



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

INCERTIDUMBRE ASOCIADA A LA EXTRACCIÓN DE
DESCRPTORES RADIÓMICOS EN IMÁGENES DE
MAMOGRAFÍA DIGITAL CON MEDIO DE CONTRASTE

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

FÍSICO

P R E S E N T A :

RICARDO MONTOYA DEL ÁNGEL

TUTOR

M. EN C. JORGE PATRICIO CASTILLO LÓPEZ



CIUDAD UNIVERSITARIA, CDMX, 2021



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Caviae Porcellae
In memoriam Ernestinae

Agradecimientos

Agradezco a la Dirección General de Asuntos del Personal Académico de la UNAM (DGAPA) que, a través del Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica PAPIIT IN-103219, me otorgó apoyo económico para la finalización de este trabajo.

Agradezco a mi tutor de tesis, el M. en C. Jorge Patricio Castillo López, quien desde el comienzo fue paciente, comprensivo y considerado conmigo. Fue él quien impulsó mi primera participación en investigación científica, y quien me guió en cada paso de este trabajo. Estoy muy agradecido de haber sido su alumno de tesis.

Vull agrair a la Dra. María Ester Brandan Siqués que em va introduir al món de la física mèdica, va confiar en mi per fer amb ella el servei social, i li va donar suport a aquest projecte fins a la seva culminació, participant com sinodal.

Quiero agradecer además a la Dra. Karla Paola García Pelagio, a la M. en C. Adriana Moreno Ramírez y a la M. en C. Itzel Militza Torres Viquez, por el tiempo que invirtieron en la revisión y corrección de este trabajo.

Enfin, je voudrais remercier mes parents José et Ignacia, qui m'ont soutenu toujours dans mes études. J'ai réalisé tous les progrès possibles grâce à eux. Je remercie également mon frère José et ma sœur Alejandra, qui sont mes exemples à suivre.

Resumen

El cáncer de mama es el sitio de cáncer con el mayor número de casos nuevos de mujeres en México y en el mundo. Existen diversas técnicas para su detección y diagnóstico, entre las cuales destaca la mamografía digital. Específicamente, la mamografía digital con medio de contraste (CEDM, *Contrast-Enhanced Digital Mammography*) ha mostrado una especial sensibilidad a la presencia de lesiones tumorales, pues resalta la angiogénesis estimulada por el crecimiento de masas.

Para apoyar a la decisión clínica, la radiómica surge como campo de investigación para la extracción de información cuantificable de imágenes médicas. Esta información puede utilizarse para la creación de modelos que predigan el estado fisiopatológico de una lesión. Existe especial interés en predecir el estado inmunohistoquímico (IHQ) pues con él se puede definir el subtipo, origen y tratamiento del cáncer.

En este trabajo se propone una técnica de evaluación de modelos de predicción de la presencia de antígenos IHQ, considerando la incertidumbre de descriptores radiómicos extraídos de imágenes CEDM, en su modalidad temporal con energía única. Los antígenos examinados fueron ER, PR, HER2 y $Ki67 > 0.20$, y se consideró el efecto en la incertidumbre del mastógrafo, el registro espacial de imagen y la delineación de la región de interés.

Se consideraron tres enfoques para la creación de modelos de predicción. Un enfoque tradicional en el que no se consideró la incertidumbre, y dos enfoques robustos. En

el primero se seleccionaron descriptores robustos a la incertidumbre, y en el segundo los modelos fueron entrenados considerando la distribución de incertidumbre.

Se encontró que el uso de incertidumbre mejoró significativamente el desempeño de los modelos al evaluarlos con la técnica propuesta. Para todos los antígenos, a excepción del HER2, los modelos del segundo enfoque robusto alcanzaron un Coeficiente de Correlación de Matthews mayor al 40 %, obteniendo un nivel de acuerdo *moderado*. La fuente de incertidumbre de mayor impacto en los descriptores fue la delineación de la región de interés.

Se espera que la metodología mostrada en este trabajo, basada en consideraciones físicas, pueda ser empleada en trabajos futuros para identificar las fuentes de incertidumbre más importantes en análisis radiómicos, para dar mayor confianza a los modelos de predicción y, así, avanzar un paso más hacia la interpretación física de los descriptores.

Índice general

Agradecimientos	II
Resumen	III
1. Introducción	1
2. Marco Teórico	9
2.1. Cáncer de mama	9
2.1.1. Diagnóstico y estado inmunohistoquímico	11
2.2. Mastógrafo	12
2.3. Mamografía digital con medio de contraste	15
2.3.1. Registro y resta de imágenes	16
2.4. Radiómica	17
2.4.1. Descriptores radiómicos	17

<i>ÍNDICE GENERAL</i>	VI
2.4.2. Flujo de trabajo	20
2.5. Incertidumbre	23
2.5.1. Definiciones	23
2.5.2. Propagación de incertidumbre	25
2.6. Diagnóstico Asistido por Computadora (CADx)	26
2.6.1. Regresión logística	27
2.6.2. Método LASSO para eliminación de descriptores	30
2.6.3. Asociación de descriptores: agrupamiento	31
2.6.4. Riqueza de los descriptores radiómicos: PCA	31
2.6.5. Evaluación de los modelos de predicción	33
2.6.6. Indicadores de desempeño	34
3. Método experimental	38
3.1. Imágenes clínicas	38
3.1.1. Adquisición de imágenes	38
3.1.2. Procesamiento de imágenes	39
3.1.3. Segmentación	41
3.2. Incertidumbre de imágenes CEDM	41

3.2.1.	Incertidumbre asociada al mastógrafo	42
3.2.2.	Incertidumbre asociada al registro de imágenes	43
3.3.	Extracción de descriptores radiómicos	46
3.3.1.	Pyradiomics	46
3.3.2.	Descriptores extraídos	48
3.4.	Distribución y presupuesto de incertidumbre	48
3.5.	CADx: Flujo de trabajo	50
3.5.1.	Análisis tradicional	51
3.5.2.	Análisis robusto: uso de DR robustos	51
3.5.3.	Análisis robusto: con entrenamiento basado en el presupuesto	52
4.	Resultados y discusión	54
4.1.	Presupuesto de incertidumbre	54
4.2.	Modelos de predicción	58
4.2.1.	Rama tradicional	59
4.2.2.	Rama robusta: uso del PI	61
4.2.3.	Comparación de métricas de desempeño	66
4.2.4.	Riqueza de los DR y su relación con el análisis IHQ	67

<i>ÍNDICE GENERAL</i>	VIII
5. Conclusiones	70
A. Presupuestos de incertidumbres	74
B. Agrupación de descriptores	79

Índice de figuras

1.1. Sitio de cáncer con la ASR de incidencia más alta por país durante el 2018. El cáncer de mama (en rosa) ocupa el primer lugar en 104 países [1].	2
2.1. Diagrama de la anatomía de una mama [2].	10
2.2. Atenuación del tejido mamario (Carcinoma ductal, glandular y adiposo) como función de la energía de fotones [3].	13
2.3. Diagrama de un mastógrafo [3].	14
2.4. Flujo de trabajo para un análisis radiómico de Lambin et al., miembros de QIN [4].	23
2.5. Función sigmoide definida como la función hipótesis en la regresión logística. Se muestra el caso $h(x) = \frac{1}{1+e^{-x}}$ para fines explicativos.	28
2.6. Función de costo J para predicción a) $y = 1$ y b) $y = 0$	29

3.1. Segmentación general (azul) y focal (roja) realizada en una misma paciente por radiólogas con a) 1 año de experiencia b) 10 años de experiencia.	42
3.2. Resta de dos pares de imágenes pre- y poscontraste del maniquí cuando las imágenes están a) alineadas y b) desalineadas.	44
3.3. Porcentaje de sesgo del descriptor <i>contraste</i> producto del desalineamiento r de las imágenes pre- y poscontraste.	45
3.4. Flujo de trabajo del CADx propuesto	50
4.1. Presupuesto de incertidumbre de las tres fuentes consideradas, con los 107 DR en orden descendente con respecto al C_V total. Los puntos verdes representan los 24 DR sensibles a la presencia de antígenos. . .	57
4.2. Diagrama de cajas y bigotes para los modelos de los tres antígenos con diferencias estadísticamente significativas entre sus medias.	64
4.3. Diagrama de cajas y bigotes para los modelos del HER2.	65
4.4. Desempeño en MCC de las 100 mejores combinaciones de DR para la predicción de presencia de HER2 de la rama tradicional.	68

Índice de tablas

2.1. Interpretación del MCC	36
4.1. Presupuesto de incertidumbre de los descriptores usados en los modelos de predicción en términos del C_V . Para facilitar la comparación con otros trabajos, se da el nombre del DR en inglés.	55
4.2. Desempeño de los modelos de predicción de ER positivo	58
4.3. Desempeño de los modelos de predicción de PR positivo	58
4.4. Desempeño de los modelos de predicción de HER2 positivo	59
4.5. Desempeño de los modelos de predicción de presencia $Ki67 > 0.20$	59
A.1. Presupuesto de incertidumbre para descriptores de forma.	75
A.2. Presupuesto de incertidumbre para descriptores de primer orden.	75
A.3. Presupuesto de incertidumbre para descriptores de la matriz GLCM	76
A.4. Presupuesto de incertidumbre para descriptores de la matriz GLSZM	77

A.5. Presupuesto de incertidumbre para descriptores de la matriz GLRLM	77
A.6. Presupuesto de incertidumbre para descriptores de la matriz NGTDM	77
A.7. Presupuesto de incertidumbre para descriptores de la matriz GLDM .	78
B.1. Presupuesto de incertidumbre para los descriptores independientes . .	80
B.2. Presupuesto de incertidumbre para los descriptores del primer grupo .	81
B.3. Presupuesto de incertidumbre para los descriptores del segundo grupo.	81
B.4. Presupuesto de incertidumbre para los descriptores del tercer grupo. .	81
B.5. Presupuesto de incertidumbre para los descriptores del cuarto grupo.	81
B.6. Presupuesto de incertidumbre para los descriptores del quinto grupo.	81
B.7. Presupuesto de incertidumbre para los descriptores del sexto grupo. .	82
B.8. Presupuesto de incertidumbre para los descriptores del séptimo grupo.	82
B.9. Presupuesto de incertidumbre para los descriptores del octavo grupo.	82
B.10. Presupuesto de incertidumbre para los descriptores del noveno grupo.	82
B.11. Presupuesto de incertidumbre para los descriptores del décimo grupo.	83
B.12. Presupuesto de incertidumbre para los descriptores del undécimo grupo.	83

1

Introducción

Un monstruo recorre el mundo; un monstruo sigiloso que destruye inadvertidamente desde el interior. Este monstruo es el cáncer.

– Breakthrough: Finding a Cancer Killer (2015)

Según el último reporte de la Agencia Internacional para la Investigación del Cáncer (IARC, *International Agency for Research on Cancer*), en el 2018 murieron alrededor de 9.6 millones de personas en el mundo debido a algún tipo de cáncer [1]. Además, se reportaron en ese mismo año cerca de 18.1 millones de casos nuevos de esta enfermedad alrededor del globo.

El cáncer de mama, en específico, representa uno de los sitios de cáncer¹ con el mayor número de incidencias globalmente. En el mismo reporte del 2018, se registró que tanto el cáncer de mama como el cáncer de pulmón representaron individualmente el 11.6% de todos los nuevos casos a nivel mundial, con más de 2 millones de casos nuevos cada uno. Más aún, el cáncer de mama surgió como el cáncer con la mayor tasa

¹Se les suele llamar tipos de cáncer al cáncer de mama, pulmón, piel, entre otros, aunque su nombre formal es sitios de cáncer pues así se indica la parte del cuerpo en donde se desarrollan.

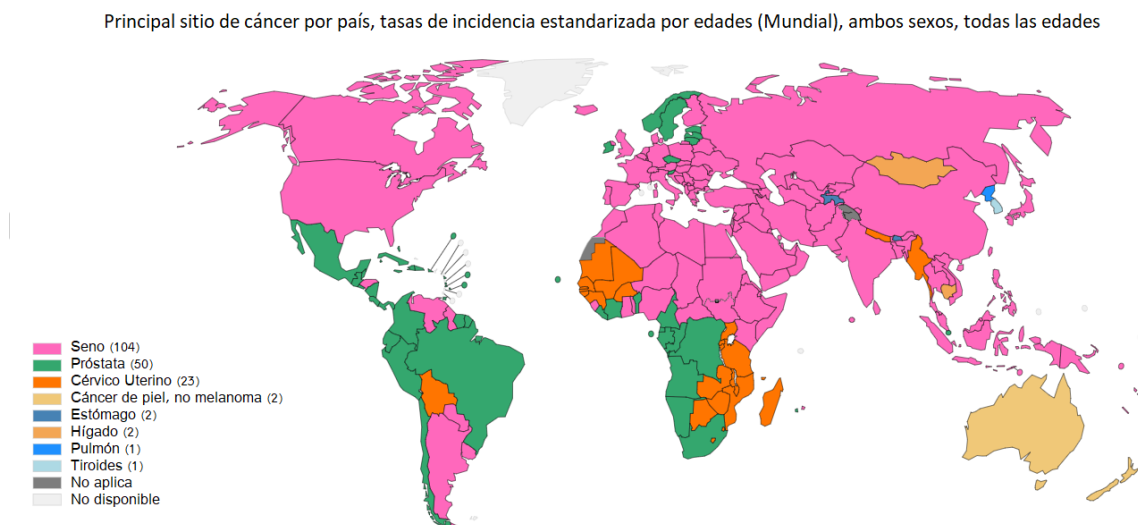


Figura 1.1: Sitio de cáncer con la ASR de incidencia más alta por país durante el 2018. El cáncer de mama (en rosa) ocupa el primer lugar en 104 países [1].

de incidencia estandarizada por edades (ASR, *Age-Standardized Rate*)² en 104 países. Considerando sólo a la población femenina, este número de países asciende hasta 156, representando cerca del 84 % de todos los países miembros de la Organización Mundial de la Salud con datos disponibles.

En México, la muerte por tumores malignos se encuentra entre las primeras 3 causas de muerte, junto con las enfermedades cardíacas y la diabetes, según datos del INEGI del 2018 [5]. De las casi 86 mil defunciones por cáncer reportadas en ese año, 7,311 fueron causadas por cáncer de mama. Además, con respecto al número de incidencias de este sitio de cáncer, se registraron más de 27 mil nuevos casos [6]. Estos datos respaldan el hecho de que las ASR de cáncer de mama, tanto de incidencia como de mortalidad, ocupen el primer lugar para mujeres en el país.[1].

Un esfuerzo multidisciplinario de distintas ramas de investigación como la medicina, la física, la informática, entre otras, se ha fomentado para el estudio, pronóstico, predicción y tratamiento de éste y otros sitios de cáncer. Esto último, combinado con la evolución tecnológica, ha permitido la implementación de nuevas técnicas y

²La tasa ASR permite comparar poblaciones que difieren en edad ya que el cáncer es una enfermedad fuertemente influida por este parámetro.

métodos que impactan la forma en la que se combate al cáncer. Específicamente, este trabajo se fundamenta en la colaboración de tres áreas de investigación: la imagenología médica, la inteligencia artificial y la física, analizadas desde el punto de vista de esta última.

En el área de imagenología médica, la creación de nuevos aparatos de adquisición de imágenes ha permitido el acceso a nueva información proveniente de ellas, antes inaccesible debido a limitaciones tecnológicas. Un ejemplo claro es la evolución del mastógrafo, que transitó de ser un equipo analógico a un sistema digital. Entre las principales diferencias destaca la sustitución de la película radiográfica por un material fotoestimulable o sensor, el cual permite la posterior digitalización de la imagen. Los mastógrafos digitales poseen receptores con un rango dinámico más amplio que el de la película radiológica, lo que permite imágenes con mayor contraste. Además, en el caso de los mastógrafos digitales con detector integrado, se utiliza menos radiación y, debido a que sus imágenes son almacenadas en una computadora, pueden ser fácilmente vistas, compartidas y editadas para una mejor visualización del personal médico en radiología. [7].

A pesar de contar con estas nuevas tecnologías, la interpretación de las imágenes médicas ha sido, tradicionalmente, hecha únicamente por el personal médico en radiología; médicos especializados y altamente entrenados para esta tarea³. Sin embargo, un problema de esta práctica es que la detección y caracterización de los objetos en una imagen médica se tornan subjetivas al conocimiento y experiencia del radiólogo, haciendo que esta información cualitativa sea variable y difícilmente reproducible [8].

Estas limitaciones pueden desembocar en impedimentos para un correcto diagnóstico. Un ejemplo específico es un estudio realizado en la población local de la provincia de Møre og Romsdal, en Noruega, para mamografía de tamizaje de cáncer de mama, la cual es altamente reconocida como el método más efectivo para la detección temprana del cáncer de mama [9, 10]. Este estudio encontró evidencia de

³De aquí en adelante, se referirá a estos médicos simplemente como radiólogos.

sensibilidad limitada⁴ de dicha mamografía pues alrededor del 30 % de los cánceres de intervalo⁵ y 20 % de los nuevos cánceres diagnosticados estaban, de hecho, presentes en la mamografía previamente tomada, pero cuyas lesiones, en la mayoría de los casos, no habían sido correctamente interpretadas por el radiólogo [11]. Este ejemplo evidencia la dificultad presente en la interpretación de algunas imágenes médicas, aún para médicos especializados.

Con el fin de apoyar a cuantificar características de las imágenes médicas para relacionarlas con el estado biomédico del paciente, entre otras cosas, surgió un área de estudio llamada radiómica. La radiómica es un campo de investigación relativamente nuevo que busca la conversión de información y características contenidas en imágenes médicas a información cuantificable, la cual posteriormente es correlacionada con el estado fisiopatológico del paciente para apoyar a la decisión clínica. [12]. La palabra nace de la unión de *radio-* (de la palabra radiación) con *-omics*, sufijo de origen griego que significa «el estudio de la totalidad de algo», usado en otras áreas de la ciencia como la genómica y la proteómica. En este caso, la radiómica puede entenderse como el estudio de grandes cantidades de imágenes radiológicas, y por extensión cualquier imagen médica, para extraer información cuantificable.

La hipótesis fundamental que rige la radiómica es que las imágenes son el producto de mecanismos a nivel genético y molecular, ligados a las características genotípicas y fenotípicas del tejido [12]. Por ende, con el fin de entender estos mecanismos, se debe de entender el principio bajo el cual la imagen se obtiene. En el caso de las imágenes mamográficas, la interacción de los rayos X con el tejido impone un modelo de física de radiaciones y dosimetría.

Es esto último lo que abre paso a un principio básico, el cual guía esta investigación: los equipos de adquisición de imágenes, en este caso particular el mastógrafo,

⁴En este caso, la sensibilidad mide cuántas veces se detectó correctamente la presencia de una lesión maligna con respecto a la muestra total.

⁵Se define al cáncer de intervalo como aquel que se presenta dentro del año posterior a la adquisición de la última mamografía.

son instrumentos de medición de cantidades físicas, y la imagen médica es un conjunto de mediciones realizadas sobre un individuo. Este principio es apoyado por uno de los objetivos centrales de la radiómica, que busca que cada imagen médica pueda ser utilizada individualmente en cada paciente, ya que contendrían información cuantitativa específica de cada individuo para ayudar en el diagnóstico de la enfermedad, apoyando así a la medicina personalizada [12].

De esta forma, el uso de descriptores de información cuantitativa, llamados descriptores radiómicos (DR), junto con otro tipo de información complementaria, como la información clínica o genómica, podría permitir la creación de herramientas de apoyo para las decisiones médicas [4, 13].

Una forma de abordar este tema, cada vez más implementada por la comunidad científica en los últimos años, es a través del uso de sistemas de inteligencia artificial (IA), específicamente sistemas pertenecientes al aprendizaje automático o *Machine Learning*. Dichos sistemas son capaces de aprender e identificar patrones a través de datos para, posteriormente, tomar decisiones con la mínima intervención humana [14]. En el caso de la imagen médica, estos datos son las imágenes.

Existen dos técnicas principales para tales sistemas: Detección Asistida por Computadora (*CAD*, *Computer-aided detection*) y Diagnóstico Asistido por Computadora (abreviado *CADx*, *Computer-aided diagnosis*, para diferenciarlo de *CAD*). La primera técnica se enfoca en utilizar patrones en las imágenes para reconocer regiones sospechosas y marcarlas, con el fin de que el radiólogo las revise e interprete con mayor detalle [11]. Por otro lado, el objetivo de *CADx* es usar la información extraída de las imágenes, es decir los descriptores, para determinar el estado fisiopatológico de las lesiones, siendo un ejemplo común la clasificación de la benignidad o malignidad de una lesión. De esta manera, la técnica *CADx* puede verse como el uso de la IA para cumplir con el objetivo de la radiómica arriba mencionado.

Dentro de la familia de técnicas denominadas *CADx* existe la llamada Radiómica

de Diseño Humano (RDH) [14], caracterizada por la extracción de descriptores con significado intuitivo, tales como la forma, o con definiciones inspiradas por cantidades físicas o matemáticas tales como energía, entropía, entre otras.

Existen trabajos precursores de RDH que han sido desarrollados dentro del grupo de Dosimetría y Física Médica (DOSIFICAME) del Instituto de Física de la UNAM, tales como la tesis de maestría de J. Mateos [15], quien propuso un método de análisis de textura⁶ aplicado sobre imágenes de mamografía digital con medio de contraste (CEDM por sus siglas en inglés), una modalidad de imagen activamente estudiada por el grupo DOSIFICAME. El modelo resultante de este estudio pudo diferenciar lesiones malignas con una precisión de 87 %.

Sin embargo, dado que la radiómica es un área relativamente nueva, la falta de protocolos y estandarización de la metodología ha hecho que su avance se ralentice. Diferencias entre estudios, tales como los parámetros de adquisición de imágenes, modelos de mastógrafos, técnicas de imagen empleadas, entre otras, obstaculizan la estandarización y armonización⁷ entre estudios, aun cuando estas diferencias no contienen información relevante que describa la biología del paciente. Como es mencionado por varios autores, tales como Mendel et al. [16], la heterogeneidad en las bases de datos causada por factores no biológicos es perjudicial para los fines de esta área de estudio, específicamente para la generalización de los resultados obtenidos. Este detrimento se debe a que la información cuantificable extraída es sensible a parámetros como los arriba mencionados, perjudicando la reproducibilidad de los hallazgos.

Diversas organizaciones como la *Quantitative Imaging Network* (QIN) y la *Quantitative Imaging Biomarkers Alliance* (QIBA) han buscado en los últimos años establecer un estándar de referencia para análisis radiómicos, proveyendo recursos probados y actualizados para, así, asegurar la repetibilidad y reproducibilidad, fortaleciendo al

⁶Un subgrupo de los descriptores radiómicos.

⁷En este caso, se entiende como armonización a la mejora en la compatibilidad entre estudios provenientes de bases de datos diferentes, o con diferentes metodologías, pero con medidas similares.

mismo tiempo a la comunidad científica de análisis cuantitativo [13].

Aún con estos esfuerzos para la estandarización y validación de resultados, el número de estudios para analizar la robustez de la información cuantitativa, es decir, su sensibilidad a factores no biológicos y su influencia en los modelos radiómicos, es relativamente bajo. Específicamente para la técnica CEDM para cáncer de mama, existe una escasez⁸ de estudios que justifiquen el origen y la influencia de la incertidumbre en los resultados con imágenes de esta modalidad.

Para apoyar en estos esfuerzos, esta investigación tiene como objetivo primario cuantificar las incertidumbres asociadas a un análisis radiómico hecho en imágenes obtenidas con modalidad CEDM. Este análisis radiómico fue diseñado para desarrollar un modelo de predicción que relacione DR con el estado inmunohistoquímico (IHQ) de la lesión mamaria. Una vez cuantificada la incertidumbre, se busca determinar su influencia en el entrenamiento y en el desempeño de los modelos de predicción, además de usar dicha incertidumbre para la elección de descriptores robustos a ella.

Para formar el modelo de predicción se extraerán DR disponibles en una plataforma de extracción de descriptores validada por QIN. Para la cuantificación de incertidumbre se utilizará una distribución de incertidumbre, representada en tablas bajo un formato denominado presupuesto de incertidumbre (PI), para tres fuentes de incertidumbre, consideradas como las más influyentes en el valor de los DR: el mastógrafo, el registro espacial de imágenes y la delineación manual de la región de interés (ROI, *Region of interest*) por parte del radiólogo.

La hipótesis formulada es que, al realizar una prueba de desempeño en los modelos considerando la influencia de la incertidumbre, los modelos creados a partir de la consideración del PI para la elección de los DR y para su entrenamiento, obtendrán un mejor desempeño de predicción que aquellos que no toman en cuenta la influencia

⁸Sólo se encontraron dos trabajos en los que se analiza la incertidumbre y su influencia en los modelos de predicción.

de la incertidumbre. Con esto, se espera identificar las fuentes de incertidumbre más significativas, cuantificarlas y proponer una forma de considerar su influencia en el entrenamiento del modelo. Esto significaría un paso hacia adelante en el aislamiento del efecto de los factores no biológicos, los cuales afectan la comparación entre investigaciones interinstitucionales, la obtención de resultados confiables, la combinación de bases de datos y la interpretación física de los DR.

2

Marco Teórico

Si se busca conocer la longitud de un objeto, por ejemplo un libro, se necesita tener el objeto a medir y un instrumento de medición adecuado, e.g. un flexómetro. El instrumento dará una medida, que en este caso será un número real positivo con su respectiva incertidumbre. Similarmente, las imágenes mamográficas, interpretadas desde el punto de vista de la física como una distribución espacial de mediciones de radiación, pueden ser entendidas usando como analogía al ejemplo anterior, donde el objeto a medir es la mama de la paciente, el instrumento de medición es el mastógrafo, y la medición es representada a través de una imagen. En las primeras tres secciones de este capítulo, se darán los principios básicos necesarios para entender cada uno de esos componentes.

2.1. Cáncer de mama

La mama está formada por tres componentes primarios: lóbulos, ductos y tejido conector. *Grosso modo*, los lóbulos son glándulas que producen la leche materna y los ductos son tubos que transportan dicha leche. Ambas partes se encuentran inmersas

en el tejido conector, constituido por tejido fibroso y adiposo [2].

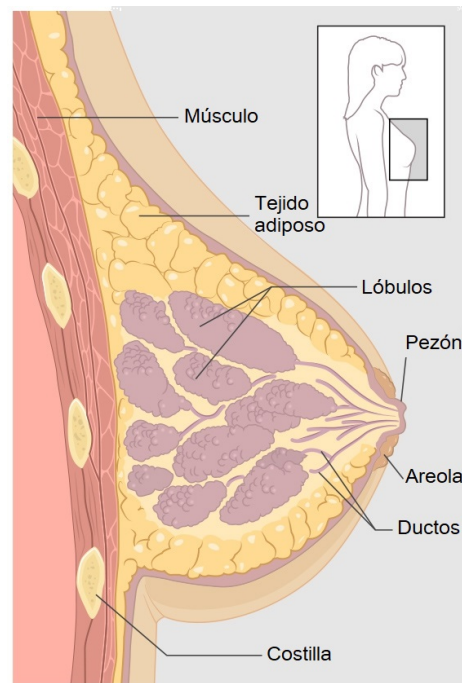


Figura 2.1: Diagrama de la anatomía de una mama [2].

El cáncer de mama es una enfermedad en la cual las células del tejido mamario crecen fuera de control [2]. Existen diferentes tipos de cáncer de mama, dependiendo del tipo de célula del tejido que se convierta en lesión. No obstante, en la mayoría de los casos, el cáncer comienza en los ductos o en los lóbulos, i.e. el tejido glandular.

Existen diversas características anatómicas y fisiológicas del tejido canceroso que las diferencian del tejido sano, las cuales dependen del tipo de cáncer, tales como su masa, forma o la presencia de microcalcificaciones. Entre los procesos fisiológicos destaca la angiogénesis, un proceso de formación de nuevos vasos sanguíneos a partir de la vasculatura preexistente, con el fin de proporcionar oxígeno y nutrientes a las células para su funcionamiento [17]. Es necesario que todas las células del cuerpo, incluidas las malignas, estén localizadas a una distancia de entre 100 y 200 μm de un vaso sanguíneo para poder sobrevivir, por lo que las células tumorales promoverán el proceso de angiogénesis para sobrevivir.

Los vasos sanguíneos creados a través de angiogénesis cancerosa son formados muy rápidamente, en comparación con los vasos de angiogénesis natural, por lo que cuentan con una porosidad elevada, es decir, son altamente permeables [18]. Esto es aprovechado por los médicos para el diagnóstico del cáncer, pues la inyección de una sustancia exógena, llamada medio de contraste, puede permitir una visualización mejorada de dichas neovascularizaciones y, por lo tanto, de la lesión maligna.

2.1.1. Diagnóstico y estado inmunohistoquímico

El diagnóstico del cáncer de mama comienza, comúnmente, con la adquisición de imágenes mamográficas, tema que será abordado en la siguiente sección. En el caso de que una mastografía muestre la presencia de una lesión sospechosa, una biopsia es requerida. La biopsia es la extracción de una muestra de tejido de un área sospechosa de tener cáncer, siendo ésta la única prueba que puede confirmar el cáncer de mama [19]. Si la muestra es considerada como maligna por el patólogo, se prosigue a realizar un análisis inmunohistoquímico.

El análisis inmunohistoquímico¹ (IHQ) es un método que tiene como fin demostrar la sobreexpresión² y localización de antígenos (en este caso, proteínas) específicos en un tejido y su posterior visualización a través de un microscopio [20], con el fin de caracterizar al tejido. En el caso de un análisis IHQ en la mama, los antígenos detectados son los receptores hormonales de estrógeno (ER) y progesterona (PR), el receptor HER2/neu (HER2) y el antígeno de proliferación celular MKI-67 (Ki67) [21]. Sólo para el antígeno Ki67 no se obtiene un resultado binario, pues se define su presencia a través de un porcentaje, determinado por el porcentaje de células en el tejido que son positivas a Ki67. Sin embargo, si el porcentaje es mayor al 20% se considera una proliferación de grado alto, permitiendo que el resultado pueda

¹La palabra puede separarse en tres partes principales: inmuno, relacionado al sistema inmunológico, histo, palabra de origen griego que significa tejido, y químico.

²Se refiere a la elaboración de demasiadas copias de una proteína u otra sustancia. La sobreexpresión de ciertas proteínas u otras sustancias puede influir en la formación de un cáncer.

abordarse como binario. Actualmente, la única forma de realizar un análisis IHQ es por medio de una biopsia.

La presencia de estas proteínas permite caracterizar subtipos de tumores, confirmar el origen de la lesión y proveer información adicional que puede resultar importante para el pronóstico, predecir la respuesta al tratamiento o evaluar tumores residuales después del tratamiento [22].

Para diferenciar entre las diferentes posibles combinaciones de la presencia de estos antígenos, se definen perfiles inmunohistoquímicos. Para el cáncer de mama destacan los perfiles tipo luminal A y B, el perfil mixto, el tipo HER2/neu y el triple negativo [21]. La importancia de estos perfiles reside en su valor pronóstico para el cáncer de mama y su tratamiento. Por ejemplo, una paciente de cáncer mamario con una respuesta positiva del receptor HER2 tendrá un riesgo mayor de recurrencia si no es tratada con una terapia anti-HER2 [23].

2.2. Mastógrafo

La mamografía³ es una examinación radiográfica de la mama, diseñada particularmente para detectar sus patologías. Debido a que existe una diferencia notablemente pequeña entre la atenuación del haz de fotones del tejido glandular sano y los carcinomas⁴ como se muestra en la figura 2.2, en términos del coeficiente lineal de atenuación, es necesario el uso de rayos X con energía baja con el fin de acentuar el contraste de la imagen [3].

Además de esto, se requieren detectores de rayos X con alta resolución espacial

³Mamografía y mastografía son palabras sinónimo cuyas etimologías provienen ultimadamente de raíces griegas. La primera (con *Mamme*) hace referencia a la madre y la segunda (con *Mastos*) hace referencia específicamente a la mama.

⁴Tipo de cáncer desarrollado a partir de células epiteliales, tales como las del tejido glandular mamario.

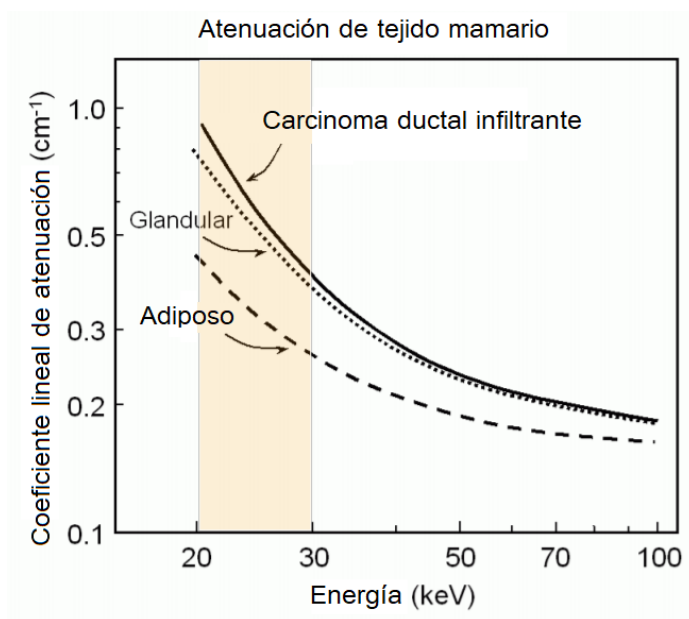


Figura 2.2: Atenuación del tejido mamario (Carcinoma ductal, glandular y adiposo) como función de la energía de fotones [3].

para detectar pequeñas microcalcificaciones, pues estas son típicamente marcadores tempranos de desarrollo de cáncer. La demanda de estas características específicas llevó a la construcción de un sistema de imagen específico, el mastógrafo [3]. A continuación se describen los componentes de un mastógrafo digital con detector integrado.

El mastógrafo está formado por las siguientes partes principales: tubo de rayos X, filtro, colimador, paleta compresora, rejilla antidispersora y detector de rayos X, mostrados en la figura 2.3.

Un espectro de rayos X es generado en el tubo de rayos X, dado un voltaje en el tubo, entre el cátodo (fuente de electrones) y el ánodo (objetivo de los electrones). Esto da lugar a radiación de frenado y característica, la cual pasará a través del filtro. El filtro tiene como función remover rayos X de muy baja y muy alta energía⁵ para modificar el espectro con el fin de transmitir sólo rayos X con la energía óptima para mastografía. El haz de este espectro pasa primero por un colimador, con el fin de limitar la incidencia de los rayos sólo a la zona de interés. Al entrar en contacto con

⁵Esto último en el caso de los filtros de absorción de borde k, como son los hechos de rodio o molibdeno.

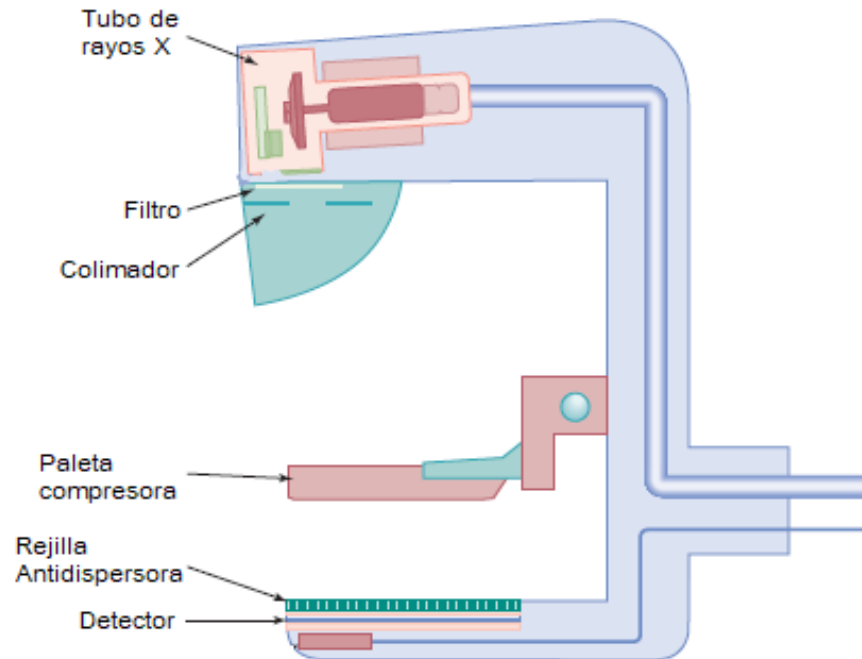


Figura 2.3: Diagrama de un mastógrafo [3].

el tejido, parte del haz atraviesa la mama, mientras que otra parte es dispersada o absorbida. La mama es inmobilizada y comprimida por la paleta compresora, con el fin de reducir la superposición, grosor del tejido e inmobilizar a la paciente. Entre la mama y el detector se posiciona la rejilla antidispersadora, cuya función es evitar la detección de fotones dispersados, producto de la dispersión de Rayleigh y del efecto Compton. Los fotones que inciden en el detector son convertidos en señales eléctricas, ya sea por conversión directa o indirecta⁶, las cuales en conjunto crean la imagen mamográfica [3].

El mastógrafo puede ser visto, desde el punto de vista de la radiómica, como un instrumento de medición de la atenuación de los rayos X a través de la mama. Siguiendo este principio, la mamografía es una distribución espacial de mediciones de atenuación de la radiación, la cual posee correlaciones espaciales e incertidumbre, reflejando

⁶Las conversiones directa e indirecta dependerán del componente utilizado en el dispositivo de lectura digital. Ejemplos de componentes usados son selenio amorfo y silicio amorfo, respectivamente.

propiedades físicas de los tejidos e información sobre la composición de la mama.

2.3. Mamografía digital con medio de contraste

La mamografía digital con medio de contraste (CEDM) es una modalidad de imagen que consiste en la obtención de imágenes mamográficas, considerando la inyección de un medio de contraste (MC) suministrado a la paciente vía intravenosa. Estas imágenes pueden ser posteriormente registradas y restadas píxel a píxel, con el fin de mejorar la visualización de los nuevos vasos sanguíneos, dependiendo de la técnica y el protocolo utilizado [24].

Para esta modalidad de imagen existen dos variantes: energía dual (DE, *Dual-energy*) y temporal a una sola energía (SET, *Single-energy temporal*).

Por un lado, la modalidad DE se refiere a la obtención de dos imágenes mamográficas, una con alta energía y una con baja energía, ambas después de la administración del MC. La dependencia no lineal del coeficiente lineal de atenuación con la energía del haz de fotones es aprovechada para resaltar la presencia del MC al realizar la resta [25].

Por otro lado, la modalidad SET consiste en la adquisición de imágenes antes y después de la administración del MC, para diferentes tiempos, usando el mismo espectro de energía para todas las imágenes [24]. El espectro de energía se obtiene al aplicar un voltaje en el tubo de rayos X de entre 45 a 49 kV. Se necesita este espectro con el fin de maximizar la sensibilidad de la imagen para bajas concentraciones del MC. Para el protocolo seguido en esta investigación, se adquiere una imagen antes de administrar el MC y una imagen cada minuto para los 5 minutos posteriores a la inyección del MC. Una característica de esta modalidad es que permite la obtención de información dinámica del MC alrededor de la lesión. La dosis absorbida por la

paciente en este tipo de estudios oscila entre 1 y 3 mGy, lo que la hace similar a la dosis de una mamografía convencional.

En este estudio se utilizó la modalidad SET, usando un medio de contraste Ultravist 300 (Bayer), aprobado para imágenes médicas con rayos X. El mismo posee una concentración de yodo de 300 mg/ml, la cual aumenta localmente el coeficiente lineal de atenuación de las neovascularizaciones con respecto al tejido glandular y adiposo.

2.3.1. Registro y resta de imágenes

Las imágenes obtenidas en cualquiera de las modalidades de CEDM deben pasar por el proceso de registro y substracción. El registro es necesario para compensar el posible movimiento de la paciente, el cual puede generar artefactos en las imágenes restadas [24], y se refiere a la aplicación de una transformación geométrica a una de las imágenes, con el fin de crear una coincidencia en algún rasgo de interés común entre ellas [26]. El método de registro depende del tipo de transformación que se utilice, tales como la transformación rígida, afín, proyectiva, entre otras.

Una vez realizado el registro, se lleva a cabo la resta de imágenes píxel a píxel. Dado que las imágenes con MC son más opacas, es decir, la presencia de yodo aumenta la absorción de rayos X, la imagen con MC es restada a la imagen sin MC (Ver sección 3.2). El objetivo es resaltar alguna región de interés en la imagen, eliminando el fondo que la rodea. En este estudio particularmente, lo que se busca resaltar es el yodo y lo que se elimina es el parénquima mamario⁷.

⁷Se hace distinción entre el parénquima mamario y el estroma mamario. El primero está compuesto por tejido glandular y el segundo por el tejido adiposo y conector.

2.4. Radiómica

Las imágenes médicas han cambiado de ser primariamente herramientas de diagnóstico para convertirse en actores principales en la medicina personalizada. Esto último ha sido impulsado a través de distintas disciplinas, entre las cuales destaca la radiómica.

La radiómica busca la conversión de información y características contenidas en imágenes médicas en información cuantificable para su subsecuente explotación, para apoyar a la decisión clínica [12]. De esta manera, para los fines de este trabajo, las imágenes médicas que contienen información relacionada con las lesiones cancerosas, son transformadas en datos con potencial para construir una relación con el estado de salud de la paciente [4]. La forma en la que esta información es extraída de las imágenes, y su posterior explotación, se conoce como análisis cuantitativo de imágenes, o análisis radiómico⁸.

Es importante enfatizar que la radiómica no busca sustituir a la información cualitativa, es decir, los descriptores semánticos cualitativos dados por los radiólogos para describir una lesión, sino la combinación de ambas para intentar describir la fisiopatología de las lesiones [27]. En este trabajo, como fue mencionado previamente, se utiliza la subtécnica de CADx llamada Radiómica de Diseño Humano (RDH) para el análisis radiómico. En la sección 2.6 se discute esta subtécnica con más detalle.

2.4.1. Descriptores radiómicos

La información cuantitativa extraída de las imágenes se obtiene sobre una región específica llamada región de interés (ROI), la cual puede ser trazada sobre la imagen

⁸Algunos autores, como Rizzo et al. [27], incluyen en la definición de radiómica también a la información cualitativa. En este trabajo sólo se considera la información cuantificable para esta definición.

por un radiólogo o a través de un algoritmo de segmentación automático o semiautomático. Dicha información es representada por descriptores radiómicos cuantitativos, o simplemente llamados descriptores radiómicos (DR). Existe una gran variedad de DR derivados de diferentes modalidades de imágenes clínicas, los cuales pueden ser extraídos de las imágenes ya sea directamente, o después de aplicar alguna transformación. En el caso específico de la subtécnica RDH, los descriptores pueden ser categorizados de la siguiente manera [27]:

1. **Descriptores de forma:** describen la forma de la ROI y sus propiedades geométricas, tales como volumen, diámetro máximo, esfericidad, entre otros.
2. **Descriptores estadísticos de primer orden:** describen la distribución individual del valor de los píxeles sin considerar su relación espacial. Entre estos se encuentran el promedio, la media, el máximo, mínimo, etc.
3. **Descriptores estadísticos de segundo orden:** toman en cuenta la distribución espacial y la relación entre píxeles vecinos.
4. **Descriptores estadísticos de órdenes mayores:** se obtienen con métodos estadísticos una vez aplicados filtros o transformaciones matemáticas en las imágenes, con el fin de determinar patrones, reducir el ruido o remarcar detalles.

De estos grupos de descriptores, los más comunes para análisis radiómicos son los de segundo orden, destacando de entre ellos el subgrupo de descriptores de textura, teniendo sus orígenes en imágenes no médicas desde 1973 [28]. La textura es una propiedad de prácticamente todas las superficies, la cual promete contener información útil para fines de caracterización de imágenes. La textura puede ser extraída usando diferentes métodos, los cuales se diferencian primariamente por la matriz bajo la cual se extraen los descriptores. Dado que la interpretación física de estos descriptores se encuentra más allá de los alcances de este trabajo, se describirán sólo dos de estas matrices, la matriz de coocurrencia de grises (GLCM, *Gray-level co-occurrence matrix*),

usada para la extracción de descriptores de textura y la matriz de tamaño de zona de grises (GLSZM, *Gray-level size-zone matrix*), usada en otros trabajos con objetivos similares, como el de Zhou et al. [23].

La GLCM describe la probabilidad de que dos niveles de intensidad de píxel aparezcan para una relación específica de ángulo θ y distancia d entre ellos [28]. Supóngase una ROI definida en una región rectangular de tamaño $N_x \times N_y$ píxeles, con $N_x, N_y \in \mathbb{N}$. Supóngase que la intensidad del tono de gris para cada píxel está cuantizado a N_g niveles, $N_g \in \mathbb{N}$. Se define la matriz de coocurrencia $P \in \mathcal{M}(\mathbb{N})_{N_g \times N_g}$ tal que el elemento $P(i, j)$ representa el número de veces que la combinación de niveles de gris i y j ocurrieron en dos píxeles de la imagen, separados una distancia d para un ángulo θ , con la distancia métrica inducida por la norma infinito $\|\dots\|_\infty$ [29], tal que si $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$ entonces $\|\mathbf{x}\|_\infty := \max(|x_1|, |x_2|)$.

Considerando que para cada d y θ se puede definir un desplazamiento $(\Delta x, \Delta y)$ de cada coordenada de píxel $(x, y) \in N_x \times N_y$, la GLCM puede definirse también como [28, 29]

$$P(i, j) = \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} \begin{cases} 1, & \text{si } I(x, y) = i \text{ e } I(x + \Delta x, y + \Delta y) = j \\ 0, & \text{otro caso,} \end{cases} \quad (2.1)$$

con $I(x, y)$ la intensidad del píxel (x, y) . Algunos autores definen la GLCM como simétrica, estableciendo que para un ángulo θ , se debe considerar también el ángulo $\theta + \pi$ en la suma de vecinos.

Por otro lado, la GLSZM es una matriz que cuantifica las zonas de nivel de gris en una imagen [29]. Una zona de nivel de gris es definida como el número de píxeles conectados que comparten la misma intensidad de gris. Dos píxeles están conectados si la distancia entre ellos es 1, siguiendo la norma infinito descrita previamente. Contrario al resto de las matrices de análisis de textura, como la GLCM, la GLSZM es

una matriz independiente de las rotaciones.

2.4.2. Flujo de trabajo

Siendo una disciplina relativamente joven, la radiómica enfrenta diferentes retos, de entre los que destacan su complejidad técnica, terminología, el diseño pobre de los estudios, y una falta de estándares y protocolos [4]. Diversos grupos y redes de investigadores han surgido con el fin de resolver estos retos, como las redes de investigadores ya mencionadas: QIBA y QIN. Ambas organizaciones urgen a la comunidad científica a establecer un flujo de trabajo riguroso y sólido [13], es decir, una estructura clara de los aspectos operacionales de la actividad de trabajo [30].

Lambin et al., miembros de QIN, desarrollaron un flujo de trabajo basado en la premisa de que la radiómica es un mapeo cuantitativo, es decir, la extracción, análisis y modelado de descriptores de imágenes médicas relacionados con objetivos de predicción, como los criterios de valoración clínicos (*clinical endpoints*) o características genómicas [4]. De esta forma, se definen 5 fases primarias de un análisis radiómico, discutidas a continuación.

1. Selección de datos

Se comienza eligiendo la modalidad de imagen y su protocolo de adquisición, la ROI y el objetivo de predicción. La reproducibilidad y compatibilidad entre estudios radiómicos solamente pueden asegurarse siguiendo un protocolo riguroso de adquisición de imagen. En este trabajo, la modalidad de imagen y su protocolo se encuentran predeterminadas tratándose de un conjunto de mamografías obtenidas dentro de las instalaciones del Instituto Nacional de Cancerología (INCan), bajo sus protocolos de obtención. También para este trabajo, el objetivo de predicción es el estado inmunohistoquímico de la lesión. El tamaño y la posición de la ROI es la primera variable a considerar para esta investigación, siendo explicada con más detalle en los capítulos

posteriores.

2. Imagenología Médica

En esta fase se determina el método de segmentación, es decir, la forma en la que la ROI es trazada. Como se ha mencionado anteriormente esta puede ser obtenida de forma manual, automática o híbrida. Además, en esta sección se consideran los estudios con maniqués, con el fin de determinar la intra- e intervariabilidad de los mastógrafos. En este trabajo, las segmentaciones son realizadas manualmente y el uso de maniqués está incluido en el análisis de incertidumbre, detallados en el siguiente capítulo. Relacionado a este último punto se encuentra la variabilidad de las imágenes con respecto al tiempo, pues existen descriptores altamente dependientes a movimientos de la paciente, o a cambios del tamaño del objeto objetivo. Para evitar esto, existen pruebas de estabilidad y robustez, como la realizada por Robinson et al. [31].

3. Extracción de descriptores

La determinación del método de extracción de descriptores es fundamental para evitar la variabilidad de extracción debido a nomenclatura, definición matemática, metodología, software y algoritmos de extracción [4]. Además, en esta fase se debe esclarecer si las imágenes fueron preprocesadas o reconstruidas, usando filtros, recuantificación de la escala de grises, etc.

4. Análisis exploratorio

En esta fase se determina el proceso de reducción o exclusión de descriptores. Es decir, se evalúa la relación entre descriptores, como los que están altamente correlacionados, para reducir un conjunto de muchos descriptores a uno solo. Esto incluye los descriptores relacionados con descriptores clínicos típicos, como la estadificación del tumor, los cuales no aportan información adicional.

Esta fase del flujo de trabajo es fundamental desde el punto de vista del análisis de robustez dado que la elección de los descriptores puede influir considerablemente en el desempeño del modelo radiómico de predicción. Además, desde el punto de vista de los modelos de aprendizaje profundo, el exceso de descriptores puede ocasionar alta varianza del modelo.

5. Modelado

En la etapa final, se determinan los descriptores seleccionados, la metodología del modelo y su validación. Para alcanzar un modelo holístico⁹ se deben incluir datos fuera de la radiómica. En esta etapa se determina el número máximo de descriptores que se van a considerar en el modelo, evitando problemas como el sobreajuste. Por otro lado, la metodología de modelado es, normalmente, dependiente de la experiencia de los investigadores, considerando para cada una sus diferentes limitaciones. La validación del modelo se evalúa normalmente en términos de la discriminación y la calibración, considerando que una validación externa otorga mayor credibilidad que una validación interna.

Para dar una calificación numérica a la calidad de un análisis radiómico, Lambin et al. propusieron el *Radiomics Quality Score* (RQS) o Nivel de Calidad Radiómico. Siguiendo el mismo flujo de trabajo previamente descrito, mostrado en la figura 2.4, se otorga una puntuación al estudio en cada uno de los rubros mencionados previamente, siempre y cuando el mismo cumpla con un criterio establecido. Si bien el RQS puede ser empleado como una guía para los estudios radiómicos, la subjetividad al determinar si un criterio se cumple, mantiene al RQS como una métrica de apoyo y no como una regla absoluta.

⁹En este caso, un modelo holístico es un modelo que busca incluir todos los aspectos médicos a través de registros clínicos, datos obtenidos de tratamientos, e información obtenida de origen genético o biológico.

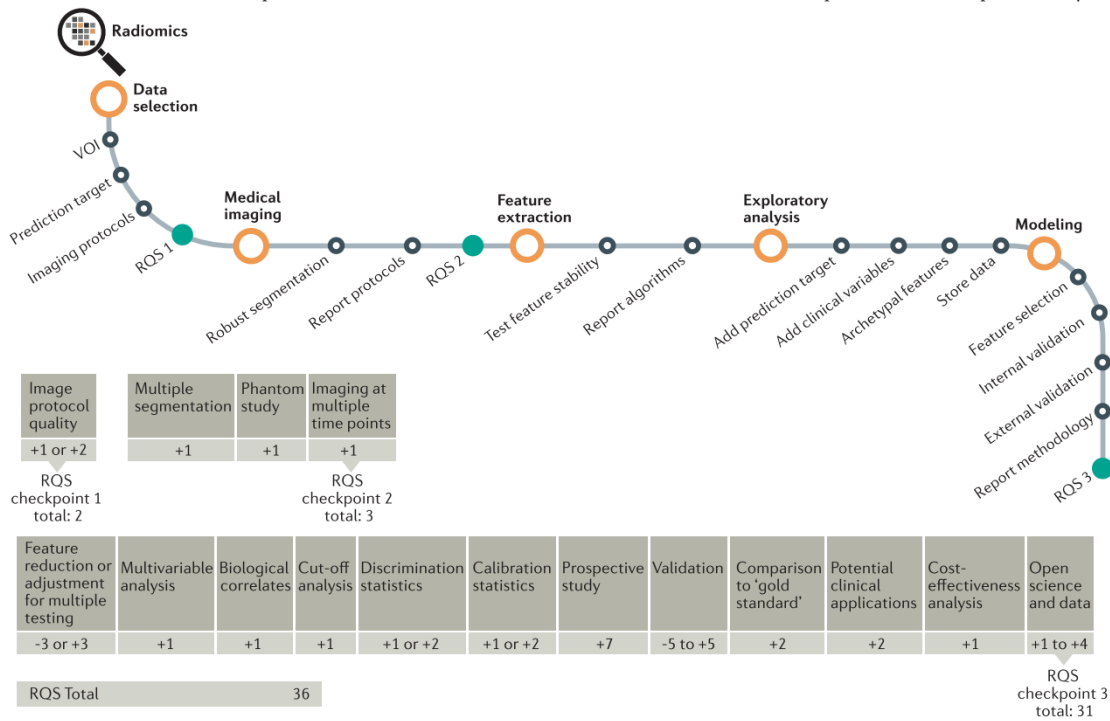


Figura 2.4: Flujo de trabajo para un análisis radiómico de Lambin et al., miembros de QIN [4].

2.5. Incertidumbre

2.5.1. Definiciones

Dado que el objetivo primario de este trabajo es cuantificar las incertidumbres asociadas a un análisis radiómico, se deben definir con más detalles algunos conceptos relacionados con su cálculo. Específicamente, se busca definir dos conceptos para la discusión de resultados: la incertidumbre tipo A y B.

La incertidumbre de una medición se define como un parámetro asociado con el resultado de dicha medición, la cual caracteriza la dispersión de los valores que podría tener el mesurando [32], y a este parámetro se le puede relacionar un método de evaluación. Cuando el método de evaluación es por medio del análisis estadístico de una serie de observaciones se le llama evaluación tipo A, y la incertidumbre se

define como incertidumbre tipo A. Por otro lado, si el método es por medios distintos al análisis estadístico, se le determina incertidumbre tipo B [32].

La incertidumbre tipo B se calcula usando el juicio científico a partir de toda la información disponible sobre la variabilidad del mesurando. Esto puede ser datos de mediciones anteriores, especificaciones del fabricante, datos provistos y certificados, incertidumbres asignadas, entre otras cosas. Para evaluaciones de tipo A, la incertidumbre es determinada por los mismos datos sin requerir otros estimados, a través de la desviación estándar de la muestra [33].

La desviación estándar puede interpretarse como la cuantificación de la incertidumbre en un proceso, relacionada con la dispersión de una variable continua, para cualquier tipo de distribución [34]. Sin embargo, en distribuciones normales, la desviación estándar toma un significado más intuitivo, además de poseer propiedades particulares, como la simetría en la dispersión de los resultados con respecto a la media.

De acuerdo con el teorema central del límite¹⁰, bajo ciertas condiciones, la distribución de probabilidad de la suma de un número grande de variables aleatorias se aproxima a una distribución normal [34]. De esta forma, dado un conjunto de variables x_1, x_2, \dots, x_n , $n \in \mathbb{N}$, las cuales son independientes e idénticamente distribuidas con valor esperado μ y varianza σ^2 , a medida que n incrementa la distribución del promedio del conjunto de variables se aproxima a una normal con media μ y varianza σ^2 [35].

En el caso de este trabajo, dado que la incertidumbre causada por cada fuente de incertidumbre se considera hecha de forma aleatoria, se espera que la distribución de las medidas de una misma cantidad sea una distribución normal.

¹⁰Comúnmente conocido como Teorema del límite central

2.5.2. Propagación de incertidumbre

Considerando a las imágenes como mediciones, estas poseen una incertidumbre, la cual es heredada a los DR extraídos de ellas. Así, cada fase del flujo de trabajo contribuye a la incertidumbre total de cada DR. En este trabajo se consideraron sólo tres fuentes de incertidumbre, bajo la premisa de que éstas influirían más en la incertidumbre total. Estas son la incertidumbre asociada al mastógrafo σ_M , al procesamiento de imágenes σ_I , y a la delineación de la ROI σ_R .

Entonces, se define la desviación estándar acumulada para cada uno de los descriptores como σ_T tal que

$$\sigma_T^2 = \sigma_M^2 + \sigma_I^2 + \sigma_R^2, \quad (2.2)$$

donde la forma de obtener cada una de las desviaciones estándares es explicada en el capítulo 3. Así, para cada DR existe una media μ y una desviación estándar σ_T , con las cuales se define un intervalo de confianza $\mu \pm \sigma_T$, en el que el valor numérico de cada DR puede ser encontrado con una probabilidad del 68% si la medición es repetida bajo las mismas condiciones.

La ecuación 2.2 puede ser utilizada para cada DR y, de esta forma, se obtiene la distribución de incertidumbre, expresada como un presupuesto de incertidumbre¹¹ (PI) definido para cada DR como la incertidumbre acumulada total σ_T . Dado que la comparación directa de la σ_T entre descriptores diferentes no es factible, pues sus valores numéricos son variables y dependen del descriptor, se utiliza el Coeficiente de Variación (C_V) definido como

¹¹El presupuesto de incertidumbre se define como una representación en tabla de la distribución de incertidumbre, con el fin de compactar y comparar información de una o más variables [36] .

$$C_V = \frac{\sigma_T}{\mu}. \quad (2.3)$$

2.6. Diagnóstico Asistido por Computadora (CADx)

El Diagnóstico Asistido por Computadora (CADx) es una técnica de aprendizaje automático cuyo objetivo es usar información extraída de imágenes médicas, en forma de descriptores cuantitativos o radiómicos (DR), para determinar el estado fisiopatológico de una lesión [11, 4]. De esta manera, CADx puede ser empleado como una técnica para desarrollar un análisis radiómico. La subtécnica de CADx utilizada en esta investigación es RDH, la cual se distingue por el uso de descriptores con significado intuitivo para el observador, y basados en definiciones de cantidades físicas y matemáticas, como los descriptores de textura [14].

La forma en la que CADx determinará el estado fisiopatológico de las lesiones es a través de un modelo de predicción. De forma general, un modelo de predicción, en el caso del aprendizaje supervisado¹², es aquel que, a través de resultados conocidos, es entrenado para predecir resultados futuros [37].

Así, en el caso de un modelo sencillo, dados m ejemplos con resultados conocidos, de los cuales se posee cierta información $x \in \mathbb{R}^n$ de cada ejemplo, se puede entrenar el modelo para así definir una función $h_\theta : \mathbb{R}^n \rightarrow \mathbb{R}^l$, la cual es conocida como la función hipótesis. h_θ es una función que, dada la información $x^{(i)}$ del ejemplo i -ésimo, es capaz de predecir su resultado $y \in \mathbb{R}^l$. En el caso de una clasificación binaria, como es el caso de este trabajo, $l = 0$ y h e y son escalares binarios, 0 o 1 [37].

El objetivo del entrenamiento es encontrar aquella variable θ para la cual el mo-

¹²El aprendizaje supervisado se diferencia del no-supervisado en que en el primero se le indica al modelo qué es lo que se quiere predecir, mientras que en el segundo el algoritmo busca por sí solo patrones entre los datos y los asocia en grupos.

delo predice con el mejor desempeño posible. Existe gran variedad de algoritmos de entrenamiento, los cuales dependen de la forma en la que el modelo entrena y lo que se busque predecir. En la siguiente sección se explica el algoritmo de entrenamiento utilizado en este trabajo.

En el caso particular de este trabajo, suponiendo que se han extraído un número n de descriptores para un grupo de m imágenes, y que se posee la información patológica de las lesiones contenidas en ellas, se pueden desarrollar análisis exploratorios como fue definido en la sección 2.4.2.

Siguiendo un orden cronológico, el primer paso a seguir sería la eliminación de descriptores. En aprendizaje automático no existe una forma estandarizada para eliminar descriptores, y su uso depende de la situación en la que vayan a utilizarse. Dado que la técnica de eliminación usada en este trabajo es un caso particular de la regresión logística, se comenzará explicando esta última.

2.6.1. Regresión logística

La regresión logística es un algoritmo de aprendizaje supervisado diseñado para resolver problemas de clasificación, es decir, problemas en los que la variable y que se quiere predecir es una variable discreta [37].

En este caso, la variable y que se quiere predecir es la presencia de las proteínas del análisis inmunohistoquímico definidas en la sección 2.1.1, y la clasificación es una clasificación binaria, es decir $y \in \{0, 1\}$. Así, para cada una de las proteínas se va a definir un modelo de regresión logística diferente.

Suponiendo un número de imágenes m y un número de DR n , se define la matriz de entrada $X \in \mathcal{M}(\mathbb{R})_{m \times n}$ tal que $X_{i,j} = x_j^{(i)}$, donde $x_j^{(i)}$ es el valor del descriptor j -ésimo para la imagen del i -ésimo paciente. Entonces, el modelo de regresión logística predice

la variable de salida $y^{(i)}$ para el i -ésimo paciente siguiendo la función de hipótesis h tal que

$$h_{\theta}(x^{(i)}) = \frac{1}{1 + e^{-\theta^T x^{(i)}}}, \quad (2.4)$$

donde $\theta \in \mathbb{R}^n$ es un vector que se encuentra minimizando una función de costo J . En este caso, la función h también es conocida como la función sigmoide y es definida de tal forma que $h \in (0, 1)$.

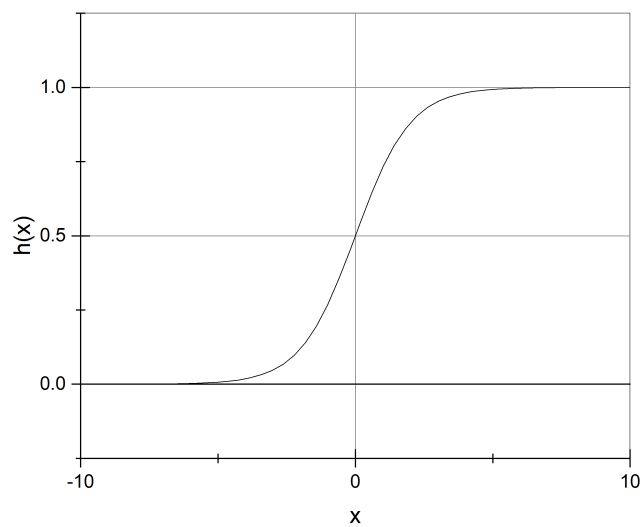


Figura 2.5: Función sigmoide definida como la función hipótesis en la regresión logística. Se muestra el caso $h(x) = \frac{1}{1+e^{-x}}$ para fines explicativos.

De esta manera, la regresión logística se reduce a un problema de minimización de la función de costo J definida como

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m y^{(i)} \log h_{\theta}(x^{(i)}) + (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)})). \quad (2.5)$$

La idea principal detrás de la función de costo J es que si la función de hipótesis

h predice correctamente a y , se le sumará a la función de costo un valor cercano a cero y, contrariamente, si se equivoca en la predicción, se le sumará a J un valor grande, castigando al modelo (ver figura 2.6). Así, la elección de θ será aquella que se equivoque menos en la predicción, teniendo el valor de $J(\theta)$ mínimo.

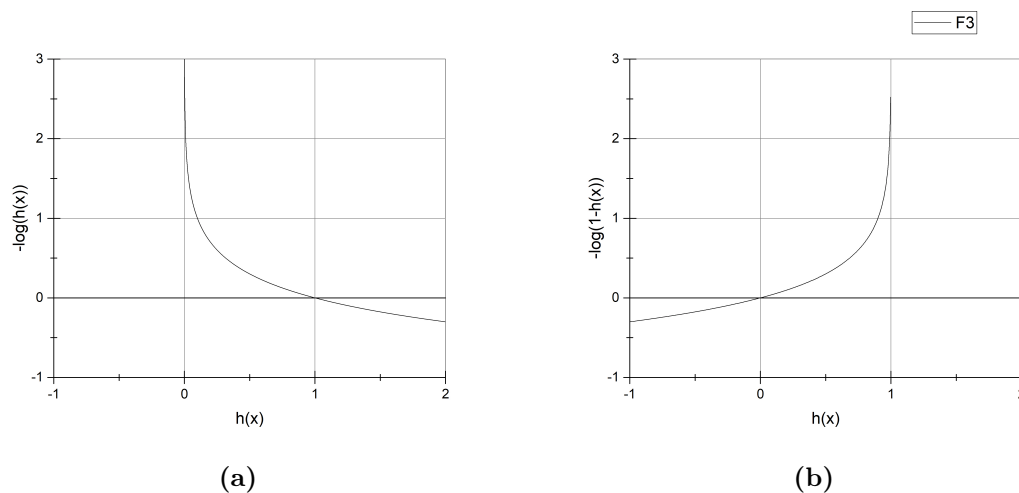


Figura 2.6: Función de costo J para predicción a) $y = 1$ y b) $y = 0$.

Existen otros algoritmos de aprendizaje que podrían ser utilizados en problemas de clasificación binaria. Entre ellos destacan dos, la máquina de vectores de soporte, (SVM, *Support Vector Machine*), y la red neuronal. Sin entrar en muchos detalles, SVM utiliza un algoritmo muy similar al descrito para regresión logística, cambiando sólo ligeramente en la definición de la función de costo J . Por otro lado, una red neuronal utiliza un algoritmo de aprendizaje mucho más complejo, el cual puede llegar a definir una gran cantidad de funciones de hipótesis, dependiendo de las características de los datos a predecir [37].

La decisión de cuál de los tres modelos utilizar se fundamenta, principalmente, en dos características de los datos disponibles: el número de muestras m y el número de descriptores n . Si n es grande con respecto a m , se recomienda utilizar un modelo de regresión logística. Si n es pequeño y m es intermedio, es decir, con una relación aproximada de $m \approx 10n$, se recomienda un modelo de SVM [37]. En el otro caso extremo, m muy grande con respecto a n , nuevamente se recomienda el uso de un modelo de regresión logística.

Por otro lado, una red neuronal podría ser adecuada para cualquiera de los casos mencionados arriba. Sin embargo, dada la complejidad del modelo, es altamente probable que el proceso de entrenamiento fuese más lento que el de los otros dos modelos [37].

2.6.2. Método LASSO para eliminación de descriptores

Cuando se tiene un número alto de descriptores n , especialmente cuando $n > m$, es probable que el modelo sufra de sobreajuste, un fenómeno común en los modelos de predicción, el cual sucede cuando el modelo predice correctamente a las imágenes dadas en el entrenamiento del modelo (las m imágenes) pero no es capaz de generalizar para otras imágenes fuera de este grupo [37].

Una primera solución a este problema es a través de la regularización, técnica que reduce la influencia de los descriptores con menor relevancia para el modelo, agregando un valor adicional a la función de costo J para cada θ de la forma

$$J_{L2}(\theta) = J(\theta) + \frac{\lambda}{2m} \sum_{j=1}^n \theta_j^2, \quad (2.6)$$

donde J_{L2} es la nueva función de costo regularizada y λ es una variable que define la penalización que se le dará a cada descriptor. Este tipo de regularización recibe el nombre de regularización L2 [37].

Si se quisiera utilizar regularización para eliminar descriptores, es necesario otorgar un mayor castigo a la función de costo para asegurar que $\theta_n = 0$ para los descriptores menos importantes. Esto se logra usando la regularización L1, también llamada LASSO, la cual difiere ligeramente de su contraparte L2 de la siguiente manera:

$$J_{L1}(\theta) = J(\theta) + \frac{\beta}{2m} \sum_{j=1}^n \theta_j, \quad (2.7)$$

donde en este caso β definirá el número de descriptores que sobrevivirán a la eliminación.

2.6.3. Asociación de descriptores: agrupamiento

Otra forma de reducir el número de descriptores, antes de realizar eliminación LASSO, es a través de la técnica de agrupamiento, o creación de grupos. Con esta técnica se juntan los descriptores que posean una relación entre ellos, definida por una métrica. Algunas métricas comúnmente utilizadas son el coeficiente de correlación de Pearson (CCP) o el de Spearman (CCS). Así, se pueden agrupar en un mismo grupo a todos los descriptores que posean entre sí un CCP tal que $CCP > \alpha$, siendo α definida por el usuario, $\alpha \in [0, 1]$

Una vez creados los grupos, se puede escoger un descriptor por cada grupo, siguiendo un criterio determinado, para eliminar la redundancia de descriptores introducidos en el modelo.

2.6.4. Riqueza de los descriptores radiómicos: PCA

Como se mencionó en la sección 2.6.2, para la eliminación LASSO, el número de descriptores seleccionados es elegido por el usuario. Para decidir este número se puede utilizar un algoritmo de aprendizaje automático llamado PCA (*Princial Components Analysis*) el cual permite conocer, entre otras cosas, la riqueza de los descriptores radiómicos de la muestra [37].

La técnica PCA busca una combinación lineal de los descriptores de entrada para definir nuevos descriptores, con el fin de reducir la dimensionalidad de la muestra al mismo tiempo que el porcentaje de la varianza de los nuevos descriptores con respecto a los viejos se mantiene alta. En otras palabras, se busca el mínimo número de nuevos descriptores que posean información similar que los descriptores originales.

El proceso seguido por PCA se basa en la descomposición en valores singulares. Si se tiene una muestra de m elementos, con r descriptores por cada elemento, y se busca reducir la dimensión a k , primero se debe normalizar a la muestra con respecto a su media, con el fin de que todos los descriptores tengan el mismo peso. Posteriormente, se define la matriz de covarianza Σ , tal que

$$\Sigma = \frac{1}{m} \sum_{i=1}^m (x^{(i)})(x^{(i)})^T, \quad (2.8)$$

donde el vector columna $x^{(i)} \in \mathbb{R}^n$ es el i -ésimo elemento de la muestra original. Nótese que los valores propios de la matriz Σ son siempre iguales o mayores que cero, por lo que se pueden calcular los valores propios de Σ y se pueden ordenar de mayor a menor.

Finalmente, dado que para cada valor propio está definido un vector propio, los primeros k vectores propios serán elegidos y agrupados en la matriz $U \in \mathbb{R}_{n \times k}$. De esta manera, los nuevos elementos z podrán ser definidos usando la matriz U de la siguiente forma,

$$z^{(i)} = U^T x^{(i)}, \quad (2.9)$$

para $z^{(i)}$ el i -ésimo nuevo elemento.

Para conocer el número de elementos k ideal, se realiza una prueba del cambio de varianza en la muestra. Primero, se reconstruye la información de los elementos

originales $x^{(i)}$ usando

$$x_{app} = Uz^{(i)}, \quad (2.10)$$

con x_{app} la reconstrucción del elemento original. Entonces, para asegurar una retención de varianza de valor $v \in (0, 1)$, se debe encontrar un k tal que

$$\frac{\sum_{i=1}^m \|x^{(i)} - x_{app}^{(i)}\|^2}{\sum_{i=1}^m \|x^{(i)}\|^2} \leq 1 - v. \quad (2.11)$$

Así, se puede utilizar la información otorgada por PCA sobre la riqueza de la muestra para decidir el número de descriptores restantes después de la eliminación LASSO.

2.6.5. Evaluación de los modelos de predicción

Independientemente del algoritmo de aprendizaje utilizado, que en este caso es regresión logística, el paso final fundamental para validar resultados es la evaluación. Cuando el tamaño de la muestra es suficientemente grande ($m > 100$), el procedimiento típico es separar los m datos de entrada en dos grupos: grupo de entrenamiento con el 70% de la muestra y grupo de prueba con el 30% de la muestra. Esto para asegurar una correcta generalización del modelo.

En el caso de muestras de tamaño reducido, como el caso de este trabajo, una técnica de evaluación sugerida es la técnica *Leave-One-Out*. Bajo esta técnica, se crean m modelos de predicción, uno por cada elemento de la muestra. Así, para el modelo k -ésimo, la matriz de entrada $X_k \in \mathcal{M}(\mathbb{R}_{(m-1) \times n})$ es idéntica a la matriz X salvo que se suprimió el renglón k . La información de la imagen de dicho renglón será

utilizada como prueba para ese modelo k -ésimo [38]. De esta forma, para cada uno de los m modelos, el renglón excluido funge como dato de prueba.

2.6.6. Indicadores de desempeño

Los indicadores de desempeño de un modelo de predicción de clasificación han variado en los últimos años. En general, todos los indicadores son definidos a partir de la matriz de confusión, también llamada matriz de decisión [3]. La matriz de confusión $C \in \mathcal{M}_{2 \times 2}(\mathbb{N})$, se define como

$$C = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad (2.12)$$

donde $a = TP$, $b = FN$, $c = FP$ y $d = TN$. El elemento TP (*true positive*) representa el número de casos predichos como verdaderos y que realmente lo eran, mientras que TN (*true negative*) es el número de casos predichos como negativos que efectivamente lo eran. Los elementos fuera de la diagonal representan los errores de predicción. Así, FN (*false negative*) es el número de casos erróneamente clasificados como negativos, y FP (*false positive*) lo análogo para los clasificados como positivos.

Los indicadores, o métricas, de desempeño más reconocidos y utilizados por la comunidad científica son la precisión y la exactitud, por ser conceptos con significado intuitivo para el lector. Ejemplos de su uso se encuentran en los trabajos de Mateos [15], Marino et al.[39] y Zhou et al.[23]. Sin embargo, existen ciertas desventajas con estos medidores, tales como la falta de confianza hacia los mismos cuando la muestra está desbalanceada, pues los medidores podrían dar resultados buenos con un modelo pobre, llamados modelos *sobreoptimistas* [40].

Otros investigadores consideran que la mejor métrica de desempeño es el valor

F1, o *F1 score*, como explican Chicco et al. [40]. Los mismos autores argumentan, sin embargo, que la definición del valor F1 es independiente de las muestras correctamente clasificadas como negativas, expresadas en la matriz de confusión como TN. Esto significa que, por ejemplo, para una muestra con pocos valores de predicción negativos, un modelo que prediga siempre que el resultado será positivo daría un valor F1 engañosamente alto.

Con el fin de resolver este problema, Chicco et al. [40] y Delgado et al. [41] sugieren el uso del Coeficiente de Correlación de Matthews (MCC). Para su definición, y suponiendo una matriz de confusión C , se define primero las variables α y β como

$$\alpha = \frac{ad - bc}{(a + b)(b + d)} \quad (2.13)$$

$$\beta = \frac{ad - bc}{(a + c)(c + d)}. \quad (2.14)$$

De estas dos variables, MCC es definido como la media geométrica de α y β , es decir,

$$MCC = \sqrt{\alpha\beta}. \quad (2.15)$$

La definición de MCC incluye los 4 elementos de la matriz de confusión lo que la hace una métrica de evaluación de desempeño más completa.

Se menciona que existe una métrica hermana del MCC, llamada la Kappa de Cohen, definida como la media armónica de α y β , es decir,

$$\kappa_C = \frac{2}{\frac{1}{\alpha} + \frac{1}{\beta}}. \quad (2.16)$$

Interpretación del MCC

A diferencia de las 3 métricas de evaluación de desempeño tradicionales, a saber, exactitud, valor F1 y precisión, el MCC es una medición poco intuitiva, dada su compleja definición dada por la ecuación 2.15. La siguiente tabla muestra una sugerencia de interpretación propuesta por Jacob Cohen [42] para su métrica de desempeño, la Kappa de Cohen, la cual puede ser extrapolada al MCC debido a su parentesco.

Tabla 2.1: Interpretación del MCC

Valor de MCC	Nivel de acuerdo
$MCC < 0$	Ninguno
0.01-0.20	Ligero
0.21 - 0.40	Suficiente
0.41-0.60	Moderado
0.61-0.80	Substancial
0.81-1.00	Casi perfecto

Se debe enfatizar que, dado que el MCC está definido en un rango de $[-1, 1]$, los intervalos de niveles de acuerdo son diferentes a los de las métricas tradicionales, definidas en un rango de $[0, 1]$. Se considera que existe un acuerdo en el sistema de predicción si el valor de MCC es positivo.

Esto contrasta con las métricas tradicionales pues, si se cuenta con una población balanceada¹³, la existencia de un acuerdo comienza a considerarse para valores mayores al 50%. Para las métricas tradicionales un desempeño de 50% significa una predicción azarosa, mientras que para el MCC esto se representa con 0%.

Se menciona que el MCC es sensible a resultados sobreoptimistas en muestras desbalanceadas. Tómese como ejemplo una población altamente desbalanceada con 100 muestras, donde todas están marcadas como positivas excepto una. Si se utiliza un modelo que siempre predice que el resultado será positivo, sin importar la información que se le dé, se obtendrá un $F1_{score} = 99\%$, aun cuando el modelo no predice

¹³Una población balanceada es aquella con 50% de casos positivos y 50% negativos.

correctamente. Por otro lado, dada su definición, el MCC no está definido, lo que evidencia el nulo poder de predicción. Algunos autores [40] consideran que en el caso de MCC no definido, es decir, cuando algún renglón o columna de la matriz C es cero, es conveniente definir el MCC como 0, para mostrar numéricamente la ausencia de acuerdo.

3

Método experimental

3.1. Imágenes clínicas

3.1.1. Adquisición de imágenes

Un conjunto de 45 estudios de pacientes fue considerado, consistiendo en la adquisición de mamografías contrastada modalidad SET y una biopsia, posterior a las mamografías, de la lesión sospechosa detectada. Estos estudios eran parte de un protocolo de investigación realizado en el INCan en población mexicana para evaluar la detección de cáncer de mama multicéntrico. El estudio fue aprobado por el comité de ética del INCan con el código de aprobación (016/064/RTI) (CEI/1137/16), y el título *Detección de lesiones secundarias de mama usando mastografía con medio de contraste*. En el estudio se incluyeron pacientes con clasificación BI-RADS 4 o 5, y con sospecha de multicentricidad o multifocalidad determinada a partir de la mastografía. La edad de las pacientes varió entre los 36 y los 66 años, con una mediana de 50 años.

De estos 45 estudios, 4 fueron excluidos debido a que la biopsia no mostró eviden-

cia de células neoplásicas, y 8 debido a artefactos en las imágenes que impedían una correcta visualización. De los 33 estudios restantes, los resultados inmunohistopatológicos se distribuyeron de la siguiente forma: 82 % ER positivo, 64 % PR positivo, 12 % HER2 positivo y 36 % con índice de proliferación alto ($Ki67 > 0.2$).

El protocolo de obtención de mamografías modalidad SET del INCa consistió en la obtención de una imagen basal, es decir, sin la inyección del MC, y otras 4 con MC inyectado. Se adquirió una mamografía de la mama derecha de 9 pacientes y de la mama izquierda de 22, todas con proyección craneocaudal. El MC aplicado fue el descrito en la sección 2.3 y la dosis administrada fue 1.5 ml/kg sin superar los 100 ml en ninguna paciente. Posteriormente, 4 imágenes con un minuto de diferencia entre ellas fueron adquiridas. Todas las imágenes se obtuvieron en un mastógrafo *Fujifilm Amulet Innovality* (46420171, Japón), usando un ánodo de tungsteno (W), un filtro de rodio (Rh), 45 kV de voltaje en el tubo de rayos X, y un filtro adicional de 5 mm de aluminio (Al). Las imágenes fueron guardadas en formato DICOM *sin procesamiento*, con un tamaño de píxel de 100 μm . Existe un preprocesamiento intrínseco del software del mastógrafo, el cual incluye una función de respuesta logarítmica y una corrección de campo plano.

3.1.2. Procesamiento de imágenes

Como se mencionó en el capítulo pasado, las imágenes de modalidad SET de este trabajo deben pasar por el proceso de registro y substracción. Además, dado que el procedimiento de adquisición de imagen toma unos minutos, es muy difícil evitar el movimiento de la paciente. La corrección de este problema se hace a través del registro de la imagen.

Se menciona además que, de los 4 tiempos poscontraste disponibles, se eligieron las imágenes poscontraste a 1 minuto del suministro del MC, dado que estas presentaron

menor presencia de artefactos por movimiento.

Registro de imagen

El registro de los pares de imágenes pre- y poscontraste se realizó bajo la aplicación de una transformación afín sobre la imagen basal para alinearla con las imágenes poscontraste. Dicha transformación afín se obtuvo para cada paciente y para cada una de las imágenes poscontraste, pues el movimiento realizado a la imagen basal para cada una fue diferente. La transformación afín se realizó en *Mathlab R2020a* usando como métrica la suma de diferencias cuadradas (SSD) para alineación de imágenes.

Substracción de imagen

Una vez aplicada la transformación afín, se efectuó una resta píxel a píxel de los pares de imágenes pre- y poscontraste. Se menciona que, debido a que el preprocesamiento de imágenes del mastógrafo incluye una función de respuesta logarítmica, la resta de imágenes se efectúa directamente sin la necesidad de una transformación logarítmica, como debe hacerse para imágenes obtenidas en otros modelos de mastógrafos [15]. Dadas dos imágenes Im_{SMC} e Im_{CMC} , imágenes sin y con MC respectivamente, la resta Im_{SET} está dada por

$$Im_{SET} = Im_{SMC} - Im_{CMC}. \quad (3.1)$$

3.1.3. Segmentación

Tres médicas radiólogas adscritas al INCan, con 10, 3 y 1 año de experiencia respectivamente, segmentaron las lesiones para las cuales se realizó una biopsia. La segmentación fue trazada sobre la imagen SET de 1 minuto. Para la segmentación de las lesiones, se definieron dos comandos de trazado:

1. Segmentar la lesión alrededor de su borde, siguiendo el contorno de la forma (Segmentación general)
2. Segmentar una región focal en donde exista una alta certidumbre de incluir únicamente lesión (Segmentación focal)

Con algunas excepciones, la segmentación general incluía a la focal en su interior. Cada radióloga repitió el trazado de las segmentaciones con una diferencia temporal de alrededor 8 semanas, con el fin de evaluar el acuerdo intraobservador. La segmentación fue realizada en el software de análisis de imágenes *3D Slicer* (4.11.0, Estados Unidos) [43, 44, 45], usando sus herramientas de segmentación y visualización.

De esta forma, para cada imagen se obtuvieron 4 segmentaciones por radióloga: dos segmentaciones generales y dos focales, dando un total de 396 segmentaciones. En la figura 3.1 se puede observar un ejemplo de la segmentación general y focal para dos radiólogas.

3.2. Incertidumbre de imágenes CEDM

Para evaluar dos de las tres fuentes de incertidumbre consideradas en este trabajo, las producidas por el mastógrafo y el registro de imágenes, se realizaron adquisiciones de imágenes usando un maniquí *CIRS phantom model 020* (1138.1-2-2, Estados

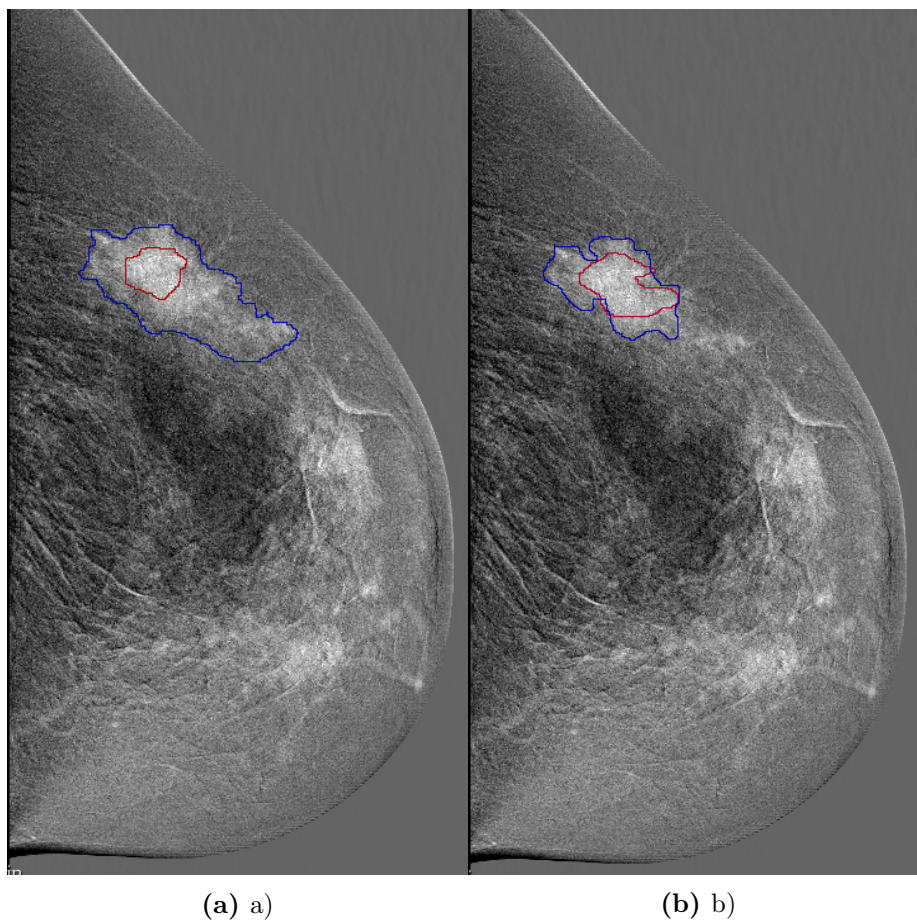


Figura 3.1: Segmentación general (azul) y focal (roja) realizada en una misma paciente por radiólogas con a) 1 año de experiencia b) 10 años de experiencia.

Unidos), el cual cuenta con una mezcla heterogénea de materiales equivalentes a tejido glandular y adiposo, diseñados para representar el tejido mamario. Para simular las imágenes precontraste y poscontraste, una resina yodada con 0.0 mg/cm^2 y 3.0 mg/cm^2 , respectivamente, fue colocada sobre el maniquí. Todas las imágenes fueron obtenidas en el mismo mastógrafo usado en los estudios clínicos previamente mencionados.

3.2.1. Incertidumbre asociada al mastógrafo

La evaluación de incertidumbre debida al mastógrafo se realizó adquiriendo 10 pares de imágenes pre- y poscontraste. Con el fin de evitar otras fuentes de incerti-

dumbre, como artefactos debidos a la resta de las imágenes, cada par de imágenes fue obtenido manteniendo la posición del maniquí y la resina inalterada.

Una posición diferente del maniquí sobre el detector del mastógrafo fue elegida para cada uno de los 10 pares, con el fin de incluir en la incertidumbre del mastógrafo la variabilidad del DR debido a ligeros cambios de posición alrededor de una posición definida sobre el maniquí.

Las imágenes resultantes fueron guardadas en formato DICOM *para procesar* y restadas usando la herramienta de análisis de imágenes *ImageJ* (1.52a, *National Institutes of Health*, Estados Unidos), para ser guardadas en formato TIFF. Respecto a la segmentación de estas regiones, se eligió una ROI rectangular que se mantuviese dentro de la región de la resina de yodo para los 10 pares de imágenes, evitando así variabilidad de los DR debido a la aparición de los bordes de la resina. La segmentación también se realizó usando la herramienta 3D Slicer.

En la figura 3.2 a) se muestra un ejemplo de resta de un par de imágenes pre- y poscontraste, incluyendo su segmentación. Debido a que la resta se realizó sin movimientos, la forma del maniquí es prácticamente imperceptible. Nótese, sin embargo, que es posible visualizar la sombra de la forma del maniquí en la misma figura. Esto es debido al cambio en la textura del ruido entre ambas zonas, a pesar de que tienen el mismo valor promedio de la señal. La fluencia de energía de los rayos X que llegan al detector tiene relación con la distribución de ruido, ya que a mayor fluencia menor ruido relativo. Debido al maniquí, la fluencia de energía es menor en esta zona por lo que es posible percibir la diferencia.

3.2.2. Incertidumbre asociada al registro de imágenes

Para aislar la incertidumbre debida a la sustracción de imágenes, se adquirieron 10 grupos de imágenes. Cada grupo estuvo compuesto por 5 imágenes precontraste,

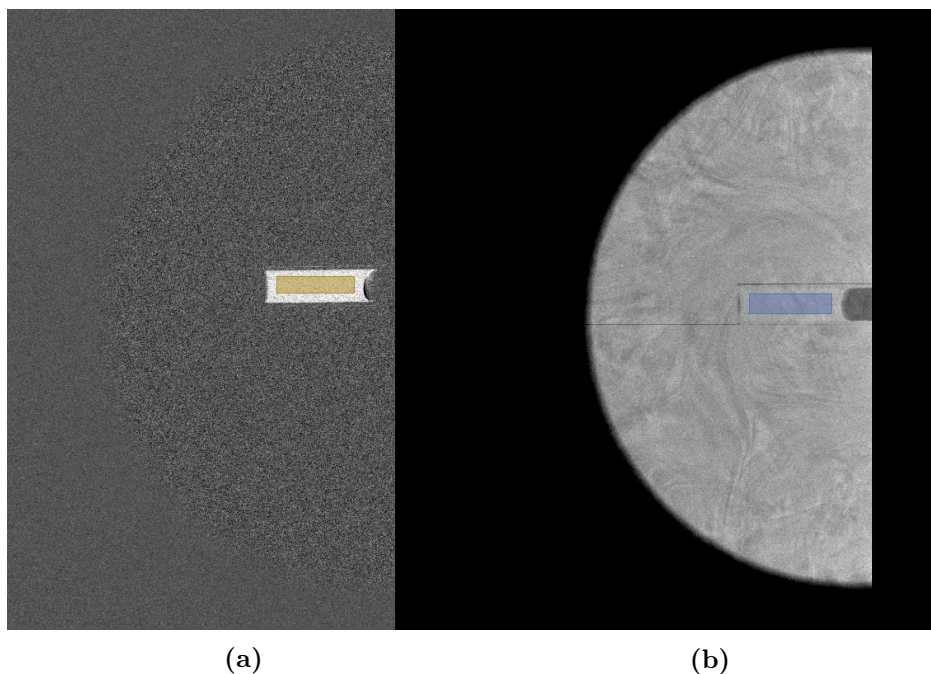


Figura 3.2: Resta de dos pares de imágenes pre- y poscontraste del maniquí cuando las imágenes están a) alineadas y b) desalineadas.

sin modificación entre ellas, y 5 imágenes poscontraste, también sin cambio entre ellas. La diferencia radicó en que después de obtener las 5 imágenes precontraste, el maniquí fue cambiado de posición sobre el mastógrafo, y posteriormente las imágenes poscontraste fueron adquiridas. Esto se repitió con los 10 grupos, obteniendo un total de 100 imágenes pre- y poscontraste.

Cada grupo de imágenes está caracterizado por poseer un sesgo del valor de los DR, definiendo sesgo como el error sistemático producido por el desfase de las imágenes basal y la contrastada. Siguiendo un método similar al de la subsección anterior, para cada grupo de imágenes se colocó el maniquí en una posición diferente alrededor de una posición definida sobre el detector.

Es importante mencionar que el comportamiento del sesgo va a depender del DR. Un ejemplo de dicho comportamiento es mostrado en la figura 3.3 para el descriptor de textura *contraste*. Para estos resultados, el desfase se realizó aplicando una traslación sobre la imagen poscontraste, con desfases en cada coordenada Δx y Δy tales que

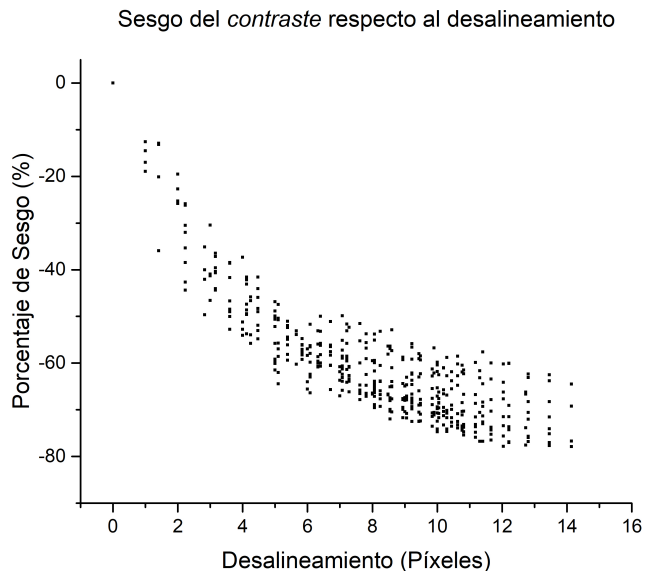


Figura 3.3: Porcentaje de sesgo del descriptor *contraste* producto del desalineamiento r de las imágenes pre- y poscontraste.

$\Delta x, \Delta y \in [-10, 10] \subset \mathbb{Z}$. En el eje de las abscisas de la figura 3.3 se marca el desfase en píxeles r entre las imágenes en función del radio de desfase $r^2 = \Delta x^2 + \Delta y^2$. El porcentaje de sesgo negativo indica una disminución del valor absoluto del contraste respecto al valor obtenido en una resta de imágenes perfecta, es decir, las imágenes sin desfase.

Una vez obtenidos los 10 grupos de imágenes, se restaron todas las combinaciones de imágenes posibles de cada grupo. Es decir, para cada grupo se obtuvieron 25 imágenes resta diferentes pertenecientes, en principio, al mismo desalineamiento de la imagen. Debido a que existen 250 combinaciones de restas posibles, se creó un programa en lenguaje *Python* sobre el entorno informático *Jupyter Notebooks* para realizar las restas. En dicho programa, se realiza la resta píxel a píxel de las imágenes y se guardan las imágenes en formato DICOM. La segmentación para estos 25 pares de imágenes siguió el mismo criterio de trazado que el de la subsección anterior.

Un ejemplo de un par de estas 250 imágenes es mostrado en la figura 3.2 b). Contrasta la diferencia con el caso de resta perfecta debido a que, en este caso, la

forma del maniquí es perceptible a simple vista, muestra del desalineamiento entre las imágenes.

3.3. Extracción de descriptores radiómicos

En la fase 3 del flujo de trabajo radiómico, presentado en la sección 2.4.2, se menciona la importancia de definir el método de extracción de los DR, con el fin de asegurar la repetibilidad y reproducibilidad de los descriptores, tal como recomienda la red de investigadores *QIN*. Esta última promueve los proyectos para la estandarización de la extracción de DR, como el realizado por investigadores de la Escuela de Medicina de Harvard y la Universidad de Maastricht. Este proyecto incluye el desarrollo de *Pyradiomics* [29], una plataforma de código libre capaz de extraer diversos descriptores radiómicos de imágenes médicas de varias modalidades.

3.3.1. Pyradiomics

Si bien los diseñadores de *Pyradiomics* la categorizan como una plataforma, la misma puede ser vista como una paquetería basada en el lenguaje de programación *Python*, que permite el procesamiento y la extracción de DR a partir de imágenes médicas. La plataforma provee dos maneras diferentes de usar la herramienta:

1. A través de un ambiente virtual de Python, como *Jupyter Notebook*. En esta opción, el usuario ingresa los comandos de la librería directamente a la línea de comando.
2. A través de 3D Slicer (previamente mencionada) la cual utiliza la biblioteca de *Pyradiomics* como una extensión llamada Slicer Radiomics.

Se utilizaron ambas modalidades en este trabajo. La modalidad de 3D Slicer se usó para grupos de imágenes pequeños como las pertenecientes al análisis de incertidumbre debida al mastógrafo. Para la extracción de imágenes en imágenes de pacientes y para incertidumbre debida a la resta de imágenes, se utilizó la modalidad de comandos directos en un ambiente de *Python*, ya que la cantidad de imágenes y segmentaciones a procesar fue considerablemente grande.

Parámetros de extracción

Dado que Pyradiomics ofrece la posibilidad de personalizar los parámetros de extracción de los DR, se debe definir un documento formato JSON llamado *Parameter file* en el que se incluya la definición de dichos parámetros. Los parámetros se eligieron siguiendo las recomendaciones hechas por Mateos [15], los cuales se describen a continuación:

1. Tipo imagen: Original (sin filtros)
2. Tipo y nombre de parámetros extraídos (determinados en el siguiente apartado)
3. Número de *Bins*: 64 (Se recuantiza a 64 niveles de gris)
4. Imagen 2D: activado (La extracción es de una imagen 2D)
5. Corrección de máscara: activado (Debe activarse para que la segmentación pueda ser compatible con la imagen sin importar su tamaño)
6. Distancia d : 3 (Distancia entre vecinos para la GLCM)
7. GLCM se define como: no simétrica

3.3.2. Descriptores extraídos

Existen estudios en los cuales se han encontrado DR sensibles a malignidad y benignidad, como en el trabajo de Mateos [15], o sensibles a la presencia del antígeno HER2, como el trabajo de Zhou et al. [23]. Específicamente para la modalidad CEDM, Marino et al. [39] realizaron un análisis radiómico para encontrar correlaciones con los receptores hormonales del análisis IHQ y el carácter invasivo del cáncer.

Sin embargo, con el fin de asegurar el máximo aprovechamiento de los descriptores disponibles en *Pyradiomics*, se decidió extraer todos los descriptores definidos por Pyradiomics para imágenes en 2D, los cuales son 107 DR. Éstos se clasifican en tres grupos: descriptores de forma, de primero, y de segundo orden.

La definición matemática de cada uno de estos descriptores, al igual que una posible interpretación de su significado se pueden encontrar en la documentación de *Pyradiomics*, disponible en su sitio web principal [29].

3.4. Distribución y presupuesto de incertidumbre

La distribución de incertidumbre, y por lo tanto el presupuesto de incertidumbre (PI), es definido a partir de σ_T , como se explicó en la sección 2.5.2. A continuación, se describe la forma en la que éste fue obtenido.

Incertidumbre asociada al mastógrafo

Para los 10 pares de imágenes descritos en la sección 3.2, correspondientes a la incertidumbre debida al mastógrafo, se utilizaron las tablas de descriptores extraídos y se obtuvo la desviación estándar σ_M y la media μ_M para cada DR a partir de estas

10 imágenes.

Incertidumbre asociada al registro

Para las imágenes de incertidumbre debida al registro, se hizo un procedimiento similar al anterior con los descriptores extraídos de las 250 imágenes restadas, obteniendo σ_A y su media μ_A . Dado que estos pares de imágenes contienen intrínsecamente la incertidumbre del mastógrafo σ_M y la asociada al registro de imágenes σ_I , siguiendo el modelo descrito por la ecuación 2.2, se cumple que

$$\sigma_A^2 = \sigma_M^2 + \sigma_I^2. \quad (3.2)$$

Esta última ecuación permite el desacoplamiento de la incertidumbre debida al mastógrafo y la debida al registro.

Incertidumbre asociada a la ROI

Respecto a las segmentaciones de las lesiones, se decidió utilizar solamente los resultados obtenidos de las segmentaciones generales, excluyendo a las focales. Las razones de esta decisión son discutidas en la sección 4. Se consideró la combinación de los resultados obtenidos por cada radióloga, obteniendo así el valor medio μ_R , y la desviación estándar de dicha muestra σ_R . Dicho de otra forma, se promediaron los resultados para las 6 segmentaciones generales disponibles y se calculó su desviación estándar. De esta forma, la incertidumbre combina la variabilidad interradiólogo e intraradiólogo.

3.5. CADx: Flujo de trabajo

Para comenzar el flujo de trabajo de la técnica CADx se consideró la matriz de entrada $X \in \mathcal{M}(\mathbb{R})_{m \times n}$, con $m = 33$ y $n = 107$, la clasificación binaria $y \in \{0, 1\}$ de cada antígeno, y los valores del PI previamente definido.

Se recuerda que el vector y contiene la información de la presencia de los antígenos del análisis IHQ para cada lesión, y que los elementos de la matriz X se definen como el promedio de los DR extraídos de las 6 segmentaciones generales, para las 33 pacientes.

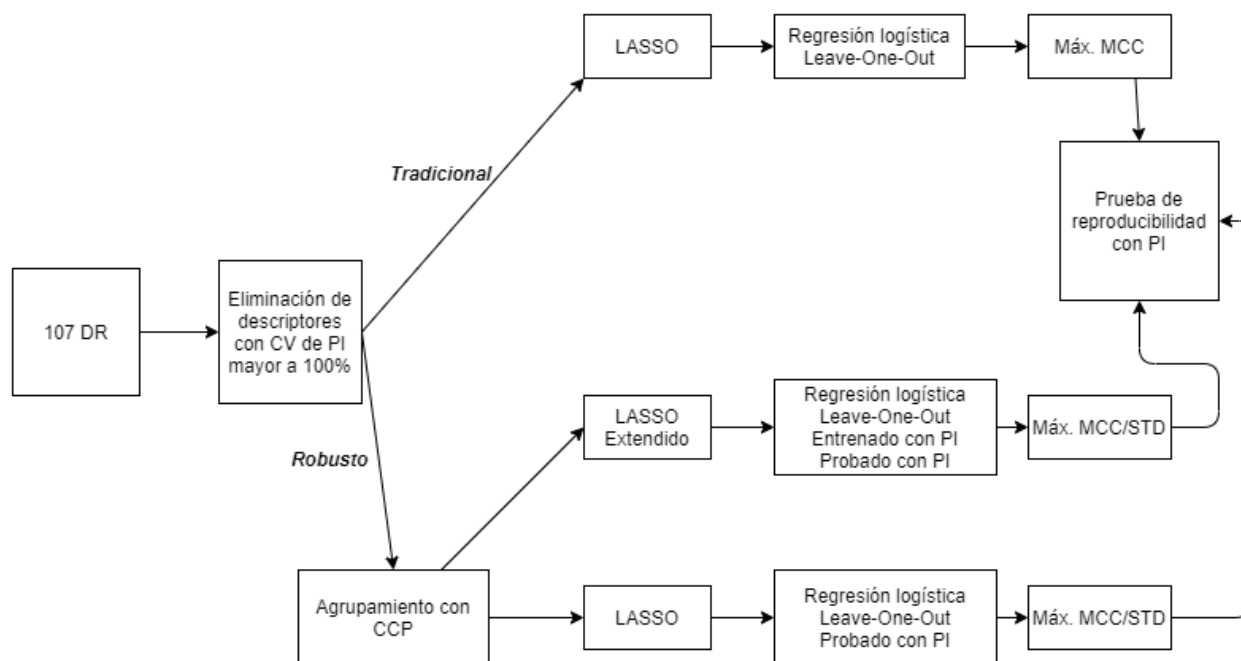


Figura 3.4: Flujo de trabajo del CADx propuesto

El flujo de trabajo propuesto en esta investigación se muestra en la figura 3.4. Se enfatiza la existencia de tres metodologías diferentes para la creación del modelo de predicción, separadas en ramas discutidas a continuación.

El primer paso del flujo, común para las 3 ramas después de la extracción de DR, fue la eliminación selectiva de descriptores a partir de su PI. El criterio fue eliminar a los DR con σ_T mayor al 100%. Esto se hizo para excluir del proceso a aquellos

descriptores incompatibles con el análisis de incertidumbre propuesto. Un ejemplo de estos descriptores incompatibles es el descriptor 35 *cluster shade* de la matriz GLCM, cuya variabilidad era tal que asemejaba a la de una distribución aleatoria (ver apéndice A). En este proceso se eliminaron 8 descriptores.

3.5.1. Análisis tradicional

En el flujo de la figura 3.4 se presenta la primera rama de análisis como análisis tradicional, pues en éste no se toma ninguna consideración con respecto a la incertidumbre.

La segunda eliminación consistió en la eliminación LASSO para reducir el número de DR a 9, de acuerdo a los resultados dados por el análisis PCA realizado sobre la muestra en MATLAB.

Posteriormente, se ejecutó una regresión logística bajo la técnica de evaluación de desempeño Leave-One-Out (LOO) para todas las combinaciones posibles de los 9 DR selectos, es decir, se ejecutaron 511 modelos diferentes de regresión logística LOO. De estos modelos, se eligió el que tuviese el mejor desempeño en función del MCC.

3.5.2. Análisis robusto: uso de DR robustos

La segunda rama de análisis entra en el grupo de análisis robusto puesto que en este se considera la incertidumbre para la elección del modelo de predicción con mejor desempeño.

El primer paso de esta segunda rama es la creación de grupos para agrupar a los descriptores con información correlacionada. Usando como métrica de correlación el CCP de pares de DR con valor $CCP > 0.99$, se crearon 51 grupos. De cada grupo se

eligió como representante al DR con C_V del PI más bajo. Los 51 descriptores selectos fueron introducidos al algoritmo de eliminación LASSO, reduciendo su número a 9.

Posteriormente, se crearon regresiones logísticas para las 511 combinaciones, con una variación a la técnica de evaluación LOO. En este caso, con el fin de tomar en cuenta la distribución de incertidumbre en la elección del modelo, se generaron valores nuevos para cada muestra de prueba de la siguiente forma. Se crearon 1000 valores nuevos de los DR seleccionados siguiendo una distribución normal, donde la media fue el valor del DR de la prueba y la desviación estándar fue el valor de σ_T para ese DR. *Id est*, se crearon 1000 nuevos valores aleatorios con distribución gaussiana a partir del valor original de la muestra de prueba.

Con estos nuevos 1001 datos de prueba, definidos para cada uno de los m modelos, se llevó a cabo la evaluación LOO y se obtuvo un valor medio de MCC μ_{MCC} y una desviación estándar σ_{MCC} . Se seleccionó como el modelo con mejor desempeño a aquel con el cociente $\frac{\mu_{MCC}}{\sigma_{MCC}}$ más alto.

3.5.3. Análisis robusto: con entrenamiento basado en el pre-supuesto

Se llevó a cabo una tercera rama de análisis, diferenciándose de la anterior en el entrenamiento del modelo. En esta rama, se agregan 50 nuevos datos por cada paciente a la matriz de entrada X , donde cada nuevo dato es generado a partir del PI y el valor original de la paciente, siguiendo una distribución gaussiana de manera similar a lo descrito en la sección pasada.

La nueva matriz de entrada Z , de dimensiones extendidas $m = 1638$ y $n = 51$, pasó por la eliminación LASSO, denominada eliminación LASSO extendida para diferenciarla de la segunda rama, y se determinaron 9 DR diferentes.

Aunque para la regresión logística se consideró la matriz aumentada Z , la evaluación de modelo LOO se realizó sólo para los 33 valores originales de las pacientes. De igual forma esto fue realizado para las 511 combinaciones posibles y se eligió a la combinación con el mejor desempeño medido a través del cociente $\frac{\mu_{MCC}}{\sigma_{MCC}}$.

4

Resultados y discusión

4.1. Presupuesto de incertidumbre

La tabla 4.1 presenta el presupuesto de incertidumbres para los 24 DR utilizados en los modelos de predicción de los 4 antígenos, en las 3 ramas del flujo de trabajo. La distribución de incertidumbre para las tres fuentes de incertidumbre consideradas en este trabajo se expresa en función del C_V , como fue descrito en la sección 2.5.2. Se incluye además la incertidumbre total propagada. La séptima columna indica el tipo de antígeno predicho por el modelo al que pertenece cada DR.

Se preserva la numeración y el nombre en inglés original de los 107 DR con el fin de mantener un orden y facilitar la localización de los DR a través de los distintos resultados. En la sección A del apéndice puede encontrarse el PI para los 107 descriptores. Los DR se agrupan de la siguiente forma a partir de su número de referencia:

1. 1-15 son DR de forma
2. 16-32 son DR de 1er orden

3. 34-107 son DR de 2do orden

Tabla 4.1: Presupuesto de incertidumbre de los descriptores usados en los modelos de predicción en términos del C_V . Para facilitar la comparación con otros trabajos, se da el nombre del DR en inglés.

	Descriptor radiómico	Mastógrafo	Registro	ROI	Total	IHQ
1	Maximum 2D Diameter Column	0.00 %	0.00 %	20.25 %	20.25 %	HER2
2	Maximum 2D Diameter Row	0.00 %	0.00 %	20.55 %	20.55 %	ER
7	Elongation	0.00 %	0.00 %	12.68 %	12.68 %	HER2
11	Minor Axis Length	0.00 %	0.00 %	18.73 %	18.73 %	ER
15	2D Sphericity	0.00 %	0.00 %	7.85 %	7.85 %	ER
27	Range	6.71 %	15.84 %	7.19 %	18.64 %	PR
30	Skewness	3.27 %	10.35 %	8.89 %	14.04 %	ER, PR, HER2, Ki67
34	Cluster Prominence	16.64 %	26.92 %	21.21 %	38.10 %	Ki67
38	Correlation	0.41 %	12.20 %	4.08 %	12.87 %	ER
47	Imc2	1.49 %	8.24 %	4.32 %	9.42 %	ER, PR, HER2, Ki67
52	MCC	2.13 %	11.80 %	4.78 %	12.91 %	PR
55	Sum Entropy	1.10 %	1.36 %	1.34 %	2.20 %	Ki67
57	Gray Level Non Uniformity (GLSZM)	3.53 %	4.33 %	39.92 %	40.31 %	PR, HER2
63	Large Area Low Gray Level Emphasis (GLSZM)	0.03 %	0.09 %	0.04 %	0.10 %	HER2
68	Small Area High Gray Level Emphasis (GLSZM)	7.23 %	10.10 %	7.91 %	14.72 %	HER2
70	Zone Entropy	0.69 %	1.98 %	0.95 %	2.30 %	Ki67
79	Long Run Low Gray Level Emphasis (GLRLM)	0.02 %	0.03 %	0.02 %	0.04 %	PR
81	Run Entropy	1.04 %	1.10 %	1.16 %	1.91 %	PR
89	Busyness	6.70 %	8.49 %	37.75 %	39.27 %	ER
91	Complexity	7.51 %	13.84 %	7.70 %	17.53 %	PR
92	Contrast (NGTDM)	2.41 %	1.10 %	1.81 %	3.21 %	ER, HER2
95	Dependence Non Uniformity (GLDM)	3.07 %	9.01 %	35.46 %	36.71 %	PR
98	Gray Level Non Uniformity (GLDM)	4.95 %	3.51 %	41.24 %	41.69 %	HER2
106	Small Dependence High Gray Level Emphasis (GLDM)	8.22 %	14.67 %	7.35 %	18.35 %	HER2

Los primeros 5 descriptores de la lista son descriptores de *forma* y no presentan incertidumbre asociada al mastógrafo y al registro dado que el trazado de la ROI fue manual y fue la misma para todos los pares de imágenes descritos en la sección 3.2. Por definición, los descriptores de forma son independientes de la información de los píxeles y sólo dependen de la forma de la región de interés. El uso de segmentación automática o semiautomática podría permitir la definición de un tipo de incertidumbre para los DR de forma, pues la ROI podría cambiar de tamaño para cada par de imágenes. Ejemplos de este tipo de segmentación pueden ser encontrados en numerosos trabajos, como el de Yuan et al. [46], donde se desarrolló un método de delineación automático para mamografía digital.

Evidencia de la sensibilidad de algunos de estos descriptores hacia receptores inmunohistoquímicos puede ser encontrada en trabajos como el de Zhou et al. [23], donde se encontró que *Maximum 2D diameter column* [1], *Elongation* [7], *Skewness* [30], *Imc2* [47] y *Large area low gray level emphasis* [63] son descriptores sensibles

a la presencia del receptor HER2 para imágenes de mamografía convencional. Otros trabajos, como el de Mao et al. [47] y el de Marino et al. [39], reportan sensibilidad de DR similares a los presentados en este trabajo, pero los DR no son reportados explícitamente.

Una representación gráfica del comportamiento de la distribución de incertidumbre de todos los DR extraídos se presenta en la figura 4.1. Los DR se ordenan en tres grupos principales: forma, 1er y 2do orden. En cada subgrupo los DR se presentan en orden descendente con referencia al valor de su C_V total. No se consideraron los DR con C_V mayor al 100 %, dado que estos no entraron en el análisis radiómico propuesto.

Se utilizan áreas con colores para mostrar el impacto de cada una de las tres fuentes de incertidumbre en la incertidumbre total. Además, se indican con puntos verdes los DR sensibles a la presencia de alguno de los 4 antígenos, es decir, los DR que fueron usados en al menos un modelo de predicción.

La figura 4.1 muestra que la fuente de incertidumbre con mayor influencia en la mayoría de los DR fue la asociada a la ROI. Como se mencionó anteriormente, los descriptores de *forma* no tienen definida incertidumbre de mastógrafo y de registro, siguiendo el método propuesto en este trabajo, por lo que toda su incertidumbre se debe al trazado de la ROI.

En el caso de los DR de *1er orden*, la incertidumbre por registro muestra una influencia mayor que en los otros dos grupos. Resalta el caso particular donde la incertidumbre asociada al registro superó el 50 %. El registro, procedimiento característico en las imágenes CEDM, tuvo un impacto considerable en algunos de los DR, superando la relacionada a la ROI en algunos casos. Finalmente, la incertidumbre asociada al mastógrafo, en general, tiene valores de C_V por debajo del 10 %, haciéndola la fuente de incertidumbre con menor impacto.

Respecto a los DR sensibles a la presencia de antígenos, se observa que estos están

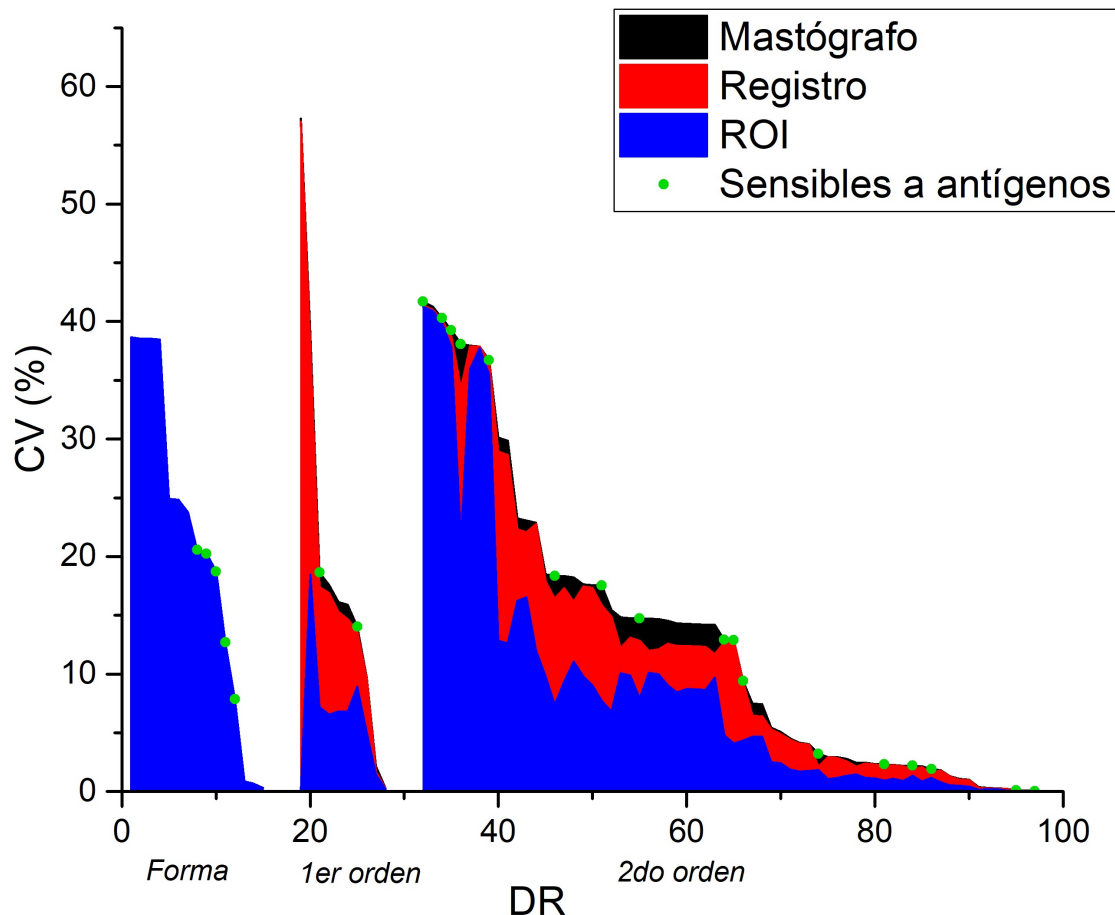


Figura 4.1: Presupuesto de incertidumbre de las tres fuentes consideradas, con los 107 DR en orden descendente con respecto al C_V total. Los puntos verdes representan los 24 DR sensibles a la presencia de antígenos.

presentes en los tres grupos de descriptores principales, con una mayoría clara en el grupo de *2do orden* posiblemente dado a que este último es el de mayor tamaño. La incertidumbre total de los DR sensibles oscila de un 40% hasta valores cercanos a 0%, por lo que no se observa una relación respecto al valor del C_V y la sensibilidad de los DR. Se menciona que, aunque idealmente es preferible el uso de DR con incertidumbres bajas, valores de C_V pequeños podrían indicar poca variabilidad de los DR hacia los antígenos y, por lo tanto, poco poder de predicción.

4.2. Modelos de predicción

A continuación, se presenta la comparación de los resultados de las tres ramas de análisis en 4 tablas diferentes, donde cada tabla contiene los modelos de predicción del respectivo antígeno del análisis IHQ. En el caso de la rama tradicional (los primeros dos renglones), se presentan dos resultados: el resultado sin considerar la evaluación de desempeño utilizando el PI y otro donde sí es considerada.

El desempeño de cada modelo es representado por el promedio de MCC, y su desviación estándar, además del promedio de las 3 mediciones tradicionales de desempeño de modelos de clasificación binaria: exactitud, precisión y valor F1. Se recuerda que se están utilizando promedios de las métricas ya que con la técnica de evaluación LOO se generan 1000 datos nuevos para cada muestra de prueba al momento de evaluar el desempeño. Sólo el primer renglón (rama tradicional sin evaluación con PI), al no contemplar una evaluación de desempeño usando incertidumbre, no presenta promedios de las métricas de desempeño y, por lo tanto, no tiene definida una desviación estándar.

Tabla 4.2: Desempeño de los modelos de predicción de ER positivo

Rama	MCC	STD	Exactitud	valor F1	Precisión	Descriptores
Tradicional sin Evaluación-PI	67.7 %	NA	90.9 %	94.6 %	92.9 %	11, 15, 47, 89
Tradicional con Evaluación-PI	24.9 %	15.1 %	72.8 %	82.4 %	87.3 %	11, 15, 47, 89
Robusta	30.1 %	16.2 %	73.0 %	82.0 %	89.3 %	11, 15, 47, 92
Robusta con Entrenamiento-PI	43.2 %	16.0 %	85.7 %	91.8 %	86.6 %	2, 11, 30, 38, 47

Tabla 4.3: Desempeño de los modelos de predicción de PR positivo

Rama	MCC	STD	Exactitud	valor F1	Precisión	Descriptores
Tradicional sin Evaluación-PI	59.8 %	NA	81.8 %	86.4 %	82.6 %	52, 57, 79, 91
Tradicional con Evaluación-PI	21.0 %	13.6 %	60.7 %	67.3 %	71.4 %	52, 57, 79, 91
Robusta	29.6 %	13.3 %	66.9 %	73.5 %	74.8 %	27, 30, 47, 57, 81
Robusta con Entrenamiento-PI	42.2 %	12.6 %	74.0 %	80.7 %	76.5 %	95, 27, 30 57, 81

Tabla 4.4: Desempeño de los modelos de predicción de HER2 positivo

Rama	MCC	STD	Exactitud	valor F1	Precisión	Descriptores
Tradicional sin Evaluación-PI	68.4 %	NA	93.9 %	66.7 %	100.0 %	1, 7, 68, 92, 98
Tradicional con Evaluación-PI	21.2 %	14.3 %	80.1 %	26.9 %	23.3 %	1, 7, 68, 92, 98
Robusta	31.9 %	15.9 %	77.3 %	40.2 %	30.7 %	1, 95, 30, 57, 92
Robusta con Entrenamiento-PI	32.2 %	18.2 %	89.0 %	29.4 %	62.4 %	30, 47, 68, 92, 106,63

Tabla 4.5: Desempeño de los modelos de predicción de presencia $Ki67 > 0.20$

Rama	MCC	STD	Exactitud	valor F1	Precisión	Descriptores
Tradicional sin Evaluación-PI	53.5 %	NA	78.8 %	69.6 %	72.7 %	30,34
Tradicional con Evaluación-PI	25.7 %	14.5 %	64.4 %	52.9 %	51.5 %	30,34
Robusta	31.7 %	13.2 %	68.3 %	56.1 %	57.0 %	30, 34, 47
Robusta con Entrenamiento-PI	42.1 %	11.8 %	74.1 %	58.8 %	70.4 %	30, 34, 47, 55, 70

4.2.1. Rama tradicional

Los dos resultados de la rama tradicional, encontrados para los 4 antígenos, representan juntos el primer resultado importante de este trabajo. Éstos evidencian que un modelo de predicción basado en DR extraídos de imágenes CEDM, que no considera el efecto de la incertidumbre en el valor de los DR, puede alcanzar indicadores de desempeño buenos, pero verá su desempeño fuertemente disminuido al realizar una prueba de desempeño usando el PI. Esto puede interpretarse como que, si el estudio con las mismas pacientes volviese a realizarse, procurando las mismas condiciones de adquisición que la primera vez, el modelo previamente definido reducirá su desempeño de predicción para las imágenes de estos nuevos estudios de las mismas pacientes, fallando en la generalización del modelo.

Así, por ejemplo, el modelo de predicción tradicional de ER que originalmente predecía con un MCC de 67.7 %, al considerar los nuevos valores de DR generados con el PI, dará en promedio un valor de MCC de $\mu_{MCC} = 24.9\%$ con desviación estándar de $\sigma_{MCC} = 15.1\%$. Para el resto de los antígenos, todos los modelos degradan

su desempeño a la categoría de acuerdo *suficiente*, en algunos casos casi llegando a nivel de acuerdo *ligero*, como en el caso de la progesterona y el HER2. Este último representa la caída de desempeño más drástica, perdiendo casi 46 % del MCC.

Determinar la influencia de la incertidumbre en los modelos de predicción es un problema que se ha trabajado en distintas modalidades de imagen, incluida CEDM. En el trabajo de Castillo et al. [48], se buscó determinar la influencia de los descriptores en el modelo de predicción de la presencia de los mismos antígenos inmunohistoquímicos a través de una figura de mérito (FOM) definida para el valor F1. Aunque el trabajo mencionado es similar al presentado en esta investigación, pues ambos se centran en determinar la influencia de la incertidumbre, una diferencia notable es que en el trabajo de Castillo et al. se emplearon descriptores que previamente habían mostrado relación en trabajos como los de Mateos y Zhou et al. [15, 23], y no se buscó determinar los DR sensibles a los antígenos, como se hizo en este trabajo.

Entre las razones por las cuales los modelos reducen tanto su valor del MCC resalta el hecho de que la elección de descriptores LASSO no contempla la inestabilidad de los DR. De esta forma, los descriptores elegidos podrían ser altamente sensibles a los cambios en su valor, haciéndolos dañinos para la reproducibilidad del modelo. Como se verá más adelante, la sustitución de estos descriptores altamente cambiantes por otro con menor C_V ayudó a mejorar el desempeño. Así, en el modelo de predicción de ER de la rama tradicional, mostrado en la tabla 4.2, al sustituir al descriptor 89 *Busyness*, con C_V cercano al 40 %, por el DR 92 *Contraste*, que tiene un C_V menor al 3 %, se mejoró el desempeño del modelo en un 5 %.

Los resultados mencionados refuerzan la conclusión de esta primera sección, que indica que los buenos desempeños mostrados por los modelos tradicionales presentados en este trabajo fueron ficticios. Esto puede deberse a problemas de sobreajuste o casualidad, y sus efectos desaparecen al evaluar su desempeño usando el efecto de la incertidumbre. Ejemplos de la búsqueda de estos efectos, y sugerencias para su

supresión, pueden ser encontrados en otros trabajos relacionados con el análisis de la robustez de los DR, como el trabajo de Mendel et al. [16], quienes analizaron el efecto de los parámetros de adquisición de la imagen en los DR y lo removieron al comparar coeficientes de correlación en una sola imagen.

4.2.2. Rama robusta: uso del PI

En esta rama del flujo del trabajo se realizó un segundo paso de eliminación de descriptores, antes de la eliminación final LASSO. Esta eliminación fue la agrupación de descriptores con correlación en grupos para reducir el número de DR redundantes. El uso de grupos con este fin es una técnica que puede encontrarse en otras investigaciones de análisis radiómico. Robinson et al. [31] realizaron agrupamiento de DR usando el CCP como métrica de distancia, similar a lo realizado en este trabajo. Una diferencia de procedimiento, sin embargo, es que en su trabajo se incluyó una variación del número de grupos basado en el tamaño de su muestra y el número total de DR, por lo que distintos análisis radiómicos se realizaron para cada número de grupos disponible. En este trabajo, se definió el número de grupos al considerar como correlacionados a los DR con un CCP entre ellos mayor a 0.99. Los 56 grupos encontrados en este trabajo se muestran en el apéndice B.

Otro término importante que se debe definir en este tipo de análisis es la robustez y la forma en la que será medida. En el caso de este trabajo, la robustez de un DR fue medida a partir de la incertidumbre, expresada en términos de la desviación estándar (ver sección 2.5). Para el caso del mastógrafo y la ROI (σ_M y σ_R respectivamente) la incertidumbre es de tipo A, mientras que en el caso del registro (σ_I) la incertidumbre es tipo B, siguiendo lo explicado en la sección 2.5.2. En el trabajo de Kalpathy-cramer et al. [49] se consideró la robustez de los DR con respecto a la segmentación a partir del coeficiente de correlación de concordancia (CCC). Así, el CCC se empleó para medir la influencia de la ROI en el valor de los DR, donde un valor de CCC alto

indica robustez con respecto a la ROI. La desviación estándar σ_R , propuesta en este trabajo, posee información similar pues valores bajos de C_V indican que el DR es robusto a cambios en la ROI.

Evaluación de desempeño usando el PI

El tercer renglón de las tablas 4.2 - 4.5 muestra las métricas de desempeño de los modelos con el mejor cociente $\frac{\sigma}{\mu}$ del MCC después de la eliminación por grupos. Se utilizó este cociente con el fin de considerar el valor de la media y su desviación estándar al mismo tiempo.

La diferencia primaria entre el segundo y el tercer renglón es que para este último se procuró la elección de descriptores con bajo C_V y, por lo tanto, una influencia menor de la incertidumbre en sus DR, lo cual mejoró el desempeño del modelo en función del MCC. En la sección anterior ya se ha dado un ejemplo de cómo el cambio de un DR mejoró el desempeño para la predicción de la presencia de ER. Para el PR y el HER2, se debe notar que, si bien existe una mejora en el desempeño, con un aumento de cerca del 10% en el valor del MCC, ambos modelos se mantienen en la misma categoría de nivel de acuerdo, determinada como *suficiente*.

Aunque para los 4 antígenos existe una mejora en la predicción, el valor del MCC se mantiene relativamente bajo. Esto puede deberse a que el modelo no fue entrenado para reconocer cambios en los valores de los DR, los cuales serían esperados si el estudio se repitiese. Dicho de otra forma, se está evaluando un modelo que desconoce la posibilidad de que sus DR de entrenamiento fluctúen. Es relativamente esperado que el modelo no mejore considerablemente.

Entrenamiento con PI

La propuesta hecha en este trabajo para incluir en el entrenamiento del modelo la existencia de dichas fluctuaciones es agregar a los datos de entrenamiento valores aleatorios alrededor del valor real de los DR, considerando una distribución gaussiana de datos aleatorios basada en el PI, tal como fue descrito en la sección 3.5.3.

Estos resultados, renglones finales de las 4 tablas, muestran un rescate en el desempeño de predicción para el receptor de estrógeno, de progesterona, y el Ki67, al entrenar a los modelos tomando en cuenta el PI, recuperando la clasificación de evaluación original, es decir, clasificación *moderada*. El receptor HER2 (tabla 4.4), sin embargo, mantiene un desempeño prácticamente similar, aún con estas consideraciones, con MCC cercano al 30 %.

La propuesta arriba mencionada es importante ya que, como se sostiene a través de los resultados, la presencia de fluctuaciones para los valores de los DR es un problema que no puede evitarse para imágenes de modalidad CEDM, dada la naturaleza de la obtención de estas imágenes.

Comparación entre modelos

En las figuras 4.2 y 4.3 se resumen los resultados de las tablas 4.2 - 4.5. En la figura 4.2 se comparan los modelos de los antígenos que mostraron una clara mejora en su desempeño, como fue discutido en las subsecciones pasadas, mientras que el HER2 es mostrado de forma separada en la figura 4.3.

Para asegurar una diferencia estadísticamente significativa entre las medias y medianas de las 9 distribuciones del primer diagrama, se realizó una prueba T de Student, usando la corrección de Welch para muestras sin homogeneidad de varianza. Esta prueba se realizó en las tres combinaciones de pares de modelos para cada antígeno.

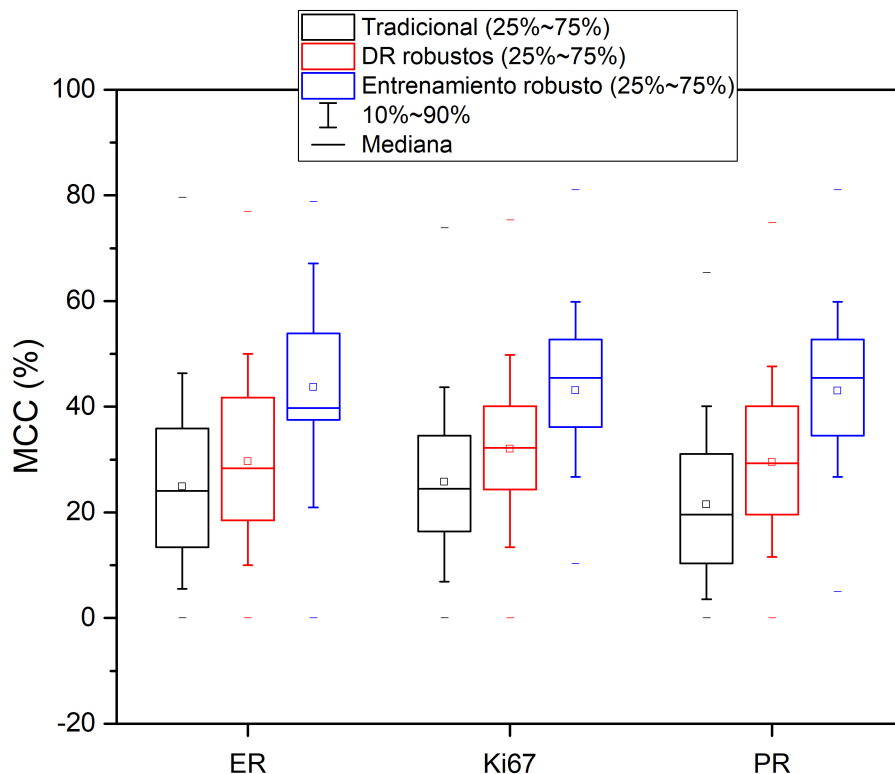


Figura 4.2: Diagrama de cajas y bigotes para los modelos de los tres antígenos con diferencias estadísticamente significativas entre sus medias.

En todos los casos, se obtuvo un valor p menor al 0.05, refutando la hipótesis nula, concluyendo que las distribuciones tienen medias μ con diferencias estadísticamente significativas. Un procedimiento similar se hizo con las medianas, usando la prueba de medianas de Mood, concluyendo que las medianas también son significativamente diferentes entre sí.

Las mismas pruebas fueron realizadas en las distribuciones de los modelos del antígeno HER2. Por un lado, el modelo tradicional y el robusto mostraron diferencias estadísticamente significativas de la media y la mediana. Sin embargo, como se anticipaba, las medias entre el modelo con DR robustos y el entrenado con PI no mostraron diferencias significativas, pues se obtuvo un $p \approx 0.106$, fallando en refutar la hipótesis nula. Sorpresivamente, la prueba de medianas de Mood sí mostró diferencias

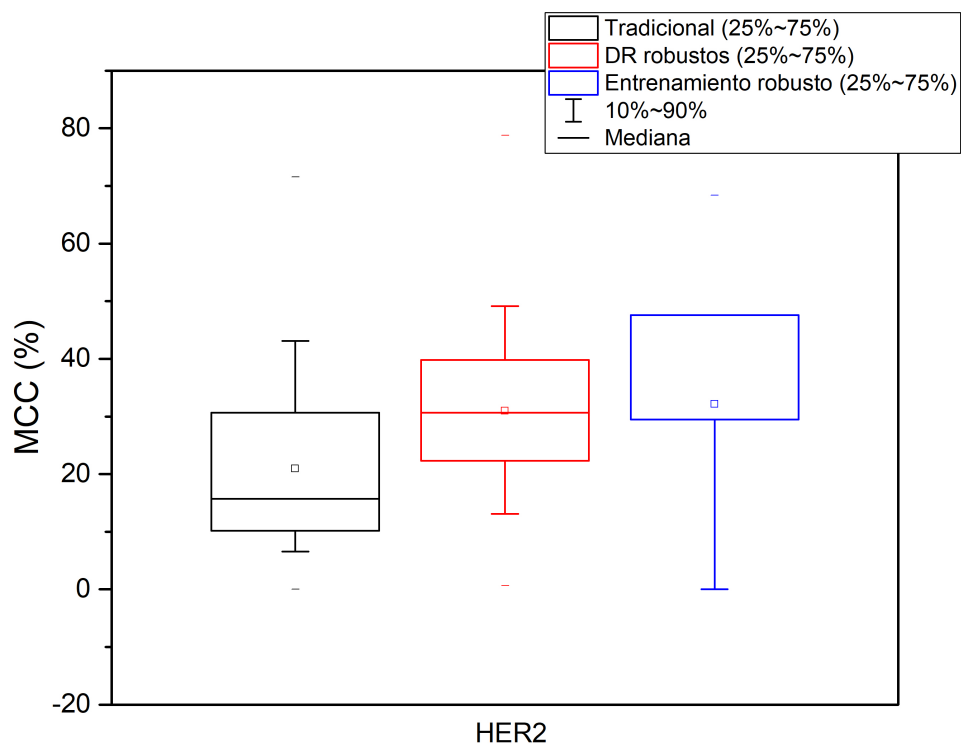


Figura 4.3: Diagrama de cajas y bigotes para los modelos del HER2.

significativas incluso para los modelos de este último antígeno.

Los desempeños de los modelos de predicción encontrados en esta investigación son acordes a los de trabajos similares. Castillo et al. [48] obtuvieron un valor F1 de 84% y 65% para los receptores hormonales ER y PR respectivamente. En este trabajo, se obtuvieron valores F1 de 91.8% y 80.7% respectivamente, mejorando el desempeño de ambos modelos. Se recuerda que considerar el valor del MCC es recomendable para una comparación más robusta de esta métrica de desempeño (ver siguiente subsección). En contraste, el desempeño del modelo de Castillo et al. para el antígeno Ki67 fue mayor en términos del valor F1 con 72%, a diferencia del 58.8% encontrado en este trabajo.

Marino et al. [39] obtuvieron una exactitud de predicción de ambos receptores

hormonales ER y PR del 78 %. En este trabajo, el ER obtuvo un mejor desempeño con cerca del 85 % de exactitud, mientras que la del PR fue de 74 %.

4.2.3. Comparación de métricas de desempeño

Se quisiera enfatizar la decisión de considerar al MCC sobre las métricas de desempeño tradicionales y su importancia para la evaluación de desempeño considerando la incertidumbre. En las 4 tablas de comparación de modelos existe una sobreestimación positiva del desempeño al considerarse las 3 métricas de desempeño tradicionales. Tómese como ejemplo la tabla 4.2, donde el valor F1 de los 4 modelos se mantiene arriba del 80 %. Sin embargo, una revisión al valor de MCC demuestra que el desempeño del modelo llega a degradarse considerablemente, al pasar de ser considerado *substancial*, con $MCC = 67.7\%$, a un desempeño casi *ligero*, con $MCC = 24.9\%$, comparando la rama tradicional con y sin evaluación de desempeño usando el PI (primer y segundo renglón).

Más aún, podría ocurrir lo contrario y el valor F1 podría no mostrar que el modelo ha mejorado su desempeño significativamente. Esto se puede observar en la tabla 4.5 del Ki67, donde el entrenamiento con PI mejoró el desempeño del MCC aproximadamente en un 10 %, pasando de clasificación *suficiente a moderada* (obsérvese tabla 2.1 para ver las clasificaciones). El valor F1, sin embargo, mantiene prácticamente el mismo valor, incapaz de ayudar a evaluar la importancia del cambio hecho en el poder de predicción del modelo.

Las ventajas del MCC son más claras para muestras de datos desbalanceadas y con número de pacientes reducido, tal como concluyen Boughorbel et al. [50]. En el trabajo de Zhou et al. [23] la presencia de pacientes con HER2 positivo es de cerca del 25 %. En esta investigación dicha presencia fue del 12 %. Esto, combinado con una población de 308 pacientes, en el caso de Zhou et al., permitió una mayor veracidad

de las métricas de desempeño tradicionales, por lo que su desempeño se considera confiable.

El hecho de que ambas muestras presenten distribuciones desbalanceadas de HER2 positivo no representa casos aislados. De acuerdo a datos registrados por Villareal-Garza et al. [51], en una población mexicana de 4316 pacientes, se encontró que aproximadamente 23 % de la población fue HER2 positivo. Debido a que el desbalance es una propiedad intrínseca de los estudios con HER2, se debe considerar una métrica de desempeño sensible a su influencia, como lo es el MCC.

4.2.4. Riqueza de los DR y su relación con el análisis IHQ

Para la construcción de estos 4 modelos de predicción de la presencia de antígenos de la mama, se asumió que dichos antígenos y la información dada por los DR de *Pyradiomics* están asociados. Sin embargo, los resultados de la tabla 4.4 del HER2 invitan a suponer que, dado que el modelo no mejoró aun considerando la variación de los DR, no existe una relación sólida entre el HER2 y los DR para la muestra usada en esta investigación, y que el modelo encontrado en la rama tradicional con MCC de 68 % fue una causalidad.

Para evaluar esta posibilidad se presenta la figura 4.4, donde se comparan el valor de MCC para las 100 mejores combinaciones de DR hechas a partir de los 9 DR seleccionados por LASSO. La primera característica por notar es la clara discretización de los valores de MCC para varias combinaciones de DR, la cual es resultado del tamaño reducido de la muestra. Si la combinación de DR es estable, se esperan «saltos pequeños» de los valores de MCC entre combinaciones, donde cada salto representa la clasificación correcta de una muestra. Sin embargo, la combinación con el mejor desempeño, MCC 68 %, se encuentra aislada del resto de las combinaciones, con una diferencia de MCC con la siguiente combinación de casi 20 %. Esto sugiere que la

combinación podría dar un desempeño alto por casualidad y que, en realidad, esos DR no están asociados al HER2 en esta muestra.

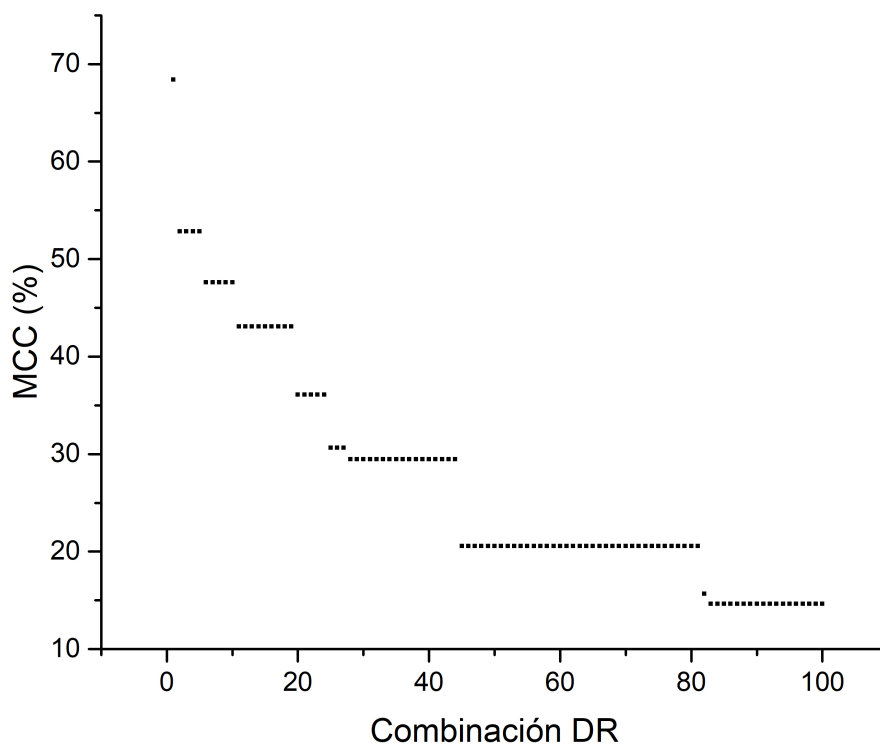


Figura 4.4: Desempeño en MCC de las 100 mejores combinaciones de DR para la predicción de presencia de HER2 de la rama tradicional.

Se debe enfatizar que esto no significa que los DR presentados no tengan relación alguna con la predicción del HER2. Como se mencionó, el trabajo de Zhou et al. [23] expuso la sensibilidad a la presencia de este receptor para 5 de los 9 DR usados en este trabajo. Resolver la falta de sensibilidad al HER2 podría lograrse aumentando el número de muestras que son entrenadas por el modelo, ya que un cambio en la metodología del experimento, como se hizo en un trabajo precursor de Castillo et al. [48], presentó el mismo problema de baja correlación.

Desviación estándar del MCC

Finalmente, se discute el papel de la desviación estándar del MCC (σ_{MCC}) en el desempeño del modelo. Representada como **STD** en las tablas de desempeño, σ_{MCC} varía entre el 11 % y el 18 % en todos los modelos. En general, estos valores de σ indican cierta inestabilidad de los modelos de predicción.

Para entender el impacto del tamaño de la muestra se simuló, considerando el PI, cómo cambiaría el desempeño del modelo si se cuenta con una muestra de cerca de 1,000 pacientes. Esto se realizó usando el método LOO (sección 2.6.5) para todos los datos generados aleatoriamente con base en el PI, y no sólo con las 33 pacientes originales. Se obtuvo una reducción de σ de cerca de la mitad de su valor. Este resultado invita a suponer que una muestra mayor disminuiría el valor de σ y, así, se obtendrían modelos más estables.

5

Conclusiones

En la física experimental, el análisis de incertidumbre es una herramienta vital pues está relacionada con la confianza en el resultado obtenido. Considerando su importancia, en este trabajo se aplicó la incertidumbre, de manera novedosa, sobre imágenes de mamografía digital con medio de contraste, considerándolas como una distribución espacial de mediciones de atenuación de la radiación al pasar a través del tejido mamario. Desde este punto de vista, la mamografía posee correlaciones espaciales y su incertidumbre se puede definir.

Específicamente, se desarrolló una técnica de evaluación de desempeño de modelos que predicen la presencia de antígenos de un análisis IHQ de la mama, considerando la incertidumbre de descriptores radiómicos (DR) extraídos a través de un análisis radiómico para 33 imágenes de mamografía digital con medio de contraste (CEDM), modalidad SET. Las fuentes de incertidumbre se escogieron a partir del proceso de adquisición de las imágenes, relativo a la modalidad CEDM, siendo éstas: mastógrafo, registro de imagen y delineación de la región de interés (ROI). Los antígenos del análisis IHQ estudiados fueron ER, PR, HER2 y $Ki67 > 0.20$. Se definió un presupuesto de incertidumbre (PI) a partir de las incertidumbres de los DR asociadas a las tres fuentes de incertidumbre.

Para estimar el valor agregado de aplicar análisis de incertidumbre se estudiaron tres enfoques de creación de modelos: un enfoque tradicional sin consideración de la incertidumbre, un primer enfoque robusto con evaluación de los modelos usando el PI, y un segundo enfoque robusto con entrenamiento y evaluación de los modelos usando incertidumbre.

Se encontró que los modelos con enfoque tradicional reducen fuertemente su poder de predicción al realizar en ellos una evaluación de desempeño usando el PI. El desempeño de predicción de todos los modelos tradicionales, en función del coeficiente de correlación de Matthews (MCC), se redujo entre un 35 % y un 45 %.

Respecto a los enfoques robustos, se registró un aumento de entre 5 % y 10 % del MCC al considerar en el análisis a los descriptores con menor variación de coeficiente de variación (C_V) con respecto al PI. Sin embargo, fueron los modelos con entrenamiento basado en el presupuesto los que obtuvieron el mejor desempeño, con valores de MCC por arriba del 40 %, considerado un nivel de acuerdo *moderado*, para todos los antígenos excepto el HER2. El antígeno con el modelo de mejor desempeño, considerando la incertidumbre, fue el ER con exactitud, precisión y valor F1 por arriba del 85 %.

Por otro lado, el PI permitió determinar que la fuente de incertidumbre más influyente en los DR es la delineación de la ROI, seguida del registro espacial de la imagen. La incertidumbre debido al mastógrafo tuvo una influencia mínima en los DR y, por lo tanto, en los modelos.

Con estos resultados se concluye que el desempeño de los modelos de predicción obtenidos del análisis radiómico tradicional disminuye significativamente al considerar la inestabilidad de los DR. Además, la incertidumbre de los DR usada como medida de robustez puede mejorar la selección de DR y optimizar el desempeño los modelos de predicción.

Así, se comprueba la importancia de considerar la incertidumbre de los DR en el desempeño de modelos de predicción basados en imágenes CEDM, proporcionando información relevante para reconocer las fuentes de incertidumbre y, en su caso, aislarlas para ayudar a alcanzar uno de los principales objetivos actuales de la radiómica: determinar la interpretación física de los descriptores.

Entre las principales limitaciones de este trabajo, se menciona la ausencia de otras fuentes de incertidumbre en el PI. La inclusión de fuentes como la variabilidad intermastógrafo (uso de varios equipos de mastografía), el uso de otros programas de extracción de DR, o la influencia del preprocesamiento de las imágenes, podría mejorar el flujo de trabajo y permitir la creación de un modelo final más completo. Otra limitación, común en trabajos de esta área, es el número bajo de pacientes de la base de datos, pues esto puede desembocar en mayor probabilidad de sobreajuste, o en la falta de muestras suficientes para determinar patrones entre las imágenes.

Además, dado que la delineación de ROI fue la fuente de incertidumbre con mayor impacto en el PI, contar con un número mayor de médicos radiólogos para segmentar podría dar una mayor certidumbre al valor promedio de los DR en lesiones con diferentes perfiles IHQ.

Para finalizar, se menciona que este trabajo aporta una técnica de evaluación de desempeño que ofrece la posibilidad de desarrollar modelos de los que se tenga una mayor confianza en los resultados, basados en análisis radiómico y no sólo para estudios de cáncer de mama. Además, la técnica ayuda a que los estudios de este tipo sean reproducibles y comparables con trabajos de índole similar. La importancia de esto último, para fines clínicos en cáncer de mama, es la capacidad no invasiva de los análisis radiómicos en imágenes CEDM de caracterizar lesiones en la mama. Si bien la biopsia es el estándar de oro para la caracterización del perfil IHQ de lesiones malignas, la misma no representa la lesión en su totalidad. Debido a esto, se presentan casos en los que existe una contradicción entre muestras patológicas debido

a la heterogeneidad de las lesiones. De esta forma, los modelos de predicción CADx complementan la información médica para la toma de decisiones.

Apéndice A

Presupuestos de incertidumbres

Se muestran los 107 DR que pueden ser extraídos para imágenes 2D usando la plataforma *Pyradiomics*. Los DR son separados en 7 grupos. Los primeros dos grupos son los DR de *forma* y de *primer orden*. Los otros 5 grupos son DR de *segundo orden*, distinguidos a partir de la matriz de grises de la cual se extraen los DR. Las 5 matrices de grises son GLCM (*Gray Level Co-ocurrence Matrix*), GLSZM (*Gray Level Size Zone Matrix*), GLRLM (*Gray Level Run-Length Matrix*), NGTDM (*Neighbouring Gray Tone Difference Matrix*) y GLDM (*Gray Level Dependence Matrix*).

A partir de la tabla A.3 los descriptores son de segundo orden.

Tabla A.1: Presupuesto de incertidumbre para descriptores de forma.

	Descriptor radiómico	Mastógrafo	Registro	ROI	Total
1	Maximum 2D Diameter Column	0.00 %	0.00 %	20.25 %	20.25 %
2	Maximum 2D Diameter Row	0.00 %	0.00 %	20.55 %	20.55 %
3	Mesh Volume	0.00 %	0.00 %	38.65 %	38.65 %
4	3D Sphericity	0.00 %	0.00 %	0.84 %	0.84 %
5	Surface Area	0.00 %	0.00 %	38.46 %	38.46 %
6	Surface Volume Ratio	0.00 %	0.00 %	0.31 %	0.31 %
7	Elongation	0.00 %	0.00 %	12.68 %	12.68 %
8	Major Axis Length	0.00 %	0.00 %	24.91 %	24.91 %
9	Maximum Diameter	0.00 %	0.00 %	23.78 %	23.78 %
10	Mesh Surface	0.00 %	0.00 %	38.55 %	38.55 %
11	Minor Axis Length	0.00 %	0.00 %	18.73 %	18.73 %
12	Perimeter	0.00 %	0.00 %	24.85 %	24.85 %
13	Perimeter Surface Ratio	0.00 %	0.00 %	0.67 %	0.67 %
14	Pixel Surface	0.00 %	0.00 %	38.55 %	38.55 %
15	2D Sphericity	0.00 %	0.00 %	7.85 %	7.85 %

Tabla A.2: Presupuesto de incertidumbre para descriptores de primer orden.

	Descriptor radiómico	Mastógrafo	Registro	ROI	Total
16	10 Percentile	5.67 %	757.78 %	0.04 %	757.80 %
17	90 Percentile	5.31 %	198.24 %	0.02 %	198.31 %
18	Energy	10.35 %	100.25 %	38.49 %	107.88 %
19	Entropy	1.32 %	0.89 %	1.46 %	2.16 %
20	Interquartile Range	6.12 %	12.98 %	6.81 %	15.89 %
21	Kurtosis	1.64 %	8.15 %	5.01 %	9.70 %
22	Maximum	5.99 %	109.17 %	0.01 %	109.33 %
23	Mean Absolute Deviation	4.94 %	15.55 %	6.56 %	17.59 %
24	Mean	5.48 %	312.65 %	0.02 %	312.69 %
25	Median	5.62 %	311.99 %	0.02 %	312.04 %
26	Minimum	27.06 %	341.48 %	0.08 %	342.55 %
27	Range	6.71 %	15.84 %	7.19 %	18.64 %
28	Robust Mean Absolute Deviation	5.09 %	13.68 %	6.80 %	16.10 %
29	Root Mean Squared	5.45 %	57.00 %	0.02 %	57.26 %
30	Skewness	3.27 %	10.35 %	8.89 %	14.04 %
31	Uniformity	0.18 %	0.14 %	0.21 %	0.31 %
32	Variance	9.38 %	32.86 %	18.43 %	38.83 %

Tabla A.3: Presupuesto de incertidumbre para descriptores de la matriz GLCM

	Descriptor radiómico	Mastógrafo	Registro	ROI	Total
33	Autocorrelation	7.26 %	8.75 %	9.02 %	14.51 %
34	Cluster Prominence	16.64 %	26.92 %	21.21 %	38.10 %
35	Cluster Shade	1260.26 %	1381.57 %	-1313.93 %	2285.47 %
36	Cluster Tendency	8.50 %	11.75 %	11.06 %	18.24 %
37	Contrast	8.43 %	25.93 %	12.81 %	30.13 %
38	Correlation	0.41 %	12.20 %	4.08 %	12.87 %
39	Difference Average	4.23 %	13.20 %	6.82 %	15.45 %
40	Difference Entropy	1.30 %	4.67 %	2.48 %	5.45 %
41	Difference Variance	8.40 %	25.72 %	12.62 %	29.86 %
42	Id	0.53 %	2.09 %	1.14 %	2.44 %
43	Idm	0.41 %	1.95 %	1.08 %	2.26 %
44	Idmn	0.22 %	0.26 %	0.17 %	0.38 %
45	Idn	0.41 %	0.89 %	0.49 %	1.09 %
46	Imc1	0.15 %	2.71 %	1.06 %	2.92 %
47	Imc2	1.49 %	8.24 %	4.32 %	9.42 %
48	Inverse Variance	0.39 %	2.02 %	1.10 %	2.34 %
49	Joint Average	3.77 %	4.41 %	4.66 %	7.44 %
50	Joint Energy	0.01 %	0.02 %	0.02 %	0.03 %
51	Joint Entropy	1.26 %	1.52 %	1.46 %	2.46 %
52	MCC	2.13 %	11.80 %	4.78 %	12.91 %
53	Maximum Probability	0.07 %	0.03 %	0.05 %	0.09 %
54	Sum Average	3.76 %	4.46 %	4.66 %	7.46 %
55	Sum Entropy	1.10 %	1.36 %	1.34 %	2.20 %
56	Sum Squares	8.54 %	6.43 %	10.11 %	14.71 %

Tabla A.4: Presupuesto de incertidumbre para descriptores de la matriz GLSZM

	Descriptor radiómico	Mastógrafo	Registro	ROI	Total
57	Gray Level Non Uniformity	3.53 %	4.33 %	39.92 %	40.31 %
58	Gray Level Non Uniformity Normalized	0.15 %	0.15 %	0.19 %	0.28 %
59	Gray Level Variance	8.03 %	6.56 %	9.68 %	14.18 %
60	High Gray Level Zone Emphasis	7.04 %	8.76 %	8.66 %	14.19 %
61	Large Area Emphasis	2.83 %	14.79 %	9.03 %	17.56 %
62	Large Area High Gray Level Emphasis	6.63 %	14.66 %	16.51 %	23.06 %
63	Large Area Low Gray Level Emphasis	0.03 %	0.09 %	0.04 %	0.10 %
64	Low Gray Level Zone Emphasis	0.02 %	0.02 %	0.02 %	0.03 %
65	Size Zone Non Uniformity	2.29 %	11.93 %	35.96 %	37.95 %
66	Size Zone Non Uniformity Normalized	1.03 %	4.01 %	1.84 %	4.53 %
67	Small Area Emphasis	0.54 %	2.62 %	1.15 %	2.91 %
68	Small Area High Gray Level Emphasis	7.23 %	10.10 %	7.91 %	14.72 %
69	Small Area Low Gray Level Emphasis	0.02 %	0.01 %	0.02 %	0.03 %
70	Zone Entropy	0.69 %	1.98 %	0.95 %	2.30 %
71	Zone Percentage	0.77 %	3.68 %	1.72 %	4.13 %
72	Zone Variance	2.18 %	19.37 %	11.97 %	22.87 %

Tabla A.5: Presupuesto de incertidumbre para descriptores de la matriz GLRLM

	Descriptor radiómico	Mastógrafo	Registro	ROI	Total
73	Gray Level Non Uniformity	4.57 %	3.09 %	40.91 %	41.28 %
74	Gray Level Non Uniformity Normalized	0.17 %	0.14 %	0.21 %	0.31 %
75	Gray Level Variance	8.33 %	6.81 %	9.98 %	14.68 %
76	High Gray Level Run Emphasis	7.07 %	8.78 %	8.70 %	14.24 %
77	Long Run Emphasis	0.67 %	3.60 %	1.74 %	4.05 %
78	Long Run High Gray Level Emphasis	6.87 %	8.55 %	9.88 %	14.77 %
79	Long Run Low Gray Level Emphasis	0.02 %	0.03 %	0.02 %	0.04 %
80	Low Gray Level Run Emphasis	0.02 %	0.02 %	0.02 %	0.03 %
81	Run Entropy	1.04 %	1.10 %	1.16 %	1.91 %
82	Run Length Non Uniformity	0.70 %	3.54 %	37.71 %	37.89 %
83	Run Length Non Uniformity Normalized	0.43 %	1.97 %	0.90 %	2.21 %
84	Run Percentage	0.23 %	1.15 %	0.53 %	1.29 %
85	Run Variance	0.26 %	1.61 %	0.77 %	1.80 %
86	Short Run Emphasis	0.18 %	0.89 %	0.40 %	0.99 %
87	Short Run High Gray Level Emphasis	7.13 %	9.06 %	8.45 %	14.29 %
88	Short Run Low Gray Level Emphasis	0.02 %	0.02 %	0.02 %	0.03 %

Tabla A.6: Presupuesto de incertidumbre para descriptores de la matriz NGTDM

	Descriptor radiómico	Mastógrafo	Registro	ROI	Total
89	Busyness	6.70 %	8.49 %	37.75 %	39.27 %
90	Coarseness	0.00 %	0.01 %	0.01 %	0.01 %
91	Complexity	7.51 %	13.84 %	7.70 %	17.53 %
92	Contrast	2.41 %	1.10 %	1.81 %	3.21 %
93	Strength	2.49 %	14.46 %	9.78 %	17.64 %

Tabla A.7: Presupuesto de incertidumbre para descriptores de la matriz GLDM

	Descriptor radiómico	Mastógrafo	Registro	ROI	Total
94	Dependence Entropy	0.56 %	1.88 %	0.87 %	2.15 %
95	Dependence Non Uniformity	3.07 %	9.01 %	35.46 %	36.71 %
96	Dependence Non Uniformity Normalized	1.06 %	2.20 %	1.33 %	2.78 %
97	Dependence Variance	6.09 %	14.56 %	9.36 %	18.34 %
98	Gray Level Non Uniformity	4.95 %	3.51 %	41.24 %	41.69 %
99	Gray Level Variance	8.44 %	6.87 %	10.07 %	14.82 %
100	High Gray Level Emphasis	7.08 %	8.78 %	8.72 %	14.26 %
101	Large Dependence Emphasis	4.72 %	14.96 %	9.79 %	18.49 %
102	Large Dependence High Gray Level Emphasis	6.52 %	15.32 %	16.22 %	23.24 %
103	Large Dependence Low Gray Level Emphasis	0.06 %	0.22 %	0.10 %	0.24 %
104	Low Gray Level Emphasis	0.02 %	0.02 %	0.02 %	0.03 %
105	Small Dependence Emphasis	1.29 %	4.27 %	2.40 %	5.07 %
106	Small Dependence High Gray Level Emphasis	8.22 %	14.67 %	7.35 %	18.35 %
107	Small Dependence Low Gray Level Emphasis	0.01 %	0.00 %	0.01 %	0.02 %

Apéndice B

Agrupación de descriptores

Se muestran los 56 grupos en los cuales fueron organizados los DR a partir del CCP, donde todos los pares de descriptores con $CCP > 0.99$ pertenecen al mismo grupo.

La primera tabla representa los 45 DR independientes de la muestra, es decir, que pertenecen a un grupo con un único miembro. Los DR son separados en cada uno de los 7 tipos de descriptores (de forma, primer orden, etc.).

Las siguientes 11 tablas representan los 11 grupos restantes, también separados por tipo de descriptores. Se escribe explícitamente el grupo al que pertenecen cuando el mismo nombre de descriptor es usado.

Tabla B.1: Presupuesto de incertidumbre para los descriptores independientes

	Descriptor radiómico	Mastógrafo	Registro	ROI	Total
1	Maximum 2D Diameter Column	0.00 %	0.00 %	20.25 %	20.25 %
2	Maximum 2D Diameter Row	0.00 %	0.00 %	20.55 %	20.55 %
4	3D Sphericity	0.00 %	0.00 %	0.84 %	0.84 %
7	Elongation	0.00 %	0.00 %	12.68 %	12.68 %
8	Major Axis Length	0.00 %	0.00 %	24.91 %	24.91 %
9	Maximum Diameter	0.00 %	0.00 %	23.78 %	23.78 %
11	Minor Axis Length	0.00 %	0.00 %	18.73 %	18.73 %
12	Perimeter	0.00 %	0.00 %	24.85 %	24.85 %
15	2D Sphericity	0.00 %	0.00 %	7.85 %	7.85 %
16	10 Percentile	5.67 %	757.78 %	0.04 %	757.80 %
17	90 Percentile	5.31 %	198.24 %	0.02 %	198.31 %
19	Entropy	1.32 %	0.89 %	1.46 %	2.16 %
21	Kurtosis	1.64 %	8.15 %	5.01 %	9.70 %
22	Maximum	5.99 %	109.17 %	0.01 %	109.33 %
26	Minimum	27.06 %	341.48 %	0.08 %	342.55 %
27	Range	6.71 %	15.84 %	7.19 %	18.64 %
30	Skewness	3.27 %	10.35 %	8.89 %	14.04 %
32	Variance	9.38 %	32.86 %	18.43 %	38.83 %
34	Cluster Prominence	16.64 %	26.92 %	21.21 %	38.10 %
35	Cluster Shade	1260.26 %	1381.57 %	-1313.93 %	2285.47 %
36	Cluster Tendency	8.50 %	11.75 %	11.06 %	18.24 %
38	Correlation	0.41 %	12.20 %	4.08 %	12.87 %
46	Imc1	0.15 %	2.71 %	1.06 %	2.92 %
47	Imc2	1.49 %	8.24 %	4.32 %	9.42 %
50	Joint Energy	0.01 %	0.02 %	0.02 %	0.03 %
51	Joint Entropy	1.26 %	1.52 %	1.46 %	2.46 %
52	MCC	2.13 %	11.80 %	4.78 %	12.91 %
53	Maximum Probability	0.07 %	0.03 %	0.05 %	0.09 %
55	Sum Entropy	1.10 %	1.36 %	1.34 %	2.20 %
62	Large Area High Gray Level Emphasis	6.63 %	14.66 %	16.51 %	23.06 %
63	Large Area Low Gray Level Emphasis	0.03 %	0.09 %	0.04 %	0.10 %
68	Small Area High Gray Level Emphasis	7.23 %	10.10 %	7.91 %	14.72 %
70	Zone Entropy	0.69 %	1.98 %	0.95 %	2.30 %
78	Long Run High Gray Level Emphasis	6.87 %	8.55 %	9.88 %	14.77 %
79	Long Run Low Gray Level Emphasis	0.02 %	0.03 %	0.02 %	0.04 %
81	Run Entropy	1.04 %	1.10 %	1.16 %	1.91 %
90	Coarseness	0.00 %	0.01 %	0.01 %	0.01 %
91	Complexity	7.51 %	13.84 %	7.70 %	17.53 %
92	Contrast	2.41 %	1.10 %	1.81 %	3.21 %
93	Strength	2.49 %	14.46 %	9.78 %	17.64 %
94	Dependence Entropy	0.56 %	1.88 %	0.87 %	2.15 %
102	Large Dependence High Gray Level Emphasis	6.52 %	15.32 %	16.22 %	23.24 %
103	Large Dependence Low Gray Level Emphasis	0.06 %	0.22 %	0.10 %	0.24 %
106	Small Dependence High Gray Level Emphasis	8.22 %	14.67 %	7.35 %	18.35 %
107	Small Dependence Low Gray Level Emphasis	0.01 %	0.00 %	0.01 %	0.02 %

Tabla B.2: Presupuesto de incertidumbre para los descriptores del primer grupo

	Descriptor radiómico	Mastógrafo	Registro	ROI	Total
3	Mesh Volume	0.00 %	0.00 %	38.65 %	38.65 %
5	Surface Area	0.00 %	0.00 %	38.46 %	38.46 %
10	Mesh Surface	0.00 %	0.00 %	38.55 %	38.55 %
14	Pixel Surface	0.00 %	0.00 %	38.55 %	38.55 %
18	Energy	10.35 %	100.25 %	38.49 %	107.88 %
65	Size Zone Non Uniformity	2.29 %	11.93 %	35.96 %	37.95 %
82	Run Length Non Uniformity	0.70 %	3.54 %	37.71 %	37.89 %
89	Busyness	6.70 %	8.49 %	37.75 %	39.27 %
95	Dependence Non Uniformity	3.07 %	9.01 %	35.46 %	36.71 %

Tabla B.3: Presupuesto de incertidumbre para los descriptores del segundo grupo.

	Descriptor radiómico	Mastógrafo	Registro	ROI	Total
6	Surface Volume Ratio	0.00 %	0.00 %	0.31 %	0.31 %
13	Perimeter Surface Ratio	0.00 %	0.00 %	0.67 %	0.67 %

Tabla B.4: Presupuesto de incertidumbre para los descriptores del tercer grupo.

	Descriptor radiómico	Mastógrafo	Registro	ROI	Total
20	Interquartile Range	6.12 %	12.98 %	6.81 %	15.89 %
23	Mean Absolute Deviation	4.94 %	15.55 %	6.56 %	17.59 %
28	Robust Mean Absolute Deviation	5.09 %	13.68 %	6.80 %	16.10 %

Tabla B.5: Presupuesto de incertidumbre para los descriptores del cuarto grupo.

	Descriptor radiómico	Mastógrafo	Registro	ROI	Total
24	Mean	5.48 %	312.65 %	0.02 %	312.69 %
25	Median	5.62 %	311.99 %	0.02 %	312.04 %
29	Root Mean Squared	5.45 %	57.00 %	0.02 %	57.26 %

Tabla B.6: Presupuesto de incertidumbre para los descriptores del quinto grupo.

	Descriptor radiómico	Mastógrafo	Registro	ROI	Total
31	Uniformity	0.18 %	0.14 %	0.21 %	0.31 %
58	Gray Level Non Uniformity Normalized (GLSZM)	0.15 %	0.15 %	0.19 %	0.28 %
74	Gray Level Non Uniformity Normalized (GLRLM)	0.17 %	0.14 %	0.21 %	0.31 %

Tabla B.7: Presupuesto de incertidumbre para los descriptores del sexto grupo.

	Descriptor radiómico	Mastógrafo	Registro	ROI	Total
33	Autocorrelation	7.26 %	8.75 %	9.02 %	14.51 %
49	Joint Average	3.77 %	4.41 %	4.66 %	7.44 %
54	Sum Average	3.76 %	4.46 %	4.66 %	7.46 %
60	High Gray Level Zone Emphasis	7.04 %	8.76 %	8.66 %	14.19 %
76	High Gray Level Run Emphasis	7.07 %	8.78 %	8.70 %	14.24 %
87	Short Run High Gray Level Emphasis	7.13 %	9.06 %	8.45 %	14.29 %
100	High Gray Level Emphasis	7.08 %	8.78 %	8.72 %	14.26 %

Tabla B.8: Presupuesto de incertidumbre para los descriptores del séptimo grupo.

	Descriptor radiómico	Mastógrafo	Registro	ROI	Total
37	Contrast	8.43 %	25.93 %	12.81 %	30.13 %
39	Difference Average	4.23 %	13.20 %	6.82 %	15.45 %
40	Difference Entropy	1.30 %	4.67 %	2.48 %	5.45 %
41	Difference Variance	8.40 %	25.72 %	12.62 %	29.86 %
42	Id	0.53 %	2.09 %	1.14 %	2.44 %
43	Idm	0.41 %	1.95 %	1.08 %	2.26 %
44	Idmn	0.22 %	0.26 %	0.17 %	0.38 %
45	Idn	0.41 %	0.89 %	0.49 %	1.09 %
48	Inverse Variance	0.39 %	2.02 %	1.10 %	2.34 %
66	Size Zone Non Uniformity Normalized	1.03 %	4.01 %	1.84 %	4.53 %
67	Small Area Emphasis	0.54 %	2.62 %	1.15 %	2.91 %
71	Zone Percentage	0.77 %	3.68 %	1.72 %	4.13 %
77	Long Run Emphasis	0.67 %	3.60 %	1.74 %	4.05 %
83	Run Length Non Uniformity Normalized	0.43 %	1.97 %	0.90 %	2.21 %
84	Run Percentage	0.23 %	1.15 %	0.53 %	1.29 %
85	Run Variance	0.26 %	1.61 %	0.77 %	1.80 %
86	Short Run Emphasis	0.18 %	0.89 %	0.40 %	0.99 %
96	Dependence Non Uniformity Normalized	1.06 %	2.20 %	1.33 %	2.78 %
97	Dependence Variance	6.09 %	14.56 %	9.36 %	18.34 %
101	Large Dependence Emphasis	4.72 %	14.96 %	9.79 %	18.49 %
105	Small Dependence Emphasis	1.29 %	4.27 %	2.40 %	5.07 %

Tabla B.9: Presupuesto de incertidumbre para los descriptores del octavo grupo.

	Descriptor radiómico	Mastógrafo	Registro	ROI	Total
56	Sum Squares	8.54 %	6.43 %	10.11 %	14.71 %
59	Gray Level Variance	8.03 %	6.56 %	9.68 %	14.18 %
75	Gray Level Variance	8.33 %	6.81 %	9.98 %	14.68 %
99	Gray Level Variance	8.44 %	6.87 %	10.07 %	14.82 %

Tabla B.10: Presupuesto de incertidumbre para los descriptores del noveno grupo.

	Descriptor radiómico	Mastógrafo	Registro	ROI	Total
57	Gray Level Non Uniformity (GLSZM)	3.53 %	4.33 %	39.92 %	40.31 %
73	Gray Level Non Uniformity (GLRLM)	4.57 %	3.09 %	40.91 %	41.28 %
98	Gray Level Non Uniformity (GLDM)	4.95 %	3.51 %	41.24 %	41.69 %

Tabla B.11: Presupuesto de incertidumbre para los descriptores del décimo grupo.

	Descriptor radiómico	Mastógrafo	Registro	ROI	Total
61	Large Area Emphasis	2.83 %	14.79 %	9.03 %	17.56 %
72	Zone Variance	2.18 %	19.37 %	11.97 %	22.87 %

Tabla B.12: Presupuesto de incertidumbre para los descriptores del undécimo grupo.

	Descriptor radiómico	Mastógrafo	Registro	ROI	Total
64	Low Gray Level Zone Emphasis	0.02 %	0.02 %	0.02 %	0.03 %
69	Small Area Low Gray Level Emphasis	0.02 %	0.01 %	0.02 %	0.03 %
80	Low Gray Level Run Emphasis	0.02 %	0.02 %	0.02 %	0.03 %
88	Short Run Low Gray Level Emphasis	0.02 %	0.02 %	0.02 %	0.03 %
104	Low Gray Level Emphasis	0.02 %	0.02 %	0.02 %	0.03 %

Bibliografía

- [1] IARC, “Latest Global Cancer Data: Cancer Burden Rises to 18.1 million new cases and 9.6 million cancer deaths in 2018,” World Health Organization, Lyon, Tech. Rep., 2018.
- [2] Centers for Disease Control and Prevention, “What Is Breast Cancer?” in *2020/02/26*, 2018. [Online]. Available: https://www.cdc.gov/cancer/breast/basic_info/what-is-breast-cancer.htm
- [3] J. T. Bushberg, J. A. Seibert, E. M. Leidholdt, and J. M. Boone, *The Essential Physics of Medical Imaging*, third edit ed. Philadelphia: Lippincott Williams & Wilkins, 2012.
- [4] P. Lambin, R. T. Leijenaar, T. M. Deist, J. Peerlings, E. E. De Jong, J. Van Timmeren, S. Sanduleanu, R. T. Larue, A. J. Even, A. Jochems, Y. Van Wijk, H. Woodruff, J. Van Soest, T. Lustberg, E. Roelofs, W. Van Elmpt, A. Dekker, F. M. Mottaghy, J. E. Wildberger, and S. Walsh, “Radiomics: The bridge between medical imaging and personalized medicine,” *Nature Reviews Clinical Oncology*, vol. 14, no. 12, pp. 749–762, 2017.
- [5] INEGI, “Características de las defunciones registradas en México 2018,” in *Comunicado de prensa No.538/19 31 de Octubre*, CDMX, 2019, pp. 1–62.

- [6] IARC, “New Global Cancer Data: GLOBOCAN 2018 | UICC,” *International Agency for Research on Cancer*, no. September, p. 1, 2018. [Online]. Available: <https://www.uicc.org/new-global-cancer-data-globocan-2018>
- [7] H. Bosmans, A. De Hauwere, K. Lemmens, F. Zanca, H. Thierens, C. Van Ongeval, K. Van Herck, A. Van Steen, P. Martens, L. Bleyen, G. Vande Putte, E. Kellen, G. Mortier, and E. Van Limbergen, “Technical and clinical breast cancer screening performance indicators for computed radiography versus direct digital radiography,” *European Radiology*, vol. 23, no. 10, pp. 2891–2898, 2013.
- [8] R. Li, L. Xing, S. Napel, and D. L. Rubin, *Radiomics and Radiogenomics: Technical Basis and Clinical Applications (Imaging in medical Diagnosis and Therapy)*. Boca Raton: Taylor & Francis Group, 2019.
- [9] S. R. Hoff, J. H. Samset, A. L. Abrahamsen, E. Vigeland, O. Klepp, and S. Hofvind, “Missed and True Interval and Screen-detected Breast Cancers in a Population Based Screening Program,” *Academic Radiology*, vol. 18, no. 4, pp. 454–460, 2011. [Online]. Available: <http://dx.doi.org/10.1016/j.acra.2010.11.014>
- [10] M. Marmot, D. G. Altman, D. A. Cameron, J. A. Dewar, S. G. Thompson, and M. Wilcox, “The benefits and harms of breast cancer screening: An independent review,” *The Lancet*, vol. 380, no. 9855, pp. 1778–1786, 2012. [Online]. Available: [http://dx.doi.org/10.1016/S0140-6736\(12\)61611-0](http://dx.doi.org/10.1016/S0140-6736(12)61611-0)
- [11] J. Katzen and K. Dodelzon, “A review of computer aided detection in mammography,” *Clinical Imaging*, vol. 52, no. August, pp. 305–309, 2018. [Online]. Available: <https://doi.org/10.1016/j.clinimag.2018.08.014>
- [12] F. Valdora, N. Houssami, F. Rossi, M. Calabrese, and A. S. Tagliafico, “Rapid review: radiomics and breast cancer,” *Breast Cancer Research and Treatment*, vol. 169, no. 2, pp. 217–229, 2018.
- [13] Quantitative Imaging Network (QIN), “The Quantitative Imaging Network Annual Report January 2018,” National Cancer Institute, Tech. Rep. January, 2018.

- [14] M. L. Giger, “Artificial Intelligence in Breast Imaging.” Leuven: The 15th International Workshop on Breast Imaging, 2020.
- [15] M. J. Mateos, A. Gastelum, J. Márquez, and M. E. Brandan, “Texture analysis of contrast-enhanced digital mammography (CEDM) images,” in *Breast Imaging. 13th International Workshop, IWDM 2016 Malmö, Sweden, June 2016, Proceedings.*, 2016, pp. 585–592.
- [16] K. R. Mendel, H. Li, L. Lan, C. M. Cahill, V. Rael, H. Abe, and M. L. Giger, “Quantitative texture analysis: robustness of radiomics across two digital mammography manufacturers’ systems,” *Journal of Medical Imaging*, vol. 5, no. 01, p. 1, 2017.
- [17] J. Martínez-Ezquerro and L. Herrera, “Angiogénesis: VEGF/VEGFRs como blancos terapéuticos en el tratamiento contra el cáncer,” *Cancerología*, vol. 1, no. 1, pp. 83–96, 2006.
- [18] M. M. Travieso Aja, M. Rodríguez Rodríguez, S. Alayón Hernández, V. Vega Benítez, and O. P. Luzardo, “Mamografía con realce de contraste mediante técnica de energía dual,” *Radiología*, vol. 56, no. 5, pp. 390–399, 2014.
- [19] Susan G. Komen, “Tumor Characteristics,” in *2020/05/15*, 2019. [Online]. Available: <https://ww5.komen.org/BreastCancer/TumorCharacteristics.html>
- [20] N. Ahmed, “Role of Integrins in Ovarian Cancer,” in *Handbook of Immunohistochemistry and in situ Hybridization of Human Carcinomas*, M. Hayat, Ed. Academic Press, 2006, p. 608.
- [21] A. Piñero-Madrona, L. Polo-García, J. L. Alonso-Romero, J. Salinas-Ramos, M. Canteras-Jordana, J. Sola-Pérez, P. J. Galindo-Fernández, J. Illana-Moreno, J. Bermejo-López, A. Navarrete-Montoya, and P. Parrilla-Paricio, “Características inmunohistoquímicas del cáncer de mama: ¿Hacia una nueva clasificación?” *Cirugía Española*, vol. 84, no. 3, pp. 138–145, 2008. [Online]. Available: [http://dx.doi.org/10.1016/S0009-739X\(08\)72155-9](http://dx.doi.org/10.1016/S0009-739X(08)72155-9)

- [22] D. C. Zaha, “Significance of immunohistochemistry in breast cancer,” *World Journal of Clinical Oncology*, vol. 5, no. 3, pp. 382–392, 2014.
- [23] J. Zhou, H. Tan, Y. Bai, J. Li, Q. Lu, R. Chen, M. Zhang, Q. Feng, and M. Wang, “Evaluating the HER-2 status of breast cancer using mammography radiomics features,” *European Journal of Radiology*, vol. 121, no. January, p. 108718, 2019. [Online]. Available: <https://doi.org/10.1016/j.ejrad.2019.108718>
- [24] C. Dromain and C. Balleyguier, “Contrast-enhanced digital mammography,” in *Digital Mammography*, U. Bick and F. Diekmann, Eds. Berlin: Springer, 2010, ch. Contrast-E, pp. 187–198.
- [25] J. M. Lewin, P. K. Isaacs, V. Vance, and F. J. Larke, “Dual-energy contrast-enhanced digital subtraction mammography: Feasibility,” *Radiology*, vol. 229, no. 1, pp. 261–268, 2003.
- [26] B. Zitová and J. Flusser, “Image registration methods: A survey,” *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, 2003.
- [27] S. Rizzo, F. Botta, S. Raimondi, D. Origgi, C. Fanciullo, A. G. Morganti, and M. Bellomi, “Radiomics: the facts and the challenges of image analysis,” *European Radiology Experimental*, vol. 2, no. 1, 2018.
- [28] R. M. Haralick, K. Shanmugam, and I. Dinstein, “Textural Features for Images Classification,” pp. 610–621, 1973.
- [29] 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, 11 2017.
- [30] I. J. Taylor, E. Deelman, D. B. Gannon, and M. Shields, *Workflows for e-Science*. Springer, London, 2007.

- [31] K. Robinson, H. Li, L. Lan, D. Schacht, and M. Giger, “Radiomics robustness assessment and classification evaluation: A two-stage method demonstrated on multivendor FFDM,” *Medical Physics*, vol. 46, no. 5, pp. 2145–2156, 2019.
- [32] J. Miranda Martín del Campo, “Evaluación de la incertidumbre en datos experimentales,” Ciudad de México, p. 42, 2000.
- [33] I. Farrance and R. Frenkel, “Uncertainty of measurement: A review of the rules for calculating Uncertainty components through functional relationships,” *Clinical Biochemist Reviews*, vol. 33, no. 2, pp. 49–75, 2012.
- [34] G. Rocha, “Notas sobre el Teorema del límite central,” Cd. de México, 2013.
- [35] A. Camargo Martínez, J. García Castro, R. D. Adriana, M. García Minjares, and R. G. Serrano Jiménez, *Estadística descriptiva e inferencial*, agosto 201 ed., SUADyED, Ed. Ciudad de México: Universidad Nacional Autónoma de México, 2017.
- [36] R. Hogan, “Sources of Uncertainty in Measurement for Every Uncertainty Budget,” pp. 1–27, 2015.
- [37] A. Ng, *Course Notes: Machine Learning*. Stanford: Stanford University, 2020.
- [38] J. Amat Rodrigo, “Validación de modelos de regresión,” in *2020/05/23*, C. d. Datos, Ed., 2016. [Online]. Available: https://www.cienciadedatos.net/documentos/30_cross-validation_oneleaveout_bootstrap
- [39] M. A. Marino, K. Pinker, D. Leithner, J. Sung, D. Avendano, E. A. Morris, and M. Jochelson, “Contrast-Enhanced Mammography and Radiomics Analysis for Noninvasive Breast Cancer Characterization: Initial Results,” *Molecular Imaging and Biology*, vol. 22, no. 3, pp. 780–787, 2020.
- [40] D. Chicco and G. Jurman, “The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation,” *BMC Genomics*, vol. 21, no. 1, pp. 1–13, 2020.

- [41] R. Delgado and X. A. Tibau, “Why Cohen’s Kappa should be avoided as performance measure in classification,” *PLoS ONE*, vol. 14, no. 9, pp. 1–26, 2019. [Online]. Available: <http://dx.doi.org/10.1371/journal.pone.0222916>
- [42] M. L. McHugh, “Lessons in biostatistics interrater reliability : the kappa statistic,” *Biochemica Medica*, vol. 22, no. 3, pp. 276–282, 2012. [Online]. Available: <https://hrcak.srce.hr/89395>
- [43] R. Kikinis, S. D. Pieper, and K. G. Vosburgh, “3D Slicer: A Platform for Subject-Specific Image Analysis, Visualization, and Clinical Support,” *Intraoperative Imaging and Image-Guided Therapy*, pp. 277–289, 2014.
- [44] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J. C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka, J. Buatti, S. Aylward, J. V. Miller, S. Pieper, and R. Kikinis, “3D Slicer as an image computing platform for the Quantitative Imaging Network,” *Magnetic Resonance Imaging*, vol. 30, no. 9, pp. 1323–1341, 2012. [Online]. Available: <http://dx.doi.org/10.1016/j.mri.2012.05.001>
- [45] S. Pieper, J. C. Fillion-Robin, and R. Kikinis, “3D Slicer,” pp. 3–7, 2015. [Online]. Available: <https://www.slicer.org/>
- [46] Y. Yuan, M. L. Giger, H. Li, K. Suzuki, and C. Sennett, “A dual-stage method for lesion segmentation on digital mammograms,” *Medical Physics*, vol. 34, no. 11, pp. 4180–4193, 2007.
- [47] N. Mao, P. Yin, Q. Li, Q. Wang, M. Liu, J. Dong, K. Che, Z. Wang, S. Duan, X. Zhang, H. Xie, and N. Hong, “Radiomics nomogram of contrast-enhanced spectral mammography for prediction of axillary lymph node metastasis in breast cancer: a multicenter study,” *EBioMedicine*, 2019.
- [48] J. P. Castillo Lopez, R. Montoya-del Angel, V. Sanchez Goytia, L. Moreno Astudillo, Y. Villaseñor Navarro, and M.-E. Brandan, “Uncertainties associated

to the extraction of texture features in single-energy contrast-enhanced mammography,” *15th International Workshop on Breast Imaging (IWBI2020)*, vol. 1151308, no. May, p. 45, 2020.

- [49] J. Kalpathy-cramer, A. Mamomov, B. Zhao, L. Lu, D. Cherezov, S. Napel, S. Echegaray, D. Rubin, M. Mcnitt-gray, P. Lo, J. C. Sieren, J. Uthoff, S. K. N. Dilger, B. Driscoll, I. Yeung, L. Hadjiiski, K. Cha, Y. Balagurunathan, R. Gillies, and D. Goldgof, “Radiomics of Lung Nodules: A Multi-Institutional Study of Robustness and Agreement of Quantitative Imaging Features,” *Tomography*, vol. 2, no. 4, pp. 430–437, 2016.
- [50] S. Boughorbel, F. Jarray, and M. El-Anbari, “Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric,” *PLoS ONE*, vol. 12, no. 6, pp. 1–17, 2017.
- [51] C. Villarreal-Garza, A. Mohar, J. E. Bargallo-Rocha, F. Lasa-Gonsebatt, N. Reynoso-Noverón, J. Matus-Santos, P. Cabrera, C. Arce-Salinas, F. Lara-Medina, A. Alvarado-Miranda, M. T. Ramírez-Ugalde, and E. Soto-Perez-de Celis, “Molecular Subtypes and Prognosis in Young Mexican Women With Breast Cancer,” *Clinical Breast Cancer*, vol. 17, no. 3, pp. e95–e102, 2017. [Online]. Available: <http://dx.doi.org/10.1016/j.clbc.2016.11.007>