

# SELECCIÓN OBJETIVA DE SECUENCIAS DE RESONANCIA MAGNÉTICA RELEVANTES PARA EL DIAGNÓSTICO DE CÁNCER DE MAMA UTILIZANDO MKL Y SVM

**Henry Jhoán Areiza Laverde**

**Maestría**  
en Automatización y Control Industrial

Directores: Ph.D. Gloria Mercedes Díaz Cabrera

M.Sc. Andrés Eduardo Castro Ospina



**Institución Universitaria**

*Instituto Tecnológico Metropolitano*  
*Facultad de Ingenierías*  
Medellín, 2019

# AGRADECIMIENTOS

Esta tesis fue desarrollada en el marco del proyecto de investigación titulado “*Protocolo Abreviado De Resonancia Magnética Asistido Por Computador Para La Detección Y Categorización De Lesiones Sospechosas De Cáncer De Mama*”, financiado por Colciencias en la convocatoria 777 de 2017. Su desarrollo fue posible gracias a la colaboración de los investigadores de este proyecto, en especial a los médicos especialistas del Instituto de Alta Tecnología Médica (IATM) y los investigadores del Laboratorio de Máquinas Inteligentes y Reconocimiento de Patrones (MIRP) del Instituto Tecnológico Metropolitano.

# RESUMEN

El cáncer de mama es el segundo más común entre todos los tipos de cáncer, y es también el más diagnosticado en mujeres, tanto en países desarrollados como en países en vías de desarrollo. En Colombia, el número de casos nuevos se incrementa sustancialmente cada año, y en muchas ocasiones sólo logra ser detectado en estadios avanzados de desarrollo. Debido a la gran importancia que ha adquirido el estudio del cáncer de mama, se han propuesto diferentes métodos de tamizaje por medio de imágenes diagnósticas, que permiten detectar y estudiar esta enfermedad, siendo las imágenes de resonancia magnética (MRI por las siglas en inglés de *Magnetic Resonance Imaging*) una de las herramientas más prometedoras en cuanto a la información útil que proporcionan para realizar diagnósticos precisos y confiables. Sin embargo, el uso de MRI encuentra un gran obstáculo para su implementación, al ser más costoso que otros métodos, debido principalmente a la gran cantidad de imágenes agrupadas por secuencias que deben ser adquiridas e interpretadas.

En los últimos años, se han desarrollado trabajos que buscan reducir el número de secuencias de MRI que deben ser implementadas para el diagnóstico de cáncer de mama, aunque estos trabajos encuentran un limitante, debido a que son completamente subjetivos y, por lo tanto, están ligados a la apreciación y criterio de los radiólogos especialistas y a los centros médicos en donde han sido realizados. Por otro lado, el uso de sistemas computacionales para apoyar los procesos de diagnóstico ha mostrado ser de gran utilidad en el área médica; además, desde el área de aprendizaje de máquina ha destacado el método de aprendizaje por múltiples kernels (MKL por las siglas en inglés de *Multiple Kernel Learning*) debido a que permite integrar múltiples fuentes de información en una sola tarea, mostrando un buen desempeño y resultados fáciles de interpretar.

En este proyecto se propone un método basado en el uso de MKL y máquinas de vectores de soporte (SVM por las siglas en inglés de *Support Vector Machines*), para seleccionar de forma objetiva y automática las secuencias de MRI más relevantes para el diagnóstico de cáncer de mama, mediante la penalización de los pesos asociados a cada una de las matrices kernel que representan las secuencias de MRI. Inicialmente, se generan características perceptuales y radiómicas a partir de cada una de las secuencias de MRI, luego se calculan y sintonizan los kernels utilizando tres algoritmos diferentes basados en la técnica de escalamiento local. Finalmente, la información de los kernels es llevada a una SVM, obteniendo como salida definitiva los pesos entrenados que son empleados como medidas de relevancia de las fuentes de información y las etiquetas predichas para los ejemplos procesados.

Los resultados obtenidos demuestran que el método propuesto es estable en cuanto a las fuentes que son seleccionadas como fuentes relevantes, también indican que las fuentes de información correspondientes a las secuencias de Fases 2, 3, 4 y 5 son relevantes para la tarea de clasificación, cuando se diferencian lesiones positivas de negativas en estudios de resonancia magnética de mama. Al comparar el desempeño del método propuesto con dos métodos clásicos de selección de características del estado del arte, teniendo en cuenta que la selección de características es la analogía más próxima a la selección de fuentes, se puede identificar que el método propuesto supera a los métodos evaluados del estado del arte respecto a la selección objetiva y eficaz de las fuentes de información relevantes.

**Palabras clave:** selección de fuentes, aprendizaje de máquina, aprendizaje por múltiples kernels, máquinas de vectores de soporte, reducción de dimensión.

# ABSTRACT

Breast cancer is the second most common among all cancer types and is also the most diagnosed in women, both in developed and in developing countries. In Colombia, the number of new cases increases substantially every year, and in many, it only can be detected in advanced stages of development. Due to the great importance that the study of breast cancer has acquired, different medical image-based screening methods that allow to detect and study this disease have been developed, being the magnetic resonance imaging (MRI) one of the most promising tools in terms of the useful information that provide to make sensible and reliable diagnoses. However, the massive use of MRI is restricted because it is more expensive than other methods, mainly due to the large number of images grouped by sequences that must be acquired and interpreted.

At recent years, some works have been developed seeking to reduce the number of MRI sequences that must be implemented for the diagnosis of breast cancer. However, these works have limitations, because the sequence selection is completely subjective, and therefore they are linked to the appreciation and criteria of the radiologists and the medical centers where they have been performed. On the other hand, the use of computer systems to support diagnostic processes has proven to be very useful in the medical field; In addition, the machine learning area has highlighted the method of multiple kernel learning (MKL) because it allows to integrate multiple information sources in a single task, showing good performance and easy to interpret results.

In this project, a method based on the use of MKL and support vector machines (SVM) is proposed to objectively and automatically select the most relevant MRI sequences for the diagnosis of breast cancer, through the penalty of the weights associated with each one of the kernel matrices that represent the MRI sequences. Initially, perceptual and radiological features are generated from each MRI sequence, then the kernels are calculated and tuned using three different algorithms based on the local scaling technique. Finally, the information of the kernels feeds an SVM, obtaining as final output the trained weights, which are used as measures of relevance of the information sources, and the predicted labels for the processed examples.

The obtained results show that the proposed method is stable in terms of the sources that are selected as relevant, they also indicate that the information sources corresponding to the sequences of Phases 2, 3, 4 and 5 are clearly relevant for the classification task, when differ positive lesions from negatives in breast MRI studies. When comparing the performance of the proposed method with two classical feature selection methods of the state of the art, taking into account that the feature selection is the closest analogy to the source selection, it can be identified that the proposed method outperform the evaluated methods of the state of the art regarding the objective and effective selection of relevant information sources.

**Keywords:** source selection, machine learning, multiple kernel learning, support vector machines, dimensionality reduction.

# ÍNDICE GENERAL

<b>Agradecimientos</b>	<b>II</b>
<b>Resumen</b>	<b>III</b>
<b>Abstract</b>	<b>IV</b>
<b>Índice General</b>	<b>VI</b>
<b>Índice de figuras</b>	<b>VII</b>
<b>Índice de tablas</b>	<b>VIII</b>
<b>Introducción</b>	<b>1</b>
Protocolos abreviados de MRI para el diagnóstico del cáncer . . . . .	2
Selección de fuentes con aprendizaje de máquina . . . . .	2
Problema de investigación . . . . .	3
Objetivos. . . . .	4
General . . . . .	4
Específicos . . . . .	4
Estructura del documento . . . . .	4
Contribución. . . . .	5
<b>1 Imágenes de resonancia magnética</b>	<b>6</b>
1.1 Fundamentos físicos de las MRI. . . . .	6
1.2 Tipos de secuencias para el diagnóstico de cáncer de mama . . . . .	8
1.2.1 Secuencia T1. . . . .	8
1.2.2 Secuencia T2. . . . .	8
1.2.3 Secuencia DWI. . . . .	8
1.2.4 Secuencia ADC. . . . .	8
1.2.5 Secuencia STIR . . . . .	9
1.2.6 Secuencias DCE . . . . .	10
1.2.7 Sustracción o Fases . . . . .	10
1.3 Categorización estandarizada de tumores en MRI . . . . .	10
1.3.1 Características morfológicas . . . . .	10
1.3.2 Características cinéticas . . . . .	11
1.3.3 Categorías BI-RADS . . . . .	11
<b>2 Representación de los datos</b>	<b>12</b>
2.1 Características perceptuales . . . . .	12
2.1.1 Saliencia visual basada en grafos. . . . .	13
2.1.2 Matriz de características perceptuales . . . . .	13
2.2 Características radiómicas . . . . .	15
2.2.1 Definición de radiómica . . . . .	15
2.2.2 Tipos de características radiómicas . . . . .	15
2.2.3 Matriz de características radiómicas . . . . .	16

<b>3</b>	<b>Método de selección de fuentes relevantes</b>	<b>18</b>
3.1	Máquinas de Vectores de Soporte . . . . .	18
3.2	Kernel y escalamiento local . . . . .	20
3.3	Aprendizaje por Múltiples Kernels (MKL) . . . . .	21
3.4	Selección de fuentes basada en MKL . . . . .	21
3.5	Método propuesto . . . . .	22
3.5.1	Resumen del método . . . . .	22
3.5.2	Predicción sin reducción de datos . . . . .	23
3.5.3	Predicción utilizando solo los vectores de soporte . . . . .	23
3.5.4	Predicción con los valores $\sigma$ medios de los vectores de soporte . . . . .	23
3.6	Optimización y evaluación del método propuesto . . . . .	24
3.6.1	Método de optimización . . . . .	24
3.6.2	Medidas de desempeño . . . . .	24
<b>4</b>	<b>Marco experimental y resultados</b>	<b>27</b>
4.1	Descripción de la base de datos . . . . .	27
4.1.1	Configuración de la base de datos . . . . .	28
4.2	Descripción del marco experimental . . . . .	29
4.2.1	Variaciones del método y número de pruebas . . . . .	29
4.2.2	Configuración de las pruebas . . . . .	29
4.3	Resultados con penalización por norma- $\ell_1$ . . . . .	30
4.3.1	Usando características perceptuales . . . . .	30
4.3.2	Usando características radiómicas . . . . .	31
4.3.3	Combinando características perceptuales y radiómicas . . . . .	32
4.4	Resultados con penalización por norma- $\ell_2$ . . . . .	33
4.4.1	Usando características perceptuales . . . . .	33
4.4.2	Usando características radiómicas . . . . .	35
4.4.3	Combinando características perceptuales y radiómicas . . . . .	36
4.5	Comparación con métodos de selección de características . . . . .	37
4.5.1	Método de selección Fisher Score . . . . .	37
4.5.2	Método de selección Gini Index . . . . .	38
<b>5</b>	<b>Conclusiones y trabajo futuro</b>	<b>41</b>
<b>A</b>	<b>Apéndice A: Parámetros óptimos</b>	<b>43</b>
	<b>Bibliografía</b>	<b>44</b>

# ÍNDICE DE FIGURAS

1.	Ejemplos de secuencias de MRI para el diagnóstico de cáncer de mama . . . . .	1
1.1.	Representación gráfica de la técnica de resonancia magnética . . . . .	7
1.2.	Visualización de las secuencias de MRI utilizadas . . . . .	9
2.1.	Características y mapa de saliencia. (a) ROI original. (b) Los tres mapas de características se agrupan de acuerdo al modelo GBVS utilizado para generar el mapa de saliencia. (c) El mapa de saliencia en sus representaciones de escala de grises y de color. . . . .	14
3.1.	Representación de una base de datos separada por una SVM . . . . .	19
3.2.	Diagrama de flujo del método propuesto . . . . .	23
4.1.	Curvas de desempeño para el conjunto de datos de prueba cuando se realiza eliminación de fuentes con características perceptuales y norma- $\ell_1$ . . . . .	30
4.2.	Curvas de desempeño para el conjunto de datos de prueba cuando se realiza eliminación de fuentes con características radiómicas y norma- $\ell_1$ . . . . .	32
4.3.	Curvas de desempeño para el conjunto de datos de prueba cuando se realiza eliminación de fuentes con características perceptuales y radiómicas con norma- $\ell_1$ . . . . .	33
4.4.	Curvas de desempeño para el conjunto de datos de prueba cuando se realiza eliminación de fuentes con características perceptuales y norma- $\ell_2$ . . . . .	34
4.5.	Curvas de desempeño para el conjunto de datos de prueba cuando se realiza eliminación de fuentes con características radiómicas y norma- $\ell_2$ . . . . .	35
4.6.	Curvas de desempeño para el conjunto de datos de prueba cuando se realiza eliminación de fuentes con características perceptuales y radiómicas con norma- $\ell_2$ . . . . .	36
4.7.	Curvas de desempeño para el conjunto de datos de prueba cuando se realiza eliminación de características con método Fisher Score . . . . .	38
4.8.	Curvas de desempeño para el conjunto de datos de prueba cuando se realiza eliminación de características con método Gini Index . . . . .	39

# ÍNDICE DE TABLAS

1.1. Categorías de evaluación BI-RADS® y recomendaciones de manejo . . . . .	11
2.1. Características radiómicas generadas . . . . .	17
4.1. Distribución de estudios de la base de datos empleada para evaluación . . . . .	28
4.2. Distribución general de las pruebas realizadas . . . . .	29
4.3. Resultados de selección de secuencias de MRI con características perceptuales y norma- $\ell_1$ , para el conjunto de datos de prueba . . . . .	31
4.4. Resultados de selección de secuencias de MRI con características radiómicas y norma- $\ell_1$ . . . . .	32
4.5. Resultados de selección de secuencias de MRI con características perceptuales y radiómicas con norma- $\ell_1$ . . . . .	33
4.6. Resultados de selección de secuencias de MRI con características perceptuales y norma- $\ell_2$ . . . . .	34
4.7. Resultados de selección de secuencias de MRI con características radiómicas y norma- $\ell_2$ . . . . .	35
4.8. Resultados de selección de secuencias de MRI con características perceptuales y radiómicas con norma- $\ell_2$ . . . . .	36
4.9. Resultados de selección de secuencias de MRI con Fisher Score . . . . .	38
4.10. Resultados de selección de secuencias de MRI con Gini Index . . . . .	39
A.1. Parámetros óptimos obtenidos para cada prueba . . . . .	43



# INTRODUCCIÓN

El cáncer es una enfermedad crónica no transmisible y uno de los principales problemas de investigación en salud pública en el mundo [1]. El cáncer de mama es el segundo más común entre todos los tipos de cáncer y es también el más diagnosticado en mujeres, con una cifra estimada de 2,1 millones de casos nuevos en el año 2018, según los datos presentados por la Agencia Internacional para la Investigación sobre el Cáncer [2]. En Colombia, el cáncer de mama ha presentado un constante incremento en las últimas décadas, presentándose en 2012 un total de 8686 casos nuevos y 2649 muertes a causa de esta enfermedad, lo cual representa el 12% de la mortalidad de mujeres en el país [3], estas cifras son verdaderamente alarmantes, a tal punto que ha sido necesario analizar su efecto directo sobre la esperanza de vida en Colombia [4]. En el año 2017 se realizó un análisis estadístico con el propósito de calcular el número de casos nuevos de cáncer de mama que se presentarán en Colombia en los próximos años, mostrando resultados poco alentadores, pues para el año 2020 se estima que el número de casos nuevos por año supere los 10000 y para el año 2050 llegue a superar incluso los 18000 [5].

Debido a la gran importancia que ha adquirido el estudio del cáncer de mama en el mundo, se han desarrollado diferentes métodos de tamizaje y diagnóstico haciendo uso de imágenes diagnósticas, algunos de los más implementados son la mamografía, la tomografía computarizada, el ultrasonido, y recientemente la resonancia magnética, entre otros [6, 7]. El uso de MRI es actualmente uno de los métodos más prometedores en cuanto a la información útil que proporciona, dado que permite observar diferentes características del tejido (morfológicas, de difusión de las moléculas de agua, de densidad, entre otras) a partir de la adquisición de diferentes secuencias de imágenes [8].

En MRI, una secuencia hace referencia al conjunto de parámetros que permiten la reconstrucción de una imagen volumétrica de los órganos de estudio. En el caso específico del estudio de las glándulas mamarias para la detección de cáncer, en la práctica clínica se ha impuesto el uso de imágenes multiparamétricas, que combinan información morfológica y funcional en un mismo estudio. Así, el protocolo de estudio de mama suele incluir secuencias morfológicas en diferentes planos anatómicos, como son las imágenes de relajación T1 y T2 y las imágenes de recuperación corta de la inversión T1 (STIR por las siglas en inglés de *Short-T1 Inversion Recovery*); y secuencias funcionales, como imágenes de difusión ponderada (DWI por las siglas en inglés de *Diffusion Weighted Imaging*), e imágenes dinámicas contrastadas (DCE por las siglas en inglés de *Dynamic Contrast Enhanced*); adicionalmente, mediante la aplicación de técnicas de procesamiento de imágenes disponibles en el equipo de adquisición, el técnico radiólogo genera imágenes adicionales, como las imágenes de sustracción del dinámico contrastado (comúnmente conocidas como fases), y los mapas de coeficiente de difusión aparente (ADC por las siglas en inglés de *Apparent Diffusion Coefficient*). Como se puede observar en la Figura 1, cada secuencia de MRI describe características diferentes del mismo órgano, y cada una aporta información detallada que es analizada por los radiólogos, para emitir una interpretación diagnóstica.

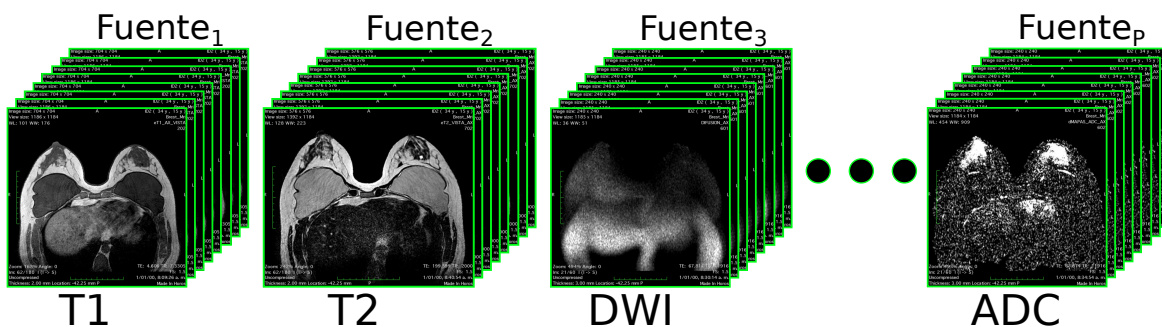


Figura 1: Ejemplos de secuencias de MRI para el diagnóstico de cáncer de mama

## PROTOS ABREVIADOS DE MRI PARA EL DIAGNÓSTICO DEL CÁNCER

A pesar de la alta sensibilidad que han demostrado las MRI, el protocolo de adquisición descrito anteriormente, conocido como protocolo completo, es comúnmente indicado sólo para la valoración de extensión y planeación quirúrgica de tumores malignos, la evaluación de respuesta a tratamiento neoadyudante, la detección precoz de recidiva; y sólo se indica como mecanismo de tamizaje en mujeres con alto riesgo de desarrollo de la enfermedad [9], debido a su alto costo, respecto a otros métodos de tamizaje, como la mamografía o el ultrasonido [10, 11]. Esto implica que su alcance esté limitado, en cuanto al número de personas que pueden acceder a este tipo de estudios para obtener un diagnóstico más preciso y acertado. Este costo está principalmente asociado al número de secuencias de imágenes que deben ser adquiridas, las cuales requieren tiempo de uso del resonador e incrementan el tiempo de lectura de los especialistas. Adicionalmente, en términos de requerimientos computacionales para el análisis de los estudios de MRI, el número de secuencias adquiridas también tiene importancia dado que un solo estudio asociado a un paciente puede tener miles de imágenes de alta resolución, de tal manera que se requieren sistemas con una elevada capacidad de almacenamiento y procesamiento para poder administrar estos estudios de forma eficiente [12]. Por lo anterior, en los últimos años se han orientado esfuerzos de investigación para definir protocolos abreviados que permitan mantener la capacidad diagnóstica de los especialistas mientras utilizan un número reducido de secuencias de MRI [13–16]. En términos generales, para definir un protocolo abreviado para el diagnóstico de cáncer de mama, se identifican las principales secuencias de MRI a partir de la experiencia de los radiólogos. Pese a que los resultados han mostrado ser efectivos en cuanto a la conservación de los niveles de sensibilidad, especificidad y exactitud [17], además de mostrar una considerable reducción de tiempos de adquisición y de lectura [18], su reproducibilidad es cuestionada dado que la mayoría de estos son desarrollados de forma subjetiva, por lo que se encuentran ligados directamente a la apreciación y conocimiento personal de los radiólogos implicados [19], y se basan en estudios retrospectivos con metodologías propias sin garantizar una estandarización de técnicas respecto a otros centros médicos [20]. Adicionalmente, y dado que la selección de las secuencias que conforman estos protocolos se orienta exclusivamente al proceso de interpretación visual del radiólogo, no se garantiza que la información que estas proveen sea la que mayor aporte tenga para el desarrollo de sistemas computacionales de apoyo al diagnóstico, un área de investigación que ha mostrado ser muy relevante en el mejoramiento de la interpretación cuando se analizan imágenes médicas.

Por lo anterior, surge la necesidad de desarrollar herramientas que contribuyan a la selección objetiva de las secuencias a emplear en estudios de este tipo, para generar una reducción de costos en el proceso de diagnóstico y mejorar el aprovechamiento de las capacidades computacionales de los sistemas implementados para esta tarea. Desde el punto de vista de la inteligencia computacional, este problema puede verse como una tarea de selección de fuentes, en donde cada una de las secuencias de MRI puede ser vista como una fuente de información independiente (ver Figura 1); de esta manera, el problema se centra en determinar cuáles de estas fuentes (secuencias) proveen la información más relevante para el diagnóstico y cuáles de ellas pueden ser ignoradas debido a la poca información que proporcionan.

## SELECCIÓN DE FUENTES CON APRENDIZAJE DE MÁQUINA

Teniendo en cuenta que el uso de sistemas computacionales para apoyar los procesos de diagnóstico ha mostrado ser de gran utilidad en el área médica [21, 22], se ha planteado con base en la literatura la posibilidad de implementar nuevos métodos de selección de fuentes relevantes de forma automática mediante el uso de algoritmos de aprendizaje de máquina o mejor conocidos como *machine learning*. Recientemente se publicó un trabajo en el cual utilizan diferentes algoritmos de aprendizaje de máquina para analizar el texto no estructurado de las indicaciones clínicas y los datos demográficos de los pacientes para protocolizar automáticamente los procedimientos de MRI a nivel de secuencias en neurología [23], aunque en dicho trabajo no realizan realmente una tarea de análisis de relevancia de las secuencias de MRI, debido a que enfocan el modelo de selección hacia una tarea de clasificación multi-etiqueta, se resalta cómo efectivamente el uso de algoritmos de aprendizaje de máquina puede optimizar los protocolos que involucran el uso de MRI y dejar de lado el problema de la subjetividad en este tipo de tareas.

La primera aproximación que se tiene al mencionar la selección automática de fuentes relevantes es la selección de características por medio de métodos de aprendizaje de máquina [24], esta aproximación se da principalmente porque una fuente de información puede ser considerada como un grupo de características de una

misma naturaleza. La selección de características es un tema ampliamente abordado desde el área de aprendizaje de máquina, es posible adaptar métodos clásicos de selección de características para seleccionar o eliminar varias características al mismo tiempo, este tema es conocido como selección de grupos de características [25, 26]. Los trabajos que aplican selección de grupos de características suelen identificar las diferentes fuentes de información entre cada área de investigación, como son los electrodos en un sistema de Interfaz Cerebro-Computadora (BCI por las siglas en inglés de *Brain-Computer Interface*) [27] o las distintas bandas de frecuencia en imágenes multiespectrales e hiperespectrales [28].

Aunque existe una gran variedad de estrategias que pueden ser implementadas con el propósito de seleccionar grupos de características, hay un desafío común entre todos estos métodos, el cual yace en el hecho de que al implementar algoritmos de selección de características como base, no se garantiza que pueda verse la fuente de información como un bloque completo e independiente, es decir, que estos métodos suelen conservar una noción individual de las características para determinar una aparente relevancia de la fuente de información, perdiendo así las propiedades globales de la fuente ya que se pueden romper algunas relaciones importantes entre las características que la componen.

Un método que ha ganado gran importancia en los últimos años desde el área de aprendizaje de máquina, referente al aprovechamiento de la información proveniente de múltiples fuentes, es el MKL [29, 30]. El MKL permite implementar una medida de similitud (kernel) asociada a cada fuente de forma independiente antes de ser integrada a la tarea de aprendizaje [31, 32], permitiendo que la información proveniente de cada fuente no solo sea aprovechada al máximo, sino que también permite obtener resultados fáciles de interpretar con relación al análisis de cada una de las fuentes de información [33].

El uso de MKL ha presentado una gran cantidad de beneficios para las diferentes tareas en que ha sido implementado, especialmente cuando es utilizado junto con el método de SVM [34], pero también al implementarse con otro tipo de algoritmos de aprendizaje de máquina [35]. El MKL además de estar muy bien documentado en el estado del arte, ha mostrado ser útil para la identificación y selección de fuentes de información relevantes [36, 37], debido a que permite tener una medida de similitud de los datos por cada fuente sin perder las posibles relaciones internas de las características que la componen, habilitando de esta forma la posibilidad de realizar estudios objetivos de análisis de relevancia muy fáciles de interpretar.

En el estado del arte se ha reportado el uso de MKL mediante la asignación de pesos a cada uno de los kernels asociados a las fuentes de información [38], lo cual es una característica importante de esta metodología ya que estos pesos pueden ser utilizados para determinar la relevancia que cada fuente representa para la solución implementada [39–41]. Se ha demostrado efectivamente que la selección de fuentes de información relevantes, mediante la penalización de los pesos de los kernels, es de gran utilidad y proporciona información fácil de interpretar [42], además de ser un método que puede ser llevado a diferentes áreas de la investigación, incluyendo el diagnóstico de cáncer de mama por MRI, área en la que hasta ahora, no ha sido implementada esta metodología.

## PROBLEMA DE INVESTIGACIÓN

La reducción del número de secuencias de MRI (fuentes de información) utilizadas para el diagnóstico de cáncer de mama es un tema de gran importancia, tanto para reducir el costo de este tipo de estudios y facilitar el acceso a los mismos por un mayor número de personas, como para optimizar el uso de recursos computacionales para su almacenamiento y procesamiento. Sin embargo, hasta ahora, este problema ha sido abordado solamente de manera subjetiva, tomando decisiones basadas en la experticia de los radiólogos involucrados, generando así una alta variabilidad en cuanto a las secuencias seleccionadas para desarrollar los protocolos abreviados. Ante este panorama, surge la pregunta de investigación que se desea abordar con el desarrollo de este trabajo:

**¿Cómo seleccionar objetivamente las secuencias de resonancia magnética más relevantes para el diagnóstico de cáncer de mama?**

Desde el punto de vista de la automatización, los métodos computacionales han mostrado ser de gran ayuda para apoyar las decisiones de los especialistas en medicina, abriendo paso así a la posibilidad de que el aprendizaje de máquina pueda ser implementado para desarrollar técnicas que apoyen la decisión de los radiólogos,

en cuanto a las secuencias que deben ser tenidas en cuenta como más relevantes para realizar un diagnóstico preciso con una metodología estandarizada.

En este trabajo se propone abordar este problema como una tarea de selección de fuentes, en la cual, cada secuencia de resonancia magnética es tomada como una fuente de información independiente. En particular, se plantea la elaboración de un esquema metodológico basado en el uso de MKL para representar cada secuencia de MRI con una medida de similitud (kernel), asociando cada una de estas medidas a un peso penalizado para entrenar un algoritmo de clasificación SVM e identificar así, de forma automática, las secuencias más relevantes con base en el valor final de los pesos entrenados. Para desarrollar esta solución, se plantean los siguientes desafíos específicos de investigación:

1. ¿Cómo se puede representar de forma compacta la información visual de las imágenes volumétricas producidas en cada secuencia de MRI, para ser analizadas con MKL?
2. ¿Qué tipo de kernel es el más apropiado para representar los datos provenientes de las secuencias de MRI caracterizadas?
3. ¿Cuál es el método de penalización adecuado para permitir que el algoritmo de aprendizaje pueda discriminar correctamente las secuencias relevantes de las que no lo son?
4. ¿Cómo se puede validar que las secuencias seleccionadas por el método computacional como relevantes, en realidad sí lo son para el proceso de diagnóstico?

## OBJETIVOS

### GENERAL

Desarrollar un método basado en el uso de MKL y SVM para seleccionar de forma objetiva y automática las secuencias de resonancia magnética más relevantes para el diagnóstico de cáncer de mama.

### ESPECÍFICOS

- Desarrollar un modelo de extracción de características para representar la información visual contenida en regiones de interés diagnóstico de secuencias de MRI que pueda ser procesado mediante MKL.
- Identificar el o los tipos de kernels que permiten representar de forma correcta la información proporcionada por cada secuencia de MRI mediante la optimización de los parámetros variables de cada kernel.
- Diseñar un algoritmo mediante el uso de MKL y SVM para seleccionar automáticamente las secuencias de MRI más relevantes, asociando pesos penalizados a cada uno de los kernels que representan las secuencias de MRI.
- Evaluar el desempeño del método propuesto en la selección de secuencias de MRI que permitan obtener una mejor sensibilidad/especificidad en un método del estado del arte para el diagnóstico automático de cáncer de mama.

## ESTRUCTURA DEL DOCUMENTO

**Capítulo 1:** En este capítulo se presentan los conceptos teóricos fundamentales sobre el diagnóstico de cáncer de mama mediante el uso de MRI y se describe en detalle la conformación de la base de datos utilizada en este trabajo.

**Capítulo 2:** En este capítulo se presenta el modelo implementado para generar las características que describen cada una de las secuencias de MRI, teniendo en cuenta tanto la percepción visual de los especialistas en radiología como la información computacional de las imágenes.

**Capítulo 3:** En este capítulo se presenta el fundamento teórico y matemático referente al uso de SVM y MKL, describiendo con detalle la metodología propuesta, desde la optimización de los parámetros de los kernels y la SVM hasta la identificación de la relevancia de las secuencias de MRI a partir de los pesos asociados a los kernels.

**Capítulo 4:** En este capítulo se presenta el marco experimental implementado para medir el desempeño de la metodología propuesta, finalizando con la evaluación de los resultados obtenidos mediante la implementación

de dos métodos clásicos de selección de características del estado del arte, los cuales se comparan con el método propuesto respecto a la selección objetiva de secuencias de MRI relevantes.

**Capítulo 5:** En este capítulo se presentan las conclusiones finales y las posibles líneas de trabajo que pueden ser abordadas en un futuro, tomando como base los resultados obtenidos en este proyecto.

## CONTRIBUCIONES

- Artículo titulado “*Breast Lesion Discrimination Using Saliency Features from MRI Sequences and MKL-Based Classification*”, publicado en *Lecture Notes in Computer Science (ISSN 0302-9743)* con categoría Q2 de Scimago (SJR). Este trabajo será presentado en el *24th Iberoamerican Congress on Pattern Recognition (CIARP 2019)* en La Habana (Cuba).
- Artículo titulado “*Feature Group Selection Using MKL Penalized with  $\ell_1$ -norm and SVM as Base Learner*”, publicado en *Communications in Computer and Information Science (ISSN 1865-0929)* con categoría Q3 de Scimago (SJR). Este trabajo fue presentado en el *Workshop Engineering Applications 2018 (WEA-2018)* en la ciudad de Medellín (Colombia) [43].
- Pasantía de investigación titulada “*Evaluación de la relevancia de conjuntos de características radiómicas para la discriminación de tumores de cuello en secuencias T1 y T2 de imágenes de resonancia magnética*”. Esta pasantía fue realizada en el *Laboratorio de Imagen Médica y Biometría (LAIMBIO)* de la Escuela Superior de Ciencias Experimentales y Tecnología de la Universidad Rey Juan Carlos de España.
- Artículo titulado “*Radiomics-Based Differentiation of Pleomorphic Adenomas and Warthin Tumors of Salivary Gland*”, aceptado para presentación oral en el *15th International Symposium on Medical Information Processing and Analysis (SIPAIM 2019)*.

# 1

## IMÁGENES DE RESONANCIA MAGNÉTICA

### Índice

---

<b>Protocolos abreviados de MRI para el diagnóstico del cáncer</b> . . . . .	2
<b>Selección de fuentes con aprendizaje de máquina</b> . . . . .	2
<b>Problema de investigación</b> . . . . .	3
<b>Objetivos</b> . . . . .	4
General . . . . .	4
Específicos . . . . .	4
<b>Estructura del documento</b> . . . . .	4
<b>Contribución</b> . . . . .	5

---

### 1.1. FUNDAMENTOS FÍSICOS DE LAS MRI

Las MRI han demostrado ser una herramienta eficaz para el diagnóstico de cáncer de mama, siendo una técnica no invasiva que ofrece excelente detalle tanto en términos de contraste como de resolución espacial de los tejidos. El uso de MRI permite obtener información funcional y estructural de los órganos que están siendo analizados, especialmente cuando se trata de tejidos blandos.

El fundamento físico del método de resonancia magnética se basa en la emisión de un campo magnético sobre la zona del cuerpo que está siendo analizada, la mayoría de protones de los átomos de hidrógeno se alinean con la misma dirección del campo magnético aplicado, mientras que unos pocos permanecen en dirección opuesta conservando un estado de alta energía, debido a esta alineación aparece una *magnetización longitudinal* que conserva la misma dirección del campo magnético externo como se ilustra en la Figura 1.1b; posteriormente se aplica un pulso de radiofrecuencia que permite que algunos protones ganen energía, equilibrando los estados de alta y baja energía, de tal manera que desaparece la magnetización longitudinal; adicionalmente, el pulso de radiofrecuencia posee un período similar al período del movimiento de precesión de los protones, provocando que estos entren en resonancia y giren de forma sincronizada, debido a esto aparece una *magnetización transversal* al campo magnético externo como se observa en la Figura 1.1c [44].

Cuando se retira la señal de radiofrecuencia, se mide el tiempo que tarda en desaparecer la magnetización transversal y el tiempo que tarda en reaparecer la magnetización longitudinal [45], para esto se utiliza una serie de bobinas que captan información tridimensional precisa sobre puntos específicos del área examinada [46], la información de todos los puntos (*vóxeles* en la imagen tridimensional) se une posteriormente para formar una secuencia de MRI, la cual contiene la información tridimensional de la parte del cuerpo que está siendo estudiada. Una secuencia de MRI difiere de otra, dependiendo de los parámetros configurados en el resonador para realizar la medición. Los principales parámetros que se tienen en cuenta para modificar una secuencia son: el tiempo de repetición de los pulsos de radiofrecuencia y el tiempo en el que se obtiene la señal emitida por los

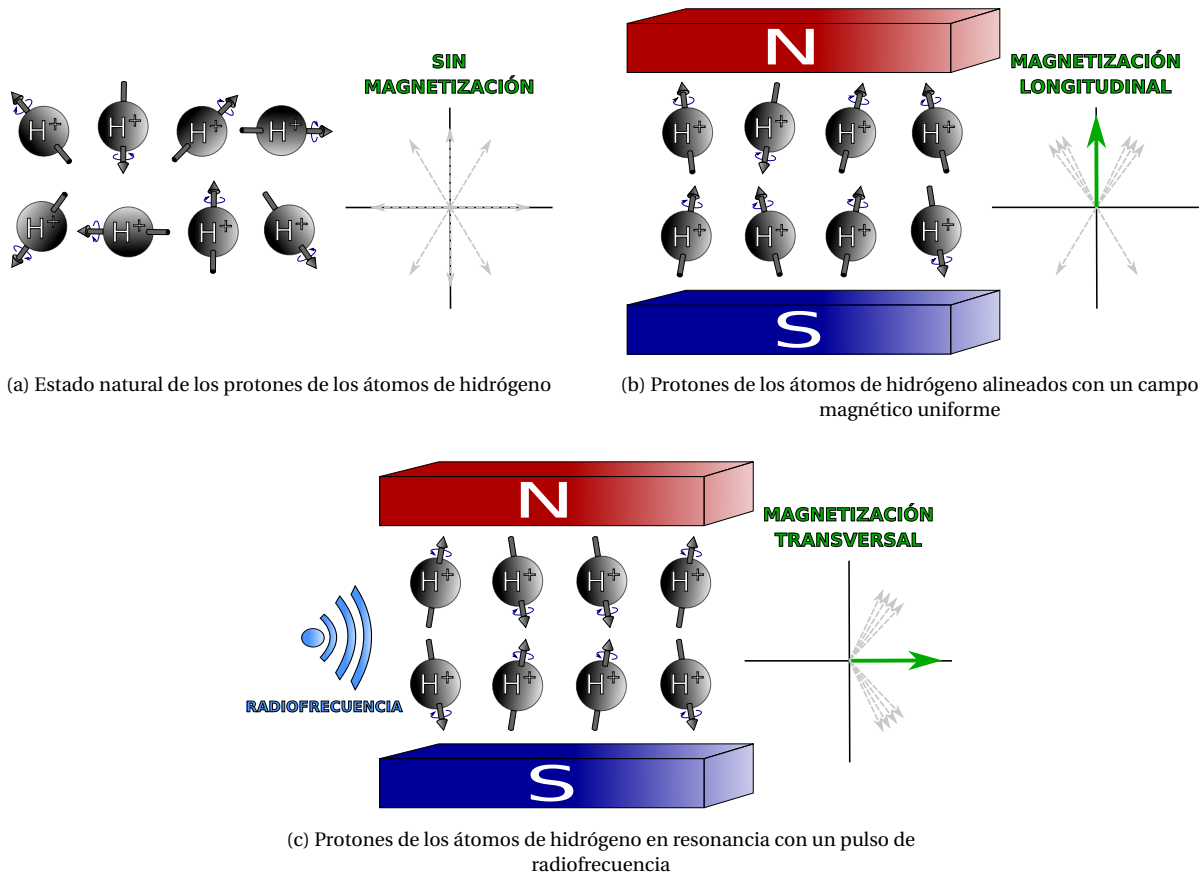


Figura 1.1: Representación gráfica de la técnica de resonancia magnética

protones, la cual también es conocida como *eco* [8]. Cada una de las secuencias de MRI resalta información diferente del tejido analizado, por lo cual se puede interpretar como que cada secuencia corresponde a una fuente de información independiente.

Respecto a otras técnicas utilizadas también para el diagnóstico de cáncer de mama, el uso de MRI presenta una significativa superioridad debido a que provee información espacial tridimensional de los tejidos con mejor nivel de contraste, esta es la principal razón para que actualmente las MRI sean propuestas como una herramienta para diagnosticar esta enfermedad [12]. Adicionalmente, el uso de MRI permite que al aplicar un medio de contraste en los tejidos se pueda mejorar la capacidad de detección del cáncer, ya que los tumores mamarios malignos demuestran tener una captación significativamente más alta de medios de contraste que los tumores benignos.

Los radiólogos son los especialistas en medicina encargados de realizar los diagnósticos de cáncer de mama utilizando MRI, estos especialistas utilizan su experiencia en el área para identificar las regiones de interés (ROI por las siglas en inglés de *Region of Interest*) en las imágenes, categorizando posteriormente las lesiones encontradas para finalmente describir y reportar sus hallazgos. Teniendo en cuenta que el personal médico especializado juega un papel protagónico para la tarea de detección y diagnóstico del cáncer, se ha considerado apoyar su labor mediante el uso de la tecnología, la cual se encuentra en constante desarrollo y perfeccionamiento. El diagnóstico asistido por computadora (CAD por las siglas en inglés de *Computer-Aided Diagnosis*) permite apoyar a los especialistas durante el proceso de diagnóstico [21, 22], siendo el área de visión por computadora una de las que se ha interesado más en esta labor debido a que permite procesar imágenes digitalmente y aplicar algoritmos de aprendizaje de máquina que realicen tareas de segmentación, reconocimiento y clasificación con resultados significativamente precisos y confiables [47], estos sistemas se encuentran en constante desarrollo y

cada vez se muestran más dispuestos a ser utilizados en procesos de diagnóstico reales.

## 1.2. TIPOS DE SECUENCIAS PARA EL DIAGNÓSTICO DE CÁNCER DE MAMA

Los protocolos médicos desarrollados para el diagnóstico de cáncer de mama con MRI hacen uso de diferentes tipos de secuencias; el principal propósito de esto es obtener suficiente información que describa el estado del tejido donde hay sospecha de la existencia de un tumor bien sea maligno o benigno. Aunque existe un gran número de secuencias de MRI que pueden ser generadas a partir de la variación de los parámetros de la medición [48], en el caso de este trabajo, las siguientes secuencias conforman el protocolo completo de estudios de MRI de mama en el Instituto de Alta Tecnología Médica (IATM), entidad con la cual se desarrolla este proyecto.

### 1.2.1. SECUENCIA T1

Esta secuencia describe el tiempo que tarda en recuperarse el 63% de la magnetización longitudinal original después de retirar el pulso de radiofrecuencia. Cuando la magnetización longitudinal se recupera, todos los protones de los átomos de hidrógeno que habían ganado energía deben liberarla en forma de calor hasta poder alinearse de nuevo con el campo magnético externo. Teniendo en cuenta que cada tipo de tejido posee un tiempo de relajación T1 diferente, esta secuencia proporciona información anatómica, que le permite al radiólogo diferenciar cambios postquirúrgicos como las cicatrices, que son importantes para la interpretación diagnóstica, cuando, por ejemplo no se cuenta con información suficiente de historia clínica o estudios previos. Esta secuencia, sin embargo, es poco sensible ante los diferentes cambios patológicos presentes en una lesión. Un ejemplo de una imagen tomada a partir de una secuencia T1 es el que se presenta en la Figura 1.2a, dicha imagen representa un solo *corte* o *slice* del volumen total que conforma la secuencia tridimensional.

### 1.2.2. SECUENCIA T2

En la secuencia T2 se mide el tiempo necesario para que la magnetización transversal disminuya hasta un 37% de su intensidad justo después de retirar el pulso de radiofrecuencia, esta reducción en la magnetización transversal se debe a que el campo magnético no es completamente homogéneo, por lo tanto, al retirar el pulso de radiofrecuencia, los protones de los átomos de hidrógeno dejan de girar en sincronía. A diferencia de la relajación T1, la relajación T2 ocurre en un tiempo considerablemente más reducido, además de que una secuencia T2 es más sensible a cambios patológicos pero no presenta una buena correspondencia anatómica como la secuencia T1; por lo cual, esta secuencia es frecuentemente utilizada para la identificación de cuadros patológicos de lesiones sólidas, teniendo en cuenta que estos tienden a generar un incremento en la acumulación de líquidos que generan hipointensidad en la secuencia T2. La Figura 1.2b muestra un ejemplo de un corte tomado de una secuencia T2.

### 1.2.3. SECUENCIA DWI

La secuencia DWI es ampliamente utilizada en radiología; el término “difusión” hace referencia al movimiento aleatorio de las moléculas en una sustancia, mejor conocido como *movimiento Browniano*; en una secuencia DWI se visualiza la intensidad de dicho movimiento dentro de los tejidos, siendo formada por imágenes derivadas de una secuencia T2 en donde el comportamiento de difusión de las moléculas de hidrógeno se determina bajo diferentes niveles de intensidad del campo magnético. Debido a que las moléculas presentan cambios de velocidad en un espacio sin limitaciones, es posible visualizar que estos cambios se restringen debido a la presencia de un medio membranoso como un tumor. Para poder adquirir esta secuencia se define un parámetro  $b$  que puede tomar valores entre 0 y 1000, este parámetro determina la sensibilidad con que se mide la difusión de las moléculas, de tal manera que se puede configurar para mostrar detalles de la atenuación del movimiento molecular [49]. En la Figura 1.2c se puede ver un corte de este tipo de secuencia.

### 1.2.4. SECUENCIA ADC

Teniendo en cuenta que la secuencia DWI se deriva de una T2, se presenta un efecto conocido como “brillo T2” sobre la secuencia DWI, el cual genera confusión entre los tejidos que presentan una restricción a la difusión y los tejidos con alto contenido de agua, debido a que ambos presentan un nivel de intensidad alto en esta secuencia. Para abordar este problema se genera una secuencia llamada ADC, a partir de dos secuencias DWI



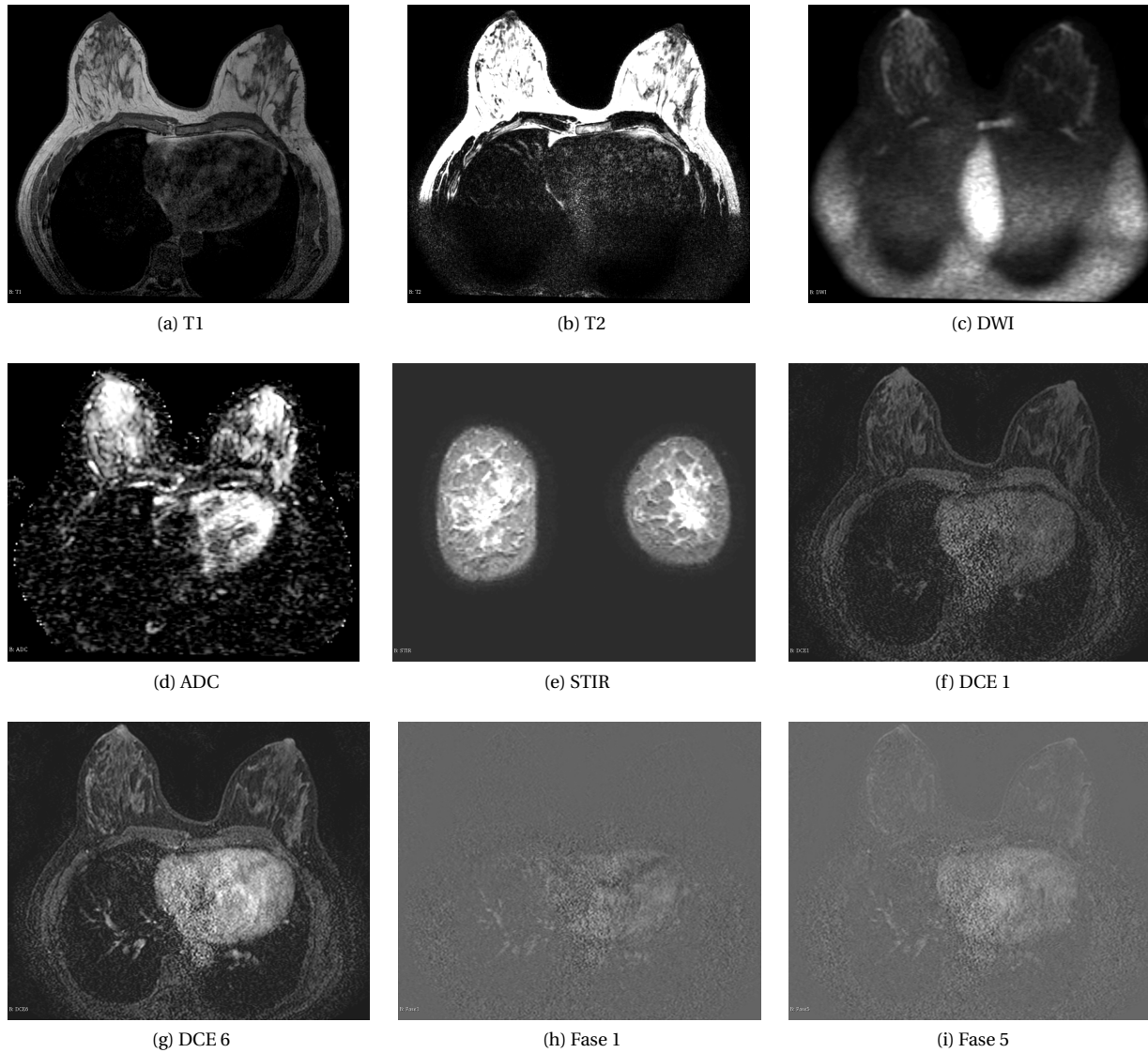


Figura 1.2: Visualización de las secuencias de MRI utilizadas

con valores de  $b$  diferentes (por lo general 0 y 1000), en esta secuencia se aplica procesamiento de imagen para filtrar el brillo T2, logrando que un tejido patológico (que presenta restricción a la difusión) se vea brillante en la secuencia DWI y opaco en la secuencia ADC. Comúnmente, las secuencias DWI y ADC se usan en conjunto para identificar patologías, la Figura 1.2d muestra un corte tomado de una secuencia ADC.

### 1.2.5. SECUENCIA STIR

De todas las secuencias que componen la base de datos, esta es la única que se toma desde el plano coronal (todas las demás se toman desde el plano axial). La secuencia STIR es utilizada para suprimir el tejido graso de las mediciones con el objetivo de disminuir las confusiones generadas por las intensidades de este tejido respecto al tejido patológico, esto se logra al identificar los tejidos que poseen un tiempo de relajación T1 entre  $200ms$  y  $300ms$ , debido a que este es el tiempo asociado al tejido graso, esta técnica también puede ser aplicada teniendo en cuenta la relajación T2. En radiología es muy común el uso de esta secuencia debido a que permite identificar regiones de interés con mayor facilidad, en la Figura 1.2e se puede ver un corte de una secuencia STIR

tomada desde el plano coronal.

### 1.2.6. SECUENCIAS DCE

Para la generación de estas secuencias se utilizan agentes de contraste con propiedades paramagnéticas que son inyectados mediante una infusión intravenosa para mejorar el contraste entre el tejido de fondo y el tejido de interés; el agente de contraste más utilizado en resonancia magnética es el *quelato de gadolinio* [50]. Las secuencias DCE corresponden a 6 secuencias T1 tomadas al aplicar un agente de contraste, la primera de ellas es adquirida antes de la administración de este, mientras que las 5 restantes se toman periódicamente luego de la administración de dicho agente, el tiempo entre cada toma puede variar dependiendo del protocolo que se implemente, pero suele ser alrededor de un minuto. Con las secuencias DCE se pueden visualizar características fisiológicas del tejido, dado que permiten conocer el comportamiento del mismo ante el agente de contraste, teniendo en cuenta variables como la cantidad captada y la velocidad con la que el contraste abandona el tejido. Comúnmente, se utilizan técnicas de supresión de tejido graso para evitar que este se confunda con el mejoramiento del contraste, las Figuras 1.2f y 1.2g muestran cortes de la primera y última secuencia DCE, respectivamente.

### 1.2.7. SUSTRACCIÓN O FASES

Las secuencias conocidas como “Fases” se obtienen a partir de las secuencias DCE, en total son 5 secuencias que corresponden a la sustracción aritmética de la primera secuencia DCE de cada una de las demás secuencias tomadas después de la administración del agente de contraste. La primera secuencia DCE es conocida como *secuencia de pre-contraste*, por lo cual, al extraer esta información del tejido después de administrar el agente de contraste, se pueden visualizar mejor los tejidos que presentan una mayor absorción de este agente; estas secuencias son muy utilizadas actualmente debido a que el tejido patológico posee una elevada actividad celular, lo que genera una alta captación del medio de contraste y una rápida liberación de este (proceso de lavado), en comparación con el tejido sano, además de poder analizar la velocidad de captación y de liberación de este medio con un mejor detalle. Las Figuras 1.2h y 1.2i muestran cortes de la primera y última secuencia de fase, respectivamente.

## 1.3. CATEGORIZACIÓN ESTANDARIZADA DE TUMORES EN MRI

El proceso de categorización de los hallazgos en MRI es llevado a cabo por un médico con especialización en el área de radiología, esta persona es la encargada de determinar la probabilidad de que un tumor o masa sea maligno o benigno. Luego de que las secuencias de MRI son adquiridas, el especialista busca y marca en ellas las regiones sospechosas asociadas a masas que resaltan entre el tejido normal, posteriormente, analiza con detalle dichas regiones teniendo en cuenta características que le permiten determinar un diagnóstico; dicho diagnóstico se realiza mediante el uso del léxico estandarizado conocido como BI-RADS, el cual proviene del inglés *Breast Imaging Reporting and Data System* [51].

La categorización BI-RADS establece un lenguaje estándar mediante el cual se analizan las masas presentes en el tejido mamario al describir características morfológicas y características de captación del contraste, estas últimas son mejor conocidas como características cinéticas. El uso de la categorización BI-RADS permite establecer protocolos para determinar de forma estandarizada la probabilidad de que una masa sea maligna o benigna. A continuación se definen con mejor detalle las características que son tenidas en cuenta para desarrollar esta tarea y las diferentes categorías que se asignan a los hallazgos al final del proceso de diagnóstico.

### 1.3.1. CARACTERÍSTICAS MORFOLÓGICAS

Este tipo de características tienen en cuenta dos variables en las masas, la primera de ellas es la forma, la cual se describe como redonda, ovalada, lobulada, o irregular; mientras que la segunda variable corresponde al margen de la masa, esta variable se enfoca en el análisis del borde, determinando si este es regular, espiculado o irregular. Un ejemplo del uso de este tipo de características se presenta al tener en cuenta que un borde regular en una masa indica que esta es probablemente benigna, mientras que la presencia de un borde espiculado en una masa es un indicador de que esta es probablemente maligna [52].

### 1.3.2. CARACTERÍSTICAS CINÉTICAS

Este tipo de características se analizan al inyectar en el tejido un agente de contraste como el quelato de gadolinio, el análisis se realiza teniendo en cuenta la velocidad y la magnitud de la captación que tiene el tejido ante el agente de contraste. Un tejido normal presenta captación del agente de contraste poco tiempo después de que este es inyectado, presentando un incremento leve y progresivo a lo largo de la prueba; por otro lado, la captación del agente de contraste en una masa puede presentarse de seis formas diferentes: realce heterogéneo, homogéneo, en el borde de la masa, de espacios internos en la masa, dejando espacios internos oscuros o finalmente, con un patrón de realce justo en el centro de la masa [53].

### 1.3.3. CATEGORÍAS BI-RADS

La Tabla 1.1 muestra las diferentes categorías que pueden ser asignadas a un hallazgo, luego de que se han analizado las diferentes características morfológicas y cinéticas mencionadas anteriormente, estas categorías se encuentran estandarizadas por el léxico BI-RADS y son implementadas globalmente.

Evaluación	Manejo	Probabilidad de Cáncer
Categoría 0: Incompleta	Requiere estudios adicionales de MRIs mamografía o Ultrasonido	N/A
Categoría 1: Negativo	Tamizaje de rutina por MRI , si el riesgo acumulado es $\geq 20\%$	0 %
Categoría 2: Benigno	Tamizaje de rutina por MRI, si el riesgo acumulado es $\geq 20\%$	0 %
Categoría 3: Probablemente benigno	Seguimiento en intervalo corto de tiempo (6-meses)	$\geq 2\%$
Categoría 4: Sospechoso de malignidad	Biopsia	$\geq 2\%$ pero $\leq 95\%$
Categoría 5: Altamente sospechoso de malignidad	Biopsia	$\geq 95\%$
Categoría 6: Malignidad comprobada con biopsia	Escisión quirúrgica si es clínicamente apropiado	N/A

Tabla 1.1: Categorías de evaluación BI-RADS® y recomendaciones de manejo

En este punto es importante señalar que la categorización BIRADS de un estudio, corresponde al de la lesión con categoría más alta identificada en el estudio. Esta categoría define la conducta o manejo a seguir. Como se observa en la misma tabla, si la categoría es 0, el estudio no es concluyente y se debe recomendar estudios adicionales, ya sea de MRI o de otras modalidades de imagen. Si es 1 o 2, la paciente deberá continuar con estudios rutinarios de tamizaje convencional, o de MRI si el riesgo acumulado de desarrollo de cáncer es  $\geq 20\%$ ; si la categoría es 3, se deberá realizar seguimiento cada seis meses, para determinar si hay cambios significativos que conlleven a una recategorización de las lesiones. Finalmente, si la categoría es 4 o 5, se deberá realizar biopsia para confirmación de malignidad. La categoría 6, se reserva cuando el estudio se realiza para valoración prequirúrgica o seguimiento a tratamiento, caso en el cual, la lesión ya ha sido biopsiada y su malignidad ha sido confirmada.

# 2

## REPRESENTACIÓN DE LOS DATOS

### Índice

---

<b>1.1 Fundamentos físicos de las MRI</b> . . . . .	<b>6</b>
<b>1.2 Tipos de secuencias para el diagnóstico de cáncer de mama</b> . . . . .	<b>8</b>
1.2.1 Secuencia T1 . . . . .	8
1.2.2 Secuencia T2 . . . . .	8
1.2.3 Secuencia DWI . . . . .	8
1.2.4 Secuencia ADC . . . . .	8
1.2.5 Secuencia STIR . . . . .	9
1.2.6 Secuencias DCE . . . . .	10
1.2.7 Sustracción o Fases . . . . .	10
<b>1.3 Categorización estandarizada de tumores en MRI</b> . . . . .	<b>10</b>
1.3.1 Características morfológicas . . . . .	10
1.3.2 Características cinéticas . . . . .	11
1.3.3 Categorías BI-RADS . . . . .	11

---

Teniendo en cuenta que la información contenida en las ROI extraídas de las secuencias de MRI debe ser llevada hacia una forma matricial que pueda ser procesada mediante algoritmos de aprendizaje de máquina, se implementaron dos tipos de características diferentes para describir la información de los hallazgos en las secuencias de imágenes, estos tipos de características son las perceptuales y las radiómicas, los fundamentos teóricos de ambas son descritos a continuación.

### 2.1. CARACTERÍSTICAS PERCEPTUALES

Las características perceptuales hacen referencia a las propiedades de una imagen que pueden ser captadas por la percepción humana, existe un área de investigación conocida como atención visual computacional (CVA por las siglas en inglés de *Computational Visual Attention*) que se enfoca en poder llevar este tipo de características a los sistemas computacionales [54, 55]; la gran ventaja de este tipo de características es que pueden describir los puntos de interés percibidos por un radiólogo al momento de analizar una imagen de resonancia magnética para determinar un diagnóstico, permitiendo que la máquina identifique la información que puede ser más relevante para la vista humana, teniendo en cuenta que la información más representativa de la imagen para la máquina no es necesariamente la misma información que es representativa para el radiólogo.

La atención visual selectiva observada en los seres vivos, en particular, la percepción de la saliencia visual (o relevancia visual) permite a los seres humanos detectar, procesar y actuar rápidamente sobre las partes que proporcionan más información en un plano visual o imagen, simplificando enormemente la función de la visión para el cerebro humano [56]; esto ha aumentado el interés en el estudio sobre sistemas visuales humanos,

debido a que la saliencia visual está relacionada con la singularidad, la rareza y la sorpresa de una escena, lo cual es determinado por características primitivas como el color, la textura, la forma, entre otras [57]. Una región visualmente saliente es típicamente rara en una imagen, y contiene información altamente discriminativa [58], por lo tanto, este concepto ha comenzado a tener una gran influencia en la identificación de lesiones de forma automática en imágenes médicas, entrando como una herramienta valiosa en el área de CAD al poder acercarse más específicamente a la forma en que un especialista en medicina analiza una imagen para determinar un diagnóstico [59].

Los modelos computacionales de saliencia toman las imágenes como entrada y generan un mapa topográfico que mide el nivel de relevancia o atracción que puede tener cada área de la imagen para un observador humano; muchos de estos modelos analizan aspectos del movimiento del ojo humano para identificar el lugar de fijación de este en las imágenes [56, 57]. Por lo tanto, es posible definir la saliencia visual como el resultado de comparar una región con su entorno, teniendo en cuenta la imprevisibilidad, el contraste y la rareza de dicha región [60, 61]. Además, estos métodos de detección de zonas salientes por lo general emplean dos tipos de estrategias, como “*bottom-up*” y “*top-down*”. La primera, consiste en detectar las regiones prominentes mediante el uso de características de bajo nivel, como el color, la intensidad y la orientación, mientras que la segunda estrategia incluye algo de aprendizaje con datos de entrenamiento que involucran la posición o la forma de un objeto destacado.

### 2.1.1. SALIENCIA VISUAL BASADA EN GRAFOS

Se ha demostrado que el modelo de saliencia visual basado en grafos (GBVS por las siglas en inglés de *Graph-Based Visual Saliency*) es uno de los que presenta mejor rendimiento de predicción en las fijaciones oculares de escenas naturales y de imágenes médicas [62]. Este modelo consiste en extraer mapas de características en múltiples escalas espaciales como intensidad, color y orientación o movimiento; luego, se aplica una transformada piramidal Gaussiana de espacio de escala que se deriva de cada característica [63]. Posteriormente, se genera un grafo completamente conectado sobre todas las ubicaciones de cuadrícula o píxel de cada mapa de características, donde los pesos de dos nodos se asignan proporcionalmente a la similitud de los valores de las características y su distancia espacial. Por lo que la disimilitud o diferencia entre dos posiciones  $(i, j)$  y  $(p, q)$  en el mapa de características, con su respectivo valor  $M(i, j)$  y  $M(p, q)$ , es definido como se muestra en la Ecuación (2.1).

$$d((i, j) \vee (p, q)) = \left| \log \frac{M(i, j)}{M(p, q)} \right| \quad (2.1)$$

A la arista dirigida del nodo  $(i, j)$  al nodo  $(p, q)$  se le asigna un peso proporcional a su disimilitud y a su distancia en el mapa de características  $M$  como se muestra en la Ecuación (2.2).

$$w((i, j), (p, q)) = d((i, j) \vee (p, q)) F(i - p, j - q) \quad (2.2)$$

donde la función  $F$  determina la medida de conectividad entre nodos y se define según la Ecuación (2.3).

$$F(a, b) = \exp\left(\frac{a^2 + b^2}{2\beta^2}\right) \quad (2.3)$$

donde  $\beta$  es un parámetro libre determinado por el usuario para ajustar la medida de conectividad a los datos que estén siendo usados [64].

Los grafos resultantes se tratan como cadenas de Márkov para normalizar los pesos de los bordes salientes de cada nodo en valores unitarios y definiendo una relación de equivalencia entre nodos y estados, así como entre los pesos de los bordes y las probabilidades de transición. Su distribución de equilibrio se adopta como los mapas de activación y de saliencia, teniendo en cuenta que los nodos que son muy diferentes a sus nodos circundantes tendrán valores grandes en esta distribución. Por último, los mapas de activación se normalizan para enfatizar los detalles conspicuos y luego se combina en un solo mapa general saliente [65, 66].

### 2.1.2. MATRIZ DE CARACTERÍSTICAS PERCEPTUALES

Luego de un proceso experimental en el que se evaluaron los diferentes grupos de características perceptuales, se pudo identificar que los mapas de características de intensidad (I), movimiento (M) y orientación (O) favorecieron considerablemente el ajuste de los niveles de saliencia con las regiones marcadas por los radiólogos.

Es importante tener en cuenta que en las secuencias de MRI no es posible generar el mapa de características de color, por lo cual se utilizó el mapa de orientación, ya que permite predecir la fijación de los ojos humanos [67].

Los mapas de características conforman el mapa de saliencia (SaM), el cual puede tener también una representación como mapa de saliencia de color (SaM-C), teniendo en cuenta que estos mapas tienen el mismo tamaño de las imágenes a partir de las cuales fueron generados, se decidió calcular finalmente las primeras 5 medidas estadísticas de primer orden sobre las ROI originales y los mapas de saliencia, teniendo en cuenta que estos últimos recogen la información de los mapas de características I, M y O. El procedimiento mencionado anteriormente se representa con detalle en la Figura 2.1.

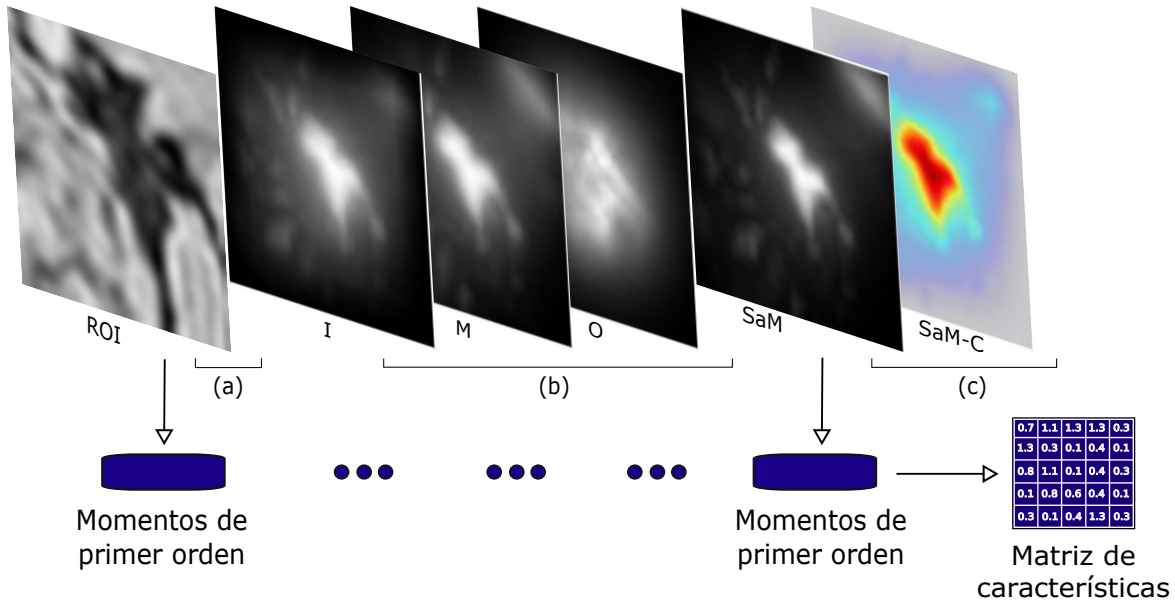


Figura 2.1: Características y mapa de saliencia. (a) ROI original. (b) Los tres mapas de características se agrupan de acuerdo al modelo GBVS utilizado para generar el mapa de saliencia. (c) El mapa de saliencia en sus representaciones de escala de grises y de color.

El primer momento estadístico es determinado por la media ( $m_1$ ), la cual se calcula utilizando en la Ecuación (2.4), mientras que los demás momentos estadísticos centrales ( $\mu_k$ ) de la variable aleatoria  $I$  que representa los niveles de intensidad en una región son dados por la Ecuación (2.5).

$$m_1 = E[I^1] = \sum_{I=0}^{N_g-1} I^1 P(I) \quad (2.4)$$

$$\mu_k = E[(I - E[I])^k] = \sum_{I=0}^{N_g-1} (I - m_1)^k P(I), \text{ con } k = 2, 3, \text{ y } 4 \quad (2.5)$$

Donde  $N_g$  es el número de los posibles niveles de intensidad.  $\mu_2$  representa la varianza, que es el momento central más común, e indica la variabilidad de los datos con respecto a la media,  $\mu_3$  es la asimetría que permite establecer el grado de simetría del histograma con respecto a la media y el cuarto momento  $\mu_4$  es la curtosis, el cual señala el grado de concentración que tienen los datos alrededor de la media en el histograma [68]. Adicionalmente, se calculó la desviación estándar, la cual corresponde a la raíz cuadrada del segundo momento  $\mu_2$ .

De esta forma, por cada una de las secuencias de MRI se generó una matriz de características perceptuales de  $146 \times 10$ , donde el número de filas corresponde al número total de regiones extraídas de los estudios de MRI y el número de columnas corresponde a los momentos estadísticos concatenados de las regiones originales y los mapas de saliencia. Es importante aclarar estas características fueron extraídas de la información bidimensional contenida en el corte específico que fue marcado por el radiólogo en cada secuencia de MRI, teniendo en cuenta que la formulación modelo GBVS solo está planteada para imágenes en  $2D$ .

## 2.2. CARACTERÍSTICAS RADIÓMICAS

Las características radiómicas representan un enfoque cuantitativo para el análisis de imágenes médicas dirigido a descifrar las propiedades morfológicas y funcionales de una lesión [69, 70], este tipo de características se enfocan en describir los puntos de interés relevantes para la máquina, de tal manera que no proporcionarían mucha información útil para el radiólogo pero pueden tener información sumamente importante para que un sistema computacional pueda interpretar la imagen médica. El principal propósito de integrar características perceptuales y radiómicas en este documento, es obtener resultados relevantes tanto para la máquina como para el radiólogo, teniendo en cuenta que los sistemas de CAD son diseñados como sistemas de apoyo, por lo cual deben tratar de emular y potenciar la forma en que los especialistas en medicina analizan las imágenes de MRI.

### 2.2.1. DEFINICIÓN DE RADIÓMICA

La radiómica es una ciencia proveniente del CAD convencional que busca básicamente generar una gran cantidad de características a partir de una imagen, esto con el objetivo de poder aplicar estrategias de minería de datos y de reconocimiento de patrones que permitan obtener hallazgos para formular una hipótesis a favor del problema en cuestión [71]. La radiómica busca principalmente apoyar los procesos de diagnóstico de los especialistas en medicina, para facilitar y robustecer la evaluación, el pronóstico y la respuesta a tratamientos de cáncer; esto se lleva a cabo mediante el desarrollo de métodos que permiten describir la información contenida en las imágenes médicas y capturan los aspectos más destacables durante cada uno de los procesos antes mencionados [72].

La radiómica se basa en tres etapas principales que corresponden a la adquisición de las imágenes, la representación de la información y el diseño de modelos de decisión. La etapa de adquisición de las imágenes tiene en cuenta diferentes métodos de generación de imágenes médicas, algunos de los más conocidos son la mamografía, la tomografía computarizada, el ultrasonido, la emisión de positrones y en el caso de este proyecto, la resonancia magnética. La etapa de representación de la información consiste en procesar las imágenes con el objetivo de obtener la mayor cantidad de información posible que pueda ser útil para la tarea en cuestión, algunas actividades derivadas de esta etapa son la segmentación de las ROI y la generación, selección y extracción de características [73]. La última etapa que corresponde al diseño de modelos de decisión, se basa principalmente en el uso de estrategias de aprendizaje de máquina que permiten identificar patrones entre la gran cantidad de datos generados en la etapa anterior, con el propósito final de poder apoyar la toma de decisiones del especialista en medicina.

### 2.2.2. TIPOS DE CARACTERÍSTICAS RADIÓMICAS

Teniendo en cuenta que las características radiómicas se derivan del análisis matemático aplicado desde la ingeniería, se han desarrollado diferentes herramientas que permiten calcular estas características cuando ya se tiene un conjunto de imágenes médicas. Una de las herramientas más utilizadas para generar características radiómicas es el paquete de librerías llamado *Pyradiomics* [74], el cual está desarrollado para ser utilizado desde el lenguaje de programación *Python* y es ampliamente soportado y verificado por la comunidad científica gracias a que cuenta con una amplia y detallada documentación (<https://pyradiomics.readthedocs.io>).

Existe una gran cantidad de características radiómicas que pueden ser calculadas a partir de una imagen médica, incluso si se trata de una representación en  $2D$  o en  $3D$  como es el caso de las MRI. Las características radiómicas son organizadas por tipos teniendo en cuenta el fundamento matemático utilizado para calcularlas. A continuación se describen los fundamentos teóricos de los diferentes tipos de características radiómicas que fueron tenidas en cuenta en este proyecto para representar la información de las MRI contenidas en la base de datos.

#### CARACTERÍSTICAS ESTADÍSTICAS DE PRIMER ORDEN (CEPO)

Este tipo de características describen la distribución de los niveles de intensidad de los vóxeles que conforman la ROI, lo cual es conocido como la generación de un histograma de la imagen para aplicar medidas estadísticas básicas.

### CARACTERÍSTICAS DE LA MATRIZ DE CO-OCURRENCIA (CMCO)

La matriz de co-ocurrencia describe la función de probabilidad conjunta de segundo orden de los niveles de intensidad de los vóxeles. Teniendo en cuenta que las características de primer orden no conservan información espacial de los vóxeles que componen la ROI, las CMCO representan una primera aproximación hacia la distribución espacial de los píxeles, revelando detalles importantes respecto a la textura de la imagen.

### CARACTERÍSTICAS DE LA MATRIZ DE NIVELES DE GRIS CONSECUTIVOS (CMNGC)

La matriz de niveles de gris consecutivos mide la longitud de las secuencias de vóxeles que tienen un mismo nivel de intensidad. Esta matriz puede ser calculada teniendo en cuenta diferentes grados de medición tanto para imágenes en  $2D$  como en  $3D$ , proporcionando también características detalladas sobre la textura de la imagen.

### CARACTERÍSTICAS DE LA MATRIZ DE TAMAÑO DE ZONAS DE GRIS (CMTZG)

La matriz de tamaño de zonas de gris mide el número de vóxeles con un mismo nivel de intensidad que se encuentran conectados, entendiendo que dos vóxeles se encuentran conectados cuando su distancia es igual a 1 de acuerdo a la norma infinito. A diferencia de la matriz de co-ocurrencia, solo se puede calcular una matriz de tamaño de zonas de gris, debido a que no depende de diferentes grados porque ya tiene en cuenta 8 posibilidades de conectividad para imágenes en  $2D$  y 26 para imágenes en  $3D$ .

### CARACTERÍSTICAS DE LA MATRIZ DE DEPENDENCIA DEL NIVEL DE GRIS (CMDNG)

La matriz de dependencia del nivel de gris cuantifica las dependencias del nivel de gris en una imagen, entendiendo que una dependencia de nivel de gris se define como el número de vóxeles conectados dentro de la distancia  $\delta$  dependientes del vóxel central. Un vóxel vecino con nivel de gris  $j$  se considera dependiente del vóxel central con nivel de gris  $i$  si  $|i - j| \leq \alpha$ . En una matriz de dependencia de nivel de gris, el elemento  $(i, j)$  describe el número de veces que un vóxel con nivel de gris  $i$  con  $j$  vóxeles dependientes en su vecindad aparece en la imagen.

### 2.2.3. MATRIZ DE CARACTERÍSTICAS RADIÓMICAS

Para el desarrollo de este proyecto se utilizó el paquete de librerías de *Pyradiomics* para generar las características radiómicas a partir de cada una de las secuencias de MRI disponibles en la base de datos. Teniendo en cuenta los tipos de características con los que dispone *Pyradiomics*, los cuales fueron descritos anteriormente, se extrajo un total de 86 características radiómicas, los nombres de estas características se presentan en la Tabla 2.1; aunque se recomienda remitirse a la documentación disponible sobre *Pyradiomics* (<https://pyradiomics.readthedocs.io>) para obtener información detallada sobre la formulación matemática que permite calcular estas características.

Finalmente, se pudo generar una matriz de características radiómicas de  $146 \times 86$  por cada una de las secuencias de MRI. Es importante aclarar que en este proyecto no se tuvieron en cuenta las características radiómicas de forma debido a que la base de datos utilizada no cuenta con una segmentación fina de las ROI. Además, buscando mantener una equivalencia entre las características perceptuales y radiómicas, se utilizó también la información bidimensional del corte marcado por el radiólogo en cada una de las secuencias para generar las características radiómicas, teniendo en cuenta también que al no tener una segmentación fina del tumor, se podía incrementar la cantidad de información irrelevante y ruido en las ROI si se tomaba un espacio tridimensional que encerrara el tumor.



Tipo de características	Características generadas		Cantidad total
CEPO	Percentil 10 Percentil 90 Energía Entropía Intercuartil Curtosis Máximo Desviación absoluta media Media	Mediana Mínimo Rango Desviación absoluta media robusta Raíz cuadrática media Asimetría Energía total Uniformidad Varianza	18
CMCO	Auto-correlación Media conjunta Prominencia de grupo Sombra de grupo Tendencia de grupo Contraste Correlación Media de diferencia Entropía de diferencia Varianza de diferencia Entropía conjunta	Energía conjunta Medida de correlación Informacional 1 Medida de correlación Informacional 2 Momento de diferencia inversa Momento de diferencia inversa normal Diferencia inversa Diferencia inversa normalizada Varianza inversa Probabilidad máxima Suma de entropía Suma de cuadrados	22
CMNGC	Énfasis de secuencia corta Énfasis de secuencia larga Heterogeneidad de niveles de gris Heterogeneidad normalizada Varianza de secuencia Entropía de secuencia Porcentaje de secuencia Varianza de niveles de gris	Heterogeneidad de longitud de secuencia Heterogeneidad de longitud normalizada Énfasis de bajo nivel de gris de secuencia Énfasis de alto nivel de gris de secuencia Énfasis de bajo nivel de gris de secuencia corta Énfasis de alto nivel de gris de secuencia corta Énfasis de bajo nivel de gris de secuencia larga Énfasis de alto nivel de gris de secuencia larga	16
CMTZG	Énfasis de área chica Énfasis de área grande Heterogeneidad de niveles de gris Heterogeneidad normalizada Varianza de zona Entropía de zona Porcentaje de zona Varianza de niveles de gris	Heterogeneidad de tamaño de zona Heterogeneidad de zona normalizada Énfasis de bajo nivel de gris de zona Énfasis de alto nivel de gris de zona Énfasis de bajo nivel de gris de área chica Énfasis de alto nivel de gris de área chica Énfasis de bajo nivel de gris de área grande Énfasis de alto nivel de gris de área grande	16
CMDNG	Énfasis de dependencia chica Énfasis de dependencia grande Entropía de dependencia Heterogeneidad de dependencia Énfasis de bajo nivel de gris Varianza de niveles de gris Varianza de dependencia	Énfasis de alto nivel de gris Heterogeneidad de niveles de gris Heterogeneidad de dependencia normal Énfasis de bajo nivel de gris de dependencia chica Énfasis de alto nivel de gris de dependencia chica Énfasis de bajo nivel de gris de dependencia grande Énfasis de alto nivel de gris de dependencia grande	14

Tabla 2.1: Características radiómicas generadas

# 3

## MÉTODO DE SELECCIÓN DE FUENTES RELEVANTES

### Índice

<b>2.1 Características perceptuales</b> . . . . .	<b>12</b>
2.1.1 Saliencia visual basada en grafos . . . . .	13
2.1.2 Matriz de características perceptuales . . . . .	13
<b>2.2 Características radiómicas</b> . . . . .	<b>15</b>
2.2.1 Definición de radiómica . . . . .	15
2.2.2 Tipos de características radiómicas . . . . .	15
2.2.3 Matriz de características radiómicas . . . . .	16

La metodología propuesta en este documento se desarrolla para poder identificar y seleccionar objetivamente las fuentes de información relevantes entre todas las que se tienen disponibles para resolver una tarea de clasificación binaria específica, como la que representa el diagnóstico de cáncer de mama utilizando MRI. El método propuesto se basa en el uso de técnicas que han sido bien estudiadas en el área de aprendizaje automático, pero que no han sido implementadas de forma conjunta en la literatura. Este capítulo contiene una explicación detallada de las técnicas involucradas en el método propuesto, comenzando por los fundamentos teóricos de cada técnica y terminando con el pseudocódigo de los algoritmos diseñados para aplicar dicho método.

### 3.1. MÁQUINAS DE VECTORES DE SOPORTE

Como se dijo anteriormente, el método propuesto solo es aplicable a problemas de clasificación binaria; esta restricción se debe al uso de una SVM como clasificador base en la tarea de aprendizaje automático. Las SVM son clasificadores discriminativos propuestos para problemas de clasificación binarios y se basan en la teoría de minimización de riesgos estructurales. La principal característica de una SVM es su categorización como clasificador de margen amplio [75]. Dada una muestra de  $N$  ejemplos de entrenamiento independientes e idénticamente distribuidos  $\{(x_i, y_i)\}_{i=1}^N$  donde  $x_i$  es un vector de entrada de dimensión  $D$  y  $y \in \{-1, +1\}$  es un vector de etiquetas de dimensión  $N$ , una SVM básicamente encuentra la línea discriminante con el margen  $M$  máximo en el espacio de características. La función de clasificación está dada por la Ecuación (3.1).

$$f(x) = \langle w, x \rangle + b \tag{3.1}$$

donde  $w$  es el vector de pesos que se comportan como coeficientes para cada ejemplo  $x_i$ ,  $b$  es el término independiente del hiperplano de separación (también conocido como “*bias term*” en inglés) y el operador  $\langle \cdot, \cdot \rangle$  representa el producto punto entre dos vectores.

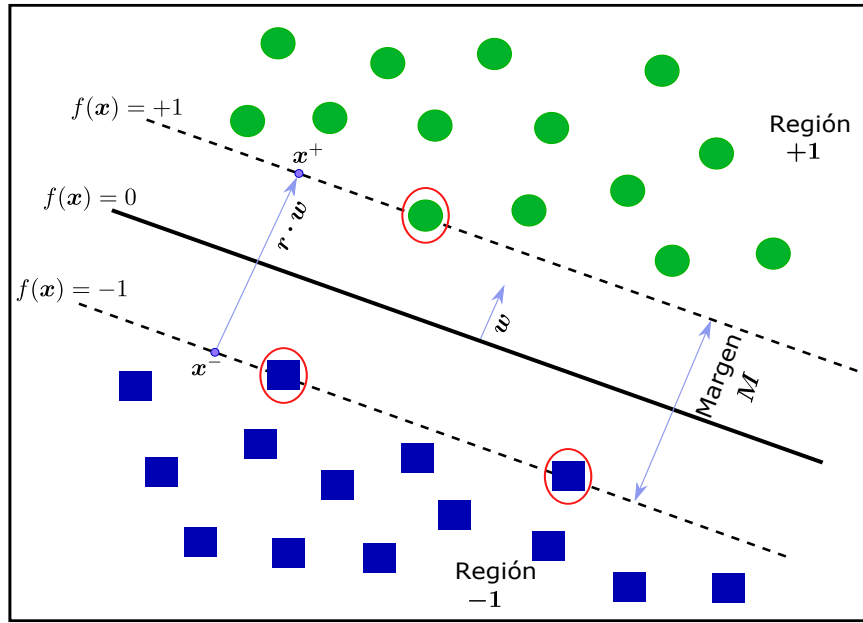


Figura 3.1: Representación de una base de datos separada por una SVM

Como se muestra en la Figura 3.1, Debe cumplirse que  $f(x) \geq 1$  en la región +1,  $f(x) \leq -1$  en la región -1 y  $f(x) = 0$  sobre la frontera. Los ejemplos encerrados en óvalos son llamados vectores de soporte ya que se encuentran sobre la frontera de decisión, además, todos los puntos sobre la frontera son designados como  $x^+$  o  $x^-$  dependiendo de la región a la que pertenezcan. La Figura 3.1 también muestra la representación del margen de separación entre fronteras  $M$ , el cual puede ser descrito por las Ecuaciones (3.2) y (3.3) al tener en cuenta un valor escalar  $r$  y el hecho de que el vector  $w$  es perpendicular a la frontera de decisión.

$$r = \frac{2}{\|w\|^2} \quad (3.2)$$

$$M = \|rw\| = \frac{2}{\|w\|^2} \|w\|$$

$$M = \frac{2}{\|w\|} \quad (3.3)$$

De esta forma es posible ver como la SVM busca maximizar el inverso de la longitud del vector  $w$ , lo cual es equivalente a minimizar  $\|w\|$  y puede ser expresado como se muestra en la Ecuación (3.4), esta ecuación es conocida como el *problema primal* de optimización de la SVM.

$$w^* = \min_w \frac{1}{2} \|w\|_2^2 + C \sum_{i=1}^N \xi_i \quad (3.4)$$

$$s.t. \quad y_i (\langle w, x \rangle + b) \geq 1 - \xi_i$$

donde  $C$  es un parámetro de regularización predefinido por el usuario y  $\xi$  es un vector de valores que permiten que la SVM tenga cierto margen de error haciendo lo que se conoce como suavizar el margen.

Es importante notar que la restricción asociada al problema de optimización resulta al determinar que cada ejemplo de entrenamiento quede en la región que le corresponde (-1 o +1) al entrenar la SVM. El método principal para resolver este problema de programación cuadrática con restricciones es mediante el uso de multiplicadores de Lagrange [76], con los cuales se llega a la definición del problema conocida como *función dual* que se muestra en la Ecuación (3.5).

$$\begin{aligned} \max_{0 \leq \alpha \leq C} \quad & \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle \\ \text{s. t.} \quad & \sum_{i=1}^N \alpha_i y_i = 0 \end{aligned} \quad (3.5)$$

donde  $\alpha$  es el vector de variables dual correspondientes a cada restricción de separación que surge al aplicar los multiplicadores de Lagrange.

La función de clasificación asociada a la *función dual* está dada por la Ecuación (3.6).

$$f(x) = \sum_{i=1}^N \alpha_i y_i \langle x_i, x \rangle + b \quad (3.6)$$

El algoritmo utilizado normalmente para resolver la función dual de la SVM es el algoritmo de optimización secuencial de mínimos (SMO por las siglas en inglés de *Sequential Minimal Optimization*) [77]. Existen diferentes librerías de programación diseñadas para aplicar con éxito el algoritmo SMO, por ejemplo, LIBSVM es una de las librerías más implementadas en la literatura porque posee un repositorio que es constantemente actualizado y se puede aplicar sobre diferentes lenguajes de programación [78], además, esta librería se encuentra bien respaldada por la comunidad científica.

### 3.2. KERNEL Y ESCALAMIENTO LOCAL

En la definición de la función dual de la SVM (Ecuación (3.5)) aparece un término de gran importancia:  $\langle x_i, x_j \rangle$ , este término es conocido como función kernel y es expresado de la forma  $K(x_i, x_j)$ , donde  $K: \mathbb{R}^D \times \mathbb{R}^D \rightarrow \mathbb{R}$ . La función kernel se define como una medida de similitud y puede tener diferentes planteamientos, algunos de los más conocidos se presentan en las Ecuaciones (3.7) a (3.9).

$$K_{\text{Lineal}}(x_i, x_j) = \langle x_i, x_j \rangle \quad (3.7)$$

$$K_{\text{Polinomial}}(x_i, x_j) = (\langle x_i, x_j \rangle + 1)^q, \quad q \in \mathbb{N} \quad (3.8)$$

$$K_{\text{Gaussiano}}(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|_2^2}{\sigma^2}\right), \quad \sigma > 0 \quad (3.9)$$

donde  $q$  es el grado del polinomio del kernel polinomial y  $\sigma$  es el ancho de banda del kernel Gaussiano.

Aunque se realizaron pruebas empíricas con los tres tipos de kernel antes mencionados, en este documento no se dan detalles de dichas pruebas, debido a que se desarrollaron de forma superficial y permitieron identificar rápidamente que el kernel Gaussiano es el tipo de kernel más apropiado para representar la información proveniente de las secuencias de MRI; por este motivo, la metodología propuesta se enfoca en evaluar detalladamente el uso del kernel Gaussiano.

Un desafío recurrente encontrado en la literatura es lograr una correcta sintonización del parámetro  $\sigma$  cuando se usa un kernel Gaussiano en la SVM [79–81], este problema se ha abordado mediante diferentes estrategias, principalmente con el uso de métodos de optimización metaheurística, pero este tipo de técnicas consumen una gran cantidad de tiempo y recursos computacionales, siendo aún un problema abierto en el área de aprendizaje automático.

En un trabajo desarrollado en el año 2005 se propuso un método eficiente e intuitivo para ajustar el parámetro  $\sigma$  en tareas de aprendizaje no supervisado sin utilizar métodos metaheurísticos [82]; además, esta estrategia permite definir un valor diferente  $\sigma_i$  relacionado con cada muestra en la base de datos en lugar de un  $\sigma$  global, esta estrategia se conoce como *escalamiento local*. Para calcular el parámetro de escalamiento local para cada  $x_i$  en la base de datos, es necesario analizar las características locales de su vecindario, porque el valor de  $\sigma_i$  está definido por la Ecuación (3.10).

$$\sigma_i = d(x_i, x_K) \quad (3.10)$$

donde  $x_K$  es el  $K$ -ésimo vecino de la muestra  $x_i$  y  $d(\cdot, \cdot)$  es una función de distancia, a menudo se utiliza la función de distancia euclidiana.

Es evidente que un nuevo parámetro debe ser sintonizado cuando se usa el escalamiento local, este parámetro es el correcto valor del vecino  $K$  para calcular la distancia que define el parámetro  $\sigma_i$ , pero cuando el método de escalamiento local fue propuesto [82], se sugirió usar  $K = 7$  porque mostró tener buenos resultados generales. Aún así, si  $K$  se mantiene como un parámetro libre, se reduce considerablemente el espacio de búsqueda para encontrar un buen valor para ser asignado, ya que estará restringido a ser un número entero y el vecindario de  $x_i$  se limitará para una pequeña cantidad de muestras porque el uso de un gran valor de  $K$  perdería la propiedad del escalamiento local; por lo tanto, el uso de la estrategia de escalamiento local reafirma la decisión de utilizar el tipo de kernel Gaussiano para poder delimitar la tarea de optimización de parámetros libres. Cuando se aplica esta estrategia, el kernel Gaussiano se calcula utilizando la Ecuación (3.11).

$$K(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|_2^2}{\sigma_i \sigma_j}\right) \quad (3.11)$$

### 3.3. APRENDIZAJE POR MÚLTIPLES KERNELS (MKL)

La función kernel es fundamental en el funcionamiento de la SVM, es en este punto donde empieza a ser útil el uso de MKL, ya que esta técnica propone básicamente que en lugar de utilizar un único kernel en el aprendizaje, se pueden utilizar diferentes tipos de combinaciones lineales o no lineales de múltiples kernels como se muestra en la Ecuación (3.12).

$$f(x_i, x_j) = \{K_m(x_i^m, x_j^m)\}_{m=1}^P \quad (3.12)$$

La función de combinación  $f_\eta : \mathbb{R}^P \rightarrow \mathbb{R}$  está compuesta por  $P$  representaciones de características de los ejemplos de entrenamiento. La gran ventaja que acarrea esta formulación es que dichas representaciones de características pueden provenir de diferentes fuentes de información, ya que la función kernel simplemente halla una medida de similitud entre los datos basada en la información de cada una de las fuentes, es así como podemos ver que pueden asignarse diferentes kernels a diferentes fuentes de información para integrarlas en una misma tarea de clasificación. Teniendo en cuenta que la función de combinación  $f_\eta$  puede ser definida a conveniencia, es posible implementar una combinación lineal en la cual se le asigne un peso a cada kernel, tal y como se expresa en la Ecuación (3.13).

$$f(x_i, x_j) = K_\eta(x_i, x_j) = \sum_{m=1}^P \eta_m K_m(x_i^m, x_j^m) \quad (3.13)$$

donde  $\eta_m$  denota los pesos de los kernels, lo cual equivale directamente a asignar un peso a cada fuente de información según el planteamiento antes descrito.

Con base en esta estructura, es posible definir una función objetivo en la cual se aplique una penalización a los pesos  $\eta_m$  de tal manera que el algoritmo pueda determinar la relevancia de cada fuente (e.g., secuencias de MRI) con base en los valores asignados a cada peso, obteniendo resultados fáciles de interpretar [36]. El tipo de penalización más utilizado para restringir los pesos asignados a los kernels es la penalización por norma- $\ell_1$  y norma- $\ell_2$  [39, 83], siendo la primera de estas normas asociada de mejor forma a la selección por relevancia, debido a que con esta se consigue que los pesos  $\eta_m$  asociados a los kernels de menor relevancia queden en cero después del entrenamiento, a lo cual se le conoce comúnmente como una penalización de tipo “sparse”.

### 3.4. SELECCIÓN DE FUENTES BASADA EN MKL

Los algoritmos de MKL se generan frecuentemente con el problema de optimización presentado en la Ecuación (3.14) [84].

$$\min_{\mathcal{H}_\eta} \frac{1}{2} \|\mathcal{H}_\eta\|^2 + \zeta \sum_{i=1}^N \mathcal{L}(y_i, f(x_i)) \quad (3.14)$$

donde  $\mathcal{L}(\cdot)$  es una función de costo regularizada por el parámetro  $\zeta$ , y  $\mathcal{H}_\eta$  es una función kernel parametrizada por  $\eta$  que corresponde a la Ecuación (3.15).

$$\mathcal{H}_\eta = K(\cdot, \cdot, \eta) = \sum_{j=1}^P \eta_j K_j(\cdot, \cdot) \quad (3.15)$$

Una función bien conocida en la literatura y especialmente práctica para este problema es la *Función de costo de bisagra* (*Hinge loss function*), que se puede usar como  $\mathcal{L}(\cdot)$  y se define en la Ecuación (3.16).

$$\mathcal{L}(y, f(x)) = (1 - yf(x))_+ \quad (3.16)$$

Cuando la función de costo es implementada, la representación dual del problema MKL puede ser calculado utilizando los multiplicadores de Lagrange para obtener la Ecuación (3.17)

$$\min_{\eta \in \Delta} \max_{\lambda \in \mathcal{Q}} \mathbf{1}^\top \lambda - \frac{1}{2} (\lambda \circ y)^\top \left( \sum_{j=1}^P \eta_j K_j \right) (\lambda \circ y) \quad (3.17)$$

donde  $\Delta$  es el dominio de  $\eta$  y  $\mathcal{Q}$  es el dominio de  $\lambda$ .  $\mathbf{1}$  es un vector con todos sus componentes iguales a uno y  $\circ$  define el producto Hadamard (punto a punto) entre dos vectores.

El dominio  $\mathcal{Q}$  se define generalmente como se muestra en la Ecuación (3.18).

$$\mathcal{Q} = \{ \lambda \in \mathbb{R}^N : \lambda^\top y = 0, 0 \leq \lambda \leq \zeta \} \quad (3.18)$$

Está claro cómo  $\mathcal{Q}$  es equivalente con las restricciones del coeficiente de Lagrange de la SVM. El análisis más interesante surge cuando se determina el dominio  $\Delta$ , porque cuando  $\eta \in \Delta$ , corresponde a una condición simplex, es decir, este dominio corresponde a la Ecuación (3.19).

$$\Delta = \left\{ \eta \in \mathbb{R}_+^P : \sum_{j=1}^P \eta_j = 1, \eta_j \geq 0 \right\} \quad (3.19)$$

Esta restricción  $\Delta$  se conoce como norma- $\ell_1$  de los pesos de los kernels, esta suma convexa puede ser resuelta utilizando el Operador de Selección y Contracción Mínima Absoluta (LASSO por las siglas en inglés de *Least Absolute Shrinkage and Selection Operator*) [85], ya que permite minimizar la suma de cuadrados residual restringida a la suma de los coeficientes haciéndola igual a una constante. La propiedad principal del método LASSO está en su naturaleza de restricción, porque este método produce coeficientes exactamente iguales a cero, lo que permite una eliminación simultánea de variables durante la tarea de entrenamiento, esto se conoce como *método sparse*. De esta manera, el algoritmo MKL penalizado por norma- $\ell_1$  y resuelto por el método LASSO, permite calcular la relevancia de los kernels basándose en los pesos  $\eta$ , que es equivalente a determinar la relevancia de las fuentes de información asociados a cada kernel.

En el año 2010 se desarrollaron dos trabajos diferentes en los que buscaban la forma correcta de entrenar los pesos  $\eta$  cuando se utiliza una penalización generalizada de las normas- $\ell_p$  con  $p \geq 1$  [84, 86], Aunque ambos trabajos realizaron el análisis de una forma diferente, el resultado fue el mismo, desarrollando una estrategia de optimización iterativa que resuelve una SVM y actualiza los pesos de los kernels con la Ecuación (3.20) hasta que se alcance una determinada medida de tolerancia.

$$\eta_m = \frac{\|w_m\|_2^{\frac{2}{p+1}}}{\left( \sum_{h=1}^P \|w_h\|_2^{\frac{2p}{p+1}} \right)^{\frac{1}{p}}} \quad (3.20)$$

Donde  $\|w_m\|_2^2 = \eta_m^2 \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K_m(x_i^m, x_j^m)$ , que se calcula a partir de la función dual de la SVM.

## 3.5. MÉTODO PROPUESTO

### 3.5.1. RESUMEN DEL MÉTODO

El método propuesto utiliza MKL para aprovechar la multimodalidad de los datos, es decir, las múltiples fuentes disponibles para la tarea de clasificación, que en este caso corresponden a las secuencias de MRI. Cada

kernel calculado a partir de las fuentes de datos se sintoniza utilizando la técnica de escalamiento local, luego, el kernel global obtenido después de aplicar la suma ponderada de todos los kernel se lleva a una SVM para realizar el proceso de entrenamiento y obtener finalmente los valores entrenados de los pesos  $\eta$  que definen la relevancia de cada una de las fuentes de información para la tarea de clasificación. La Figura 3.2 muestra una representación sencilla del método propuesto como un diagrama de flujo.

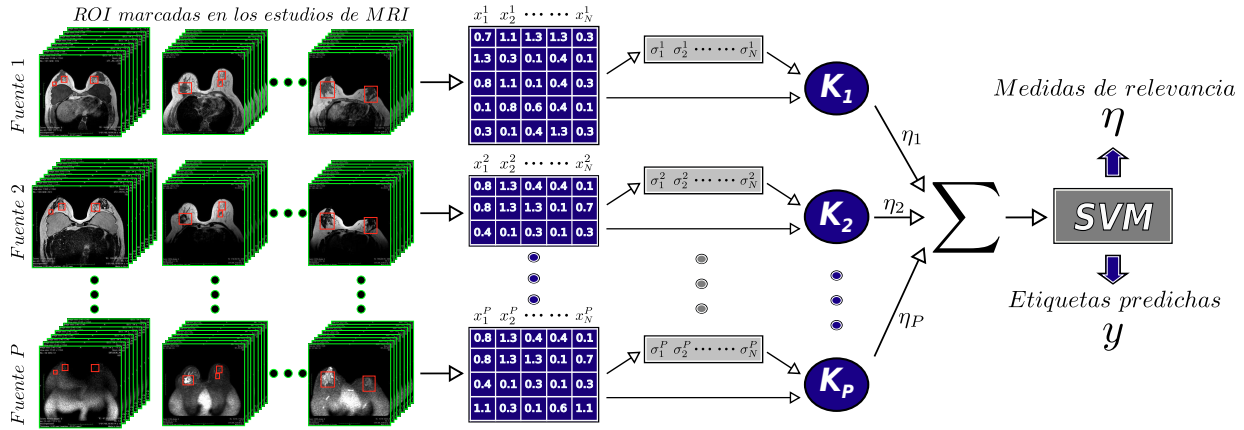


Figura 3.2: Diagrama de flujo del método propuesto

Aunque la técnica de escalamiento local fue propuesta inicialmente para ser aplicada en tareas de aprendizaje no supervisado [82], en este documento se proponen tres algoritmos diferentes para expandir su aplicación a tareas de aprendizaje supervisado, estos algoritmos se describen a continuación.

### 3.5.2. PREDICCIÓN SIN REDUCCIÓN DE DATOS

El principal problema encontrado para aplicar la estrategia de escalamiento local sobre una tarea de aprendizaje supervisado es el cálculo de los valores  $\sigma$  asociados a las muestras de prueba, la primera intuición que se tiene al querer abordar este problema es calcular la distancia entre las muestras de prueba y entrenamiento y asignar los valores de  $\sigma$  a las muestras de prueba utilizando la distancia del K-ésimo vecino, una clara desventaja de esta solución es que todas las muestras de entrenamiento serán necesarias en la etapa de predicción para calcular correctamente los valores de  $\sigma$  de las muestras de prueba. El Algoritmo 0 describe cómo se puede aplicar esta solución.

### 3.5.3. PREDICCIÓN UTILIZANDO SOLO LOS VECTORES DE SOPORTE

Aunque el primer algoritmo representa el enfoque más preciso para calcular los valores  $\sigma$  de las muestras de prueba, desperdicia una de las características más importantes de la SVM, que es la capacidad de requerir solo algunas muestras de entrenamiento conocidas como *vectores de soporte* en la etapa de predicción. El segundo algoritmo se diseñó con el objetivo de evitar este problema, utilizando solo los valores  $\sigma$  de los vectores de soporte. En este algoritmo se calcula la distancia entre las muestras de prueba y los vectores de soporte, asignando a cada muestra de prueba el  $\sigma$  correspondiente de su vector de soporte más cercano. El Algoritmo 1 describe cómo se puede aplicar esta estrategia.

### 3.5.4. PREDICCIÓN CON LOS VALORES $\sigma$ MEDIOS DE LOS VECTORES DE SOPORTE

El segundo algoritmo presenta la mejor solución respecto al procesamiento de datos y el consumo de memoria, pero parece desperdiciar la información que pertenece a las muestras de entrenamiento que no se seleccionaron como vectores de soporte, la cual podría ser importante para determinar los valores de  $\sigma$  de los datos de prueba. El tercer algoritmo fue diseñado para evitar este problema, en este algoritmo se agrega un nuevo paso en la etapa de entrenamiento, el cual consiste en identificar todas las muestras de entrenamiento que no se han seleccionado como vectores de soporte y asignar cada una de ellas al vector de soporte más cercano, luego se debe calcular para cada vector de soporte un nuevo valor  $\sigma$  compuesto por la media de los valores  $\sigma$  que pertenecen a las muestras de entrenamiento más cercanas a ellos. En la etapa de predicción se calcula la distancia

**Etapas de entrenamiento**

**Data:** datos de entrenamiento  $X_{tr}$ , Etiquetas de entrenamiento  $Y_{tr}$ , K-ésimo vecino  $K$ , Parámetro de regularización  $C$ , Parámetro para el algoritmo SMO  $\tau$

**Result:** modelo ajustado  $(\eta, \sigma, \alpha, b, X_{tr}, Y_{tr}, K)$ , medidas de relevancia  $\eta$   
calcular el número de fuentes de información ( $P$ );

**Para cada fuente de información en  $X_{tr}$ :**

**Para cada muestra de entrenamiento:**

- | calcular el valor de  $\sigma$  usando la Ecuación (3.10);
- | calcular el *kernel* para la fuente actual utilizando la Ecuación (3.11);

inicializar  $\eta = 1/P$ ;

calcular el *kernel global* usando la Ecuación (3.13);

calcular el *error*,  $\alpha$  y  $b$  resolviendo la SVM con el algoritmo SMO [78];

**Mientras *error* > *tolerancia*:**

**Para cada fuente de información en  $X_{tr}$ :**

- | actualizar los pesos  $\eta$  usando la Ecuación (3.20);
- | calcular el *error*,  $\alpha$  y  $b$  resolviendo la SVM con el algoritmo SMO [78];

asignar  $\eta$  como las medidas de relevancia de las fuentes de información;

**Etapas de predicción**

**Data:** datos de prueba  $X_{ts}$ , etiquetas de prueba  $Y_{ts}$ , modelo ajustado

**Result:** etiquetas predichas  $Y_p$

**Para cada fuente de información en  $X_{ts}$ :**

**Para cada muestra de prueba:**

- | calcular el valor de  $\sigma_{ts}$  usando la Ecuación (3.10) respecto a  $X_{tr}$ ;
- | calcular el *kernel* entre  $X_{ts}$  y  $X_{tr}$  usando la Ecuación (3.11);

calcular el *kernel global* usando la Ecuación (3.13);

calcular  $Y_p$  usando el *kernel global* y el modelo ajustado;

**Algoritmo 0:** Algoritmo sin reducción de datos

entre las muestras de prueba y los vectores de soporte, asignando a cada muestra de prueba el valor  $\sigma$  medio asociado a su vector de soporte más cercano, este proceso se describe en el Algoritmo 2.

## 3.6. OPTIMIZACIÓN Y EVALUACIÓN DEL MÉTODO PROPUESTO

### 3.6.1. MÉTODO DE OPTIMIZACIÓN

Teniendo en cuenta que el método requiere que se optimicen algunos parámetros libres (el K-ésimo vecino  $K$ , el parámetro de regularización  $C$  y el parámetro para el algoritmo SMO  $\tau$ ), se decidió utilizar el algoritmo de optimización por enjambre de partículas (PSO por las siglas en inglés de *Particle Swarm Optimization*) para encontrar el valor óptimo de estos parámetros. El PSO es uno de los algoritmos de optimización más utilizados y ampliamente conocidos en el estado del arte [87–89], este algoritmo utiliza métodos cooperativos y estocásticos para encontrar el punto de trabajo óptimo de la función a optimizar, que en este caso corresponde al desempeño del clasificador. Además, el algoritmo PSO generalmente es capaz de encontrar un óptimo global, siendo menos susceptible que otros algoritmos a caer en óptimos locales [90].

### 3.6.2. MEDIDAS DE DESEMPEÑO

Aunque el objetivo principal del método propuesto está enfocado a la selección objetiva de las secuencias de MRI relevantes, es necesario evaluar la eficacia del método a partir de su desempeño en cuanto a la tarea de clasificación, por lo cual se decidió implementar cuatro medidas de desempeño bien conocidas para tareas de clasificación, estas medidas corresponden a la exactitud, la media geométrica, la sensibilidad y la especificidad [91].

#### EXACTITUD

Esta medida busca determinar la proporción de ejemplos que fueron clasificados correctamente entre el número total de ejemplos positivos y negativos, es una de las medidas más conocidas y utilizadas en el estado del arte, aunque puede ser poco eficaz cuando se tiene una base de datos con alto desbalance entre clases. Esta



**Etapa de entrenamiento**

**Data:** datos de entrenamiento  $X_{tr}$ , etiquetas de entrenamiento  $Y_{tr}$ , K-ésimo vecino  $K$ , parámetro de regularización  $C$ , parámetro para el algoritmo SMO  $\tau$

**Result:** modelo ajustado  $(\eta, \sigma_{sp}, \alpha_{sp}, b, X_{sp}, Y_{sp})$ , medidas de relevancia  $\eta$   
calcular el número de fuentes de información ( $P$ );

**Para cada fuente de información en  $X_{tr}$ :**

**Para cada muestra de entrenamiento:**

- calcular el valor de  $\sigma$  usando la Ecuación (3.10);
- calcular el *kernel* para la fuente actual utilizando la Ecuación (3.11);

inicializar  $\eta = 1/P$ ;

calcular el *kernel global* usando la Ecuación (3.13);

calcular el *error*,  $\alpha$  y  $b$  resolviendo la SVM con el algoritmo SMO [78];

**Mientras  $error > tolerancia$ :**

**Para cada fuente de información en  $X_{tr}$ :**

- actualizar los pesos  $\eta$  usando la Ecuación (3.20);
- calcular el *error*,  $\alpha$  y  $b$  resolviendo la SVM con el algoritmo SMO [78];
- identificar los *vectores de soporte* a través de los valores  $\alpha$ ;

retornar solo los  $\sigma_{sp}$ ,  $\alpha_{sp}$ ,  $X_{sp}$  y  $Y_{sp}$  asociados a los *vectores de soporte*;  
asignar  $\eta$  como las medidas de relevancia de las fuentes de información;

**Etapa de predicción**

**Data:** datos de prueba  $X_{ts}$ , etiquetas de prueba  $Y_{ts}$ , modelo ajustado

**Result:** etiquetas predichas  $Y_p$

**Para cada fuente de información en  $X_{ts}$ :**

**Para cada muestra de prueba:**

- asignar  $\sigma_{ts}$  igual al valor  $\sigma_{sp}$  del *vector de soporte* más cercano;
- calcular el *kernel* entre  $X_{ts}$  y  $X_{sp}$  usando la Ecuación (3.11);

calcular el *kernel global* usando la Ecuación (3.13);

calcular  $Y_p$  usando el *kernel global* y el modelo ajustado;

**Algoritmo 1:** Algoritmo con reducción de datos usando solo los vectores de soporte

medida es calculada como se muestra en la Ecuación (3.21).

$$\text{Exactitud} = \frac{\text{Ejemplos predichos correctamente}}{\text{Número total de ejemplos}} \quad (3.21)$$

**SENSIBILIDAD**

Esta medida indica la capacidad que tiene el clasificador para predecir correctamente los ejemplos positivos, es decir, la capacidad para detectar sujetos enfermos en el caso de aplicaciones médicas. Esta medida es calculada como se muestra en la Ecuación (3.22).

$$\text{Sensibilidad} = \frac{\text{Verdaderos positivos}}{\text{Verdaderos positivos} + \text{Falsos negativos}} \quad (3.22)$$

**ESPECIFICIDAD**

Esta medida es opuesta a la sensibilidad, debido a que indica la capacidad que tiene el clasificador para predecir correctamente los ejemplos negativos, es decir, la capacidad para detectar sujetos sanos. Esta medida es calculada como se muestra en la Ecuación (3.23).

$$\text{Especificidad} = \frac{\text{Verdaderos negativos}}{\text{Verdaderos negativos} + \text{Falsos positivos}} \quad (3.23)$$

**MEDIA GEOMÉTRICA (MED-Geo)**

Esta medida relaciona la sensibilidad y la especificidad, buscando que se mantenga un equilibrio entre ambas, debido a que la única forma de obtener una media geométrica alta es que tanto la sensibilidad como la especificidad sean altas también; por esta razón es una medida sumamente útil cuando se tienen bases de datos desbalanceadas. Esta medida es calculada como se muestra en la Ecuación (3.24).

$$\text{Med-Geo} = \sqrt{\text{Sensibilidad} \cdot \text{Especificidad}} \quad (3.24)$$

**Etapa de entrenamiento**

**Data:** datos de entrenamiento  $X_{tr}$ , etiquetas de entrenamiento  $Y_{tr}$ , K-ésimo vecino  $K$ , parámetro de regularización  $C$ , parámetro para el algoritmo SMO  $\tau$

**Result:** modelo ajustado  $(\eta, \sigma_{mn}, \sigma_{sp}, \alpha_{sp}, b, X_{sp}, Y_{sp})$ , medidas de relevancia  $\eta$   
calcular el número de fuentes de información ( $P$ );

**Para cada fuente de información en  $X_{tr}$  :**

**Para cada muestra de entrenamiento :**

- | calcular el valor de  $\sigma$  usando la Ecuación (3.10);
- | calcular el *kernel* para la fuente actual utilizando la Ecuación (3.11);

inicializar  $\eta = 1/P$ ;

calcular el *kernel global* usando la Ecuación (3.13);

calcular el *error*,  $\alpha$  y  $b$  resolviendo la SVM con el algoritmo SMO [78];

**Mientras  $error > tolerancia$  :**

**Para cada fuente de información en  $X_{tr}$  :**

- | actualizar los pesos  $\eta$  usando la Ecuación (3.20);
- | calcular el *error*,  $\alpha$  y  $b$  resolviendo la SVM con el algoritmo SMO [78];
- | identificar los *vectores de soporte* a través de los valores  $\alpha$ ;

**Para cada muestra de entrenamiento :**

- | identificar los *vectores de soporte* más cercanos;

**Para cada vector de soporte :**

- | calcular los  $\sigma_{mn}$  medios usando los valores  $\sigma$  de las muestras de entrenamiento más cercanas;
- asignar  $\eta$  como las medidas de relevancia de las fuentes de información;

**Etapa de predicción**

**Data:** datos de prueba  $X_{ts}$ , etiquetas de prueba  $Y_{ts}$ , modelo ajustado

**Result:** etiquetas predichas  $Y_p$

**Para cada fuente de información en  $X_{ts}$  :**

**Para cada muestra de prueba :**

- | asignar  $\sigma_{ts}$  igual al valor  $\sigma_{mn}$  del *vector de soporte* más cercano;
- | calcular el *kernel* entre  $X_{ts}$  y  $X_{sp}$  usando la Ecuación (3.11);

calcular el *kernel global* usando la Ecuación (3.13);

calcular  $Y_p$  usando el *kernel global* y el modelo ajustado;

**Algoritmo 2:** Reducción de datos usando la media de los valores  $\sigma$  relacionados con las muestras más cercanas a los vectores de soporte

# 4

## MARCO EXPERIMENTAL Y RESULTADOS

### Índice

---

<b>3.1 Máquinas de Vectores de Soporte.</b>	<b>18</b>
<b>3.2 Kernel y escalamiento local</b>	<b>20</b>
<b>3.3 Aprendizaje por Múltiples Kernels (MKL)</b>	<b>21</b>
<b>3.4 Selección de fuentes basada en MKL</b>	<b>21</b>
<b>3.5 Método propuesto.</b>	<b>22</b>
3.5.1 Resumen del método	22
3.5.2 Predicción sin reducción de datos.	23
3.5.3 Predicción utilizando solo los vectores de soporte	23
3.5.4 Predicción con los valores $\sigma$ medios de los vectores de soporte	23
<b>3.6 Optimización y evaluación del método propuesto.</b>	<b>24</b>
3.6.1 Método de optimización	24
3.6.2 Medidas de desempeño	24

---

### 4.1. DESCRIPCIÓN DE LA BASE DE DATOS

La base de datos de MRI utilizada en este proyecto, fue desarrollada en un trabajo colaborativo con el Instituto de Alta Tecnología Médica (IATM) de Medellín, Colombia. En un estudio retrospectivo, se seleccionaron 84 pacientes mujeres con un promedio de edad de 50,36 años, en un rango de entre 23 y 87 años, quienes asistieron al IATM para la realización de un estudio de resonancia magnética de mama, quienes contaban con imágenes completas del protocolo definido por la institución para este tipo de estudio, descrito en el Capítulo 1. Se excluyeron pacientes con mastectomía radical sin reconstrucción, con estudios incompletos, o con imágenes categorizadas como de baja calidad. Con el propósito de garantizar la privacidad y confidencialidad de la información de los pacientes, se realizó la anonimización de cada estudio.

La adquisición de los estudios fue realizada con un scanner Phillips achieve de 1.5 T, y consiste en secuencias en el siguiente orden de realización: T1 VISTA sin saturación grasa plano axial, FOV 280x368x180, matriz 352x459, TR 7.5 TE 4.6, DWI axial factor B0 y B 800, FOV 300x400x198 matriz 120x157, TR 7112, TE 70, T1 SPIR volumétrica (THRIVE) dinámica post gadolinio 300x337x156, matriz 252x280, TR6.8 TE3.3, con secuencia inicial T1 supresión grasa y posterior a la inyección de gadoterato de meglumina 376.9 mg/mL (Dotarem® Guerbert) dosis 0.2 mL/kg (0,1 mmol/kg) de peso corporal. La inyección con inyector a tasa de 2.5 ml/seg con venoclis yelco n° 20, 5 dinámicos posteriores de 60.1 segundos de duración cada uno, T2 VISTA sin saturación grasa plano axial, FOV 280x370x180, matrix 351x463, TR 2000 TE 211, coronal STIR FOV 299x372x200 matriz 232x241 TR 10000 TE 80; cuando la paciente tiene implantes mamarios se sustituye la secuencia STIR por T2 VISTA supresión grasa coronal para lograr mejor contraste con la silicona FOV 300x364x200, matriz 252x 256, TR 7010, TE

80. El tecnólogo de imágenes en el equipo realiza la sustracción de las secuencias dinámicas, y genera el mapa de ADC. Cada estudio en la base de datos está representado por 16 fuentes de información, donde cada fuente corresponde a una de las secuencias de MRI descritas anteriormente

Cada uno de los estudios anonimizados, fue interpretado por un médico especialista con experiencia en el diagnóstico de imagen mamaria. En este proceso participaron dos médicos radiólogos con 6 y 10 años de experiencia, respectivamente. Los estudios fueron desplegados en el software Horos (<https://horosproject.org/>), el cual permite visualizar imágenes médicas almacenadas en formato DICOM. La densidad mamaria de cada estudio según las categorías definidas por ACR, con la siguiente distribución: 16 estudios ACR A, 34 ACR B, 21 ACR C, y 13 ACR D. El software Horos incluye una función de anotación de Regiones de Interés (ROIs), que permite marcar una región (un punto o un rectángulo), asignarle un nombre y realizar una descripción sobre su contenido. En el proceso de lectura los radiólogos realizaban marcación de los hallazgos, en la secuencia donde mejor los observaban y asignaban una categoría BIRADS a cada uno de ellos. Posteriormente, un médico residente de radiología propagaba la marcación en las demás secuencias, haciendo uso de la función de triangulación del mismo software. Estas marcaciones fueron exportadas para implementar un algoritmo que permitiera extraer las regiones marcadas en imágenes independientes, para ser procesadas por los algoritmos de extracción de características radiómicas. La distribución específica de ROIs por estudio fue de 46, 24, 7, 6 y 1 estudios, conteniendo 1, 2, 3, 4 y 7 ROIs, respectivamente.

En total se extrajeron 146 regiones, las cuáles fueron clasificadas utilizando el sistema BI-RADS descrito anteriormente. Teniendo en cuenta que el sistema BI-RADS cuenta con 5 categorías distintas, se decidió realizar una nueva asignación de etiquetas a las ROI buscando que la tarea de clasificación fuera de tipo binario, de tal forma que una clase esté asociada a la detección de tumores positivos o probablemente malignos, es decir, la detección de lesiones que requieran seguimiento o biopsia (clase 1), mientras que la otra se asocie a la detección de hallazgos benignos (clase -1). Las ROIs se dividieron de la siguiente manera: la clase -1 para todas las ROI con BI-RADS menor a 3 y la clase 1 para las ROI con BI-RADS mayor o igual a 3; un total de 61 ROI se asociaron con la clase -1 (tumores benignos) y 85 con la clase 1 (tumores malignos) aplicando dicha configuración. La Tabla 4.1 muestra detalladamente la conformación de la base de datos respecto al número de ROIs por cada categoría BI-RADS y por cada clase definitiva. Es importante mencionar que, aun cuando los estudios se obtuvieron de forma retrospectiva, no se cuenta con la confirmación por biopsia para estos hallazgos, lo que constituye una limitación para este trabajo. Sin embargo, y dado que el objetivo es identificar la relevancia de la información para la interpretación diagnóstica, se toma como “*ground truth*” la categorización asignada por el especialista.

Categoría BI-RADS	Número de ROI	Clase definitiva
0	0	
1	1	-1
2	60	-1
3	26	1
4	15	1
5	44	1
6	0	1

Tabla 4.1: Distribución de estudios de la base de datos empleada para evaluación

#### 4.1.1. CONFIGURACIÓN DE LA BASE DE DATOS

La base de datos completa está conformada por 16 secuencias de MRI diferentes, pero para el desarrollo de las pruebas se omitieron las 6 secuencias de imágenes de DCE y la secuencia STIR. Las secuencias DCE se omitieron debido a que las secuencias de sustracción (Fases) son construidas a partir de las DCE, por lo cual pueden entenderse como fuentes redundantes para la tarea de clasificación; mientras que la secuencia STIR se omitió porque esta está tomada desde un plano coronal que difiere de todas las demás secuencias en la base de datos, de tal manera que al retirarse de las pruebas se logra que el análisis de relevancia se haga de forma equivalente para todas las fuentes de información.

Teniendo en cuenta que se propusieron dos tipos de características para ser generadas a partir de cada una de las secuencias de MRI en la base de datos, se decidió mantener cada tipo de características como fuentes de

información independientes para poder aprovechar de una mejor manera la información que estas contienen; es decir, que por cada secuencia de MRI se obtuvieron dos fuentes de información, una por cada tipo de característica. A partir de las características perceptuales se obtuvieron 9 fuentes de información ( $S_{T1}$ ,  $S_{T2}$ ,  $S_{ADC}$ ,  $S_{DWI}$ ,  $S_{FAS1}$ ,  $S_{FAS2}$ ,  $S_{FAS3}$ ,  $S_{FAS4}$  y  $S_{FAS5}$ ) y de igual forma se obtuvieron otras 9 fuentes a partir de las características radiómicas ( $R_{T1}$ ,  $R_{T2}$ ,  $R_{ADC}$ ,  $R_{DWI}$ ,  $R_{FAS1}$ ,  $R_{FAS2}$ ,  $R_{FAS3}$ ,  $R_{FAS4}$  y  $R_{FAS5}$ ), obteniendo así un total de 18 fuentes de información independientes.

## 4.2. DESCRIPCIÓN DEL MARCO EXPERIMENTAL

### 4.2.1. VARIACIONES DEL MÉTODO Y NÚMERO DE PRUEBAS

Las pruebas desarrolladas para evaluar el desempeño del método propuesto tienen en cuenta las diferentes variaciones que este puede presentar, las cuales están asociadas principalmente a los tres algoritmos propuestos para adaptar el método de escalamiento local a la tarea de aprendizaje supervisado, los tipos de características generadas a partir de las secuencias de MRI y los dos principales tipos de penalización que pueden ser aplicados a los pesos  $\eta$  asociados a los kernels, los cuales corresponden a la norma- $\ell_1$  y la norma- $\ell_2$ .

Inicialmente, las pruebas se dividieron en dos grandes grupos determinados por los tipos de penalización por norma- $\ell_1$  y norma- $\ell_2$ ; luego, cada uno de estos grupos se dividió en tres subgrupos determinados por los tipos de características disponibles, en donde el primer subgrupo incluye únicamente características perceptuales, el segundo incluye solo características radiómicas y el tercero incluye ambos tipos de características; finalmente, sobre cada subgrupo se aplicó una prueba con cada uno de los tres algoritmos propuestos para adaptar el método de escalamiento local, dando como resultado un total de 18 pruebas diferentes tenidas en cuenta para evaluar el desempeño del método propuesto, esta distribución de las pruebas se presenta con detalle en la Tabla 4.2.

Tipo de penalización	Tipos de características	Tipo de algoritmo	Número de prueba
Penalización por norma- $\ell_1$	Características Perceptuales	Algoritmo 0	Prueba 1
		Algoritmo 1	Prueba 2
		Algoritmo 2	Prueba 3
	Características Radiómicas	Algoritmo 0	Prueba 4
		Algoritmo 1	Prueba 5
		Algoritmo 2	Prueba 6
	Características Perceptuales y Radiómicas	Algoritmo 0	Prueba 7
		Algoritmo 1	Prueba 8
		Algoritmo 2	Prueba 9
Penalización por norma- $\ell_2$	Características Perceptuales	Algoritmo 0	Prueba 10
		Algoritmo 1	Prueba 11
		Algoritmo 2	Prueba 12
	Características Radiómicas	Algoritmo 0	Prueba 13
		Algoritmo 1	Prueba 14
		Algoritmo 2	Prueba 15
	Características Perceptuales y Radiómicas	Algoritmo 0	Prueba 16
		Algoritmo 1	Prueba 17
		Algoritmo 2	Prueba 18

Tabla 4.2: Distribución general de las pruebas realizadas

### 4.2.2. CONFIGURACIÓN DE LAS PRUEBAS

En cada una de las pruebas se dividió la base de datos, manteniendo un 20% (equivalente a 29 ejemplos) de los ejemplos por fuera de la etapa de entrenamiento y validación con el propósito de utilizarlos como ejemplos de prueba, mientras que el 80% restante (equivalente a 117 ejemplos) se utilizó para entrenar el método. Para el proceso de entrenamiento y validación se implementó el método de validación cruzada conocido como *K-Fold*

con un total de 10 divisiones, esto se hizo con el propósito de poder entrenar el método de tal forma que pudiera generalizar correctamente las predicciones sobre los datos de prueba.

Para todas las pruebas se utilizó el método de PSO con 40 partículas de búsqueda y coeficientes de aprendizaje individual y social iguales a 1,1931, el valor de inercia se estableció de forma dinámica entre 0,1 y 1,1 con un valor máximo de 60 iteraciones; la función de costo utilizada para el proceso de optimización fue la media geométrica, debido a que fue la medida de desempeño que demostró tener una mejor estabilidad y rápida convergencia respecto a los valores óptimos encontrados. Una vez terminado el proceso de optimización mediante el uso del PSO y la validación cruzada en cada prueba, se utilizaron los parámetros óptimos encontrados (ver Apéndice A) para entrenar un modelo global utilizando el total de los datos de entrenamiento (80% de toda la base de datos), este modelo se utilizó finalmente para realizar la predicción sobre los datos de prueba que se habían separado inicialmente, obteniendo así los valores entrenados de los pesos  $\eta$  que determinan la relevancia de las fuentes de información y las predicciones sobre los datos de prueba que permitirían medir el desempeño del método como clasificador.

### 4.3. RESULTADOS CON PENALIZACIÓN POR NORMA- $\ell_1$

#### 4.3.1. USANDO CARACTERÍSTICAS PERCEPTUALES

En este conjunto de pruebas se utilizaron únicamente las 9 fuentes de información generadas a partir de las características perceptuales. Una vez obtenidos los parámetros óptimos que determinan el desempeño del método, se procedió a analizar las medidas de relevancia asociadas a cada una de las fuentes de información, las cuales corresponden a los valores  $\eta$  entrenados; teniendo en cuenta estos valores, se realizó un proceso de eliminación de fuentes, en donde se retiraba de la base de datos cada una de las fuentes desde la menos relevante hasta la más relevante; es decir, eliminando primero las fuentes que obtuvieron un valor  $\eta$  más pequeño, generando así las curvas de desempeño mostradas en la Figura 4.1. Las fuentes seleccionadas como las más relevantes son aquellas con las cuales se alcanza el máximo desempeño manteniendo el mínimo número de fuentes posibles. Para el caso de la Figura 4.1, el Algoritmo 0 encuentra su máximo desempeño con 2 fuentes, Algoritmo 1 con 4 fuentes y el Algoritmo 2 con 3 fuentes.

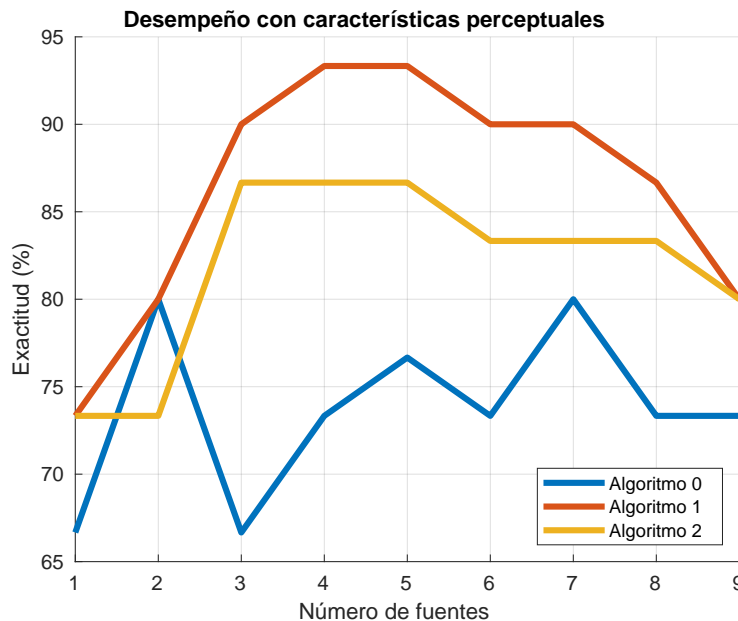


Figura 4.1: Curvas de desempeño para el conjunto de datos de prueba cuando se realiza eliminación de fuentes con características perceptuales y norma- $\ell_1$

En la Tabla 4.3 se muestran con detalle los resultados obtenidos en este conjunto de pruebas. Para cada uno de los algoritmos se muestran las diferentes medidas de desempeño obtenidas cuando se implementaron todas

las fuentes de información y cuando solo se implementaron las fuentes más relevantes, también se especifica cuáles son las fuentes seleccionadas como relevantes y además se indica la tasa de reducción obtenida al quitar las fuentes menos relevantes, la tasa de reducción se calcula utilizando la Ecuación (4.1).

$$\text{Tasa de reducción} = 1 - \frac{\text{Número de fuentes seleccionadas}}{\text{Número total de fuentes}} \quad (4.1)$$

Adicionalmente, se puede ver en la Tabla 4.3 el tiempo implementado por el algoritmo para completar la etapa de entrenamiento del modelo global y la etapa de predicción sobre los datos de entrenamiento y de prueba; cabe mencionar que todas las pruebas fueron desarrolladas en un equipo de cómputo que cuenta con un procesador Intel de 16 núcleos a 3,00GHz y 16GB de memoria RAM. La última columna de la Tabla 4.3 muestra el número de ejemplos de entrenamiento que se necesitaron para construir el modelo global, esta columna es importante para resaltar el uso de los Algoritmos 1 y 2, con los cuales precisamente se busca aprovechar una de las ventajas de la SVM que es la necesidad de utilizar solamente los vectores de soporte del modelo.

	Medidas de desempeño	Con todas las fuentes (%)	Con reducción de fuentes (%)	Tasa de reducción	Secuencias relevantes	Tiempo empleado (s)	Ejemplos conservados
<b>Algoritmo 0</b>	Exactitud	73.33	80.00	0.78	S_ADC, S_FAS4	0.013	117 (100.0%)
	Med-Geo	65.91	79.59				
	Sensibilidad	94.12	82.35				
	Especificidad	46.15	76.92				
<b>Algoritmo 1</b>	Exactitud	80.00	93.33	0.56	S_FAS2, S_FAS4 S_FAS3, S_FAS5	0.021	97 (82.9%)
	Med-Geo	76.10	93.21				
	Sensibilidad	94.12	94.12				
	Especificidad	61.54	92.31				
<b>Algoritmo 2</b>	Exactitud	80.00	86.67	0.67	S_FAS2, S_FAS5 S_FAS3	0.020	97 (82.9%)
	Med-Geo	76.10	85.09				
	Sensibilidad	94.12	94.12				
	Especificidad	61.54	76.92				

Tabla 4.3: Resultados de selección de secuencias de MRI con características perceptuales y norma- $\ell_1$ , para el conjunto de datos de prueba

Es evidente tanto en la Figura 4.1 como en la Tabla 4.3 que al poder identificar cuáles de las fuentes de información son relevantes para la tarea de clasificación, el desempeño puede mejorar considerablemente; también se evidencia un mejor desempeño del Algoritmo 1 respecto a los otros dos, obteniendo los mejores resultados en todas las medidas de desempeño cuando se ha aplicado la reducción de fuentes. Es de notar la coincidencia de las fuentes de información seleccionadas como relevantes por los Algoritmos 1 y 2, los cuales muestran tener un desempeño muy similar y al mismo tiempo superior al del Algoritmo 0.

#### 4.3.2. USANDO CARACTERÍSTICAS RADIÓMICAS

En este conjunto de pruebas se utilizaron únicamente las fuentes de información obtenidas a partir de la generación de las características radiómicas sobre las secuencias de MRI. La Figura 4.2 muestra las curvas de desempeño obtenidas mediante el proceso de eliminación de fuentes, evidenciando nuevamente que el Algoritmo 1 tiene el mejor comportamiento, resaltando que alcanzó el máximo desempeño con el menor número de fuentes; además, el Algoritmo 0 se muestra nuevamente en desventaja ante los otros dos en términos del máximo desempeño alcanzado.

La Tabla 4.4 muestra con detalle los resultados obtenidos para este conjunto de pruebas, en este caso se hace notar que la tasa de reducción es considerablemente menor para todos los algoritmos respecto al conjunto de pruebas anterior. Nuevamente se puede identificar un buen nivel de estabilidad entre algoritmos respecto a las fuentes que fueron seleccionadas como relevantes, además de notarse nuevamente la ventaja de los Algoritmos 1 y 2, especialmente por hacer una reducción considerable en cuanto a la cantidad de datos que requieren para la construcción del modelo definitivo.

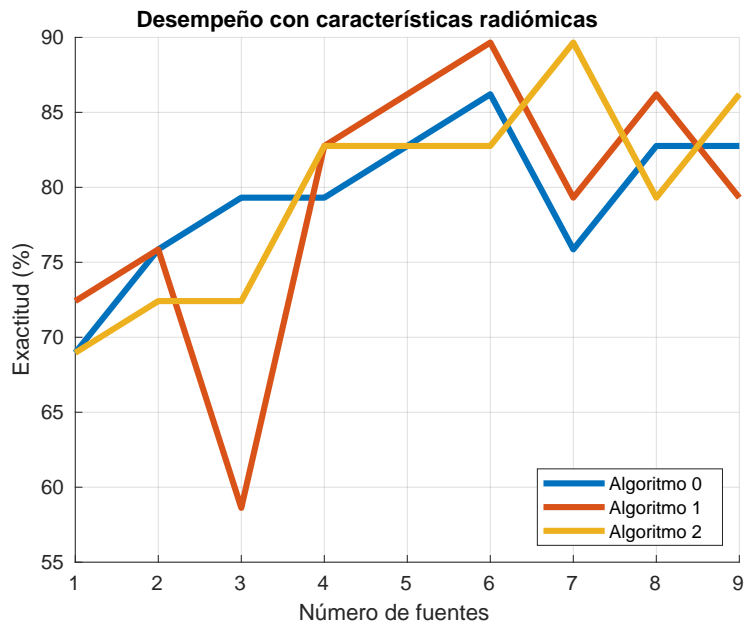


Figura 4.2: Curvas de desempeño para el conjunto de datos de prueba cuando se realiza eliminación de fuentes con características radiómicas y norma- $\ell_1$

	Medidas de desempeño	Con todas las fuentes (%)	Con reducción de fuentes (%)	Tasa de reducción	Secuencias relevantes	Tiempo empleado (s)	Ejemplos conservados
<b>Algoritmo 0</b>	Exactitud	82.76	86.21	0.33	R_T2, R_T1 R_DWI, R_FAS3 R_FAS5, R_FAS4	0.073	117 (100.0%)
	Med-Geo	84.02	87.45				
	Sensibilidad	70.59	76.47				
	Especificidad	100.00	100.00				
<b>Algoritmo 1</b>	Exactitud	79.31	89.66	0.33	R_T1, R_DWI R_FAS3, R_T2 R_FAS4, R_FAS5	0.063	81 (69.2%)
	Med-Geo	80.44	90.75				
	Sensibilidad	64.71	82.35				
	Especificidad	100.00	100.00				
<b>Algoritmo 2</b>	Exactitud	86.21	89.66	0.22	R_T1, R_DWI R_FAS3, R_T2 R_FAS5, R_FAS4 R_FAS2	0.113	92 (78.6%)
	Med-Geo	87.45	90.75				
	Sensibilidad	76.47	82.35				
	Especificidad	100.00	100.00				

Tabla 4.4: Resultados de selección de secuencias de MRI con características radiómicas y norma- $\ell_1$

### 4.3.3. COMBINANDO CARACTERÍSTICAS PERCEPTUALES Y RADIÓMICAS

Este conjunto de pruebas utiliza un número mayor de fuentes de información debido a que combina las fuentes obtenidas a partir de la generación de características perceptuales y radiómicas. La Figura 4.3 muestra las curvas de desempeño obtenidas mediante el proceso de eliminación de fuentes de información, en este caso se puede observar un comportamiento similar entre los Algoritmos 1 y 2, que nuevamente presentan un mejor desempeño que el Algoritmo 0 pero un poco menos notorio en este caso; se puede observar una reducción considerable de las fuentes de información con los tres algoritmos, lo cual indica un buen desempeño general del método propuesto.

La Tabla 4.5 muestra en este caso, resultados muy similares entre los tres algoritmos, evidenciando una tasa de reducción de más de la mitad de las fuentes de información con todos los algoritmos y una coincidencia casi perfecta de las fuentes de información seleccionadas como relevantes. En este punto se hace evidente la



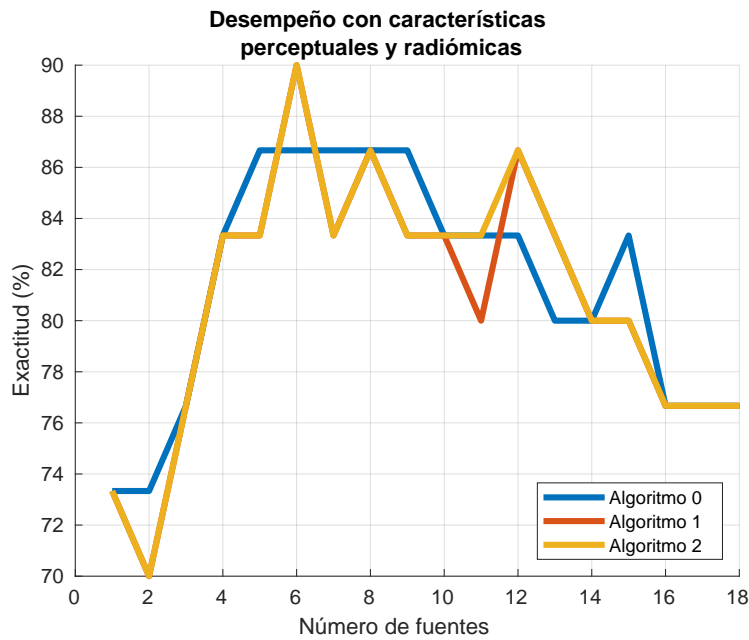


Figura 4.3: Curvas de desempeño para el conjunto de datos de prueba cuando se realiza eliminación de fuentes con características perceptuales y radiómicas con norma- $\ell_1$

	Medidas de desempeño	Con todas las fuentes (%)	Con reducción de fuentes (%)	Tasa de reducción	Secuencias relevantes	Tiempo empleado (s)	Ejemplos conservados
<b>Algoritmo 0</b>	Exactitud	76.67	86.67	0.72	S_FAS2, S_FAS5 S_FAS4, S_FAS3 R_FAS3	0.033	117 (100.0%)
	Med-Geo	75.51	85.09				
	Sensibilidad	82.35	94.12				
	Especificidad	69.23	76.92				
<b>Algoritmo 1</b>	Exactitud	76.67	90.00	0.67	S_FAS2, S_FAS5 S_FAS4, S_FAS3 R_FAS3, R_FAS5	0.039	97 (82.9%)
	Med-Geo	75.51	89.24				
	Sensibilidad	82.35	94.12				
	Especificidad	69.23	84.62				
<b>Algoritmo 2</b>	Exactitud	76.67	90.00	0.67	S_FAS2, S_FAS5 S_FAS4, S_FAS3 R_FAS3, R_FAS5	0.051	97 (82.9%)
	Med-Geo	75.51	89.24				
	Sensibilidad	82.35	94.12				
	Especificidad	69.23	84.62				

Tabla 4.5: Resultados de selección de secuencias de MRI con características perceptuales y radiómicas con norma- $\ell_1$

identificación en la mayoría de las pruebas de una alta relevancia para las secuencias de MRI correspondientes a las Fases, tanto con las características perceptuales como con las radiómicas.

## 4.4. RESULTADOS CON PENALIZACIÓN POR NORMA- $\ell_2$

### 4.4.1. USANDO CARACTERÍSTICAS PERCEPTUALES

En este conjunto de pruebas se implementa la norma- $\ell_2$  en lugar de la norma- $\ell_1$ , aunque la penalización por norma- $\ell_2$  puede asignar valores mayores que 1 a los pesos  $\eta$ , se conserva el mismo concepto mediante el cual se determina una alta relevancia para una fuente de información si el peso que tiene asociado tiene un valor elevado, respecto a los demás pesos asociados a las otras fuentes de información.

La Figura 4.4 muestra las curvas de desempeño obtenidas al eliminar secuencialmente las fuentes de información desde la menos relevante hasta la más relevante, en este caso se observa un comportamiento muy

diferente del Algoritmo 2 respecto a los demás algoritmos, especialmente teniendo en cuenta que el Algoritmo 2 conserva la mayoría de fuentes como relevantes en esta prueba, mientras que los otros dos algoritmos reducen al máximo el número de fuentes al presentar su mejor resultado con una sola fuente de información.

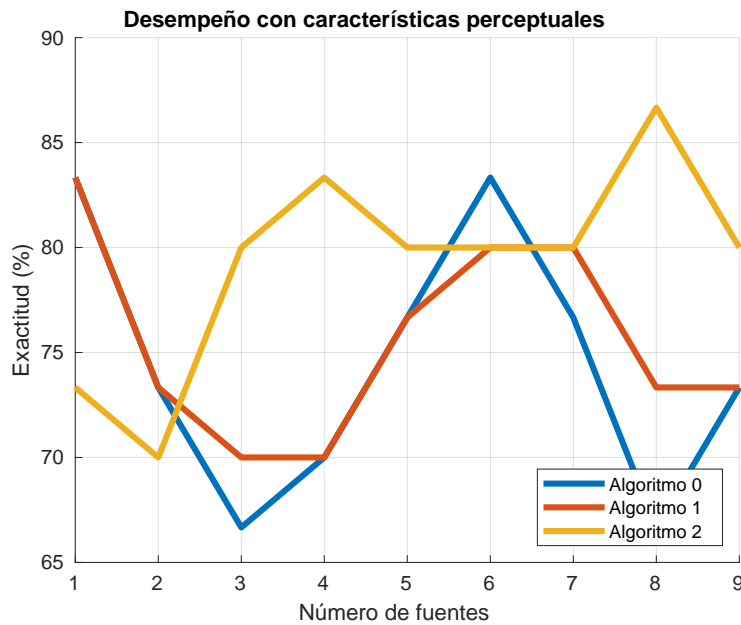


Figura 4.4: Curvas de desempeño para el conjunto de datos de prueba cuando se realiza eliminación de fuentes con características perceptuales y norma- $\ell_2$

	Medidas de desempeño	Con todas las fuentes (%)	Con reducción de fuentes (%)	Tasa de reducción	Secuencias relevantes	Tiempo empleado (s)	Ejemplos conservados
<b>Algoritmo 0</b>	Exactitud	73.33	83.33	0.89	S_FAS4	0.007	117 (100.0%)
	Med-Geo	65.91	83.48				
	Sensibilidad	94.12	82.35				
	Especificidad	46.15	84.62				
<b>Algoritmo 1</b>	Exactitud	73.33	83.33	0.89	S_FAS4	0.006	66 (56.4%)
	Med-Geo	68.93	83.48				
	Sensibilidad	88.24	82.35				
	Especificidad	53.85	84.62				
<b>Algoritmo 2</b>	Exactitud	80.00	86.67	0.11	S_FAS2, S_FAS4 S_FAS5, S_FAS3 S_DWI, S_T2 S_FAS1, S_T1	0.050	97 (82.9%)
	Med-Geo	78.16	85.09				
	Sensibilidad	88.24	94.12				
	Especificidad	69.23	76.92				

Tabla 4.6: Resultados de selección de secuencias de MRI con características perceptuales y norma- $\ell_2$

La Tabla 4.6 muestra que el método se desempeñó mejor sobre el conjunto de datos de esta prueba cuando se utilizó la norma- $\ell_1$  (ver Tabla 4.3), debido a que no solo se logró alcanzar un mejor desempeño por parte del Algoritmo 1, sino que también se logró una mejor estabilidad respecto a las fuentes que fueron seleccionadas como relevantes entre los tres algoritmos; pero aún así, se puede resaltar que en dos de las pruebas se logró la mayor tasa de reducción y coincidieron perfectamente con la fuente que fue seleccionada como la más relevante.

#### 4.4.2. USANDO CARACTERÍSTICAS RADIÓMICAS

Para este conjunto de pruebas se utilizaron las 9 fuentes de información generadas a partir de las características radiómicas y se aplicó la norma- $\ell_2$  como estrategia de penalización del método. La Figura 4.5 muestra un comportamiento similar entre los tres algoritmos al realizar la eliminación secuencial de fuentes de información, obteniendo incluso el mismo número de fuentes de información en los puntos máximos de desempeño de los tres algoritmos.

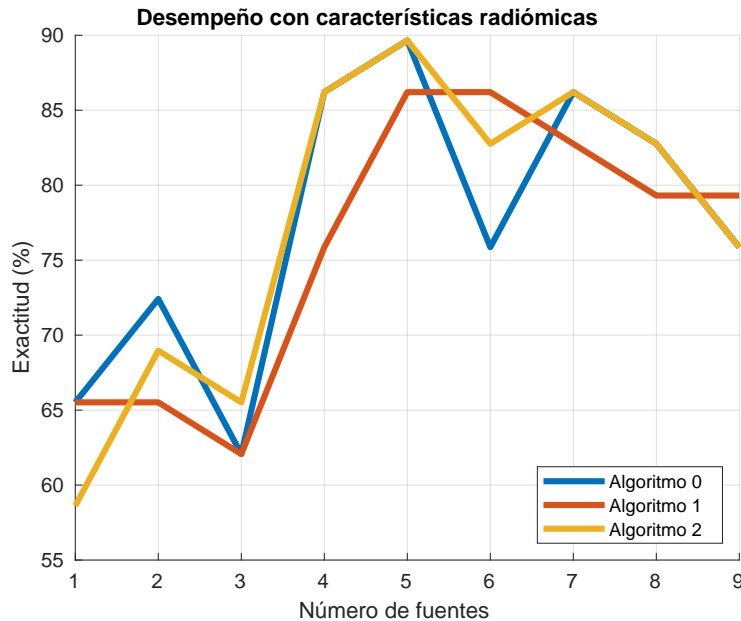


Figura 4.5: Curvas de desempeño para el conjunto de datos de prueba cuando se realiza eliminación de fuentes con características radiómicas y norma- $\ell_2$

	Medidas de desempeño	Con todas las fuentes (%)	Con reducción de fuentes (%)	Tasa de reducción	Secuencias relevantes	Tiempo empleado (s)	Ejemplos conservados
<b>Algoritmo 0</b>	Exactitud	75.86	89.66	0.44	R_T1, R_FAS3 R_DWI, R_FAS4 R_T2	0.060	117 (100.0%)
	Med-Geo	76.70	90.75				
	Sensibilidad	58.82	82.35				
	Especificidad	100.00	100.00				
<b>Algoritmo 1</b>	Exactitud	79.31	86.21	0.44	R_T1, R_FAS3 R_FAS4, R_DWI R_T2	0.050	82 (70.1%)
	Med-Geo	80.44	87.45				
	Sensibilidad	70.59	76.47				
	Especificidad	91.67	100.00				
<b>Algoritmo 2</b>	Exactitud	75.86	89.66	0.44	R_T1, R_FAS3 R_DWI, R_FAS4 R_T2	0.070	85 (72.6%)
	Med-Geo	76.70	90.75				
	Sensibilidad	58.82	82.35				
	Especificidad	100.00	100.00				

Tabla 4.7: Resultados de selección de secuencias de MRI con características radiómicas y norma- $\ell_2$

La Tabla 4.7 presenta detalles de los resultados obtenidos en este grupo de pruebas, estos resultados muestran una significativa mejora en el desempeño del método al conservar únicamente las fuentes relevantes, además de presentar una completa estabilidad entre las fuentes de información seleccionadas como relevantes, manteniendo únicamente 44% de las fuentes y reduciendo hasta el 70% el número de ejemplos de entrena-

miento requeridos por el modelo global. Al igual que para las características perceptuales, el mejor desempeño se obtiene para las fuentes seleccionadas cuando se emplea la norma- $\ell_1$ .

#### 4.4.3. COMBINANDO CARACTERÍSTICAS PERCEPTUALES Y RADIÓMICAS

Finalmente, en este conjunto de pruebas se utilizan 18 fuentes de información correspondientes a las generadas por el uso de características perceptuales y radiómicas sobre las secuencias de MRI. La Figura 4.6 muestra que contrario a los resultados obtenidos al utilizar la norma- $\ell_1$ , cuando se utiliza la norma- $\ell_2$  el Algoritmo 1 no es el que mejor se desempeña en las pruebas, ya que en este caso los Algoritmos 0 y 2 demuestran alcanzar el mejor desempeño respecto a las medidas de evaluación.

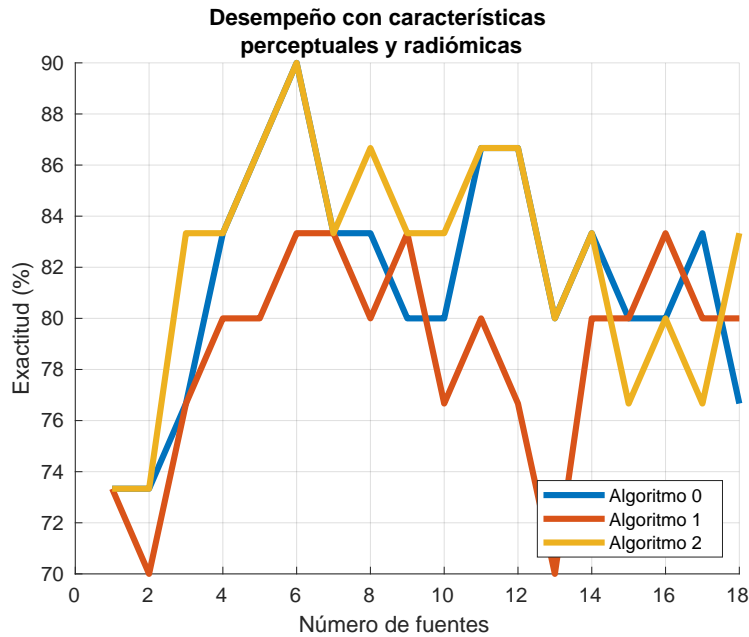


Figura 4.6: Curvas de desempeño para el conjunto de datos de prueba cuando se realiza eliminación de fuentes con características perceptuales y radiómicas con norma- $\ell_2$

	Medidas de desempeño	Con todas las fuentes (%)	Con reducción de fuentes (%)	Tasa de reducción	Secuencias relevantes	Tiempo empleado (s)	Ejemplos conservados
<b>Algoritmo 0</b>	Exactitud	76.67	90.00	0.67	S_FAS2, S_FAS5 S_FAS4, S_FAS3 R_FAS3, R_FAS5	0.043	117 (100.0%)
	Med-Geo	75.51	89.24				
	Sensibilidad	82.35	94.12				
	Especificidad	69.23	84.62				
<b>Algoritmo 1</b>	Exactitud	80.00	83.33	0.67	R_DWI, R_T1 S_FAS4, R_FAS2 S_FAS2, S_FAS5	0.038	70 (59.8%)
	Med-Geo	78.16	82.39				
	Sensibilidad	88.24	88.24				
	Especificidad	69.23	76.92				
<b>Algoritmo 2</b>	Exactitud	83.33	90.00	0.67	S_FAS2, S_FAS5 S_FAS4, S_FAS3 R_FAS3, R_FAS5	0.051	97 (82.9%)
	Med-Geo	82.39	89.24				
	Sensibilidad	88.24	94.12				
	Especificidad	76.92	84.62				

Tabla 4.8: Resultados de selección de secuencias de MRI con características perceptuales y radiómicas con norma- $\ell_2$

La Tabla 4.8 muestra nuevamente que los resultados son estables entre los algoritmos, respecto a la tasa de reducción y las fuentes que fueron seleccionadas como relevantes, mostrando nuevamente que las secuencias

correspondientes a las Fases son las más relevantes para la tarea de diagnóstico de cáncer de mama, siendo este un comportamiento que se repitió a lo largo de todas las pruebas.

## 4.5. COMPARACIÓN CON MÉTODOS DE SELECCIÓN DE CARACTERÍSTICAS

Teniendo en cuenta que hasta ahora no se han desarrollado métodos para seleccionar de forma objetiva las fuentes de información relevantes para una tarea de clasificación, y más aún, aplicados a la identificación de las secuencias de MRI relevantes para el diagnóstico de cáncer de mama, se decidió utilizar dos métodos clásicos de selección de características para ser comparados con el método propuesto respecto a la tarea de selección de fuentes. Esto, teniendo en cuenta que la primera aproximación que se tiene hacia la selección automática de fuentes con aprendizaje de máquina es la selección de características desarrollada desde esta misma área de investigación [92]. Se puede asociar la selección de fuentes de información con la selección de características, si se tiene en cuenta que es posible seleccionar características individuales esperando que se eliminen todas las características de una fuente específica y de esta manera se pueda eliminar la fuente completa.

### 4.5.1. MÉTODO DE SELECCIÓN FISHER SCORE

Fisher Score es un método de selección de características supervisado, este método busca conservar las características que presentan una distribución más uniforme en una clase y más dispersa en otras [93]. Este método de selección de características es representado mediante la Ecuación (4.2).

$$FS(f_i) = \frac{\sum_{j=1}^l n_j (\mu_{i,j} - \mu_i)^2}{\sum_{j=1}^l n_j s_{i,j}^2} \quad (4.2)$$

Donde  $l$  es el número de clases,  $\mu_i$  es la media de la característica  $f_i$ ,  $n_j$  es el número de muestras en la  $j$ -ésima clase,  $\mu_{i,j}$  y  $s_{i,j}$  son la media y la varianza de  $f_i$  en la  $j$ -ésima clase, respectivamente [94].

Al aplicar el método de selección de características Fisher Score sobre la base de datos, se obtiene un orden por relevancia de todas las características concatenadas de todas las fuentes de información (864 en total). Para poder evaluar el desempeño de esta selección se implementó una SVM clásica con un kernel Gaussiano en el cual se concatenaron las características de todas las fuentes; al igual que en las pruebas anteriores, se utilizó el algoritmo PSO para optimizar los parámetros de este clasificador (el ancho de banda del kernel  $\sigma$ , el parámetro de regularización  $C$  y el parámetro para el algoritmo SMO  $\tau$ ) y se dividió la base de datos en prueba y entrenamiento manteniendo las mismas proporciones de las pruebas anteriores. Teniendo los parámetros óptimos del clasificador, se procedió a realizar una prueba de eliminación de características desde la menos relevante hasta la más relevante, los resultados de esta prueba son presentados en la Figura 4.7.

La Tabla 4.9 muestra el orden en el cual las fuentes de información se van eliminando a medida que se eliminan todas las características que las componen, es decir, que solamente cuando quedan 286 características se elimina por completo la primera fuente (S\_T1). Adicionalmente la Tabla 4.9 muestra las diferentes medidas de desempeño obtenidas con todas las características y en el punto de mayor desempeño luego de empezar la reducción de características, esta información es sumamente importante debido a que aunque el desempeño en cuanto la tarea de clasificación es superior al que se alcanzó con el método propuesto, se puede observar que al momento de alcanzar dicho desempeño no se había eliminado ninguna fuente de información, es decir, que aunque se haya eliminado un gran número de características, realmente no se ha cumplido el objetivo principal que corresponde a la selección de fuentes relevantes.

Al observar detalladamente la Figura 4.7 se puede identificar que el punto de mayor desempeño representa apenas un pico en la gráfica, el cual difiere considerablemente del resto de puntos. Si se tienen en cuenta los siguientes puntos de desempeño más altos que requieren un menor número de características, se puede observar que con 110 características se obtuvo una exactitud de 93,10% y con 60 características se obtuvo 89,66% de exactitud, lo cual corresponde al uso de 12 y 10 fuentes de información, respectivamente; esto representa la necesidad de utilizar más fuentes de información de las que se necesitaron en cualquiera de los resultados obtenidos con el método propuesto.

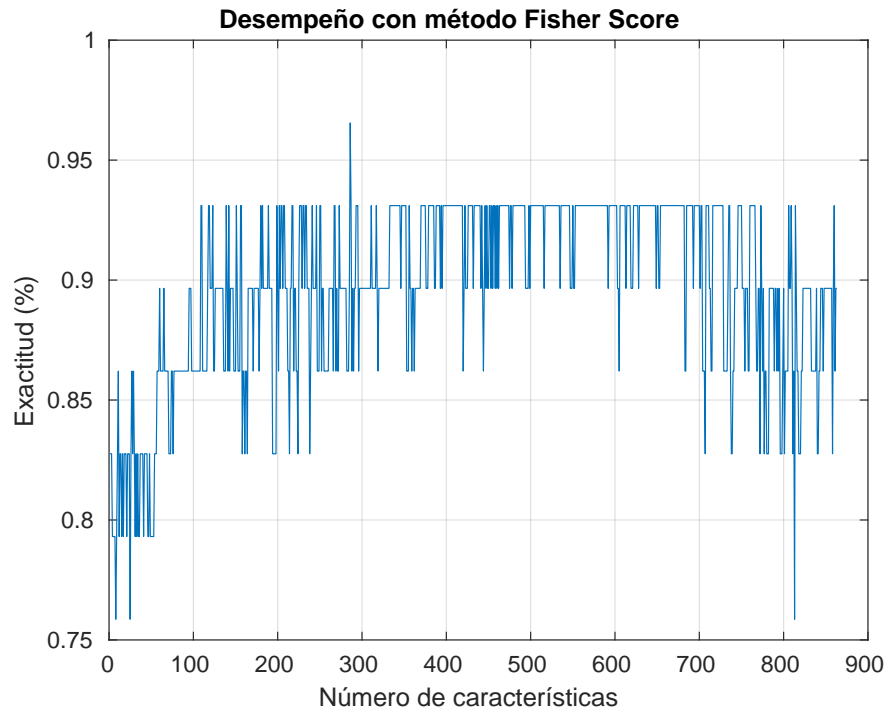


Figura 4.7: Curvas de desempeño para el conjunto de datos de prueba cuando se realiza eliminación de características con método Fisher Score

Secuencia a eliminar	S_T1	S_FAS1	S_ADC	R_T1	S_T2	S_DWI	R_T2	R_ADC	R_FAS1	R_DWI	S_FAS4	S_FAS2	S_FAS3	S_FAS5	R_FAS2	R_FAS3	R_FAS4	R_FAS5
Características retenidas	286	256	221	153	145	137	136	98	91	55	53	39	32	24	4	3	2	1

	Exactitud (%)	Med-Geo (%)	Sensibilidad (%)	Especificidad (%)	Número de características	Número de fuentes
Desempeño sin reducción	89.66	90.75	82.35	100.00	864	18
Desempeño con reducción	96.55	97.01	94.12	100.00	286	18 Reducción = 0.0

Tabla 4.9: Resultados de selección de secuencias de MRI con Fisher Score

#### 4.5.2. MÉTODO DE SELECCIÓN GINI INDEX

El método de selección de características supervisado Gini Index es utilizado mayormente en clasificadores basados en árboles de decisión. Este método de selección busca conservar las características que mejor distinguen muestras entre diferentes tipos de clases, teniendo en cuenta que entre menor sea el valor Gini Index de una característica, mejor será esta para la clase [94]. El Gini Index puede ser calculado utilizando la Ecuación (4.3).

$$GI(f_i) = 1 - \sum_{j=1}^l [p(j|f_i)]^2 \quad (4.3)$$

Donde la función  $p$  determina una medida de probabilidad.

La prueba con el método de selección Gini Index se realizó de la misma forma que con el método Fisher Score. Inicialmente se obtuvo el orden de relevancia generado por el método de selección y luego se pasó a la etapa de optimización de los parámetros, seguido de la etapa de eliminación secuencial de características obteniendo con estos resultados la Figura 4.8, en la cual se observa un comportamiento muy similar al presentado por el anterior método de selección de características.

En la Tabla 4.10 se puede observar unos resultados muy similares a los obtenidos en la Tabla 4.9, nuevamen-

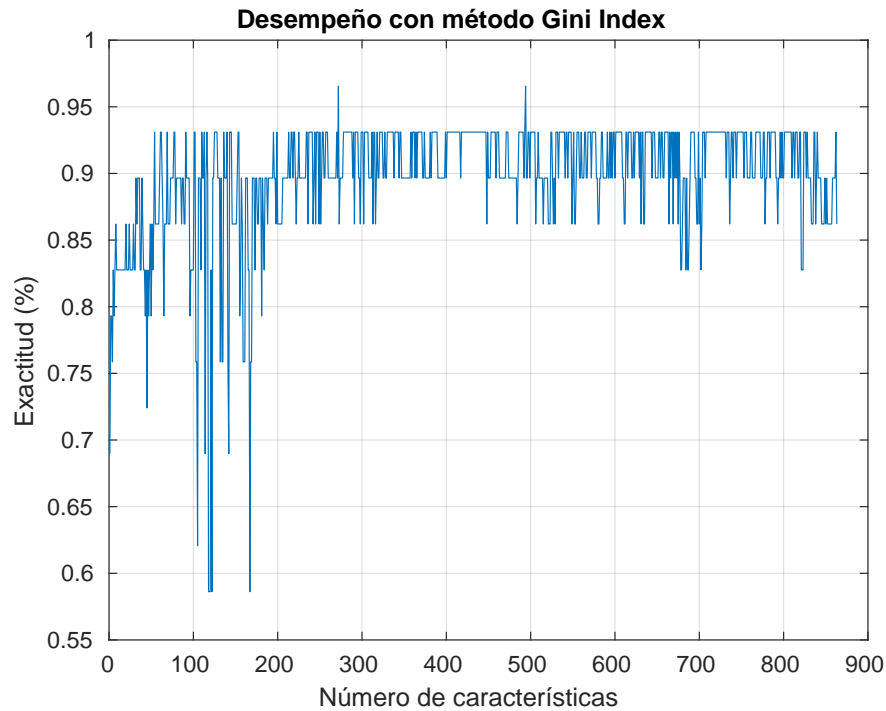


Figura 4.8: Curvas de desempeño para el conjunto de datos de prueba cuando se realiza eliminación de características con método Gini Index

te el desempeño alcanza un punto muy alto pero sin lograr eliminar ninguna de las fuentes de información por completo; esto representa un resultado interesante si se tiene en cuenta que el método propuesto puede ser utilizado inicialmente para seleccionar de forma objetiva las fuentes más relevante, para posteriormente utilizar un algoritmo de selección de características dentro de cada una de las fuentes seleccionadas como relevantes y así incrementar el desempeño en cuanto a la tarea de clasificación. Adicionalmente, se puede observar que el orden de relevancia asociado a las fuentes a partir de los dos métodos de selección de características no son estables, debido a que se hace mucho más complejo poder coincidir en la selección cuando se trata de características individuales.

<b>Secuencia a eliminar</b>	S_T1	S_T2	S_ADC	R_T2	S_FAS4	R_ADC	S_FAS1	R_T1	R_FAS1	S_FAS3	S_DWI	R_DWI	S_FAS5	S_FAS2	R_FAS2	R_FAS4	R_FAS5	R_FAS3
<b>Características retenidas</b>	250	249	218	157	89	86	72	63	38	36	34	31	21	11	8	3	2	1

	Exactitud (%)	Med-Geo (%)	Sensibilidad (%)	Especificidad (%)	Número de características	Número de fuentes
<b>Desempeño sin reducción</b>	89.66	90.75	82.35	100.00	864	18
<b>Desempeño con reducción</b>	96.55	97.01	94.12	100.00	272	18 Reducción = 0.0

Tabla 4.10: Resultados de selección de secuencias de MRI con Gini Index

Al igual que se observó en los resultados obtenidos con el método Fisher Score, la Figura 4.8 muestra que el punto de mayor desempeño representa un pico en la gráfica, de tal forma que si se tienen en cuenta los siguientes puntos de desempeño más altos que requieren un menor número de características, se puede observar que con 54 características se obtuvo una exactitud de 93,10% y con 32 características se obtuvo 89,66% de exactitud, lo cual corresponde al uso de 10 y 7 fuentes de información, respectivamente; esto representa la necesidad de utilizar más fuentes de información de las que se necesitaron en los resultados obtenidos con el método propuesto cuando se utilizaron las características perceptuales y radiómicas al mismo tiempo, es decir, que el

método propuesto demuestra obtener los mejores resultados si se tiene en cuenta la necesidad de mantener un alto desempeño mientras se utiliza el menor número de fuentes posible.

Un detalle importante para resaltar respecto a los resultados obtenidos con los métodos de selección de características es que ambos coincidieron en mantener una mayor relevancia para las fuentes asociadas a las características radiómicas, lo cual es totalmente opuesto a los resultados obtenidos con el método propuesto, el cual asignó en la mayoría de las pruebas, una mayor relevancia a las características perceptuales. Este resultado se debe principalmente a que las fuentes de información conformadas por características radiómicas tienen muchas más características (86 características por cada fuente) que las conformadas por características perceptuales (10 características por cada fuente), de tal manera que las fuentes con un mayor número de características tendrán ventaja respecto a las que tienen menos características. Este problema no afecta de ninguna manera al método propuesto, porque al calcular un kernel por cada una de las fuentes de información antes de determinar la relevancia que estas tienen, el número de características que conforman la fuente no afectará las medidas de relevancia, porque el kernel tendrá la misma dimensión para todas las fuentes ( $N \times N$ ), independientemente del número de características que lo generen.

Finalmente, es posible validar el hecho de que las secuencias correspondientes a las Sustracciones o Fases son relevantes para el proceso diagnóstico, teniendo en cuenta que estas secuencias fueron destacadas en todas las pruebas realizadas con el método propuesto, y que además, fueron las características provenientes de estas secuencias las que representaron una mayor relevancia al momento de utilizar los métodos de selección de características. Estos resultados son concordantes con algunos trabajos encontrados en el estado del arte, en donde se destaca el uso de las secuencias dinámicas contrastadas (DCE) debido a que proporcionan información altamente discriminativa respecto a la clasificación de tumores posiblemente malignos o benignos [95, 96], en estos trabajos se destaca el gran interés que ha representado el uso reciente de las secuencias DCE descritas a partir de diferentes tipos de características, que permiten realizar diagnóstico automático con base en métodos de clasificación, con resultados acordes a los obtenidos en este documento (96% en sensibilidad y 85% en especificidad para [95] y 85% en sensibilidad; 89% en especificidad para [96]).



# 5

## CONCLUSIONES Y TRABAJO FUTURO

En este trabajo se propuso un método automático, que hace uso de diferentes tipos de algoritmos de aprendizaje de máquina, para seleccionar de forma objetiva las secuencias de MRI más relevantes para el diagnóstico de cáncer de mama. Inicialmente, se generan descriptores a partir de cada una de las secuencias de MRI utilizando características perceptuales y radiómicas, las primeras intentan capturar la información visual que es usada por los radiólogos para realizar la interpretación diagnóstica, mientras que las segundas intentan capturar la información cuantitativa que puede ser extraída de las lesiones, y que puede ser empleada por un modelo computacional para apoyar el proceso diagnóstico. Luego, se calcula un kernel Gaussiano por cada una las fuentes de información obtenidas; estos kernels corresponden a medidas de similitud entre los datos que representan adecuadamente la información proporcionada por las secuencias de MRI, teniendo en cuenta que los parámetros de los kernels son sintonizados utilizando tres algoritmos diferentes basados en la técnica de escalamiento local, la cual es adaptada para ser utilizada en la tarea de aprendizaje supervisado implementada sobre una SVM. El primer algoritmo utiliza todos los datos de entrenamiento para calcular el parámetro de ancho de banda del kernel ( $\sigma$ ) correspondiente a cada muestra que se desea predecir, el segundo algoritmo utiliza únicamente los vectores de soporte de la SVM para calcular dicho parámetro, reduciendo así la cantidad de información requerida por el modelo entrenado, mientras que el tercer algoritmo recolecta la información del vecindario de los vectores de soporte para calcular el valor  $\sigma$  de las muestras que se van a predecir, intentado capturar la información de los datos de entrenamiento que no son seleccionados como vectores de soporte al mismo tiempo que reduce la información requerida por el modelo entrenado.

El método propuesto utiliza MKL sobre una SVM para sacar el mayor provecho de la información proveniente de las múltiples fuentes generadas a partir de las secuencias de MRI; la ecuación fundamental de MKL implementada corresponde a una suma ponderada de kernels en donde se aplica un peso  $\eta$  a cada kernel, con el propósito de que dichos pesos puedan ser utilizados como medidas de relevancia asociadas a las fuentes de información al aplicar una penalización sobre ellos, esta penalización puede estar basada en la norma- $\ell_1$  o en la norma- $\ell_2$ . La salida definitiva entregada por el método corresponde a los pesos  $\eta$  entrenados, los cuales funcionan como medidas de relevancia de las fuentes de información; adicionalmente, el método entrega las etiquetas predichas para los ejemplos procesados, las cuales permiten establecer el desempeño de la clasificación al hacer uso de dichas fuentes de información, manteniendo la función fundamental de la SVM como clasificador binario.

El método propuesto fue evaluado sobre una base de datos de estudios de resonancia magnética de mama, de los cuales fueron extraídas regiones de interés, que fueron interpretadas como hallazgos positivos o negativos, por médicos especialistas en radiología con más de quince años de experiencia en la lectura de este tipo de estudios y familiarizados con el protocolo de imágenes empleado. Los resultados obtenidos demuestran que el método propuesto es estable en cuanto a las fuentes que son seleccionadas como fuentes relevantes, incluso cuando se observan los resultados de forma transversal entre los tres tipos de algoritmos propuestos. Aunque los tres algoritmos presentaron desempeños considerablemente similares y cumplieron con la función esperada de selección de fuentes, se presentó un desempeño relativamente superior del Algoritmo 1 cuando se utiliza la penalización por norma- $\ell_1$  y de igual forma un comportamiento sobresaliente del Algoritmo 2 cuando se utiliza

la penalización por norma- $\ell_2$ , estos dos algoritmos presentan la ventaja especial de que permiten reducir el número de ejemplos que se requieren para construir el modelo entrenado, debido a que solamente requieren la información de los ejemplos seleccionados por la SVM como vectores de soporte. Los resultados obtenidos muestran también que las fuentes de información correspondientes a las secuencias de Fases 2, 3, 4 y 5 son contundentemente relevantes para la tarea de clasificación, especialmente cuando son descritas a través de características perceptuales; se estima que una de las causas de este alto nivel de relevancia en estas secuencias se debe a que son tomadas al aplicar un agente de contraste en el tejido, que permite obtener mejores detalles de la información morfológica y cinética de las ROI que están siendo analizadas.

El desempeño del método propuesto fue comparado con aproximaciones que hacían uso de métodos clásicos de selección de características como son Gini Index y Fisher Score, teniendo en cuenta que la selección de características es la analogía más próxima a la selección de fuentes. En este caso, una fuente era eliminada, cuando el proceso de selección eliminaba todas las características asociadas a esta; de esta manera, las fuentes que fueron eliminadas más tardíamente serían las más relevantes para la tarea. Aunque los resultados obtenidos, en términos de las secuencias "*seleccionadas*", presentó coincidencias con los resultados obtenidos con el método propuesto, esta estrategia no resulta eficaz, dado que estos métodos tienden a conservar hasta el final todas las fuentes de información, aunque sea con pocas características en cada una de estas. Adicionalmente, los métodos de selección de características presentan un sesgo respecto a la asignación de la relevancia de las fuentes de información cuando existen algunas fuentes que poseen más características que otras, asociando incorrectamente una alta relevancia a las fuentes con mayor número de características; cabe resaltar que este problema no se presenta con el método propuesto debido a que cuando se calcula el kernel que determina la similitud entre los datos, este no tiene en cuenta si la fuente tiene muchas o pocas características, siendo por lo tanto, completamente invariable ante el número de características que conforman cada fuente de información.

El método propuesto puede ser implementado en otro tipo de tareas de selección de fuentes de información. Particularmente, el Laboratorio de Máquinas Inteligentes y Reconocimiento de Patrones, está explorando su aplicación en la selección de electrodos relevantes para aplicaciones de interfaces cerebro computador basadas en electroencefalografía (EEG) y para la selección de bandas de frecuencia relevantes en imágenes multiespectrales e hiperespectrales, con resultados prometedores.

Como trabajo futuro se propone investigar estrategias mediante las cuales se pueda ampliar el uso del método propuesto hacia tareas de clasificación con múltiples clases, debido a que el principal limitante del método está en que por ahora solo puede ser utilizado en tareas de clasificación binaria; también se propone la realización de un análisis de relevancia entre secuencias de MRI que no requieran de la administración de un agente de contraste, con el propósito de desarrollar protocolos abreviados mínimamente invasivos o no invasivos que mejoren la experiencia de los pacientes durante el proceso de diagnóstico. Finalmente, y dado que este estudio estuvo limitado a estudios adquiridos retrospectivamente, en los que no se contó con confirmación por biopsia de la malignidad o benignidad de las lesiones, en un futuro cercano se estará evaluando el método propuesto con estudios confirmados por biopsia y se tendrá en cuenta una evaluación detallada de los resultados a nivel de paciente con el objetivo de realizar un análisis más profundo en cuanto a sus resultados clínicos.

# A

## APÉNDICE A: PARÁMETROS ÓPTIMOS

En la Tabla A.1 se presentan los diferentes valores de los parámetros utilizados para cada prueba en la que se evaluó el método propuesto, después de implementar el algoritmo de optimización PSO para definirlos.

Tipo de penalización	Tipos de características	Tipo de algoritmo	Parámetros óptimos		
			$K$	$C$	$\tau$
Penalización por norma-L1	Características Perceptuales	Algoritmo 0	17	8.730536	0.323197
		Algoritmo 1	3	0.061205	0.875159
		Algoritmo 2	7	0.005000	0.379566
	Características Radiómicas	Algoritmo 0	7	9.717202	1.000000
		Algoritmo 1	20	4.832142	0.931730
		Algoritmo 2	19	6.636269	0.491576
	Características Perceptuales y Radiómicas	Algoritmo 0	15	0.005000	0.472143
		Algoritmo 1	17	0.005000	0.739945
		Algoritmo 2	17	0.005000	0.562117
Penalización por norma-L2	Características Perceptuales	Algoritmo 0	15	4.166470	0.689963
		Algoritmo 1	15	4.703700	0.632108
		Algoritmo 2	12	0.005000	0.738110
	Características Radiómicas	Algoritmo 0	20	6.396270	0.681070
		Algoritmo 1	19	9.431793	0.854049
		Algoritmo 2	20	6.159603	0.693639
	Características Perceptuales y Radiómicas	Algoritmo 0	10	0.013763	0.508150
		Algoritmo 1	17	2.181308	0.916434
		Algoritmo 2	10	0.005000	0.484126

Tabla A.1: Parámetros óptimos obtenidos para cada prueba

Adicionalmente, los valores óptimos obtenidos al utilizar el algoritmo PSO durante las pruebas desarrolladas con los métodos clásicos de selección de características fueron 1.007490, 14.608329 y 0.792028 para el ancho de banda del kernel  $\sigma$ , el parámetro de regularización  $C$  y el parámetro para el algoritmo SMO  $\tau$ , respectivamente.

## BIBLIOGRAFÍA

- [1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2017," *CA: a cancer journal for clinicians*, vol. 67, no. 1, pp. 7–30, 2017.
- [2] 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.
- [3] J. Martínez and V. Martínez, "Protocolo de vigilancia en salud pública: Cáncer de mama y cuello uterino," *Bogotá: INS*, 2016.
- [4] E. d. Vries, M. X. Meneses, and M. Piñeros, "Years of life lost as a measure of cancer burden in colombia, 1997-2012," *Biomédica*, vol. 36, no. 4, pp. 547–555, 2016.
- [5] G. Henríquez and E. de Vries, "El efecto del envejecimiento para la carga de cáncer en colombia: proyecciones para las primeras cinco localizaciones por departamento y sexo en colombia, 2020 y 2050," *Revista Colombiana de Cancerología*, vol. 21, no. 2, pp. 104–112, 2017.
- [6] K. C. Oeffinger, E. T. Fontham, R. Etzioni, A. Herzig, J. S. Michaelson, Y.-C. T. Shih, L. C. Walter, T. R. Church, C. R. Flowers, S. J. LaMonte *et al.*, "Breast cancer screening for women at average risk: 2015 guideline update from the american cancer society," *Jama*, vol. 314, no. 15, pp. 1599–1614, 2015.
- [7] B. Lauby-Secretan, C. Scoccianti, D. Loomis, L. Benbrahim-Tallaa, V. Bouvard, F. Bianchini, and K. Straif, "Breast-cancer screening—viewpoint of the iarc working group," *New England journal of medicine*, vol. 372, no. 24, pp. 2353–2358, 2015.
- [8] M. A. Brown and R. C. Semelka, "Mr imaging abbreviations, definitions, and descriptions: a review," *Radiology*, vol. 213, no. 3, pp. 647–662, 1999.
- [9] P. G. Balboa, M. S. Criville, M. V. Fernández, L. T. Giménez, and M. A. T. Meneses, "Indicaciones para la resonancia magnética de mama," in *Anales de Radiología, México*, vol. 3, no. 3, 2004, pp. 209–212.
- [10] T. Onega, A. N. Tosteson, J. Weiss, J. Alford-Teaster, R. A. Hubbard, L. M. Henderson, K. Kerlikowske, M. E. Goodrich, C. O'Donoghue, K. J. Wernli *et al.*, "Costs of diagnostic and preoperative workup with and without breast mri in older women with a breast cancer diagnosis," *BMC health services research*, vol. 16, no. 1, p. 76, 2016.
- [11] D. Saslow, C. Boetes, W. Burke, S. Harms, M. O. Leach, C. D. Lehman, E. Morris, E. Pisano, M. Schnall, S. Sener *et al.*, "American cancer society guidelines for breast screening with mri as an adjunct to mammography," *CA: a cancer journal for clinicians*, vol. 57, no. 2, pp. 75–89, 2007.
- [12] B. M. Dale, M. A. Brown, and R. C. Semelka, *MRI: basic principles and applications*. John Wiley & Sons, 2015.
- [13] S.-Q. Chen, M. Huang, Y.-Y. Shen, C.-L. Liu, and C.-X. Xu, "Abbreviated mri protocols for detecting breast cancer in women with dense breasts," *Korean journal of radiology*, vol. 18, no. 3, pp. 470–475, 2017.
- [14] Y. Machida, A. Shimauchi, Y. Kanemaki, T. Igarashi, M. Harada, and E. Fukuma, "Feasibility and potential limitations of abbreviated breast mri: an observer study using an enriched cohort," *Breast Cancer*, vol. 24, no. 3, pp. 411–419, 2017.

- [15] G. Oldrini, B. Fedida, J. Poujol, J. Felblinger, I. Trop, P. Henrot, E. Darai, and I. Thomassin-Naggara, "Abbreviated breast magnetic resonance protocol: Value of high-resolution temporal dynamic sequence to improve lesion characterization," *European journal of radiology*, vol. 95, pp. 177–185, 2017.
- [16] B. Panigrahi, L. Mullen, E. Falomo, B. Panigrahi, and S. Harvey, "An abbreviated protocol for high-risk screening breast magnetic resonance imaging: impact on performance metrics and bi-rads assessment," *Academic radiology*, vol. 24, no. 9, pp. 1132–1138, 2017.
- [17] S. C. Harvey, P. A. Di Carlo, B. Lee, E. Obadina, D. Sippo, and L. Mullen, "An abbreviated protocol for high-risk screening breast mri saves time and resources," *Journal of the American College of Radiology*, vol. 13, no. 11, pp. R74–R80, 2016.
- [18] M. Moschetta, M. Telegrafo, L. Rella, A. A. S. Ianora, and G. Angelelli, "Abbreviated combined mr protocol: a new faster strategy for characterizing breast lesions," *Clinical breast cancer*, vol. 16, no. 3, pp. 207–211, 2016.
- [19] C. M. Chhor and C. L. Mercado, "Abbreviated mri protocols: wave of the future for breast cancer screening," *American Journal of Roentgenology*, vol. 208, no. 2, pp. 284–289, 2017.
- [20] L. Heacock, A. N. Melsaether, S. L. Heller, Y. Gao, K. M. Pysarenko, J. S. Babb, S. G. Kim, and L. Moy, "Evaluation of a known breast cancer using an abbreviated breast mri protocol: correlation of imaging characteristics and pathology with lesion detection and conspicuity," *European journal of radiology*, vol. 85, no. 4, pp. 815–823, 2016.
- [21] M. D. Dorrius, M. C. Jansen-van der Weide, P. M. van Ooijen, R. M. Pijnappel, and M. Oudkerk, "Computer-aided detection in breast mri: a systematic review and meta-analysis," *European radiology*, vol. 21, no. 8, pp. 1600–1608, 2011.
- [22] C. D. Lehman, J. D. Blume, W. B. DeMartini, N. M. Hylton, B. Herman, and M. D. Schnall, "Accuracy and interpretation time of computer-aided detection among novice and experienced breast mri readers," *American Journal of Roentgenology*, vol. 200, no. 6, pp. W683–W689, 2013.
- [23] A. D. Brown and T. R. Marotta, "Using machine learning for sequence-level automated mri protocol selection in neuroradiology," *Journal of the American Medical Informatics Association*, vol. 25, no. 5, pp. 568–571, 2017.
- [24] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu, "Feature selection: A data perspective," *ACM Computing Surveys (CSUR)*, vol. 50, no. 6, p. 94, 2017.
- [25] G. Gan and M. K.-P. Ng, "Subspace clustering with automatic feature grouping," *Pattern Recognition*, vol. 48, no. 11, pp. 3703–3713, 2015.
- [26] D. Pir and T. Brown, "Acoustic group feature selection using wrapper method for automatic eating condition recognition," in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.
- [27] T. N. Lal, M. Schroder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer, and B. Scholkopf, "Support vector channel selection in bci," *IEEE transactions on biomedical engineering*, vol. 51, no. 6, pp. 1003–1010, 2004.
- [28] J. M. Sotoca, F. Pla, and J. S. Sanchez, "Band selection in multispectral images by minimization of dependent information," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 2, pp. 258–267, 2007.
- [29] Y. Li, F.-X. Wu, and A. Ngom, "A review on machine learning principles for multi-view biological data integration," *Briefings in bioinformatics*, vol. 19, no. 2, pp. 325–340, 2016.
- [30] Y. Ren, L. Zhang, and P. N. Suganthan, "Ensemble classification and regression-recent developments, applications and future directions," *IEEE Comp. Int. Mag.*, vol. 11, no. 1, pp. 41–53, 2016.

- [31] Y. Gu, T. Liu, X. Jia, J. A. Benediktsson, and J. Chanussot, "Nonlinear multiple kernel learning with multiple-structure-element extended morphological profiles for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 6, pp. 3235–3247, 2016.
- [32] S. Althloothi, M. H. Mahoor, X. Zhang, and R. M. Voyles, "Human activity recognition using multi-features and multiple kernel learning," *Pattern recognition*, vol. 47, no. 5, pp. 1800–1812, 2014.
- [33] C. Xu, D. Tao, and C. Xu, "A survey on multi-view learning," *arXiv preprint arXiv:1304.5634*, 2013.
- [34] S. Qiu and T. Lane, "A framework for multiple kernel support vector regression and its applications to sirna efficacy prediction," *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, vol. 6, no. 2, pp. 190–199, 2009.
- [35] M. Gönen and A. A. Margolin, "Localized data fusion for kernel k-means clustering with application to cancer biology," in *Advances in Neural Information Processing Systems*, 2014, pp. 1305–1313.
- [36] M. Gönen and E. Alpaydm, "Multiple kernel learning algorithms," *Journal of machine learning research*, vol. 12, no. Jul, pp. 2211–2268, 2011.
- [37] G. R. Lanckriet, M. Deng, N. Cristianini, M. I. Jordan, and W. S. Noble, "Kernel-based data fusion and its application to protein function prediction in yeast," in *Biocomputing 2004*. World Scientific, 2003, pp. 300–311.
- [38] D. P. Lewis, T. Jebara, and W. S. Noble, "Support vector machine learning from heterogeneous data: an empirical analysis using protein sequence and structure," *Bioinformatics*, vol. 22, no. 22, pp. 2753–2760, 2006.
- [39] L. Foresti, D. Tuia, V. Timonin, and M. F. Kanevski, "Time series input selection using multiple kernel learning." in *ESANN*, 2010.
- [40] D. Tuia, G. Camps-Valls, G. Matasci, and M. Kanevski, "Learning relevant image features with multiple-kernel classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 10, pp. 3780–3791, 2010.
- [41] N. Subrahmanya and Y. C. Shin, "Sparse multiple kernel learning for signal processing applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, p. 788, 2010.
- [42] M. Gönen, "Bayesian efficient multiple kernel learning," *arXiv preprint arXiv:1206.6465*, 2012.
- [43] H. J. Areiza-Laverde, G. M. Díaz, and A. E. Castro-Ospina, "Feature group selection using mkl penalized with  $\ell_1$ -norm and svm as base learner," in *International Workshop on Experimental and Efficient Algorithms*. Springer, 2018, pp. 136–147.
- [44] S. C. Bushong and G. Clarke, *Magnetic Resonance Imaging-E-Book: Physical and Biological Principles*. Elsevier Health Sciences, 2013.
- [45] R. Damadian, "Tumor detection by nuclear magnetic resonance," *Science*, vol. 171, no. 3976, pp. 1151–1153, 1971.
- [46] P. Lauterbur, "Image formation by induced local interactions: Examples employing nuclear magnetic resonance," *Nature*, vol. 242, no. 5394, pp. 190–191, 1973. [Online]. Available: <https://doi.org/10.1038/242190a0>
- [47] A. M. Lakhwani, K. H. Shah, A. S. Vaghela, D. S. Panchal, and S. R. Rathod, "Review on basics of computer vision and its applications," *Research & Reviews: Journal of Computational Biology*, vol. 6, no. 2, pp. 33–40, 2018.
- [48] E. M. Haacke, R. W. Brown, M. R. Thompson, R. Venkatesan *et al.*, *Magnetic resonance imaging: physical principles and sequence design*. Wiley-Liss New York, 1999, vol. 82.

- [49] P. Soffia, "Difusión por resonancia magnética: bases y aplicaciones oncológicas en órganos extracraneales," *Revista chilena de radiología*, vol. 15, pp. 17–24, 2009.
- [50] K. N. Raymond and V. C. Pierre, "Next generation, high relaxivity gadolinium mri agents," *Bioconjugate chemistry*, vol. 16, no. 1, pp. 3–8, 2005.
- [51] L. Liberman and J. H. Menell, "Breast imaging reporting and data system (bi-rads)," *Radiologic Clinics*, vol. 40, no. 3, pp. 409–430, 2002.
- [52] 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.
- [53] 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.
- [54] A. Borji, D. N. Sihite, and L. Itti, "Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study," *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 55–69, 2013.
- [55] S. Frintrop, E. Rome, and H. I. Christensen, "Computational visual attention systems and their cognitive foundations: A survey," *ACM Transactions on Applied Perception (TAP)*, vol. 7, no. 1, p. 6, 2010.
- [56] V. Leboran, A. Garcia-Diaz, X. R. Fdez-Vidal, and X. M. Pardo, "Dynamic whitening saliency," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 39, no. 5, pp. 893–907, 2017.
- [57] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li, "Salient object detection: A discriminative regional feature integration approach," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 2083–2090.
- [58] S. Mitra, S. Banerjee, and Y. Hayashi, "Volumetric brain tumour detection from mri using visual saliency," *PloS one*, vol. 12, no. 11, p. e0187209, 2017.
- [59] K. S. Deepak, A. Chakravarty, J. Sivaswamy *et al.*, "Visual saliency based bright lesion detection and discrimination in retinal images," in *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*. IEEE, 2013, pp. 1436–1439.
- [60] R. Achanta, F. Estrada, P. Wils, and S. Süsstrunk, "Salient region detection and segmentation," in *International conference on computer vision systems*. Springer, 2008, pp. 66–75.
- [61] Y.-F. Ma and H.-J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," in *Proceedings of the eleventh ACM international conference on Multimedia*. ACM, 2003, pp. 374–381.
- [62] V. Jampani, J. Sivaswamy, V. Vaidya *et al.*, "Assessment of computational visual attention models on medical images," in *Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing*. ACM, 2012, p. 80.
- [63] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.
- [64] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Advances in neural information processing systems*, 2007, pp. 545–552.
- [65] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 1, pp. 185–207, 2013.
- [66] R. A. Khan, "Detection of emotions from video in non-controlled environment," Ph.D. dissertation, 2013.
- [67] L. Itti, N. Dhavale, and F. Pighin, "Realistic avatar eye and head animation using a neurobiological model of visual attention," in *Applications and Science of Neural Networks, Fuzzy Systems, and Evolutionary Computation VI*, vol. 5200. International Society for Optics and Photonics, 2003, pp. 64–79.

- [68] N. Aggarwal and R. Agrawal, "First and second order statistics features for classification of magnetic resonance brain images," *Journal of Signal and Information Processing*, vol. 3, no. 02, p. 146, 2012.
- [69] F. A. Shaikh, B. J. Kolowitz, O. Awan, H. J. Aerts, A. von Reden, S. Halabi, S. A. Mohiuddin, S. Malik, R. B. Shrestha, and C. Deible, "Technical challenges in the clinical application of radiomics," *JCO Clinical Cancer Informatics*, vol. 1, pp. 1–8, 2017.
- [70] S. S. Yip and H. J. Aerts, "Applications and limitations of radiomics," *Physics in Medicine & Biology*, vol. 61, no. 13, p. R150, 2016.
- [71] 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.
- [72] 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.
- [73] 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.
- [74] J. J. Van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R. G. Beets-Tan, J.-C. Fillion-Robin, S. Pieper, and H. J. Aerts, "Computational radiomics system to decode the radiographic phenotype," *Cancer research*, vol. 77, no. 21, pp. e104–e107, 2017.
- [75] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.
- [76] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer Science & Business Media, 1999.
- [77] R.-E. Fan, P.-H. Chen, and C.-J. Lin, "Working set selection using second order information for training support vector machines," *Journal of machine learning research*, vol. 6, no. Dec, pp. 1889–1918, 2005.
- [78] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [79] S. Lessmann, R. Stahlbock, and S. F. Crone, "Genetic algorithms for support vector machine model selection," in *The 2006 IEEE International Joint Conference on Neural Network Proceedings*. IEEE, 2006, pp. 3063–3069.
- [80] T. A. Gomes, R. B. Prudêncio, C. Soares, A. L. Rossi, and A. Carvalho, "Combining meta-learning and search techniques to select parameters for support vector machines," *Neurocomputing*, vol. 75, no. 1, pp. 3–13, 2012.
- [81] J. Liu and E. Zio, "Svm hyperparameters tuning for recursive multi-step-ahead prediction," *Neural Computing and Applications*, vol. 28, no. 12, pp. 3749–3763, 2017.
- [82] L. Zelnik-Manor and P. Perona, "Self-tuning spectral clustering," in *Advances in neural information processing systems*, 2005, pp. 1601–1608.
- [83] G. B. Gönen, M. Gönen, and F. Gürgen, "Probabilistic and discriminative group-wise feature selection methods for credit risk analysis," *Expert Systems with Applications*, vol. 39, no. 14, pp. 11 709–11 717, 2012.
- [84] 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.



- [85] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.
- [86] M. Kloft, U. Brefeld, S. Sonnenburg, and A. Zien, "Non-sparse regularization and efficient training with multiple kernels," *arXiv preprint arXiv:1003.0079*, vol. 186, pp. 189–190, 2010.
- [87] H. J. Areiza-Laverde, A. E. Castro-Ospina, and D. H. Peluffo-Ordóñez, "Voice pathology detection using artificial neural networks and support vector machines powered by a multicriteria optimization algorithm," in *Workshop on Engineering Applications*. Springer, 2018, pp. 148–159.
- [88] M. A. Dávila-Guzmán, W. Alfonso-Morales, and E. F. Caicedo-Bravo, "Heterogeneous architecture to process swarm optimization algorithms," *TecnoLógicas*, vol. 17, no. 32, pp. 11–20, 2014.
- [89] J. E. González-Pérez and D. F. García-Gómez, "Electric field relaxing electrodes design using particle swarm optimization and finite elements method," *TecnoLógicas*, vol. 20, no. 38, pp. 27–39, 2017.
- [90] M. Clerc, "Beyond standard particle swarm optimisation," in *Innovations and Developments of Swarm Intelligence Applications*. IGI Global, 2012, pp. 1–19.
- [91] A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, and F. Herrera, "Performance measures," in *Learning from Imbalanced Data Sets*. Springer, 2018, pp. 47–61.
- [92] J. S. Rincón, A. E. Castro-Ospina, F. R. Narváez, and G. M. Díaz, "Machine learning methods for classifying mammographic regions using the wavelet transform and radiomic texture features," in *International Conference on Technology Trends*. Springer, 2018, pp. 617–629.
- [93] Q. Gu, Z. Li, and J. Han, "Generalized fisher score for feature selection," *arXiv preprint arXiv:1202.3725*, 2012.
- [94] S. Vora and H. Yang, "A comprehensive study of eleven feature selection algorithms and their impact on text classification," in *2017 Computing Conference*. IEEE, 2017, pp. 440–449.
- [95] R. Fusco, M. Sansone, S. Filice, G. Carone, D. M. Amato, C. Sansone, and A. Petrillo, "Pattern recognition approaches for breast cancer dce-mri classification: a systematic review," *Journal of medical and biological engineering*, vol. 36, no. 4, pp. 449–459, 2016.
- [96] H. Cai, Y. Peng, C. Ou, M. Chen, and L. Li, "Diagnosis of breast masses from dynamic contrast-enhanced and diffusion-weighted mr: a machine learning approach," *PloS one*, vol. 9, no. 1, p. e87387, 2014.