización, lo cual es un limitante en aplicaciones médicas, debido a que generalmente se cuenta con bases de datos pequeñas [72]. Para abordar esta limitación, se han planteado diferentes estrategias de caracterización de imágenes en sistemas de apoyo al diagnóstico, conocidas como transferencia de aprendizaje, dado que se usan modelos pre-entrenados en otras bases de datos e incluso para otras tareas de clasificación, con el fin de usar los parámetros aprendidos ya sea para ajustarlos en un nuevo proceso de entrenamiento para la tarea de clasificación deseada o para utilizar los mismos filtros de la red como caracterizadores. La transferencia de aprendizaje en la cual se realiza un reajuste de los parámetros de la red a partir de los aprendidos en otras tareas, puede ser llevada a cabo en dos formas diferentes, ajustando todos los parámetros de la red o manteniendo los parámetros de algunas de las capas (generalmente las convolucionales) para ajustar las capas restantes [73]. En la actualidad, tanto el reajuste de parámetros como la utilización directa de redes pre-entrenadas han sido ampliamente usadas en los sistemas de apoyo diagnóstico, obteniendo en algunas aplicaciones resultados comparables e incluso mejores a los obtenidos con las características diseñadas a mano; sin embargo, no se ha establecido cuál de las dos estrategias es mejor en términos generales.

3 Metodología

La Figura **3-1** ilustra la metodología propuesta para la clasificación de hallazgos identificados en estudios de resonancia magnética, para el apoyo al diagnóstico de cáncer. Inicialmente se deben identificar o seleccionar las secuencias que se van a emplear, lo cual puede responder a un análisis por conocimiento a priori, a la disponibilidad de las mismas o a necesidades propias del estudio. Sobre estas secuencias se identifican las regiones de interés, a partir de las cuales se construyen volúmenes compuestos por imágenes de cortes alrededor de la región marcada. Estos volúmenes son luego representados por dos tipos de características: radiómicas y profundas. En este punto es importante mencionar que en esta metodología no se hace uso de imágenes segmentadas o con tumores delineados, con el fin de reducir la dependencia de marcaciones precisas del tumor, por esta razón las características radiómicas no incluyen características geométricas. Adicionalmente, se propone el uso de la técnica de transferencia de aprendizaje, usando la red ResNet50 pre-entrenada, para la extracción de características profundas, que se combinan con las características radiómicas para entrenar un modelo de aprendizaje que clasifica las regiones de interés en categorías diagnósticas.

Un aspecto importante de esta estrategia es que el uso de la salida de los filtros de la red como características genera un vector de altísima dimensión; por ejemplo, si se usa la salida de la capa del primer bloque con 64 filtros, el vector de características profundas tendría 193600 dimensiones, mientras que si es la capa de 2048 filtros del último bloque, sería de 100352 dimensiones. Se proponen entonces dos mecanismos para reducir la dimensionalidad de estos vectores, en primer lugar, se obtiene una representación compacta de la respuesta a los filtros convolucionales, a partir del cálculo de los momentos estadísticos de primer orden para cada uno de ellos. Luego, se realiza un proceso de selección, que permite identificar el conjunto de filtros que mayor información aporta a la clasificación, mediante una estrategia basada en el método de aprendizaje por múltiples kernels (MKL), el cuál además permite la clasificación de las regiones a través de una SVM.

A continuación se describen los detalles de los métodos de extracción e integración de características y la estrategia de clasificación empleada.

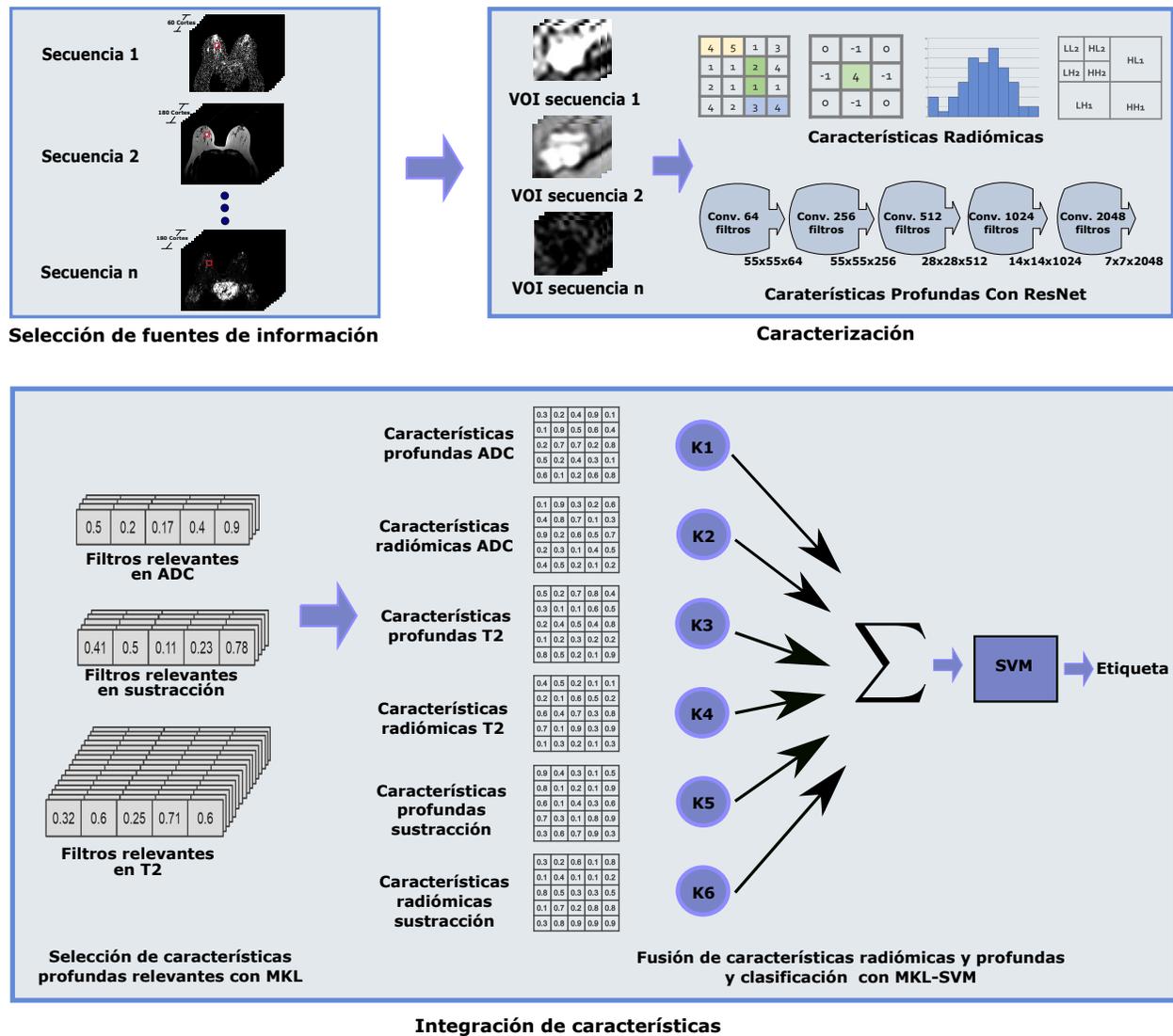


Figura 3-1: Metodología propuesta para la fusión de características radiómicas y profundas en sistemas de diagnóstico automático

3.1. Extracción de características de volúmenes de MRI

Con el objetivo de encontrar detalles propios de tejido normal o patológico por medio de una amplia y detallada descripción de la información visual contenida en las imágenes, se realiza una etapa de caracterización, en la cual, se emplean múltiples descriptores de características con los que se genera información relevante que posibilita la identificación de patrones que permitan generar hipótesis acerca del pronóstico o diagnóstico de una patología. Por lo cual, en este trabajo se planteó el uso de dos técnicas de representación de información volumétrica de hallazgos en MRI, con el objetivo de aprovechar estos dos tipos de características

generadas a través de un proceso de integración de las mismas, el cual permita a un modelo de clasificación la categorización automática de los hallazgos.

Teniendo en cuenta la revisión de trabajos previos presentada en el capítulo anterior, y con el objetivo de obtener información de los hallazgos y los cambios en el tejido circundante en 3 dimensiones. En este caso, asumiendo que se tiene la marcación de una región de interés en un corte específico del volumen, se construye un volumen conformado por tres cortes: aquel en el que se encuentra marcado el hallazgo y los dos cortes anterior y posterior a este, como se observa en el bloque de representación de información de la Figura **3-1**. A continuación, se presenta la metodología utilizada en la caracterización de VOI utilizando la tecnología radiómica y aprendizaje profundo.

3.1.1. Características radiómicas

Teniendo en cuenta el flujo de trabajo del análisis radiómico, en el cual se diseña un espacio de características de alta dimensionalidad con el objetivo de encontrar en este toda la información posible que permita la toma de decisiones en cuanto a diagnóstico y pronóstico de cáncer, se realizó una revisión del estado del arte, con el objetivo de identificar los descriptores de imagen utilizados en este contexto.

Como se presentó en el capítulo anterior, los descriptores de imagen utilizados en el marco de trabajo de radiómica se dividen en dos grandes ramas, aquellos que cuantifican la morfometría de un hallazgo con base en el contorno, y los que describen la apariencia interna del mismo. Teniendo en cuenta que este trabajo estuvo orientado al análisis de volúmenes de interés de hallazgos que no fueron finamente segmentados, los descriptores morfológicos que dependen de este tipo de segmentación no fueron implementados en el proceso, por lo cual, el análisis estuvo basado en la información cuantificada a partir de descriptores que detallan la apariencia interna de la región en la que se localiza el hallazgo. En concreto, fueron implementados descriptores de primer orden, segundo orden y de orden superior, como se describe a continuación. En este trabajo, estos descriptores son implementados en su extensión 3D, empleando la implementación provista por una herramienta desarrollada en python llamada Pyradiomics en su versión 1.3.0 (<https://www.radiomics.io/pyradiomics.html>), la cual es ampliamente utilizada por la comunidad científica para la extracción de características en imágenes médicas.

Descriptores de primer orden

Los descriptores de imagen de primer orden o basados en histograma, son descriptores de textura de bajo nivel, que brindan información de la frecuencia de cada nivel de intensidad i

Tabla 3-1: Características radiómicas de primer orden

Características de primer orden		
Nombre	Formula	Descripción
Percentil 10	10^{th} Percentil de X	
Percentil 90	10^{th} Percentil de X	
Energía	$\sum_{i=1}^{N_p} (X(i) + c)^2$	Cuantifica la uniformidad de la intensidad de $P(i)$
Entropía	$-\sum_{i=1}^{N_g} p(i) \log_2(p(i) + \epsilon)$	Relacionada con la aleatoriedad de los valores de gris en el ROI
Rango intercuartil	$P_{75} - P_{25}$	Diferencia entre el mayor y menor cuartil
Curtosis	$\frac{\sqrt{\frac{1}{N_p} \sum_{i=1}^{N_p} (X(i) - \bar{X})^4}}{\sigma^4}$	Mide el pico de la de la distribución de valores
Máximo	$máximo(X)$	Es el nivel de gris con mayor intensidad en la imagen
Desviación absoluta media	$\frac{1}{N_p} \sum_{i=1}^{N_p} X(i) - \bar{X} $	Distancia media de todos los valores de intensidad al valor medio
Media	$\frac{1}{N_p} \sum_{i=1}^{N_p} X(i)$	Intensidad media de niveles de gris en el VOI
Mediana	Valor gris	Mediana de los niveles de intensidad de gris en el VOI
Mínimo	$mínimo(X)$	Es el nivel de gris con menor intensidad en la imagen
Rango	$máximo(X) - múnimo(X)$	El rango de los valores de intensidad de gris en el VOI
Media absoluta robusta	$\frac{1}{N_{10-90}} \sum_{i=1}^{N_{10-90}} X_{10-90}(i) - \bar{X}_{10-90} $	Desviación media absoluta es la distancia media de todos los niveles de gris con el valor medio del VOI, con niveles de gris intermedios o iguales al percentil 10 y 90
Raíz media cuadrada	$\sqrt{\frac{1}{N_p} \sum_{i=1}^{N_p} (X(i) + c)^2}$	Representa la magnitud de los valores de VOI
Oblicuidad	$\frac{\sqrt{\frac{1}{N_p} \sum_{i=1}^{N_p} (X(i) - \bar{X})^3}}{\sigma^3}$	Mide la asimetría de la distribución de los datos sobre la media
Energía total	$V_{voxel} \sum_{i=1}^{N_p} (X(i) + c)^2$	Energía calculada por el volumen del vóxel en mm cúbicos
Uniformidad	$\sum_{i=1}^{N_g} p(i)^2$	Relacionada con la homogeneidad de la imagen
Varianza	$\frac{1}{N_p} \sum_{i=1}^{N_p} (X(i) - \bar{X})^2$	Medida de dispersión de los datos
<p>X= Conjunto de N_p voxeles incluidos en el VOI P(i)= Histograma de primer orden con N_g valores de intensidad discretos, siendo N_g el número de contenedores igualmente espaciados en el histograma diferentes de cero. p(i) = Histograma de primer orden normalizado, igual a $P(i)/N_p$</p>		

presente en la imagen, por lo cual, partiendo del histograma generado para cada volumen de interés, se realizaron medidas estadísticas, con el fin de establecer tendencias discriminantes entre imágenes con hallazgos patológicos o normales para el proceso de clasificación. A partir del histograma normalizado de intensidades de niveles de gris, denotado en la anterior tabla como $p(i)$, son estimadas las características de primer orden, para el cual se define un ancho del los contenedores del histograma igual a 64.

La Tabla **3-1** presenta las medidas estadísticas de primer orden estimadas para cada uno de los hallazgos.

Descriptores de segundo orden

Dentro de los descriptores de textura ampliamente usados en el análisis de imágenes médicas, se encuentran los basados en la matriz de co-ocurrencia, también conocidos como característi-

cas de segundo orden o Haralick, las cuales han mostrado ser relevantes para el diagnóstico de cáncer [50, 74, 75].

Las características extraídas de las matrices de co-ocurrencia brindan información de la distribución espacial de los niveles de gris presentes en una imagen [76]. Esta relación puede ser hallada de diferentes maneras, teniendo en cuenta una dirección y distancia entre los niveles de intensidad. Por ejemplo, se busca relacionar cuantas parejas de voxels en una dirección y distancia determinada existen en la imagen, por lo cual, el tamaño de la matriz de co-ocurrencia es proporcional a la cantidad de niveles de gris presentes en la imagen, como se observa en la Figura 3-2.

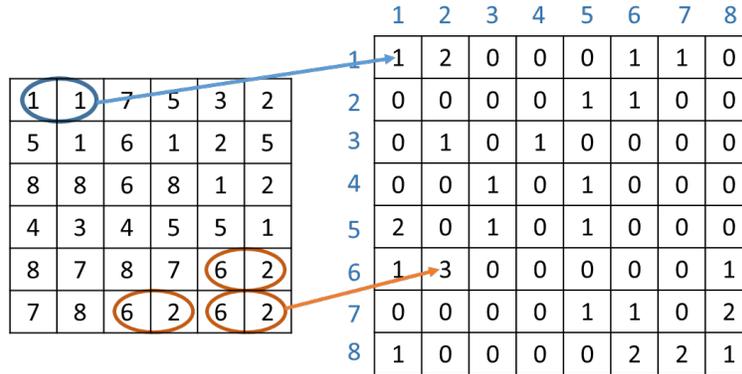


Figura 3-2: Matriz de co-ocurrencia con dirección 0° y distancia 1

La matriz de co-ocurrencia I generada, describe la función de probabilidad conjunta de segundo orden del VOI, y está dada como $P(i, j|\delta, \theta)$. Donde, (i, j) se refiere a el número de veces que se encuentra en la región la combinación de los niveles de intensidad i y j , separados por una distancia de δ voxels en una dirección θ .

Finalmente, la matriz de Co-ocurrencia normalizada está dada por:

$$p(i, j) = \frac{P(i, j)}{\sum P(i, j)} \quad (3-1)$$

Para la metodología propuesta fueron calculadas 13 matrices, una por cada dirección posible en una imagen 3D, considerando una distancia $\delta = 1$ para todas. Finalmente, las 13 matrices fueron promediadas para generar una matriz final, de la cual se calcularon las 22 características presentadas en la Tabla 3-2.

Tabla 3-2: Características radiómicas de segundo orden

Características de segundo orden		
Nombre	Formula	Descripción
Autocorrelación	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j)ij$	Magnitud de finura y tosquedad de la textura
Promedio Conjunto	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j)ij$	Intensidad de gris promedio de la distribución i
Prominencia del grupo	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i + j - \mu_x - \mu_y)^4 p(i, j)$	Medida de la oblicuidad y asimetría de la matriz
Sombra del grupo	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i + j - \mu_x - \mu_y)^3 p(i, j)$	Medida de oblicuidad y uniformidad de la matriz
Tendencia del grupo	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i + j - \mu_x - \mu_y)^2 p(i, j)$	Medida de agrupamientos de voxeles con valores de gris similares
Contraste	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - j)^2 p(i, j)$	Medida de variación local de intensidad
Correlación	$\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j)ij - \mu_x \mu_y}{\sigma_x(i) \sigma_y(j)}$	Indica la dependencia lineal de los niveles de gris a sus respectivos voxeles en la matriz
Diferencia promedio	$\sum_{k=0}^{N_g-1} k p_{x-y}(k)$	Mide la diferencia entre la aparición de pares de intensidades similares y de pares con intensidades diferentes
Diferencia de entropía	$\sum_{k=0}^{N_g-1} p_{x-y}(k) \log_2(p_{x-y}(k) + \varepsilon)$	Medida de aleatoriedad en las diferencias de intensidad en un vecindario
Diferencia de varianza	$\sum_{k=0}^{N_g-1} (k - DA)^2 p_{x-y}(k)$	Medida de heterogeneidad que asigna pesos altos pares de intensidad que se desvían más de la media
Energía conjunta	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (p(i, j))^2$	Medida de patrones homogéneos en la imagen
Entropía conjunta	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) \log_2(p(i, j) + \varepsilon)$	Mide la aleatoriedad de un vecindario de niveles de gris
Información de medida de correlación 1	$\frac{HXY - HXY1}{\max(HX, HY)}$	Evalúa la correlación entre la distribuciones de probabilidad de i y j indicando la complejidad de la textura
Información de medida de correlación 2	$\sqrt{1 - e^{-2(HXY2 - HXY)}}$	Evalúa la correlación entre la distribuciones de probabilidad de i y j
Momento de diferencia inversa	$\sum_{k=0}^{N_g-1} \frac{p_{x-y}(k)}{1+k^2}$	Medida de homogeneidad local
Diferencia inversa	$\sum_{k=0}^{N_g-1} \frac{p_{x-y}(k)}{1+k}$	Medida de homogeneidad local
Diferencia inversa normalizada	$\sum_{k=0}^{N_g-1} \frac{p_{x-y}(k)}{1 + \frac{k}{N_g}}$	Normaliza la diferencia entre los valores de intensidad vecinos
Varianza inversa	$\sum_{k=1}^{N_g-1} \frac{p_{x-y}(k)}{k}$	Varianza inversa de los niveles de gris
Probabilidad máxima	$\max(p(i, j))$	Ocurrencia de las parejas predominantes de niveles de gris
Suma entropía	$\sum_{k=2}^{2N_g} p_{x-y}(k) \log_2(p_{x+y}(k) + \varepsilon)$	Suma de la diferencia de los valores de intensidad vecinos
Suma de cuadrados	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - \mu_x)^2 p(i, j)$	Mide la distribución de una vecindad de niveles de gris respecto a la media

$\varepsilon =$ valor arbitrario $\approx 2,2 \times 10^{-16}$
 $P(i, j) =$ Matriz de co-ocurrencia en dirección δ y ángulo θ
 $N_g =$ número de niveles de intensidad discretos en la imagen
 $p_x(i) = \sum_{j=1}^{N_g} P(i, j)$ es la probabilidad marginal de la fila
 $p_{-}\{y\}(j) = \sum_{i=1}^{N_g} P(i, j)$ es la probabilidad marginal de la columna
 μ_x intensidad de gris promedio de p_x , definida como $\sum_{i=1}^{N_g} p_x(i)i$
 μ_y intensidad de gris promedio de p_y , definida como $\sum_{j=1}^{N_g} p_y(j)j$
 σ_x es la desviación estándar de p_x
 σ_y es la desviación estándar de p_y
 $HX =$ entropía de p_x
 $HY =$ entropía de p_y
 $HXY =$ entropía de $p(i, j)$
 $HXY1 = - \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) \log_2(p_x(i)p_y(j) + \varepsilon)$
 $HXY2 = - \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p_x(i)p_y(j) \log_2(p_x(i)p_y(j) + \varepsilon)$

Descriptores de orden superior

Los descriptores de orden superior, al igual que los de segundo orden, también evalúan la distribución de los niveles de intensidad en la imagen analizada como medida de textura. Sin embargo, las matrices que se generan en este caso, y de las cuales se calculan las diferentes características como la Matriz de Longitud de Niveles de Gris o GLRLM (del inglés Grey Level Run Length Matrix), la Matriz de tamaño de niveles de gris en una zona o GLSZM (del inglés Grey Level Size Zone Based Matrix) y la Matriz de Dependencia de Niveles de Gris o GLDM (del inglés Gray Level Dependence Matrix), dan cuenta de los tamaños de las zonas de gris, la cantidad de las zonas de gris y el número de voxeles conectados en una distancia específica que dependen de un voxel central, como se explica a continuación.

Matriz de Longitud de Niveles de Gris o GLRLM: Esta matriz cuantifica las corridas de niveles de gris, las cuales son definidas como la longitud en píxeles, dando cuenta de los píxeles consecutivos con el mismo nivel de gris. Esta matriz es denotada como $P(i, j|\theta)$, de donde el $(i, j)^{th}$ elemento indica el número de corridas con intensidad de gris i y longitud j presentes en la imagen a lo largo de la dirección θ .

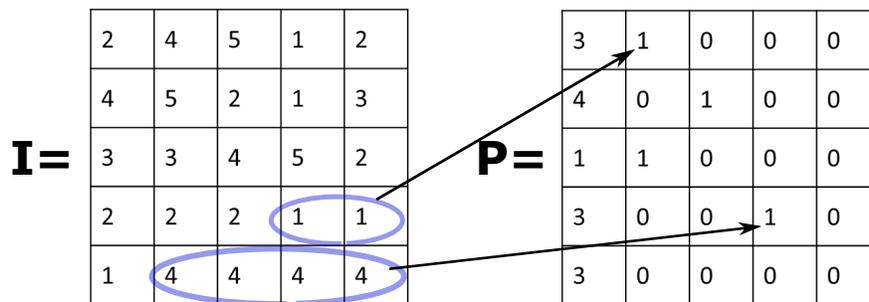


Figura 3-3: Matriz de Longitud de Niveles de Gris en dirección 0°

En la metodología propuesta, se generan 13 matrices, una por dirección y se promedian sus resultados para calcular las 16 características presentadas en la Tabla **3-3**

Matriz de tamaño de niveles de gris en una zona o GLSZM: Esta matriz cuantifica las zonas de nivel de gris en una imagen, la cual se define como el número de voxeles conectados que comparten la misma intensidad de gris. En la matriz GLSZM, definida como $P(i, j)$ el $(i, j)^{th}$ elemento indica la cantidad de zonas con intensidad i y tamaño j que tiene la imagen. En la metodología propuesta, a partir de la matriz GLSZM generada se calculan las 16 características presentadas en la Tabla **3-4**

Tabla 3-3: Características radiómicas de orden superior provenientes de GLRLM

Características de orden superior GLRLM		
Nombre	Formula	Descripción
No uniformidad de niveles de gris	$\frac{\sum_{i=1}^{N_g} (\sum_{j=1}^{N_r} P(i,j \theta))^2}{N_r(\theta)}$	Mide la similitud de los niveles de intensidad en la imagen
No uniformidad de niveles de gris normalizada	$\frac{\sum_{i=1}^{N_g} (\sum_{j=1}^{N_r} P(i,j \theta))^2}{N_r(\theta)}$	Mide la similitud de los niveles de intensidad en la imagen
Varianza de niveles de gris	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i,j \theta)(i - \mu)^2$	Mide la varianza de las intensidades de la imagen
Énfasis en niveles de gris altos	$\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} P(i,j \theta)i^2}{N_r(\theta)}$	Mide la distribución de los niveles de gris altos
Énfasis en longitud	$\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} P(i,j \theta)j^2}{N_r(\theta)}$	Medida de la distribución del largo de las longitudes
Longitud de niveles de gris altos	$\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} P(i,j \theta)i^2j^2}{N_r(\theta)}$	Mide la distribución de los valores de intensidad altos
Longitud de niveles de gris bajos	$\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{P(i,j \theta)j^2}{i^2}}{N_r(\theta)}$	Mide la distribución de los valores de intensidad bajos
Énfasis en niveles de gris bajos	$\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{P(i,j \theta)}{i^2}}{N_r(\theta)}$	Mide la distribución de los niveles de gris bajos
Entropía	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} P(i,j \theta) \log_2(p(i,j \theta) + \epsilon)$	Mide la aleatoriedad en la distribución de longitudes de niveles de gris
Longitud de no uniformidad	$\frac{\sum_{j=1}^{N_r} (\sum_{i=1}^{N_g} P(i,j \theta))^2}{N_r(\theta)}$	Mide la similitud de las longitudes a través de la imagen
Longitud de no uniformidad normalizada	$\frac{\sum_{j=1}^{N_r} (\sum_{i=1}^{N_g} P(i,j \theta))^2}{N_r(\theta)^2}$	Mide la similitud de las longitudes a través de la imagen
Porcentaje	$\frac{N_r(\theta)}{N_p}$	Mide la aspereza de la textura tomando el número de corridas con el de voxels en el VOI
Varianza	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i,j \theta)(j - \mu)^2$	Medida de varianza de las corridas
Énfasis en longitudes cortas	$\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{P(i,j \theta)}{j^2}}{N_r(\theta)}$	Mide la distribución de las longitudes cortas
Niveles de gris altos con longitud corta	$\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{P(i,j \theta)i^2}{j^2}}{N_r(\theta)}$	Mide la distribución conjunta de longitudes cortas con los valores de gris más altos
Niveles de gris bajos con longitud corta	$\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{P(i,j \theta)}{i^2j^2}}{N_r(\theta)}$	Mide la distribución conjunta de longitudes cortas con los valores de gris más bajos
<p>N_g es el número discreto de valores de intensidad en la imagen N_r es el número discreto de longitudes en la imagen N_p es el número de voxels en la imagen $N_r(\theta)$ es el número de longitudes en la imagen a lo largo de θ calculado como $\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} P(i,j \theta)$ $P(i,j \theta)$ es la matriz en una dirección θ $p(i,j \theta)$ es la matriz normalizada calculada como $\frac{P(i,j \theta)}{N_r(\theta)}$</p>		

Tabla 3-4: Características radiómicas de orden superior provenientes de GLSZM

Características de orden superior GLSZM		
Nombre	Formula	Descripción
No uniformidad de niveles de gris	$\frac{\sum_{i=1}^{N_g} (\sum_{j=1}^{N_s} P(i,j))^2}{N_z}$	Mide la variabilidad de los valores de las intensidades de gris
No uniformidad de niveles de gris normalizada	$\frac{\sum_{i=1}^{N_g} (\sum_{j=1}^{N_s} P(i,j))^2}{N_z^2}$	Mide la variabilidad de los valores de las intensidades de gris
Varianza de niveles de gris	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} p(i,j)(i - \mu)^2$	Mide la varianza de las intensidades de gris en la zona
Énfasis en zonas de niveles de gris altos	$\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} p(i,j)i^2}{N_z}$	Mide la distribución de las intensidades con mayor nivel de gris
Énfasis en grandes áreas	$\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} P(i,j)j^2}{N_z}$	Mide la distribución de zonas de gran tamaño de área
Énfasis en áreas grandes de niveles de gris altos	$\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} P(i,j)i^2j^2}{N_z}$	Mide la proporción en la imagen de la distribución conjunta de zonas grandes con niveles de gris altos
Énfasis en áreas grandes de niveles de gris bajos	$\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} \frac{P(i,j)j^2}{N_z}}{N_z}$	Mide la proporción en la imagen de la distribución conjunta de zonas grandes con niveles de gris bajos
Énfasis en zonas con niveles de gris bajos	$\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} \frac{P(i,j)}{i^2}}{N_z}$	Mide la distribución del tamaño de las zonas de niveles de gris bajos
Tamaño de zona no uniforme	$\frac{\sum_{j=1}^{N_s} (\sum_{i=1}^{N_g} P(i,j))^2}{N_z}$	Mide la variabilidad de los volúmenes de tamaño de zona en la imagen
Tamaño de zona no uniforme normalizado	$\frac{\sum_{j=1}^{N_s} (\sum_{i=1}^{N_g} P(i,j))^2}{N_z^2}$	Mide la variabilidad de los volúmenes de tamaño de zona en la imagen
Énfasis en áreas pequeñas	$\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} \frac{P(i,j)}{j^2}}{N_z}$	Medida de distribución de zonas de tamaño pequeñas
Énfasis en áreas pequeñas de niveles de gris bajos	$\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} \frac{P(i,j)}{i^2j^2}}{N_z}$	Mide la proporción en la imagen de la distribución conjunta de zonas pequeñas con intensidades bajos
Énfasis en áreas pequeñas de niveles de gris altos	$\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} \frac{P(i,j)i^2}{j^2}}{N_z}$	Mide la proporción en la imagen de la distribución conjunta de zonas pequeñas con intensidades altos
Entropía de la zona	$-\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} p(i,j)\log_2(p(i,j) + \epsilon)$	Mide la aleatoriedad del tamaño de las zonas y niveles de gris
Porcentaje de la zona	$\frac{N_z}{N_p}$	Mide la aspereza de la textura
Varianza de la zona	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} p(i,j)(i - \mu)^2$	Mide la varianza en volúmenes del tamaño de la zona

N_g = es el número de valores de intensidades de gris en la imagen
 N_s = es el número de tamaños de zonas discretas en la imagen
 N_p = es el número de voxels en la imagen
 N_z = es el número de zonas en el VOI definido como $\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} P(i,j)$
 $P(i,j)$ es la matriz GLSZ
 $\frac{P(i,j)}{N_z}$ es la matriz normalizada

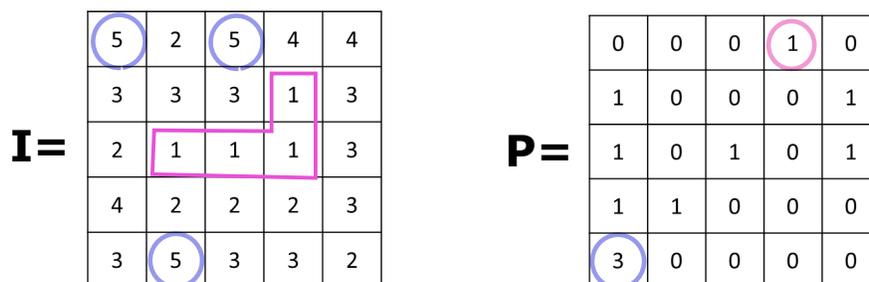


Figura 3-4: Matriz de tamaño de niveles de gris en una zona

Matriz de Dependencia de Niveles de Gris o GLDM: Esta matriz cuantifica las dependencias de niveles de gris en una imagen, definidas como el número de voxeles conectados en una distancia δ que son dependientes del voxel central. Los voxeles vecinos j , se consideran dependientes del voxel central con nivel de intensidad i si $|i - j| \leq \alpha$ definido por defecto y para esta metodología $\alpha = 0$. Esta matriz es denotada como $P(i, j)$, donde el $(i, j)^{th}$ elemento denota el número de veces que aparece un voxel con intensidad i con j voxeles dependientes en su vecindario.

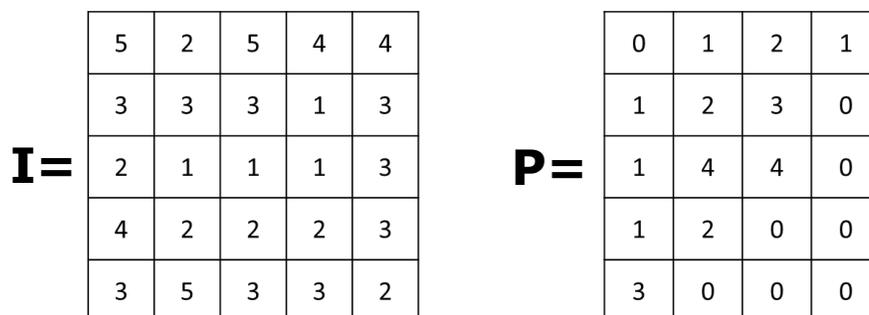


Figura 3-5: Matriz de Dependencia de Niveles de Gris

En la metodología propuesta fueron utilizadas las 14 características medidas sobre la GLDM y presentadas en la Tabla 3-5 para una distancia $\delta = 1$.

Características sobre la imagen filtrada

Con el objetivo de resaltar atributos en la imagen que permitan una mejor cuantificación de las características propias de cada clase de hallazgo, fueron implementadas dos técnicas de filtrado de imagen ampliamente utilizadas en el marco de trabajo de la radiómica, la Transformada Wavelet Discreta (DWT) y el Filtro Laplaciano del Gaussiano (LoG). Teniendo en cuenta que con la DWT se descompone la imagen de entrada en diferentes frecuencias, según los niveles de descomposición elegidos, para este trabajo se realizan 3 niveles de descompo-

Tabla 3-5: Características radiómicas de orden superior provenientes de GLDM

Características de orden superior GLDM		
Nombre	Formula	Descripción
Entropía de dependencia	$-\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} p(i, j) \log_2(p(i, j) + \epsilon)$	Medida de aleatoriedad de la dependencia
No uniformidad de dependencia	$\frac{\sum_{j=1}^{N_d} (\sum_{i=1}^{N_g} P(i, j))^2}{N_z}$	Medida de similitud de la dependencia
No uniformidad de dependencia normalizada	$\frac{\sum_{j=1}^{N_d} (\sum_{i=1}^{N_g} P(i, j))^2}{N_z^2}$	Medida de similitud de la dependencia
Varianza de la dependencia	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} p(i, j) (j - \mu)^2$	Varianza en el tamaño de la dependencia
No uniformidad de niveles de gris	$\frac{\sum_{i=1}^{N_d} (\sum_{j=1}^{N_g} P(i, j))^2}{N_z}$	Mide la similitud de los niveles de gris
Varianza de niveles de gris	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} p(i, j) (i - \mu)^2$	Varianza de niveles de gris
Énfasis en niveles de gris altos	$\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} p(i, j) i^2}{N_z}$	Medida de distribución de los valores de intensidad de gris altos
Énfasis en grandes dependencias	$\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} P(i, j) j^2}{N_z}$	Mide la distribución de las dependencias grandes
Énfasis en grandes dependencias con niveles de gris altos	$\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} P(i, j) i^2 j^2}{N_z}$	Mide la distribución conjunta de grandes dependencias con niveles de gris altos
Énfasis en grandes dependencias con niveles de gris bajos	$\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} \frac{P(i, j) j^2}{i^2}}{N_z}$	Mide la distribución conjunta de grandes dependencias con niveles de gris bajos
Énfasis en niveles de gris bajos	$\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} \frac{P(i, j)}{i^2}}{N_z}$	Mide la distribución de los niveles de gris bajos
Énfasis en dependencias pequeñas	$\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} \frac{P(i, j)}{i^2}}{N_z}$	Mide la distribución de los niveles de gris bajos
Énfasis en dependencias pequeñas con niveles de gris altos	$\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} P(i, j) i^2 j^2}{N_z}$	Mide la distribución conjunta de dependencias grandes con niveles de gris altos
Énfasis en dependencias pequeñas con niveles de gris bajos	$\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} \frac{P(i, j)}{i^2 j^2}}{N_z}$	Mide la distribución conjunta de dependencias grandes con niveles de gris bajos
<p>N_g = es el número de valores de intensidades de gris en la imagen N_d = es el número de tamaños de dependencias discretas en la imagen N_z = es el número de zonas de dependencia en la imagen definido como $\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} P(i, j)$ $P(i, j)$ es la matriz GLDM $\frac{P(i, j)}{N_z}$ es la matriz normalizada</p>		

sición utilizando la función madre Daubechies 4 y se toma el coeficiente de aproximación del tercer nivel para ser caracterizado, debido a que es este coeficiente el que retiene la información relevante de la imagen después de la descomposición. Al igual que sobre la imagen original, en el coeficiente de aproximación son estimadas las características radiómicas de primer, segundo y tercer orden descritas.

EL filtrado LoG es utilizado para realizar un proceso de suavizado en la imagen y consiste en la aplicación de una convolución entre esta y la segunda derivada de un kernel Gaussiano. El valor de sigma asociado al ancho del kernel Gaussiano fue establecido en 2, con el fin de realizar un énfasis en los cambios finos en la imagen. Al igual que las imágenes procesadas con DWT, para las imágenes resultantes de LoG se calcularon las características radiómicas de primer, segundo y tercer orden descritas anteriormente.

3.1.2. Obtención de características profundas mediante transferencia de aprendizaje

Teniendo en cuenta la capacidad de las CNN profundas de estimar de forma automática características óptimas de las imágenes para el proceso de separación de clases, en esta metodología se propuso la fusión de las características radiómicas presentadas anteriormente y de características profundas provenientes de los filtros de una CNN que mejor describen cada una de las secuencias utilizadas.

Considerando la dependencia del desempeño en las arquitecturas de aprendizaje profundo con el tamaño de la base de datos utilizada en el proceso de entrenamiento, y teniendo en cuenta que en problemas del ámbito clínico, en particular en el caso del diagnóstico del cáncer por IRM, no se suele contar con datos suficientes para llevar a cabo un proceso de entrenamiento de una arquitectura de aprendizaje profundo, en este trabajo se ha implementado transferencia de aprendizaje, utilizando una red neuronal pre-entrenada para extraer características de las imágenes en nuestra base de datos. Específicamente, en este trabajo se hizo uso de la red ResNet50, una CNN residual pre-entrenada con la base de datos de imágenes RGB llamada ImageNet (Figura**3-6**), para extraer características que representan la información de los VOI en las diferentes secuencias de MRI [77]. Esta arquitectura fue elegida debido a su buen desempeño en la tarea de clasificación original, comparada con otros modelos propuestos para el mismo fin, ya que requiere un bajo número de operaciones en el proceso de propagación hacia adelante por su naturaleza residual, es decir, al uso de atajos que permiten usar funciones identidad para sumar activaciones de dos capas no contiguas, evitando la cantidad de operaciones realizadas en el flujo de trabajo convencional de la propagación hacia adelante, y reduciendo los requerimientos de computo para la tarea a ejecutar [78].

La arquitectura de ResNet50, está diseñada para recibir como entrada tres imágenes, cada

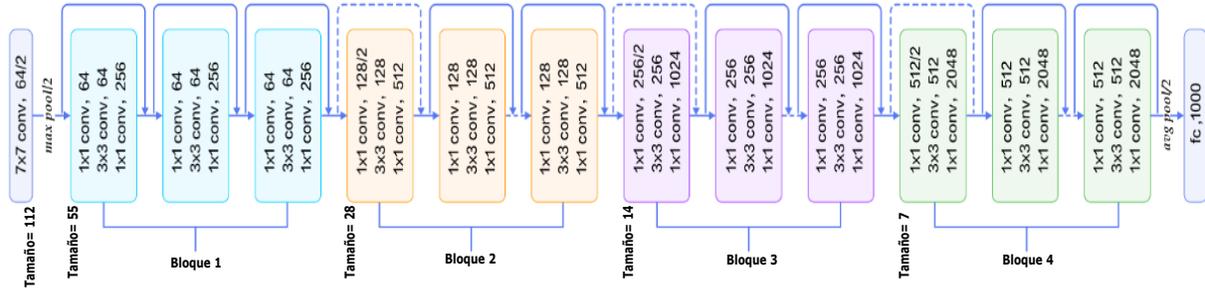


Figura 3-6: Arquitectura de ResNet50

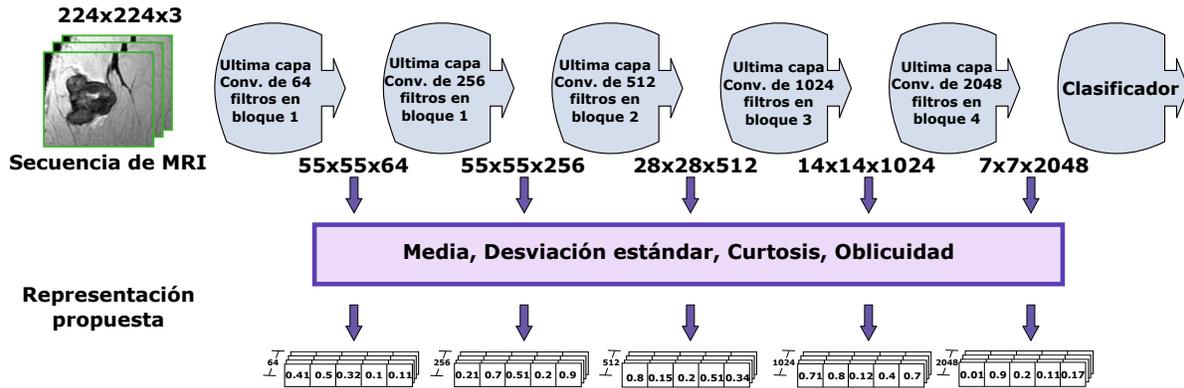


Figura 3-7: Representación de Filtros ResNet50

una de ellas correspondiente a un componente de la representación de la imagen en el modelo de color RGB, esta propiedad fue usada en este trabajo para introducir las tres imágenes que conforman los volúmenes de interés que se emplearon en la caracterización con radiomics, pero en este caso fueron redimensionados a $224 \times 224 \times 3$, por ser las dimensiones requeridas por la red. De esta manera, características profundas pueden ser obtenidas a la salida de cada una de las capas convolucionales de la arquitectura pre-entrenada, pasando a través de la red todos los VOI de cada secuencia de MRI.

Finalmente, se genera una representación compacta de estas características mediante el cálculo de los primeros momentos estadísticos (5), con el fin de mitigar problemas relacionados con la altísima dimensión de los vectores resultantes. Como se observa en la Figura 3-7, en este trabajo, se tomaron características provenientes de cinco diferentes capas convolucionales a lo largo de la arquitectura, con 64, 256, 512, 1024 y 2048 filtros, respectivamente.

Selección de filtros relevantes para la clasificación usando MKL

Una vez se obtiene la respuesta a los filtros, al pasar cada uno de los VOI por la red ResNET, se procede a seleccionar los filtros de mayor relevancia para cada una de las secuencias utilizadas, empleando el algoritmo MKL, el cual ha sido propuesto como selector de grupos de características en [79]. MKL fue implementado para generar una medida de relevancia de la información de cada filtro para la tarea de clasificación de hallazgos positivos y negativos, con el objetivo de conocer en cuál de las 5 capas se representaba mejor la información para cada secuencia, y específicamente con cuáles de los P filtros de dicha capa. Teniendo en cuenta lo anterior, para cada secuencia fue aplicado el análisis de relevancia de filtros independientemente en cada una de las 5 capas, siguiendo el flujo de trabajo de MKL presentado a continuación.

MKL permite integrar información de múltiples fuentes (en nuestro caso filtros) por medio de una medida de similitud entre todos los datos, utilizando las características provenientes de cada fuente en forma independiente. Esta medida es calculada a partir de un kernel particular asignado a cada fuente para representar su información, es decir, por cada grupo de características provenientes de un filtro es asignado un kernel que permite generar una medida de similitud entre los datos, respecto a ese grupo de características. En la metodología propuesta fue asignado el kernel Gaussiano presentado en la Ecuación (3-2), por su conocido desempeño en tareas de clasificación de imágenes.

$$K_{Gaussian}(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|_2^2}{\sigma^2}\right), \quad \sigma > 0 \quad (3-2)$$

donde, x_i y x_j son muestras o VOIs en la base de datos y σ es el ancho de banda del kernel Gaussiano, definidos en esta metodología por medio de la Optimización por enjambre de partículas o PSO [80].

La capacidad de MKL para representar e integrar información de múltiples fuentes es aprovechada por algoritmos de clasificación basados en kernel, como las SVM, dado que utilizan un kernel como medida de similitud entre datos, el cual puede ser implementado como una combinación lineal de los múltiples kernels asignados para representar cada fuente, como se muestra en la Ecuación (3-3).

$$K_{\eta}(x_i, x_j) = \sum_{m=1}^P \eta_m K_m(x_i^m, x_j^m) \quad (3-3)$$

donde, P es el número de fuentes de información, que en nuestro caso corresponde al grupo de características de cada filtro en una capa convolucional.

Como se puede observar en la Ecuación (3-3), un peso η se asigna a cada uno de los kernels, lo que implica que este también esté directamente relacionado con cada filtro. Los pesos η están en el dominio Δ mostrado en la Ecuación (3-4), y son penalizados durante el entrenamiento tomando valores iguales o cercanos a cero, cuando la relevancia del filtro al cual están asociados es baja [81], lo que permite determinar cuáles filtros en la respectiva capa analizada son relevantes para discriminar hallazgos negativos de positivos en cada secuencia.

$$\Delta = \left\{ \eta \in R_+^P : \sum_{j=1}^P \eta_j = 1, \eta_j \geq 0 \right\} \quad (3-4)$$

3.2. Clasificación

Con el fin de integrar en la tarea de clasificación las características profundas y radiómicas estimadas para cada ROI en las diferentes secuencias, fue aprovechado el algoritmo de clasificación MKL-SVM y su habilidad de representar de forma independiente la información proveniente cada grupo de características a través de una medida de similitud basada en kernel. En este sentido, fueron establecidos 6 grupos de características o fuentes de información, tres de características radiómicas en ADC, T2 y Sustracción, y tres de característica profundas para las mismas secuencias, teniendo en cuenta que como características profundas fueron utilizados los cinco momentos estadísticos de los filtros seleccionados como relevantes para cada secuencia.

3.3. Evaluación

Con el objetivo de medir el desempeño de la metodología propuesta, los resultados fueron presentados en términos de varias medidas de desempeño utilizadas en el estado del arte para algoritmos de aprendizaje de máquina, esto con el fin de evaluar el comportamiento de la metodología en la clasificación de ambos tipos de hallazgos, dada la importancia de tener sistemas CAD que sean tanto específicos como sensible. A continuación, se describen las medidas de evaluación utilizadas.

Exactitud: Esta medida indica la proporción de ejemplos que fueron predichos correctamente en el proceso de clasificación, y está denotada por la Ecuación (3-5)

$$Exactitud = \frac{VN + VP}{VN + VP + FN + FP} \quad (3-5)$$

De donde: VN = verdaderos negativos, VP = verdaderos positivos, FN = falsos negativos y FP = falsos positivos

Sensibilidad: La sensibilidad es una medida que indica la capacidad del clasificador en la predicción de los ejemplos positivos, y está denotada por la Ecuación (3-6)

$$\text{Sensibilidad} = \frac{VP}{VP + FN} \quad (3-6)$$

Especificidad: La especificidad es una medida que indica la capacidad del clasificador en la predicción de los ejemplos negativos, y está denotada por la Ecuación (3-7)

$$\text{Especificidad} = \frac{VN}{VN + FP} \quad (3-7)$$

Área bajo la curva ROC: La curva ROC o Característica Operativa del Receptor, es una curva que nos permite conocer la relación entre sensibilidad y especificidad para un clasificador binario a medida que se varía el umbral desde el cual se dice que un ejemplo es positivo. Para graficar la curva ROC son utilizadas las razones de falsos positivos y de verdaderos positivos en los ejes x e y respectivamente, o lo que es igual la sensibilidad y 1-especificidad. Para medir el desempeño de un clasificador a partir de esta curva, es calculada el área, también conocida como AUC, la cual puede tomar valores entre 0,5 y 1, lo que indica una prueba sin capacidad discriminatoria o un diagnóstico perfecto respectivamente.

4 Resultados

La metodología propuesta fue evaluada en la tarea de clasificación de hallazgos en estudios de resonancia magnética de mama, en el marco del proyecto titulado “protocolo abreviado de resonancia magnética asistido por computador para la detección y categorización de lesiones sospechosas de cáncer de mama”, el cual es desarrollado en colaboración entre el Instituto Tecnológico Metropolitano de Medellín (ITM) y el Instituto de Alta Tecnología Médica de Antioquia (IATM).

4.1. Construcción de base de datos

4.1.1. Búsqueda, selección y preparación de estudios

A partir de un estudio retrospectivo, aprobado por el comité de ética del Instituto de Alta Tecnología Médica (Acta 37 de junio 01 de 2017), se seleccionaron estudios de imágenes de resonancia magnética de mama de pacientes que acudieron al servicio de imagen de la mujer del IATM, entre enero de 2015 y diciembre de 2018, quienes por protocolo, firman consentimiento informado en el momento de la realización de la ayuda diagnóstica. Los estudios fueron seleccionados, teniendo en cuenta que cumplieran los siguientes criterios de inclusión.

- El estudio debe contener las secuencias T1 axial, T2 axial, STIR Coronal o T2 spair, sustracciones, mapas de ADC, Difusión Axial B800 o B máxima y Dinámico axial.
- La sustracción total y el dinámico deben contener una cantidad de imágenes mayor o igual a 700 y 900 respectivamente.

Un total de 100 estudios fueron incluidos en la base de datos, lo cuales fueron anonimizados para garantizar la privacidad de los pacientes, reemplazando el nombre y el número de identificación en cada imagen de cada secuencia por Breast_MRI e ID respectivamente. Cada estudio fue sometido a un proceso de interpretación por parte de dos especialistas en radiología con 10 y 6 años de experiencia en la lectura de estudios de resonancia magnética de mama, respectivamente. Durante este proceso de lectura, los especialistas marcaron cada uno de los hallazgos identificados en cada estudio utilizando un cuadro para encerrar la masa en el corte central y asignaron un nombre al hallazgo con su respectiva categoría BIRADS.

A partir de los estudios leídos, un total de 88 estudios de pacientes que presentan hallazgos entre las categorías BIRADS 1 a 5 fueron extraídos para la evaluación de la metodología propuesta.

4.2. Selección de secuencias

El protocolo de estudio de MRI de mama del IATM está conformado por diez secuencias, entre las que se incluyen T1, T2, DWI, STIR Coronal, y 6 dinámicos adquiridos tras la inyección de gadoterato de meglumina, a partir de las cuales se generan 5 secuencias de sustracción y una imagen de ADC. Sin embargo, y dado que el objetivo del proyecto señalado es el diseño de un protocolo abreviado para la detección de cáncer de mama, en esta evaluación se emplearon las secuencias que conforman tal protocolo, estas son: ADC, T2 y la segunda sustracción.

4.3. Preparación de los volúmenes de interés

Durante el proceso de lectura de los estudios para la identificación y categorización de las lesiones, los especialistas utilizan la aplicación de generación de ROIs del software Osirix para encerrar con un rectángulo cada hallazgo en la secuencia en la que es más visible. Tomando como referencia esta marcación, el ROI es proyectado a todas las secuencias por medio de una triangulación con el mismo software, la cual permite obtener la localización de una región específica del tejido analizado en todas las secuencias, para posteriormente almacenar de forma automática las coordenadas y longitud en x, y, z de cada ROI identificado en el estudio.

Con el objetivo de utilizar Pyradiomics para caracterizar solo los hallazgos identificados por los especialistas en cada estudio, se almacenaron las 3 secuencias de cada uno en formato de almacenamiento de datos crudos (nearly raw raster data or .nrrd) utilizando el software 3D Slicer y adicionalmente, se generaron las respectivas máscaras binarias 3D en este mismo formato, indicando la localización del ROI por medio de las coordenadas del centro en x, y, z de cada uno y definiendo las dimensiones del cubo en estas mismas direcciones. En la metodología propuesta, la información 3D tomada para caracterizar tanto con radiómica como con la arquitectura de red neuronal, se tomó para cada hallazgo un cuadrado con la longitud máxima entre x e y más la información en esta misma región del corte anterior y posterior al marcado. Esta aproximación para la obtención de un ROI que contenga información 3D del hallazgo es propuesta teniendo en cuenta que la resolución espacial en x, y de un voxel en una imagen de resonancia es alta, comúnmente en el orden de 0.5mm y 1mm; mientras que el espacio entre los cortes está en el orden de los 2mm y 3mm, por lo cual, para abarcar el hallazgo en z basta con dos o tres cortes del estudio.

4.4. Distribución de la base de datos

Un total de 146 volúmenes de interés fueron identificados como hallazgos en los 88 estudios de resonancia magnética de mama incluidos en la base de datos. Cada hallazgo fue inicialmente categorizado usando el léxico BIRADS; pero, considerando que el modelo de aprendizaje basado en MKL-SVM está inicialmente diseñado para resolver problemas de clasificación binarios, se decidió realizar una nueva asignación de etiquetas a las ROI, de tal forma que una clase esté asociada a la detección de tumores probablemente malignos, es decir, hallazgos positivos (clase 1), mientras que la otra se asocia a la detección de hallazgos probablemente benignos o hallazgos negativos (clase -1). Así, la base de datos incluyó 85 hallazgos positivos (BIRADS 3, 4 y 5) y 61 negativos (BIRADS 1 y 2), como lo muestra la Tabla 4-1.

Tabla 4-1: Distribución de estudios de la base de datos empleada para evaluación

Categoría BI-RADS	Número de ROI	Clase definitiva
0	0	
1	1	-1
2	60	
3	26	1
4	15	
5	44	

Es importante mencionar que, aun cuando los estudios se obtuvieron de forma retrospectiva del IATM, esta institución no cuenta con los resultados de confirmación por biopsia para estos hallazgos, lo que constituye una limitación para este trabajo. Por lo anterior, para este estudio, se tomará como "*ground truth*" la categorización asignada por el especialista.

4.5. Configuración experimental

La metodología propuesta fue evaluada para la tarea de clasificación de hallazgos en las categorías de positivo o negativo, como se describió en la distribución de la base de datos. El modelo de clasificación basado en MKL-SVM fue evaluado en tres contextos, usando sólo características radiómicas, sólo características profundas e integrando los dos grupos de características. Adicionalmente, como línea base de comparación, se evaluó el desempeño de una SVM para los mismos tres contextos, dado que este modelo de clasificación es ampliamente utilizado en el estado del arte en el diseño de sistemas de apoyo al diagnóstico; pero en este caso se usa como vector de características el vector resultante de concatenar todas las características en evaluación. En todos los experimentos fueron usados kernels gaussianos,

con los parámetros σ de cada kernel y el parámetro C de la SVM optimizados mediante el algoritmo de optimización por enjambre de partículas (PSO), usando como objetivo la Media Geométrica, la cual logra un buen ajuste entre la sensibilidad y la especificidad del método (Ecuación(4-1)).

En todos los casos se realizó una partición estratificada de los datos, en un esquema de validación cruzada de 5 iteraciones (5-fold-cross-validation). En todo caso, el desempeño de los métodos fue calculado como la media de las medidas reportadas en cada uno.

Todos los experimentos fueron implementados en un equipo de cómputo con procesador Intel de 16 núcleos a 3,00 GHz, 16 Gb de memoria RAM y una unidad de procesamiento gráfico o GPU (por sus siglas en inglés) Geforce GTX 1080 ti.

$$\text{Media geométrica} = \sqrt{\text{Sensibilidad} * \text{Especificidad}} \quad (4-1)$$

4.6. Selección de filtros para representación basada en características profundas

Antes de proceder con la clasificación de los hallazgos, se seleccionaron los filtros que serían empleados para la extracción de las características profundas para cada una de las secuencias de MRI, aplicando el siguiente procedimiento para cada una de las 5 capas definidas en la metodología. El proceso aplicado consistió en tomar las características de los VOIs provenientes de una de las capas para una secuencia, con el objetivo de realizar el análisis de relevancia sobre los P filtros de dicha capa. En la primera ejecución del algoritmo, es obtenida la relevancia de cada filtro, la cual es tenida en cuenta para eliminar posteriormente uno a uno los filtros menos relevantes, midiendo cada vez el desempeño de la tarea de clasificación respecto a la media geométrica para conocer el menor número de filtros que puede ser utilizado para obtener el mejor desempeño para cada capa. La Tabla 4-2 presenta el número de filtros que reportan el mejor desempeño para cada una de las capas evaluadas. Para cada secuencia se seleccionó la capa que reporto la mejor relación desempeño-número de filtros, teniendo en cuenta que lo que se busca con este proceso es reducir la dimensión del descriptor.

Con el objetivo de evaluar la representación compacta propuesta, respecto a la representación completa, basada en la vectorización de la respuesta a cada filtro, se evaluó el desempeño del proceso de selección de filtros para esta representación. Así, 30 pruebas fueron ejecutadas, considerando las 3 secuencias, las 5 capas y los dos tipos de características.

La Figura 4-1 presenta los resultados de las dos pruebas para cada secuencia en la capa con mejor desempeño, mostrando el desempeño en la clasificación respecto a la cantidad de filtros utilizados organizados de mayor a menor relevancia. Como se puede observar, cuando son utilizados las características completas vectorizadas, se obtiene en general un desempeño menor respecto al de los momentos y el algoritmo no se estabiliza y no permite la determinación de la cantidad de filtros relevantes en la caracterización de cada secuencia. Sin embargo, cuando se utiliza la representación propuesta, se puede observar para cada secuencia que a partir de un punto de corte x el desempeño no mejora si se adicionan durante la clasificación características provenientes de más filtros, lo que permite establecer para cada secuencia el número de filtros a utilizar en el proceso de fusión de características radiómicas y profundas.

En la Tabla 4-4 se indican los filtros seleccionados como relevantes en la representación de hallazgos positivos y negativos en cada secuencia de MRI, utilizando la representación propuesta.

4.7. Clasificación con características profundas

Antes de la implementación del método propuesto, se intentó realizar transferencia de aprendizaje por ajuste de pesos de algunas de las capas de la arquitectura de ResNet-50, por lo cual se mantuvieron los pesos de todas las capas de la red excepto las del perceptrón, para las cuales se realizó el reajuste para cada secuencia de forma independiente, obteniendo un desempeño en validación del 60 %, 50 %, y 66.7 % en ADC, T2 y Sustracción respectivamente. Para los tres casos, el desempeño de entrenamiento fue alrededor de 96 %, lo que indica un sobre ajuste de la red. Adicionalmente, se intentó ajustar los pesos de la red completa para las mismas tres secuencias, tomando como pesos iniciales los obtenidos para ResNet-50 durante el entrenamiento con ImageNet, Al igual que en el método anterior fue aplicada validación cruzada con 5 folds y fue tomada como medida de desempeño la exactitud, que para ninguna de las tres secuencias superó en 50 %.

Como se mencionó anteriormente, una de las pruebas de clasificación de hallazgos positivos

Tabla 4-2: Desempeño respecto a la media geométrica de la selección de filtros para todas las capas. N indica la cantidad de filtros relevantes con los cuales de obtuvo el desempeño reportado.

Desempeño de la selección de filtros con representación basada en momentos estadísticos										
Secuencia	64 filtros	N	256 filtros	N	512 filtros	N	1024 filtros	N	2048 filtros	N
ADC	66.52 %	4	63.9 %	4	67.18 %	489	67.69 %	163	69.94 %	84
T2	74.54 %	8	70.07 %	81	83.54 %	16	64.87 %	847	76.54 %	6
Sustracción	84.15 %	5	75.11 %	75	83.54 %	7	83.89 %	7	78.71 %	7

Tabla 4-3: Desempeño respecto a la media geométrica de la selección de filtros para todas las capas, utilizando el filtro vectorizado. N indica la cantidad de filtros relevantes con los cuales se obtuvo el desempeño reportado.

Desempeño de la selección de filtros con activaciones vectorizadas										
Secuencia	64 filtros	N	256 filtros	N	512 filtros	N	1024 filtros	N	2048 filtros	N
ADC	61.20 %	64	58.18 %	256	57.19 %	32	54.75 %	190	56.79 %	55
T2	71.49 %	64	59.33 %	3	68.57 %	245	75.61 %	15	80.94 %	44
Sustracción	56.12 %	48	46.10 %	34	56.00 %	28	79.39 %	132	82.45 %	19

Tabla 4-4: Filtros seleccionados para cada secuencia

Secuencia	Índices de filtros seleccionados	Layer
ADC	22-30-44-57	Capa convolucional con 64 filtros
T2	49-66-146-148-150-160 185-189-216-222-240-287 314-331-383-426	Capa convolucional con 512 filtros
Sustracción 2	9-24-37-39-41	Capa convolucional con 64 filtros

y negativos estuvo orientada a la utilización de MKL-SVM y una SVM convencional como modelos de clasificación, tomando solo las características profundas estimadas sobre las tres secuencias. Para el primer modelo, los momentos estadísticos provenientes de los filtros seleccionados como relevantes para cada secuencia fueron concatenados de forma independiente, obteniendo así 3 fuentes de información para la clasificación con MKL-SVM con 20, 25 y 80 características profundas de ADC, sustracción 2 y T2 respectivamente. En el caso de la línea base, estas mismas características fueron utilizadas concatenando en una sola fuente de información las características de todos los filtros relevantes de las 3 secuencias, dado que con la SVM convencional la similitud entre los datos es cuantificada con un único kernel que relaciona todas las características, con lo cual se obtuvo una matriz de 146 datos por 125 características. En este sentido, es importante aclarar que con la representación con los 5 momentos estadísticos de las características provenientes de la red se conservan los niveles de abstracción obtenidos a través de las capas, dado que lo que permiten es representar las tendencias de estas características. El desempeño promedio en los 5 folds para la clasificación con MKL-SVM respecto a exactitud, sensibilidad, especificidad y AUC fue de 80,78 %, 82,35 %, 78,72 % y 81,10 % y para la línea base planteada con SVM, el desempeño promedio presentado fue de 72,67 %, 81,18 %, 60,77 % y 76,20 % para las mismas medidas.

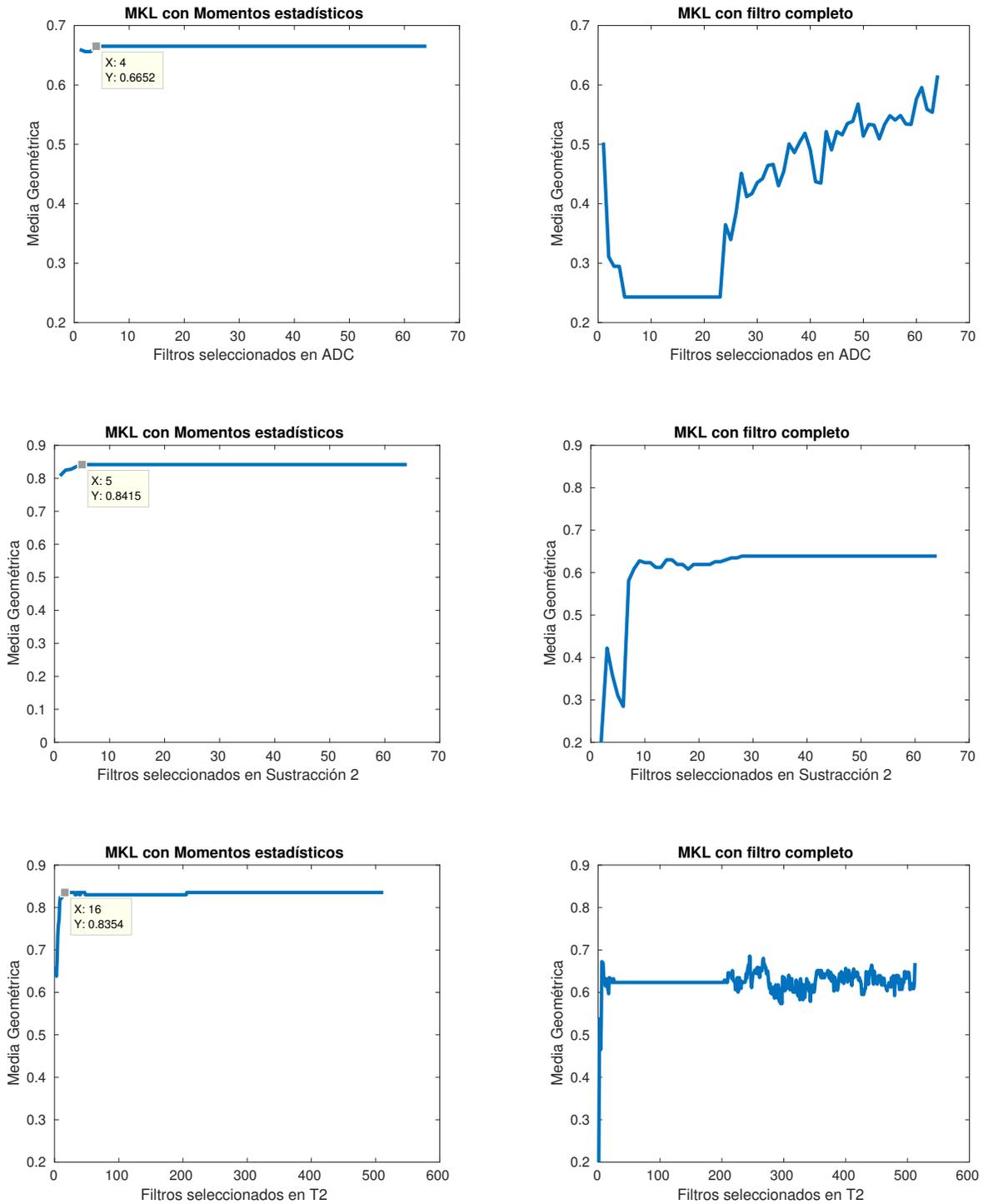


Figura 4-1: Comparación entre la selección de filtros con la representación propuesta respecto y la salida del filtro completa

4.8. Clasificación con características Radiómicas

En la tarea de clasificación de hallazgos positivos y negativos a partir de características radiómicas, fueron concatenadas las características generadas sobre la imagen original, el coeficiente de aproximación de la descomposición Wavelet y la imagen filtrada con LoG, obteniendo así tres fuentes de información para la clasificación con MKL-SVM, una por cada secuencia con 258 características. Para la clasificación con SVM, al igual que en la prueba con las características profundas, las características de las tres secuencias fueron concatenadas obteniendo una matriz con 774 características radiómicas. El desempeño promedio en los 5 folds para la clasificación con MKL-SVM respecto a exactitud, sensibilidad, especificidad y AUC fue de 75,98 %, 78,82 %, 72,31 % y 76,55 % y para SVM los desempeños respecto a estas mismas medidas fueron 69,18 %, 75,29 %, 61,15 % y 70,01 % respectivamente.

4.9. Clasificación con la integración de características profundas y radiómicas

Para la integración de características radiómicas y profundas en las tareas de clasificación asociadas a cáncer en tejidos blandos, se propuso la utilización de MKL-SVM con el objetivo de tomar ventaja de su capacidad en el uso y representación de información de diferentes fuentes, teniendo en cuenta tanto del tipo de características, como de la información aportada por la secuencia de la que provienen. En este orden de ideas, en el proceso de clasificación de hallazgos positivos y negativos con la metodología propuesta, fueron utilizadas 6 fuentes de información una por cada tipo de característica en cada secuencia. Adicionalmente, también fue evaluada con la línea base realizando la integración de las características radiómicas y profundas por medio de la concatenación de la información utilizada en cada uno de los enfoques independientes, para obtener así matriz de 146 datos por 899 características. El desempeño promedio en los 5 folds para la clasificación con MKL-SVM respecto a exactitud, sensibilidad, especificidad y AUC fue de 82,85 %, 88,24 %, 75,51 % y 81,74 % y para SVM los desempeños respecto a estas mismas medidas fueron 72,55 %, 87,06 %, 52,44 % y 72,51 % respectivamente.

En las Tablas 4-5 y 4-6 es presentado un consolidado del desempeño en la clasificación de hallazgos positivos y negativos en MRI de seno, utilizando respectivamente la línea base de clasificación con SVM y el método de integración propuesto con MKL-SVM. De estos resultados se puede observar que en general, el desempeño tanto utilizando las características de forma independiente según el tipo o integrándolas para la clasificación es notablemente superior cuando se utiliza MKL-SVM que cuando se utiliza la línea base propuesta, teniendo en cuenta que la aproximación basada en concatenación de las características de diferentes

Tabla 4-5: Desempeño de las estrategias de clasificación evaluadas con SVM

Tipo de Características	SVM			
	Exactitud (%)	Sensibilidad (%)	Especificidad (%)	AUC (%)
Radiómicas	69,18 ±10,91	75,29 ±20,55	61,15 ±18,34	70,01
Profundas	72,67 ±9,43	81,18 ±13,41	60,77 ±13,09	76,20
Radiómicas + profundas	72,55 ±7,48	87,06 ±9,67	52,44 ±6,95	72,51

Tabla 4-6: Desempeño de las estrategias de clasificación evaluadas con MKL-SVM

Tipo de Características	MKL			
	Exactitud (%)	Sensibilidad (%)	Especificidad (%)	AUC (%)
Radiómicas	75,98 ±5,08	78,82 ±12,89	72,31 ±13,23	76,55
Profundas	80,78 ±7,24	82,35 ±7,20	78,72 ±13,86	81,10
Radiómicas + profundas	82,85 ±7,37	88,24 ±8,32	75,51 ±11,38	81,74

secuencias es la más utilizada para realizar clasificación en sistemas CAD.

Por otro lado, a partir de estos resultados se evidencia que el mejor desempeño en términos de exactitud, sensibilidad y área bajo la curva ROC es obtenido utilizando la metodología propuesta. A pesar de que el desempeño en términos de área bajo la curva de la metodología propuesta no supera con una diferencia marcada la clasificación utilizando solo características profundas con MKL, es importante destacar que la integración de ambas características con la metodología propuesta si permite un aumento considerable en la predicción de hallazgos positivos, es decir, permite que el sistema sea más sensible, lo cual es importante y tiene una mayor prioridad en la realización de diagnósticos en la práctica clínica.

Adicionalmente, para la evaluación de la metodología propuesta se consideraron también las características de la última capa, teniendo en cuenta que son estas las ingresadas a la etapa de clasificación o perceptrón de la red. En ResNet-50 la última capa es de agrupamiento y su salida tiene una resolución de 1x1, dado que la dimensión de los datos de entrada es de 7x7 al igual que el kernel definido para el agrupamiento. De este modo, con la integración de las características profundas y radiómicas de cada secuencia usando MKL-SVM, se obtiene un desempeño en Exactitud de $79,45 \pm 7,72$, en Media Geométrica de $77,68 \pm 8,35$, Sensibilidad de $85,88 \pm 7,89$ y Especificidad de $70,64 \pm 11,96$

En la Figura 4-2 son presentadas las curvas ROC obtenidas en cada una de las estrategias de clasificación evaluadas para las características radiómicas y profundas de forma independiente y para la integración de estas, usando SVM y MKL-SVM.

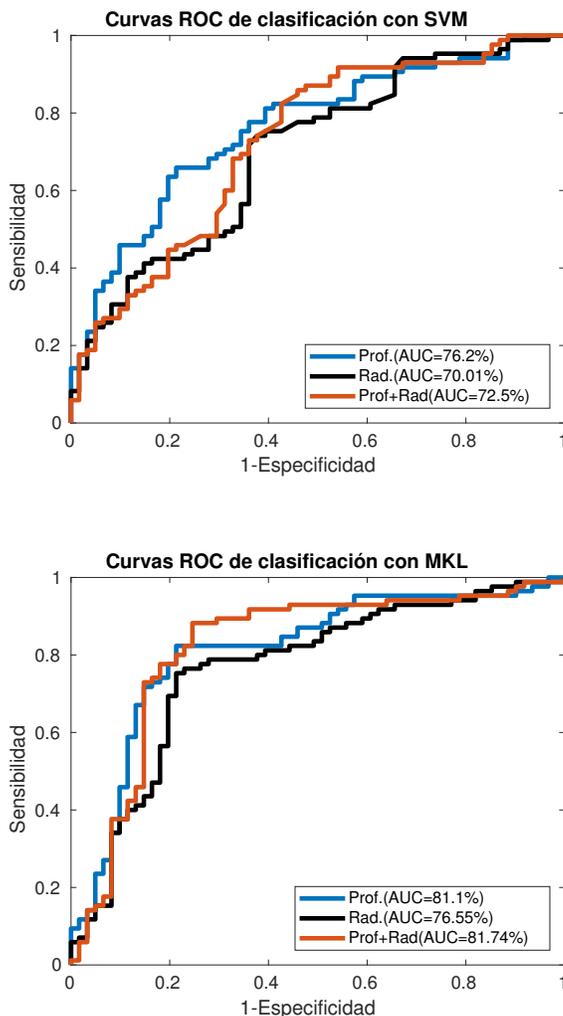


Figura 4-2: Desempeño de la clasificación de hallazgos positivos y negativos en MRI de mama respecto al AUC, evaluando características radiómicas y profundas por separado y la integración de ambas, tanto en la clasificación con SVM (arriba) como en la clasificación con MKL (abajo)

4.9.1. Significancia estadística de los resultados

Con el objetivo de determinar la contribución de la combinación de características radiómicas y profundas en el desempeño de la clasificación de hallazgos malignos y benignos en MRI de seno, fue implementada una prueba T-student para los 5-folds, dado que este test permite determinar la diferencia de las varianzas muestrales entre las partes de dos poblaciones y su respectivo intervalo de confianza en muestras pequeñas [82].

En este trabajo, los desempeños del algoritmo en cada fold, respecto a cada medida de desempeño, son tomados como muestras para realizar la comparación de los métodos de clasifi-

cación MKL-SVM usando características profundas y MKL-SVM usando las características profundas y radiómicas integradas, Tabla 4-7. Esta comparación puede ser realizada debido a que los datos durante la validación cruzada fueron los mismos en cada fold para ambos clasificadores.

Tabla 4-7: Desempeños en cada fold para MKL-SVM

Medida	Características Profundas					Características Profundas y radiómicas				
	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Exactitud	0.8965	0.7241	0.7931	0.7586	0,8667	0,9311	0,7586	0,8276	0,7586	0,8667
Sensibilidad	0.8235	0.7647	0.7647	0.8235	0.9412	0.9411	0.8235	0.8235	0.8235	1
Especificidad	1	0.6667	0.8333	0.6667	0.7692	0.9167	0.6667	0.8333	0.6667	0.6923

Para comparar los dos clasificadores, se aplica una prueba de hipótesis simple en la que se asume como hipótesis nula (H_0) que las características radiómicas no tienen efecto sobre el desempeño del clasificador, lo que indica que las medias de rendimiento entre los métodos comparados son iguales. Como hipótesis alternativa (H_1), se propone la existencia de una diferencia en estos rendimientos cuando se integran las características radiómicas y profundas.

Hipótesis

$$H_0 : \mu = 0$$

$$H_1 : \mu <> 0$$

Nivel de confianza

$$1 - \alpha = 0,95$$

$$\alpha = 0,05$$

La Tabla 4-8 muestra los resultados de los T-test realizados para la exactitud, sensibilidad y especificidad. Para un nivel de confianza de 95 %, la hipótesis nula es rechazada con un *p-value* menor a 0,05. Como se puede observar, la integración de las características radiómicas y profundas mejora la sensibilidad diagnóstica respecto al uso de las características profundas, para la estrategia propuesta. Respecto a la especificidad, no se puede establecer que los resultados sean estadísticamente significativos, lo cual es evidente al analizar los resultados para cada fold. En cuanto a la exactitud, si bien el resultado de la prueba estadística no permite descartar la hipótesis nula, los resultados del test no son totalmente concluyentes si se considera el tamaño de la muestra.

Tabla 4-8: Resultados del T-test para las tres medidas de desempeño

$H_0 =$ Las características radiómicas no tienen efecto en el desempeño		
$H_1 =$ Las características radiómicas tienen efecto en el desempeño		
Medida	Hipótesis aprobada	$p - value$
Exactitud	H_0	0.0705
Sensibilidad	H_1	0.0341
Especificidad	H_0	0.1783

5 Conclusiones, recomendaciones y trabajo futuro

La metodología propuesta hace uso de la información volumétrica que ofrece las imágenes por resonancia magnética, para lo cual se utilizan tres cortes en el proceso de análisis; el central, donde se identifica el hallazgo sobre el cual se requiere soporte de decisión y los dos contiguos (anterior y posterior). El cubo que conforman estos cortes es procesado sin realizar una segmentación fina del hallazgo, de esta manera se logra considerar la información de cambios en el tejido propio del tumor y del tejido circundante, relacionados con la presencia de diferentes tipos de masas. Por otro lado, el uso de estos tres cortes, permitió tomar ventaja de la arquitectura de red neuronal profunda ResNet50 pre-entrenada en ImageNet50, con el propósito de realizar transferencia de aprendizaje durante la estimación de características profundas en los volúmenes de interés extraídos.

Un aspecto importante al realizar aprendizaje por transferencia empleando redes neuronales convolucionales profundas, es el de definir cuáles filtros serán los empleados para la extracción de las características, lo cual implica tener que lidiar con el problema de la alta dimensionalidad que se genera al usar la respuesta a los filtros como vector de características. En este trabajo se propusieron dos estrategias para abordar estos problemas, en primer lugar, se propuso el uso de una representación compacta de la respuesta a los filtros mediante el cálculo de sus momentos estadísticos, reduciendo así la dimensión del vector de características a 5 para todos los casos. Por otro lado, en lugar de usar una capa completa de filtros, lo que multiplicaría también la dimensión del vector de representación, se realiza una selección de los filtros en los que se concentra la información más relevante para la tarea de clasificación; para esto se propuso emplear el método de aprendizaje por múltiples kernels, el cual permite abordar el problema de selección de filtros como un problema de selección de grupos de características. Los parámetros del modelo son seleccionados empleando optimización por enjambre de partículas, utilizando como medida de desempeño la media geométrica, la cual permite establecer un balance entre la medida de sensibilidad y especificidad. Los resultados de este proceso muestran que el uso de la representación compacta no sólo reduce la dimensionalidad de la representación, lo que implica un menor uso de recursos computacionales, sino que mejora el desempeño de la clasificación basada en MKL, siendo esta además mucho más estable en la selección de filtros relevantes.

La estrategia propuesta fue evaluada en el problema de clasificación automática de lesiones extraídas de estudios de resonancia magnética de mama, adquiridos retrospectivamente en el Instituto de Alta Tecnología Médica (Medellín-Colombia). Las regiones fueron localizadas y anotadas por radiólogos con 6 y 10 años de experiencia en la interpretación de este tipo de estudios, quienes asignaron a cada lesión una categoría de acuerdo al léxico BIRADS. Para este estudio, las regiones fueron agrupadas en dos categorías, esto es positivas o negativas, según la lesión fuera sospechosa o no de malignidad.

Los resultados obtenidos muestran que la integración de información de diferentes secuencias de resonancia magnética basada en MKL-SVM permite aprovechar mejor esta información en el proceso de clasificación, dado que se representa cada una de las fuentes de forma independiente, conservando así las relaciones internas entre las características provenientes de cada una. Adicionalmente, se evidenció que a pesar de que con las características profundas utilizadas se obtiene una exactitud general similar a la obtenida con la integración de características radiómicas y profundas, este proceso de integración permite mejorar la sensibilidad en el proceso de clasificación automática, lo que tiene mayor relevancia en la práctica clínica. Este resultado puede incluso ser mejorado si se realiza la optimización de los parámetros de los kernels del modelo de aprendizaje con respecto a la sensibilidad y no con respecto a la media geométrica, caso en el que se sacrificaría la especificidad, pero que puede ser útil en tareas específicas.

A pesar de los resultados obtenidos, este trabajo, y la metodología propuesta, presentan algunas limitaciones que pueden ser abordadas en trabajos futuros. En primer lugar, la estrategia de integración basada en MKL-SVM está limitada a problemas de clasificación de dos clases o categorías, la extensión a múltiples clases será posible mediante la implementación de estrategias de clasificación uno Vs todos o uno Vs uno en el modelo de aprendizaje; aunque también se puede explorar la implementación de modelos de clasificación en cascada, para problemas de interpretación diagnóstica en los que la tarea puede verse como una cascada de clasificadores binarios.

Otra limitación importante de este estudio es el conjunto de datos empleado para la evaluación del método propuesto, el cual estuvo compuesto por un número pequeño de instancias para el entrenamiento y validación. Adicionalmente, las lesiones no fueron delineadas por los médicos especialistas, debido al alto consumo de tiempo que requiere esta tarea, y tampoco se contó con resultados confirmados por biopsia, dado que los estudios fueron adquiridos retrospectivamente en un centro de ayudas diagnósticas. Estas características del conjunto de datos condicionaron algunas decisiones del diseño de este estudio, que pueden ser replanteadas en el futuro, cuando se cuente con un conjunto de datos mayor y con otras condiciones

de adquisición.

Respecto al proceso de análisis radiómico, la extracción de características estuvo limitada a descriptores globales, es decir que consideran toda la región, esto puede ser una limitante del método para afrontar problemas donde la geometría del tumor es importante, caso en el cual se deberá realizar una delineación previa del tumor (ya sea manual o automática), e incluir características de forma en el proceso; o explorar otros métodos de extracción de características que, procesando la imagen de manera global, consigan una mejor representación de los bordes tenues de la lesión, como pueden ser los momentos de alto orden como los momentos de Zernike o de Krawtchouk. Adicionalmente, se puede evaluar el uso de técnicas de preprocesamiento, que mejoren el contraste entre el tumor y el tejido mamario, lo cual puede mejorar también el desempeño de estos descriptores.

Con relación a las características profundas, el limitado número de instancias para entrenamiento, no permiten realizar un proceso de ajuste fino de los pesos de la red o entrenar la arquitectura específica para la tarea, para tomar ventaja de la capacidad de las redes neuronales para integrar las etapas de caracterización y clasificación. Por lo que, en trabajos futuros se pueden explorar estas alternativas al incrementar el número de imágenes de entrenamiento, ya sea por adquisición de las mismas o empleando estrategias de generación de datos, aun las mismas basadas en aprendizaje profundo. Por otro lado, a pesar de que en este trabajo se seleccionó la arquitectura ResNet 50, debido a su buen desempeño en la tarea de clasificación para la cual fue entrenada originalmente y por las características de alto nivel que se extraen, en el estado del arte se han propuesto diversas arquitecturas CNN para la misma tarea que han tenido resultados comparables. Por esta razón, es importante considerar como trabajo futuro la evaluación de la metodología propuesta utilizando características profundas provenientes de otras arquitecturas. Adicionalmente, puede realizarse la evaluación de la metodología propuesta realizando el análisis de relevancia de filtros en el total de las capas de la arquitectura o incluso con la combinación de características profundas de primer nivel y nivel superior obtenidas a través de la arquitectura.

Finalmente, con relación a la tarea específica de clasificación de lesiones asociadas al cáncer de mama, el no contar con lesiones con confirmación por biopsia introduce una incertidumbre alta en la tarea, especialmente para las lesiones con menor probabilidad de malignidad (BIRADS 4 y 5). En este sentido, en trabajo futuro se propone la implementación de la misma estrategia, entrenando el modelo con datos confirmados en que se conozca con precisión si la lesión es maligna o benigna. Incluso, el uso de MKL, puede ser útil en tal caso, para determinar las características que mayor contribución tienen para diferenciar este tipo de lesiones y reducir la incertidumbre en el diagnóstico y manejo del paciente.

Bibliografía

- [1] A. C. P. Nazário, G. Facina, and J. R. Filassi, “Breast cancer: news in diagnosis and treatment,” *Revista da Associação Médica Brasileira*, vol. 61, no. 6, pp. 543–552, 2015.
- [2] A. Kreso and J. E. Dick, “Evolution of the cancer stem cell model,” *Cell stem cell*, vol. 14, no. 3, pp. 275–291, 2014.
- [3] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, “Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries,” *CA: a cancer journal for clinicians*, vol. 68, no. 6, pp. 394–424, 2018.
- [4] L. A. Torre, R. L. Siegel, E. M. Ward, and A. Jemal, “Global cancer incidence and mortality rates and trends—an update,” *Cancer Epidemiology and Prevention Biomarkers*, vol. 25, no. 1, pp. 16–27, 2016.
- [5] L. Caplan, “Delay in breast cancer: implications for stage at diagnosis and survival,” *Frontiers in public health*, vol. 2, p. 87, 2014.
- [6] C. C. Riedl, N. Luft, C. Bernhart, M. Weber, M. Bernathova, M.-K. M. Tea, M. Rudas, C. F. Singer, and T. H. Helbich, “Triple-modality screening trial for familial breast cancer underlines the importance of magnetic resonance imaging and questions the role of mammography and ultrasound regardless of patient mutation status, age, and breast density,” *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, vol. 33, no. 10, p. 1128, 2015.
- [7] B. Hochegger, E. Marchiori, O. Sedlacek, K. Irion, C. Heussel, S. Ley, J. Ley-Zaporozhan, A. Soares Souza Jr, and H. Kauczor, “Mri in lung cancer: a pictorial essay,” *The British journal of radiology*, vol. 84, no. 1003, pp. 661–668, 2011.
- [8] Y. Ohno, H. Koyama, H. Y. Lee, T. Yoshikawa, and K. Sugimura, “Magnetic resonance imaging (mri) and positron emission tomography (pet)/mri for lung cancer staging,” *Journal of thoracic imaging*, vol. 31, no. 4, pp. 215–227, 2016.
- [9] M. A. Marino, P. Clauser, R. Woitek, G. J. Wengert, P. Kapetas, M. Bernathova, K. Pinker-Domenig, T. H. Helbich, K. Preidler, and P. A. Baltzer, “A simple scoring

- system for breast mri interpretation: does it compensate for reader experience?" *European radiology*, vol. 26, no. 8, pp. 2529–2537, 2016.
- [10] R. M. Hayward, N. Patronas, E. H. Baker, G. Vézina, P. S. Albert, and K. E. Warren, "Inter-observer variability in the measurement of diffuse intrinsic pontine gliomas," *Journal of neuro-oncology*, vol. 90, no. 1, pp. 57–61, 2008.
- [11] M. Lunkiewicz, S. Forte, B. Freiwald, G. Singer, C. Leo, and R. A. Kubik-Huch, "Inter-observer variability and likelihood of malignancy for fifth edition bi-rads mri descriptors in non-mass breast lesions," *European Radiology*, pp. 1–10, 2019.
- [12] L. H. Eadie, P. Taylor, and A. P. Gibson, "A systematic review of computer-assisted diagnosis in diagnostic cancer imaging," *European journal of radiology*, vol. 81, no. 1, pp. e70–e76, 2012.
- [13] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, "Medical image classification with convolutional neural network," in *2014 13th International Conference on Control Automation Robotics & Vision (ICARCV)*. IEEE, 2014, pp. 844–848.
- [14] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.
- [15] L. J. Grimm, A. L. Anderson, J. A. Baker, K. S. Johnson, R. Walsh, S. C. Yoon, and S. V. Ghatge, "Interobserver variability between breast imagers using the fifth edition of the bi-rads mri lexicon," *American Journal of Roentgenology*, vol. 204, no. 5, pp. 1120–1124, 2015.
- [16] M. J. Stoutjesdijk, J. J. Fütterer, C. Boetes, L. E. van Die, G. Jager, and J. O. Barentsz, "Variability in the description of morphologic and contrast enhancement characteristics of breast lesions on magnetic resonance imaging," *Investigative radiology*, vol. 40, no. 6, pp. 355–362, 2005.
- [17] L. Dickinson, H. U. Ahmed, C. Allen, J. O. Barentsz, B. Carey, J. J. Fütterer, S. W. Heijmink, P. Hoskin, A. P. Kirkham, A. R. Padhani *et al.*, "Scoring systems used for the interpretation and reporting of multiparametric mri for prostate cancer detection, localization, and characterization: could standardization lead to improved utilization of imaging within the diagnostic pathway?" *Journal of Magnetic Resonance Imaging*, vol. 37, no. 1, pp. 48–58, 2013.

-
- [18] A. Jalalian, S. B. Mashohor, H. R. Mahmud, M. I. B. Saripan, A. R. B. Ramli, and B. Karasfi, “Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review,” *Clinical imaging*, vol. 37, no. 3, pp. 420–426, 2013.
- [19] K. Doi, “Computer-aided diagnosis in medical imaging: historical review, current status and future potential,” *Computerized medical imaging and graphics*, vol. 31, no. 4-5, pp. 198–211, 2007.
- [20] V. M. Rao, D. C. Levin, L. Parker, B. Cavanaugh, A. J. Frangos, and J. H. Sunshine, “How widely is computer-aided detection used in screening and diagnostic mammography?” *Journal of the American College of Radiology*, vol. 7, no. 10, pp. 802–805, 2010.
- [21] R. E. Hendrick and W. R. Hendee, “Breast mri: Fundamentals and technical aspects,” *Medical Physics*, vol. 35, p. 1163, 2008.
- [22] P. Dang and C. D. Lehman, “Breast mri overview: Imaging techniques and equipment, clinical uses, image interpretation,” *Breast Imaging*, p. 81, 2018.
- [23] C. P. Slichter, *Principles of magnetic resonance*. Springer Science & Business Media, 2013, vol. 1.
- [24] P. Soffia, “Difusión por resonancia magnética: bases y aplicaciones oncológicas en órganos extracraneanos,” *Revista chilena de radiología*, vol. 15, pp. 17–24, 2009.
- [25] L. Liberman and J. H. Menell, “Breast imaging reporting and data system (bi-rads),” *Radiologic Clinics*, vol. 40, no. 3, pp. 409–430, 2002.
- [26] B. Erguvan-Dogan, G. J. Whitman, A. C. Kushwaha, M. J. Phelps, and P. J. Dempsey, “Bi-rads-mri: a primer,” *American Journal of Roentgenology*, vol. 187, no. 2, pp. W152–W160, 2006.
- [27] G. Agrawal, M.-Y. Su, O. Nalcioglu, S. A. Feig, and J.-H. Chen, “Significance of breast lesion descriptors in the acr bi-rads mri lexicon,” *Cancer*, vol. 115, no. 7, pp. 1363–1380, 2009.
- [28] H. Lee and Y.-P. P. Chen, “Image based computer aided diagnosis system for cancer detection,” *Expert Systems with Applications*, vol. 42, no. 12, pp. 5356–5365, 2015.
- [29] N. I. Yassin, S. Omran, E. M. El Houbay, and H. Allam, “Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review,” *Computer methods and programs in biomedicine*, vol. 156, pp. 25–45, 2018.

- [30] A. R. M. Al-shamasneh and U. H. B. Obaidallah, "Artificial intelligence techniques for cancer detection and classification: review study," *European Scientific Journal*, vol. 13, no. 3, pp. 342–370, 2017.
- [31] I. Arel, D. C. Rose, T. P. Karnowski *et al.*, "Deep machine learning-a new frontier in artificial intelligence research," *IEEE computational intelligence magazine*, vol. 5, no. 4, pp. 13–18, 2010.
- [32] R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: images are more than pictures, they are data," *Radiology*, vol. 278, no. 2, pp. 563–577, 2015.
- [33] E. Florez, A. Fatemi, P. Claudio, and C. Howard, "Emergence of radiomics: Novel methodology identifying imaging biomarkers of disease in diagnosis, response, and progression," *SM Journal of Clinical and Medical Imaging*, vol. 4, no. 1, p. 1019, 2018.
- [34] R. F. Brem, M. J. Lenihan, J. Lieberman, and J. Torrente, "Screening breast ultrasound: past, present, and future," *American Journal of Roentgenology*, vol. 204, no. 2, pp. 234–240, 2015.
- [35] M. Hatt, F. Tixier, D. Visvikis, and C. C. Le Rest, "Radiomics in pet/ct: more than meets the eye?" *Journal of Nuclear Medicine*, vol. 58, no. 3, pp. 365–366, 2017.
- [36] J. Wang, F. Kato, N. Oyama-Manabe, R. Li, Y. Cui, K. K. Tha, H. Yamashita, K. Kudo, and H. Shirato, "Identifying triple-negative breast cancer using background parenchymal enhancement heterogeneity on dynamic contrast-enhanced mri: a pilot radiomics study," *PloS one*, vol. 10, no. 11, p. e0143308, 2015.
- [37] V. Kumar, Y. Gu, S. Basu, A. Berglund, S. A. Eschrich, M. B. Schabath, K. Forster, H. J. Aerts, A. Dekker, D. Fenstermacher *et al.*, "Radiomics: the process and the challenges," *Magnetic resonance imaging*, vol. 30, no. 9, pp. 1234–1248, 2012.
- [38] P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R. G. Van Stiphout, P. Granton, C. M. Zegers, R. Gillies, R. Boellard, A. Dekker *et al.*, "Radiomics: extracting more information from medical images using advanced feature analysis," *European journal of cancer*, vol. 48, no. 4, pp. 441–446, 2012.
- [39] Z.-C. Li, Y. Chen, Q. Li, Q. Sun, and R. Luo, "Automatic extraction of mri radiomics features in glioblastoma multiforme: A reproducibility evaluation," in *2017 3rd IEEE International Conference on Cybernetics (CYBCONF)*. IEEE, 2017, pp. 1–4.
- [40] C. Parmar, E. R. Velazquez, R. Leijenaar, M. Jermoumi, S. Carvalho, R. H. Mak, S. Mitra, B. U. Shankar, R. Kikinis, B. Haibe-Kains *et al.*, "Robust radiomics feature quantification using semiautomatic volumetric segmentation," *PloS one*, vol. 9, no. 7, p. e102107, 2014.

- [41] R. Stoyanova, M. Takhar, Y. Tschudi, J. C. Ford, G. Solórzano, N. Erho, Y. Balagurunathan, S. Punnen, E. Davicioni, R. J. Gillies *et al.*, “Prostate cancer radiomics and the promise of radiogenomics,” *Translational cancer research*, vol. 5, no. 4, p. 432, 2016.
- [42] E. Gohariyan, M. Esmaeilpour, and M. M. Shirmohammadi, “The combination of mammography and mri for diagnosing breast cancer using fuzzy nn and svm.” *International Journal of Interactive Multimedia & Artificial Intelligence*, vol. 4, no. 5, 2017.
- [43] S. Agliozzo, M. De Luca, C. Bracco, A. Vignati, V. Giannini, L. Martincich, L. Carbonaro, A. Bert, F. Sardanelli, and D. Regge, “Computer-aided diagnosis for dynamic contrast-enhanced breast mri of mass-like lesions using a multiparametric model combining a selection of morphological, kinetic, and spatiotemporal features,” *Medical physics*, vol. 39, no. 4, pp. 1704–1715, 2012.
- [44] A. K. Dhara, S. Mukhopadhyay, A. Dutta, M. Garg, and N. Khandelwal, “A combination of shape and texture features for classification of pulmonary nodules in lung ct images,” *Journal of digital imaging*, vol. 29, no. 4, pp. 466–475, 2016.
- [45] L. J. Grimm, J. Zhang, and M. A. Mazurowski, “Computational approach to radiogenomics of breast cancer: luminal a and luminal b molecular subtypes are associated with imaging features on routine breast mri extracted using computer vision algorithms,” *Journal of Magnetic Resonance Imaging*, vol. 42, no. 4, pp. 902–907, 2015.
- [46] S. C. Agner, M. A. Rosen, S. Englander, J. E. Tomaszewski, M. D. Feldman, P. Zhang, C. Mies, M. D. Schnall, and A. Madabhushi, “Computerized image analysis for identifying triple-negative breast cancers and differentiating them from other molecular subtypes of breast cancer on dynamic contrast-enhanced mr images: a feasibility study,” *Radiology*, vol. 272, no. 1, pp. 91–99, 2014.
- [47] D. V. Fried, S. L. Tucker, S. Zhou, Z. Liao, O. Mawlawi, G. Ibbott, and L. E. Court, “Prognostic value and reproducibility of pretreatment ct texture features in stage iii non-small cell lung cancer,” *International Journal of Radiation Oncology* Biology* Physics*, vol. 90, no. 4, pp. 834–842, 2014.
- [48] J. T. Kwak, S. Xu, B. J. Wood, B. Turkbey, P. L. Choyke, P. A. Pinto, S. Wang, and R. M. Summers, “Automated prostate cancer detection using t2-weighted and high-b-value diffusion-weighted magnetic resonance imaging,” *Medical physics*, vol. 42, no. 5, pp. 2368–2378, 2015.
- [49] V. Pomponiu, H. Hariharan, B. Zheng, and D. Gur, “Improving breast mass detection using histogram of oriented gradients,” in *Medical Imaging 2014: Computer-Aided Diagnosis*, vol. 9035. International Society for Optics and Photonics, 2014, p. 90351R.

- [50] K. Gnep, A. Fargeas, R. E. Gutiérrez-Carvajal, F. Commandeur, R. Mathieu, J. D. Ospina, Y. Rolland, T. Rohou, S. Vincendeau, M. Hatt *et al.*, “Haralick textural features on t2-weighted mri are associated with biochemical recurrence following radiotherapy for peripheral zone prostate cancer,” *Journal of Magnetic Resonance Imaging*, vol. 45, no. 1, pp. 103–117, 2017.
- [51] U. R. Acharya, W. L. Ng, K. Rahmat, V. K. Sudarshan, J. E. Koh, J. H. Tan, Y. Hagiwara, C. H. Yeong, and K. H. Ng, “Data mining framework for breast lesion classification in shear wave ultrasound: a hybrid feature paradigm,” *Biomedical Signal Processing and Control*, vol. 33, pp. 400–410, 2017.
- [52] E. Oczeretko, M. Borowska, A. Kitlas, A. Borusiewicz, and M. Sobolewska-Siemieniuk, “Fractal analysis of medical images in the irregular regions of interest,” in *2008 8th IEEE International Conference on BioInformatics and BioEngineering*. IEEE, 2008, pp. 1–6.
- [53] Z. Ning, J. Luo, Y. Li, S. Han, Q. Feng, Y. Xu, W. Chen, T. Chen, and Y. Zhang, “Pattern classification for gastrointestinal stromal tumors by integration of radiomics and deep convolutional features,” *IEEE journal of biomedical and health informatics*, vol. 23, no. 3, pp. 1181–1191, 2018.
- [54] U. R. Acharya, Y. Hagiwara, V. K. Sudarshan, W. Y. Chan, and K. H. Ng, “Towards precision medicine: from quantitative imaging to radiomics,” *Journal of Zhejiang University-Science B*, vol. 19, no. 1, pp. 6–24, 2018.
- [55] Y. Bengio *et al.*, “Learning deep architectures for ai,” *Foundations and trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.
- [56] J. H. Cole, R. P. Poudel, D. Tsagkrasoulis, M. W. Caan, C. Steves, T. D. Spector, and G. Montana, “Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker,” *NeuroImage*, vol. 163, pp. 115–124, 2017.
- [57] M. Ghafoorian, N. Karssemeijer, T. Heskes, M. Bergkamp, J. Wissink, J. Obels, K. Keizer, F.-E. de Leeuw, B. van Ginneken, E. Marchiori *et al.*, “Deep multi-scale location-aware 3d convolutional neural networks for automated detection of lacunes of presumed vascular origin,” *NeuroImage: Clinical*, vol. 14, pp. 391–399, 2017.
- [58] E. Hosseini-Asl, R. Keynton, and A. El-Baz, “Alzheimer’s disease diagnostics by adaptation of 3d convolutional network,” in *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2016, pp. 126–130.

- [59] S. Korolev, A. Safiullin, M. Belyaev, and Y. Dodonova, "Residual and plain convolutional neural networks for 3d brain mri classification," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, 2017, pp. 835–838.
- [60] J. Li, M. Fan, J. Zhang, and L. Li, "Discriminating between benign and malignant breast tumors using 3d convolutional neural network in dynamic contrast enhanced-mr images," in *Medical Imaging 2017: Imaging Informatics for Healthcare, Research, and Applications*, vol. 10138. International Society for Optics and Photonics, 2017, p. 1013808.
- [61] X. Huang, J. Shan, and V. Vaidya, "Lung nodule detection in ct using 3d convolutional neural networks," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, 2017, pp. 379–383.
- [62] S. Hamidian, B. Sahiner, N. Petrick, and A. Pezeshk, "3d convolutional neural network for automatic detection of lung nodules in chest ct," in *Medical Imaging 2017: Computer-Aided Diagnosis*, vol. 10134. International Society for Optics and Photonics, 2017, p. 1013409.
- [63] Q. Dou, H. Chen, L. Yu, J. Qin, and P.-A. Heng, "Multilevel contextual 3-d cnns for false positive reduction in pulmonary nodule detection," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1558–1567, 2016.
- [64] W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, "Multi-scale convolutional neural networks for lung nodule classification," in *International Conference on Information Processing in Medical Imaging*. Springer, 2015, pp. 588–599.
- [65] A. Teramoto, H. Fujita, O. Yamamuro, and T. Tamaki, "Automated detection of pulmonary nodules in pet/ct images: Ensemble false-positive reduction using a convolutional neural network technique," *Medical physics*, vol. 43, no. 6Part1, pp. 2821–2827, 2016.
- [66] F. Ciompi, B. de Hoop, S. J. van Riel, K. Chung, E. T. Scholten, M. Oudkerk, P. A. de Jong, M. Prokop, and B. van Ginneken, "Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2d views and a convolutional neural network out-of-the-box," *Medical image analysis*, vol. 26, no. 1, pp. 195–202, 2015.
- [67] H. Chen, L. Yu, Q. Dou, L. Shi, V. C. Mok, and P. A. Heng, "Automatic detection of cerebral microbleeds via deep learning based 3d feature representation," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2015, pp. 764–767.

- [68] R. Paul, S. H. Hawkins, L. O. Hall, D. B. Goldgof, and R. J. Gillies, "Combining deep neural network and traditional image features to improve survival prediction accuracy for lung cancer patients from diagnostic ct," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2016, pp. 002 570–002 575.
- [69] J. Lao, Y. Chen, Z.-C. Li, Q. Li, J. Zhang, J. Liu, and G. Zhai, "A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme," *Scientific reports*, vol. 7, no. 1, p. 10353, 2017.
- [70] N. Antropova, B. Q. Huynh, and M. L. Giger, "A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets," *Medical physics*, vol. 44, no. 10, pp. 5162–5171, 2017.
- [71] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.
- [72] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.
- [73] H. Ravishankar, P. Sudhakar, R. Venkataramani, S. Thiruvenkadam, P. Annangi, N. Babu, and V. Vaidya, "Understanding the mechanisms of deep transfer learning for medical images," in *Deep Learning and Data Labeling for Medical Applications*. Springer, 2016, pp. 188–196.
- [74] N. Zayed and H. A. Elnemr, "Statistical analysis of haralick texture features to discriminate lung abnormalities," *Journal of Biomedical Imaging*, vol. 2015, p. 12, 2015.
- [75] A. Ahmed, P. Gibbs, M. Pickles, and L. Turnbull, "Texture analysis in assessment and prediction of chemotherapy response in breast cancer," *Journal of Magnetic Resonance Imaging*, vol. 38, no. 1, pp. 89–101, 2013.
- [76] R. M. Haralick, K. Shanmugam *et al.*, "Textural features for image classification," *IEEE Transactions on systems, man, and cybernetics*, no. 6, pp. 610–621, 1973.
- [77] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [78] A. Canziani, A. Paszke, and E. Culurciello, "An analysis of deep neural network models for practical applications," *arXiv preprint arXiv:1605.07678*, 2016.

-
- [79] H. J. Areiza-Laverde, G. M. Díaz, and A. E. Castro-Ospina, “Feature group selection using mkl penalized with ℓ_1 -norm and svm as base learner,” in *Workshop on Engineering Applications*. Springer, 2018, pp. 136–147.
- [80] Y. Shi *et al.*, “Particle swarm optimization: developments, applications and resources,” in *Proceedings of the 2001 congress on evolutionary computation (IEEE Cat. No. 01TH8546)*, vol. 1. IEEE, 2001, pp. 81–86.
- [81] Z. Xu, R. Jin, H. Yang, I. King, and M. R. Lyu, “Simple and efficient multiple kernel learning by group lasso,” in *Proceedings of the 27th international conference on machine learning (ICML-10)*. Citeseer, 2010, pp. 1175–1182.
- [82] J. C. De Winter, “Using the student’s t-test with extremely small sample sizes.” *Practical Assessment, Research & Evaluation*, vol. 18, no. 10, 2013.