

EVALUACIÓN DE RIESGO DE CÁNCER DE SENO MEDIANTE EL ANÁLISIS
PARENQUIMATOSO DE DIFERENTES VISTAS MAMOGRÁFICAS

LIGIA PAOLA MARTIN SUÁREZ

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICOMECÁNICAS
ESCUELA DE INGENIERÍAS ELÉCTRICA, ELECTRÓNICA Y DE
TELECOMUNICACIONES
BUCARAMANGA
2020

EVALUACIÓN DE RIESGO DE CÁNCER DE SENO MEDIANTE EL ANÁLISIS
PARENQUIMATOSO DE DIFERENTES VISTAS MAMOGRÁFICAS

LIGIA PAOLA MARTIN SUÁREZ

Trabajo de Grado para optar al título de
Ingeniera Electrónica

Director

Said Pertuz

Ph.D, en Ciencias de la Computación

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICOMECÁNICAS
ESCUELA DE INGENIERÍAS ELÉCTRICA, ELECTRÓNICA Y DE
TELECOMUNICACIONES
BUCARAMANGA

2020

Dedicado

A Dios por permitirme el logro de esta meta.

A mis padres, Ligia y Emiliano, mis hermanos Leidy y Juan David por su apoyo e incondicionalidad desde el principio del camino.

A mis primas Katy, Karen y Clara por siempre tener una palabra de aliento en el momento indicado.

A Yasú, por acompañarme cuando debía trasnochar.

A mis amigos, quienes me acompañaron en el trayecto de cada semestre e hicieron del paso por la universidad un camino más bonito y ameno.

CONTENIDO

	pág.
INTRODUCCIÓN	9
1. OBJETIVOS	13
2. MATERIALES Y METODOLOGÍA	14
2.1. BASE DE DATOS	15
2.2. MÉTODOS PROPUESTOS EN EL ESTADO DEL ARTE	17
2.3. ESTRATEGIAS MULTI-VISTA	18
2.3.1. Misma vista - mismo seno	18
2.3.2. Diferente vista - mismo seno	19
2.3.3. Misma vista - diferente seno	20
2.3.4. Diferente vista - diferente seno	21
2.4. IMPLEMENTACIÓN DE ALGORITMOS	22
2.5. SELECCIÓN DE CARACTERÍSTICAS	25
2.5.1. Selección secuencial de características	26
2.6. MODELO PROPUESTO A PARTIR DE ESTRATEGIAS MULTI-VISTA	29
3. EXPERIMENTOS Y RESULTADOS	30
3.1. MÉTODOS MULTI-VISTA	30
3.2. COMPARACIÓN DE RENDIMIENTOS	31
3.3. MODELO PROPUESTO A PARTIR DE ESTRATEGIAS MULTI-VISTA	33
4. CONCLUSIONES	36
BIBLIOGRAFÍA	37

LISTA DE FIGURAS

	pág.
Figura 1. Casos (LCC) - controles (LCC)	19
Figura 2. Casos (LMLO) - controles (LCC)	20
Figura 3. Casos (LMLO) - controles (RMLO)	21
Figura 4. Casos (LCC) - controles (RMLO)	22

LISTA DE TABLAS

	pág.
Tabla 1. Estado del arte en evaluación de riesgo de cáncer de seno	12
Tabla 2. Recopilación de los rendimientos por cada método	32
Tabla 3. Rendimiento de los modelos propuestos a partir de estrategias multi-vista	34

RESUMEN

TÍTULO: EVALUACIÓN DE RIESGO DE CÁNCER DE SENO MEDIANTE EL ANÁLISIS PARENQUIMATOSO DE DIFERENTES VISTAS MAMOGRAFÍAS *

AUTOR: LIGIA PAOLA MARTIN SUÁREZ **

PALABRAS CLAVE: CÁNCER DE SENO, PROCESAMIENTO DE IMÁGENES, MAMOGRAFÍAS, EVALUACIÓN DE RIESGO.

DESCRIPCIÓN:

El cáncer de seno es un problema de gran interés en la comunidad científica y médica debido a su creciente incidencia a nivel mundial. La mamografía es una de las principales herramientas de diagnóstico y se utiliza cada vez más para la evaluación de riesgo de cáncer de seno. Típicamente, los exámenes mamográficos implican la captura de cuatro vistas: las vistas cráneo-caudal (CC) y medio-lateral oblicua (MLO) de los senos derecho e izquierdo. Recientemente, el análisis parenquimatoso ha surgido como una herramienta computacional de análisis de mamografías que funciona mediante la extracción de características de las imágenes para la construcción de modelos de predicción de riesgo.

Actualmente en la literatura hay pocos estudios sobre si el uso de una vista particular, o combinaciones entre ellas, genere mejores resultados para el análisis parenquimatoso. Esta propuesta de trabajo de investigación busca estudiar, implementar y comparar diferentes estrategias para el uso simultáneo de distintas vistas mamográficas para la evaluación de riesgo de cáncer de seno mediante el análisis parenquimatoso.

* Trabajo de grado

** Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingenierías Eléctrica, Electrónica y telecomunicaciones. Director: Said Pertuz, Ph.D, en Ciencias de la Computación.

ABSTRACT

TITLE: BREAST CANCER RISK ASSESSMENT USING PARENCHYMAL ANALYSIS OF DIFFERENT MAMMOGRAPHIC VIEWS *

AUTHOR: LIGIA PAOLA MARTIN SUÁREZ **

KEYWORDS: BREAST CANCER, IMAGE PROCESSING, MAMMOGRAMS, RISK ASSESSMENT.

DESCRIPTION:

Breast cancer is a problem of great interest in the scientific and medical community due to its increasing incidence worldwide. Mammography is one of the main diagnostic tools and is increasingly used for breast cancer risk assessment. Typically, mammographic exams involve the capture of four views: the cranio-caudal (CC) and medio-lateral oblique (MLO) views of the right and left breasts. Recently, parenchymal analysis has emerged as a computational mammography analysis tool that works by extracting features from images for the construction of risk prediction models.

Currently, there are few studies in the literature on whether the use of a particular view, or combinations thereof, generates better results for parenchymal analysis. This research proposal seeks to study, implement, and compare different strategies for the simultaneous use of different mammographic views for breast cancer risk assessment through parenchymal analysis.

* Bachelor Thesis

** Faculty of Physical-Mechanical Engineering; School of Electrical, Electronic and Telecommunications Engineering. Director: Said Pertuz, Ph.D, Computer Science.

INTRODUCCIÓN

El cáncer de seno viene siendo una problemática mundial. Es el segundo cáncer más común en el mundo y el más frecuente entre las mujeres, con cerca de 1.67 millones de casos nuevos diagnosticados en 2012 (25 % de todos los cánceres). Es el cáncer más común en las mujeres, tanto en países desarrollados (794.000 casos) como en los menos desarrollados (883.000 casos) ¹. Las tasas de incidencia varían según las regiones del mundo, con tasas que van desde 27 por 100.000 en el África Central y Asia Oriental, hasta 96 por 100.000 en Europa Occidental. Las tasas de mortalidad también varían siendo mayor la supervivencia en regiones desarrolladas ¹.

Aunque tradicionalmente se investiga sobre el diagnóstico y la detección, la capacidad de estimar el riesgo de una mujer de desarrollar cáncer de seno es cada vez más importante en la práctica clínica. La evaluación del riesgo de cáncer de seno se usa como criterio para formar pautas, ofrecer recomendaciones de detección personalizadas², adaptar tratamientos individuales de cáncer de seno³ y formar estrategias preventivas⁴, especialmente para mujeres asociadas con un riesgo más

¹ Jacques FERLAY y col. "Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012". En: *International journal of cancer* 136.5 (2015), E359-E386.

² Constance D LEHMAN y col. "Screening women at high risk for breast cancer with mammography and magnetic resonance imaging". En: *Cancer* 103.9 (2005), págs. 1898-1905.

³ Constance D LEHMAN y col. "Cancer yield of mammography, MR, and US in high-risk women: prospective multi-institution breast cancer screening study". En: *Radiology* 244.2 (2007), págs. 381-388.

⁴ Claudine SMITH Karen Lisa y ISAACS. "Management of women at increased risk for hereditary breast cancer". En: *Breast disease* 27.1 (2007), págs. 51-67.

alto. La mamografía es una de las principales herramientas de diagnóstico y se utiliza cada vez más para la evaluación de riesgo⁵.

Recientemente el análisis parenquimatoso ha surgido como una herramienta computacional de análisis de mamografía que funciona mediante la extracción de características de las imágenes para la construcción de modelos de predicción de riesgo⁶⁷. Típicamente, los exámenes mamográficos implican la captura de cuatro vistas: las vistas cráneo-caudal (CC) y medio-lateral oblicua (MLO) de los senos derecho e izquierdo⁸. Sin embargo, en la literatura hay pocos trabajos de investigación que determinen si el uso de una vista particular, o combinaciones entre vistas generan mejores resultados en el análisis parenquimatoso. Por ello, este trabajo de investigación busca estudiar, implementar y comparar diferentes estrategias para el uso simultáneo de distintas vistas mamográficas para la evaluación de riesgo de cáncer de seno mediante el análisis parenquimatoso.

La tabla 1 contiene información recopilada de distintos trabajos realizados en los que se hace uso de una vista o múltiples vistas. A continuación se hará una breve

-
- ⁵ Oscar ARAQUE y col. "Selecting the Mammographic-View for the Parenchymal Analysis-Based Breast Cancer Risk Assessment". En: mayo de 2019.
- ⁶ Said PERTUZ y col. "Open Framework for Mammography-based Breast Cancer Risk Assessment". En: *IEEE-EMBS International Conference on Biomedical and Health Informatics* (2019). accepted, págs. 1-4.
- ⁷ Said y SASSI Antti y HOLLI-HELENIUS Kirsi y KAMARAINÉ Joni y RINTA-KIKKA Irina y LAAPERI Anna-Leena y ARPONEN Otso PERTUZ. "Clinical Evaluation of a Fully-automated Parenchymal Analysis Software for Breast Cancer Risk Assessment: a Pilot Study in a Finnish Sample". En: *European Journal of Radiology* (2019), pág. 108710.
- ⁸ PU J. CHENG S. LIU H. ZHENG B. TAN M. "Assessment of a four-view mammographic image feature based fusion model to predict near-term breast cancer risk..". En: *Annals of biomedical engineering*. 2015 ().

clasificación de como emplean las vistas los distintos trabajos del estado del arte.

- **Única vista:** Diversos trabajos realizan sus experimentos usando una de las dos vistas MLO o CC, del mismo seno o de diferente seno.
- **Multi-vista:** En la literatura, numerosos estudios hacen uso de las dos vistas mamográficas de ambos senos. En algunos de los trabajos realizan una combinación entre las vistas y en otros casos, usan vistas por separado de ambos senos.
- **Meta-modelo:** Estos estudios realizan los experimentos teniendo en cuenta la asimetría de los senos o usando redes neuronales.

Los estudios presentados en el estado del arte se tendrán en cuenta para probar las distintas combinaciones y experimentos. Además, serán usados para llegar a una propuesta de modelo multi-vista que mejore la predicción de riesgo.

Tabla 1. Estado del arte en la evaluación de riesgo de cáncer de seno. En estos trabajos son utilizadas las vistas CC y MLO en diferentes métodos.

Análisis	Estudio	Vistas	
		CC	MLO
Única Vista	Byng <i>et al.</i> (1997) ⁹	X	
	Wu <i>et al.</i> (2008) ¹⁰	X	
	Wei <i>et al.</i> (2011) ¹¹	X	
	Häberle <i>et al.</i> (2012) ¹²	X	
	Li <i>et al.</i> (2012) ¹³	X	
	Sun <i>et al.</i> (2015) ¹⁴	X	
	Huo <i>et al.</i> (2000) ¹⁵	X	
	Huo <i>et al.</i> (2002) ¹⁶	X	
	Li <i>et al.</i> (2004) ¹⁷	X	
	Li <i>et al.</i> (2005) ¹⁸	X	
	Li <i>et al.</i> (2007) ¹⁹	X	
	Li <i>et al.</i> (2008) ²⁰	X	
	Li <i>et al.</i> (2012) ²¹	X	
	Li <i>et al.</i> (2014) ²²	X	
	Gierach <i>et al.</i> (2014) ²³	X	
	Chen <i>et al.</i> (2014) ²⁴		X
	Nielsen <i>et al.</i> (2011) ²⁵		X
	Brandt <i>et al.</i> (2011) ²⁶		X
	Karemore <i>et al.</i> (2014) ²⁷		X
	Zheng <i>et al.</i> (2015) ²⁸		X
Multi-vista	Nielsen <i>et al.</i> (2014) ²⁹	X	X
	Pertuz <i>et al.</i> (2019) ³⁰	X	X
Meta-modelo	Torres Mejia <i>et al.</i> (2005) ³¹	X	X
	Manduca <i>et al.</i> (2009) ³²	X	X
	Tan <i>et al.</i> (2015) ³³	X	X
	Tan <i>et al.</i> (2015) ³⁴	X	X

Fuente: Elaboración propia

1. OBJETIVOS

Objetivo general

- Implementar y comparar diferentes estrategias de análisis de imágenes mamográficas a partir de múltiples vistas.

Objetivos específicos

- Implementar diferentes algoritmos del estado del arte que permitan usar combinaciones de las vistas mamográficas para el análisis parenquimatoso.
- Diseñar y llevar a cabo un conjunto de experimentos para evaluar las diferentes estrategias implementadas en una base de datos de imágenes mamográficas.
- Analizar y documentar los resultados para identificar ventajas y desventajas en las diferentes estrategias estudiadas.

2. MATERIALES Y METODOLOGÍA

En el presente proyecto de investigación se estudia el estado del arte con el objetivo de ver como se aborda la utilización de las vistas en cada trabajo. Posteriormente, se propone probar distintas combinaciones de vistas posibles a partir de las cuatro existentes por mamografía⁸.

Para verificar el rendimiento de las combinaciones, y poder concluir si en realidad existe diferencia en el rendimiento de la evaluación de riesgo de cáncer de seno según las vistas usadas en cada estudio, se implementa cada modelo sobre la misma base de datos.

La base de datos usada es la TAYS-I, conformada por 7710 imágenes mamográficas de 277 pacientes que padecen cáncer de seno y 330 pacientes sanas (sección 2.1). Estas imágenes serán utilizadas para extraer las 33 características mediante el software *OpenBreast v1.5*^{7 35} y analizar su rendimiento en las distintas combinaciones implementadas.

Para mejorar el rendimiento en cada método se utilizará la técnica de selección de características (sección 2.5), en este caso se opta por la selección secuencial de características (sección 2.5.1). Para ver si las diferencias entre los métodos propuestos son estadísticamente significativas, se usa la función *AUC COMPARE CORRELATED* de MATLAB, la cual proporciona el valor p, junto con el intervalo de confianza

³⁵ Said PERTUZ y col. "Open Framework for Mammography-based Breast Cancer Risk Assessment". En: *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*. IEEE. 2019, págs. 1-4.

del 95 % por cada par de combinaciones, a partir del valor p se eliminan métodos que no superen el valor del umbral propuesto.

Finalmente, para obtener el rendimiento de la evaluación de riesgo de cáncer de seno en cada estudio que ha superado las pruebas anteriores, se analizan los valores de AUC, con sus respectivos intervalos de confianza del 95 %.

2.1. BASE DE DATOS

Para realizar las pruebas de las diferentes propuestas de combinaciones de vistas, se cuenta con una base de datos conformada por 7710 imágenes mamográficas que corresponden a 607 pacientes, de los cuales 330 son pacientes sanas (controles) y 277 pacientes han sido diagnosticadas con cáncer de seno (casos). Al contar con una base de datos que se puede considerar extensa, se procede a realizar una reducción mediante un proceso de selección, en donde fueron eliminados los casos que cumplían con al menos uno de los siguientes criterios de restricción⁷:

1. **Intervención o tumores malignos previos (n=50):** Los casos que presentaron estas condiciones fueron rechazados, debido a que suelen tener afecciones en el tejido parenquimatoso, por lo cual se verán afectados los resultados.
2. **Síntomas en el momento del cribado (n=30):** Al no saber con seguridad cuál es la condición del paciente, se rechazan estos casos.
3. **Lesión en el seno contralateral (n=6):** El seno contralateral a menudo es utilizado en estudios para el análisis mamográfico, por lo cual fueron rechazados.

4. **No disponibilidad de la anterior ronda de cribado (n=41):** Estos casos fueron rechazados ya que cuando el problema a tratar es la evaluación de riesgo, se necesita de la ronda de cribado anterior para la comparación entre casos y controles.
5. **Sistema mamográfico con poca disponibilidad (n=38):** La base de datos TAYS-I cuenta con variedad en sistemas mamográficos, a pesar de ello, muchos de estos cuentan con poca disponibilidad, por ende se decidió tener en cuenta el sistema con mayor disponibilidad, este es: Senographe Essential (General Electrics).

Finalmente, el número de casos rechazados en los 5 puntos anteriores da un total de n=165, lo cual reduce el número de casos de la base de datos original a 112. Cada caso fue emparejado con un único control, con el objetivo de poder realizar estudios de comparación entre un paciente sano con un paciente enfermo. Este emparejamiento fue realizado bajo los siguientes criterios:

- Mismo año de nacimiento.
- Mismo sistema mamográfico.
- Mismo año de cribado y día más cercado posible de cribado.

Tras el proceso de selección anterior queda como resultado un total de 224 pacientes, de los cuales 112 corresponden a casos y 112 a controles. Cada paciente cuenta con 4 vistas mamográficas (cráneo-caudal [CC] y medio-lateral oblicua [MLO] de los senos derecho e izquierdo⁸), al tratarse de un estudio de evaluación de riesgo de cáncer de seno a partir de múltiples vistas, se hará uso de las 4 vistas por cada paciente. La región de interés (ROI) seleccionada es la región retroareolar (RA), usada en múltiples estudios del estado del arte contemplado en el presente trabajo.

Posteriormente, se presentarán los métodos usados en los trabajos del estado del arte (sección 2.2) y los propuestos en este trabajo (sección 2.3).

2.2. MÉTODOS PROPUESTOS EN EL ESTADO DEL ARTE

Para abordar el tema multi-vista en la evaluación de riesgo de cáncer de seno, se analizaron los documentos reunidos en el estado del arte con el objetivo de estudiar la manera en que cada trabajo emplea las vistas por método.

A continuación se enlista la forma en que han utilizado las vistas algunos de los trabajos:

- Controles (LCC) casos (CC contralateral).³⁶
- Controles (CC del seno afectado), casos (CC seno contralateral).³⁷
- Controles (LCC), casos (RCC).^{38 39}
- Controles y casos (LCC).^{40 41}

³⁶ Jun WEI y col. "Association of computerized mammographic parenchymal pattern measure with breast cancer risk: a pilot case-control study". En: *Radiology* 260.1 (2011), págs. 42-49.

³⁷ Lothar HÄBERLE y col. "Characterizing mammographic images by using generic texture features". En: *Breast Cancer Research* 14.2 (2012), R59.

³⁸ Zhimin HUO y col. "Computerized analysis of digitized mammograms of BRCA1 and BRCA2 gene mutation carriers". En: *Radiology* 225.2 (2002), págs. 519-526.

³⁹ Tzu Liang y QIAN Wei y ZHANG Jianying y SALTZSTEIN-Edward C y ZHENG Bin y LURE Fleming y YU Hui y ZHOU Shi SUN Wenqing y TSENG. "Using multiscale texture and density features for near-term breast cancer risk analysis". En: *Medical physics* 42.6Part1 (2015), págs. 2853-2862.

⁴⁰ Zhimin HUO y col. "Computerized analysis of mammographic parenchymal patterns for breast cancer risk assessment: feature selection". En: *Medical Physics* 27.1 (2000), págs. 4-12.

⁴¹ Hui LI y col. "Computerized Texture Analysis of Mammographic Parenchymal Patterns of Digitized

- Controles (LCC), casos (LMLO).⁴²
- Controles (LMLO), casos (RMLO).^{43 44}

Después del análisis de esta sección, se propone probar 4 grupos de métodos con distintas combinaciones que incluyen las 4 vistas (sección 2.3).

2.3. ESTRATEGIAS MULTI-VISTA

Con el fin de incluir las 4 vistas mamográficas en las estrategias propuestas, se alternaron las vistas CC y MLO tanto de un mismo seno como de diferentes senos. Esto con el fin de abarcar suficientes combinaciones para la posterior implementación sobre la misma base de datos definida en la (sección 2.1).

A continuación se presentan los 4 grupos con sus respectivas combinaciones.

2.3.1. Misma vista - mismo seno Cada propuesta hace uso del mismo seno y la misma vista tanto en casos como en controles. Dentro del grupo se analizan las 4 combinaciones, las cuales están conformadas por ambos senos y ambas vistas.

En este grupo se tienen las siguientes 4 combinaciones:

Mammograms¹". En: *Academic Radiology* 12.7 (2005), págs. 863-873.

⁴² Mads NIELSEN y col. "Mammographic texture resemblance generalizes as an independent risk factor for breast cancer". En: *Breast Cancer Research* 16.2 (2014), R37.

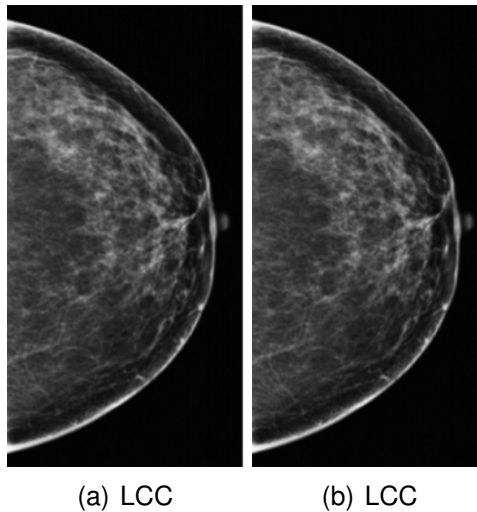
⁴³ Gopal KAREMORE y col. "A method to determine the mammographic regions that show early changes due to the development of breast cancer". En: *Physics in Medicine & Biology* 59.22 (2014), pág. 6759.

⁴⁴ Sami S BRANDT y col. "An anatomically oriented breast coordinate system for mammogram analysis". En: *IEEE transactions on medical imaging* 30.10 (2011), págs. 1841-1851.

- Casos (LCC) - controles (LCC).
- Casos (LMLO) - controles (LMLO).
- Casos (RCC) - controles (RCC).
- Casos (RMLO) - controles (RMLO).

A manera de ejemplo se muestra la primera combinación en la Fig 1.

Figura 1. Casos (LCC) - controles (LCC)



Fuente: Replicada de ⁴⁵

2.3.2. Diferente vista - mismo seno Este grupo de combinaciones usa el mismo seno tanto en casos como en controles y alterna entre casos y controles las vistas CC y MLO.

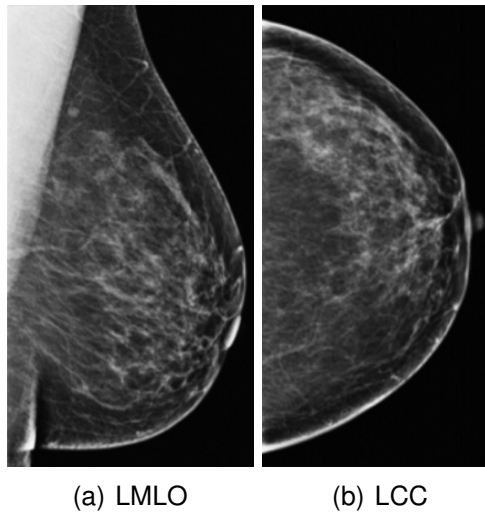
Las 4 combinaciones resultantes son:

- Casos (LMLO) - controles (LCC).

- Casos (LCC) - controles (LMLO).
- Casos (RMLO) - controles (RCC).
- Casos (RCC) - controles (RMLO).

En la Fig 2, se pueden observar las mamografías usadas en la primer combinación de este grupo para casos y controles respectivamente.

Figura 2. Casos (LMLO) - controles (LCC)



Fuente: Replicada de ⁴⁶

2.3.3. Misma vista - diferente seno Las combinaciones de este grupo mantiene estable la vista en los casos y controles, mientras varían el seno para la comparación entre el paciente sano y el paciente con cáncer.

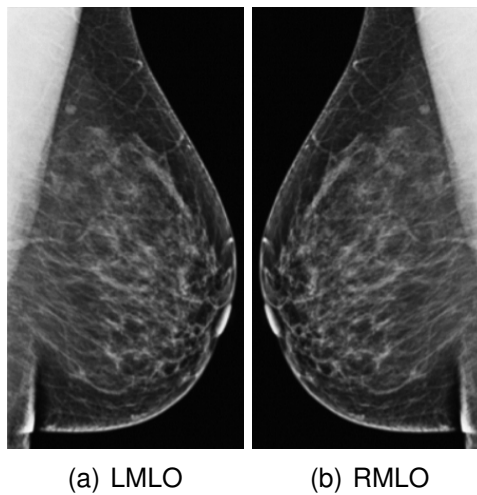
Las combinaciones bajo los criterios de este grupo son:

- Casos (LMLO) - controles (RMLO).
- Casos (RMLO) - controles (LMLO).

- Casos (LCC) - controles (RCC).
- Casos (RCC) - controles (LCC).

Las mamografías de la primera combinación de éste grupo se pueden observar en la Fig 3.

Figura 3. Casos (LMLO) - controles (RMLO)



Fuente: Replicada de ⁴⁷

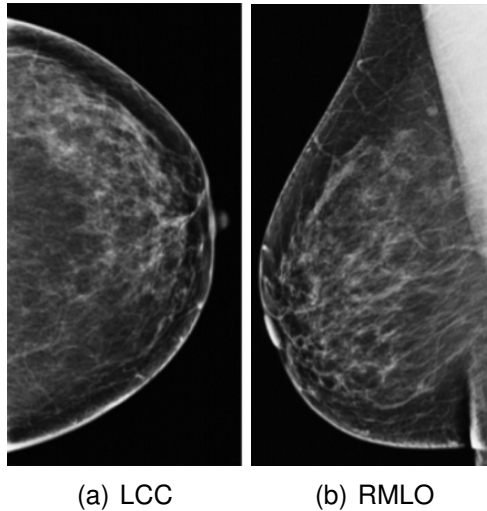
2.3.4. Diferente vista - diferente seno Este grupo de combinaciones hace uso de diferente vista y diferente seno en cada caso-control.

Las combinaciones resultantes son:

- Casos (LCC) - controles (RMLO).
- Casos (RMLO) - controles (LCC).
- Casos (RCC) - controles (LMLO).
- Casos (LMLO) - controles (RCC).

La combinación casos (LCC) controles (RMLO) se presenta en la Fig 4.

Figura 4. Casos (LCC) - controles (RMLO)



Fuente: Replicada de ⁴⁸

2.4. IMPLEMENTACIÓN DE ALGORITMOS

En esta sección se presenta la implementación del algoritmo para el análisis computarizado de imágenes mamográficas. Este proceso es realizado con el software *OpenBreast v1.5* ^{7 35}.

La implementación es conformada por 5 pasos esenciales los cuales se detallarán a continuación:

1. Segmentación del seno

La segmentación del seno se realiza con el objetivo de detectar la región mamaria, esto debido a que los estudios se deben realizar sobre el seno y no en toda la mamografía, pues existen regiones que afectan el análisis.

Los elementos dentro de la mamografía que se buscan eliminar con la segmentación del seno son: El fondo, la pared pectoral (en caso de tratarse de una mamografía de vista MLO), el pezón, entre otros. Para la eliminación de los elementos se usan diversos algoritmos^{49 50 51}.

2. Detección de la región de interés (ROI)

Seguido de la segmentación del seno se selecciona una región de interés (ROI). Estudios anteriores sugieren la detección de una ROI en lugar de la utilización del seno en totalidad, debido a que los análisis sobre una ROI pueden tener mejoras estadísticamente significativas en la evaluación de riesgo de cáncer de seno³⁵. La región de interés detectada será utilizada para la posterior extracción de características.

La ROI seleccionada para los métodos multi-vista será la región situada justo detrás del pezón, la región retroareolar (RA)⁵¹.

Para la elección de la ROI se tuvieron en cuenta los trabajos contemplados en el estado del arte, estudios previos han demostrado que la asociación de biomarcadores basados en imágenes es mayor en la zona inmediatamente detrás

⁴⁹ Dana H BALLARD. "Generalizing the Hough transform to detect arbitrary shapes". En: *Pattern recognition* 13.2 (1981), págs. 111-122.

⁵⁰ Brad M KELLER y col. "Estimation of breast percent density in raw and processed full field digital mammography images via adaptive fuzzy c-means clustering and support vector machine segmentation". En: *Medical physics* 39.8 (2012), págs. 4903-4917.

⁵¹ Said TORRES Germán F y PERTUZ. "Automatic detection of the retroareolar region in x-ray mammography images". En: *VII Latin American Congress on Biomedical Engineering CLAIB 2016, Bucaramanga, Santander, Colombia, October 26th-28th, 2016*. Springer. 2017, págs. 157-160.

del pezón⁵². Varios investigadores también han desarrollado métodos para el análisis de tejido parenquimatoso en la región (RA)³⁶, además la región (RA) puede adaptarse a una amplia variedad de formas y tamaños del seno⁵¹. En su mayoría los trabajos del estado del arte realizan la implementación de los métodos sobre la región (RA)^{40 38 53 54 55 56}.

3. Extracción de características

La extracción de características, tiene por objetivo describir los patrones de textura visual en la ROI seleccionada previamente en términos de mediciones cuantitativas y reproducibles³⁵.

Los métodos de extracción de características se pueden clasificar en 5 grupos³⁵: *statistical features* (STA), *gray-level co-occurrence features* (GLC), *gray-level run-length features* (GLR), *gradient-based features* (GRA) and *spatial-frequency analysis* (SFA), de los 5 grupos a partir de *OpenBreast v1.5* se obtienen 33 características en cada una de las 4 vistas, las cuales serán ne-

⁵² Hui LI y col. "Computerized analysis of mammographic parenchymal patterns for assessing breast cancer risk: effect of ROI size and location". En: *Medical Physics* 31.3 (2004), págs. 549-555.

⁵³ Hui LI y col. "Computerized analysis of mammographic parenchymal patterns on a large clinical dataset of full-field digital mammograms: robustness study with two high-risk datasets". En: *Journal of digital imaging* 25.5 (2012), págs. 591-598.

⁵⁴ Hui LI y col. "Power spectral analysis of mammographic parenchymal patterns for breast cancer risk assessment". En: *Journal of digital imaging* 21.2 (2008), págs. 145-152.

⁵⁵ Hui LI y col. "Fractal analysis of mammographic parenchymal patterns in breast cancer risk assessment". En: *Academic radiology* 14.5 (2007), págs. 513-521.

⁵⁶ Yi-Ta WU y col. "Comparison of mammographic parenchymal patterns of normal subjects and breast cancer patients". En: *Medical Imaging 2008: Computer-Aided Diagnosis*. Vol. 6915. International Society for Optics y Photonics. 2008, pág. 691520.

cesarias para aplicar las combinaciones de las vistas y finalmente obtener el rendimiento de cada método.

4. **Métodos multi-vista**

Una vez realizada la extracción de características se implementan las combinaciones multi-vistas expuestas anteriormente (sección 2.3).

Este paso es esencial para corroborar o descartar la relevancia que tiene el usar ciertas combinaciones de vistas en los casos y controles.

5. **Puntuación de riesgo**

La puntuación de riesgo como paso final en la implementación del algoritmo es sumamente importante, ya que estima el nivel de riesgo ($r \in [0, 1]$) de la imagen de entrada de prueba, donde 0 representa el riesgo más bajo y 1 es el riesgo más alto³⁵.

Esta medida también será hallada mediante el software *OpenBreast v1.5*.

2.5. SELECCIÓN DE CARACTERÍSTICAS

La selección de características tiene por objetivo reducir las entradas de los datos a un tamaño apropiado para su procesamiento y análisis. La capacidad de aplicar la selección de características es esencial para un análisis eficiente, ya que los conjuntos de datos suelen contener mucha más información de la necesaria para la generación del modelo, ocasionando degradar la calidad de los patrones a detectar, esto es debido a la existencia de variables correlacionadas, redundantes, irrelevantes o ruidosas, las cuales dificultan la detección de patrones significativos a partir

de los datos⁵⁷.

La selección consiste en obtener el subconjunto formado por las características más relevantes del conjunto original de características⁵⁸.

Entre las razones más comunes para realizar el proceso de selección de características a un conjunto de datos están:

- **Evitar el *overfitting*:** El *overfitting* o sobreajuste ocurre al entrenar un modelo de aprendizaje automático con una gran cantidad de variables, lo que ocasiona que el algoritmo se ajuste al aprendizaje de casos particulares y pierda poder predictivo ante nuevos datos de entrada.
- **Reducir el tiempo de entrenamiento:** En la literatura existen diversos métodos de selección de características, sin embargo, en este trabajo se realizará el proceso mediante el método de selección secuencial de características.

2.5.1. Selección secuencial de características La selección secuencial de características utiliza un algoritmo de búsqueda para intentar hallar el mejor subconjunto de características⁵⁸, por tanto este método es comúnmente utilizado para la

⁵⁷ Augusto PEREIRA. "Selección de características para el reconocimiento de patrones con datos de alta dimensionalidad en fusión nuclear". Tesis doct. UNED. Universidad Nacional de Educación a Distancia (España), 2015.

⁵⁸ Maria Lucia VIOLINI. "Selección de características. Su aplicación a texturas". Tesis doct. 2014, págs. 8-25.

reducción de variables⁵⁹.

El método se compone de dos aspectos principales: el primero busca la reducción de características en el subconjunto de selección y el segundo es el algoritmo de búsqueda secuencial cuyo objetivo es eliminar o agregar características de un subconjunto mientras se está evaluando el criterio. Las búsquedas secuenciales se mueven en una sola dirección, siempre incrementando o reduciendo el conjunto de candidatos⁶⁰.

El método tiene dos variantes:

1. **Selección secuencial hacia adelante (SFS):** Es un algoritmo de búsqueda secuencial hacia adelante. Parte del conjunto vacío y adiciona sucesivamente la característica más significativa en relación al subconjunto obtenido en la etapa anterior.

Este algoritmo selecciona un subconjunto de tamaño d de un conjunto original de tamaño D , donde $d < D$. La selección se realiza mediante una búsqueda secuencial del “mejor” subconjunto.

En cada una de las iteraciones siguientes se agrega la “mejor” característica (entendiendo por “mejor” característica a aquella que en combinación con las otras características ya seleccionadas maximiza la función criterio).

Cuando una característica es agregada al subconjunto ya no puede ser removida del mismo. El algoritmo finaliza cuando el subconjunto seleccionado

⁵⁹ Christian y VAN-Patrick RÜCKSTIEB Thomas y OSENDORFER. “Sequential feature selection for classification”. En: *Australasian Joint Conference on Artificial Intelligence*. Springer. 2011, págs. 132-141.

⁶⁰ F J FERRI y col. *Comparative Study of Techniques for Large-Scale Feature Selection*. Inf. téc.

posee d características ⁵⁸.

2. **Selección secuencial hacia atrás (SBS):** Es un algoritmo de búsqueda secuencial hacia atrás. Realiza el proceso inverso al SFS. Parte del conjunto completo de características y remueve sucesivamente las características menos significativas en relación al subconjunto obtenido en la etapa anterior.

Este algoritmo realiza un proceso contrario al anterior, en cada una de las iteraciones remueve la “peor” característica (entendiendo por “peor” característica a aquella que siendo removida del subconjunto seleccionado maximiza la función criterio).

Una vez es removida una característica del subconjunto, ya no puede volver a formar parte del mismo. El algoritmo finaliza cuando el subconjunto seleccionado posee d características⁵⁸.

Los métodos SFS y SBS sufren el problema llamado efecto de anidamiento⁶¹. Esto significa que en el caso de SFS las características seleccionadas no pueden ser descartadas más tarde, y en el caso de SBS las características descartadas no pueden ser nuevamente seleccionadas. Esto resulta en subconjuntos de características anidados sin posibilidad de realizar correcciones en pasos posteriores, haciendo que el resultado final generalmente sea lejano al óptimo.

Una manera de abordar este problema es combinar SFS y SBS durante la ejecución, haciendo adiciones o eliminaciones de características. Esta técnica es conocida como regresión escalonada, la cual consiste en la adición de características al subconjunto candidato, luego se analiza el subconjunto buscando si alguna de

⁶¹ Pavel PUDIL, Jana Novovičová y Josef Kittler. “Floating search methods in feature selection”. En: *Pattern recognition letters* 15.11 (1994), págs. 1119-1125.

las características ha reducido su significancia en el modelo por debajo del criterio especificado⁶².

El método precisa del cumplimiento de dos criterios: el primero para añadir variables y el segundo para eliminarlas, el primero debe ser menor al segundo para evitar que el proceso entre en un ciclo infinito. Este algoritmo tiene mayor probabilidad de encontrar la solución óptima o próxima en relación a las versiones más simples de búsqueda secuencial.

2.6. MODELO PROPUESTO A PARTIR DE ESTRATEGIAS MULTI-VISTA

Una vez estudiadas las combinaciones presentadas por el estado del arte y probadas las combinaciones multi-vistas de la (sección 2.3), se propone implementar un modelo a partir de las combinaciones de los grupos que otorguen mejor rendimiento.

⁶² Leland WILKINSON. "Tests of significance in stepwise regression." En: *Psychological bulletin* 86.1 (1979), pág. 168.

3. EXPERIMENTOS Y RESULTADOS

Esta sección mostrará los experimentos realizados para la implementación de los métodos multi-vista y la comparación entre los rendimientos de estos métodos. La técnica utilizada para evaluar los resultados del análisis estadístico que se usó fue validación cruzada (CV) por sus siglas en inglés (*cross-validation*) de 5 iteraciones y el desempeño de cada método fue medido en términos del área bajo la curva (ROC) AUC.

Para determinar si las diferencias entre los métodos son estadísticamente significativas se utilizó la prueba de DeLong⁶³. El valor de umbral α para la prueba DeLong fue definido como $\alpha = 0,05$ comúnmente utilizado en estudios bioestadísticos⁶⁴.

Además, se mostrarán los resultados del metamodelo propuesto a partir de los métodos con mejor desempeño, los cuales serán determinados al realizar la comparación de rendimientos entre las estrategias multi-vistas en los dos escenarios estudiados.

3.1. MÉTODOS MULTI-VISTA

Las combinaciones de estos métodos son el resultado de variar las vistas existentes o alternar los senos para cada estrategia.

⁶³ David M y CLARKE PEARSON Daniel L DELONG Elizabeth R y DELONG. "Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach". En: *Biometrics* (1988), págs. 837-845.

⁶⁴ Magdalena CASTRO. "Bioestadística aplicada en investigación clínica: conceptos básicos". En: *Revista Médica Clínica Las Condes* 30.1 (2019), págs. 50-65.

Para cada combinación de los 4 grupos se realizaron dos variantes: en la primera se usaron las 33 características obtenidas a través del software *OpenBreast v1.5* y en la segunda se usó un grupo de características seleccionado con la técnica de selección secuencial de características (sección: 2.5.1).

Para medir los rendimientos se usó un mismo experimento. En ambos casos se utilizó el modelo de regresión logística, este proceso consiste en pasar los arreglos con los datos de las características (el primero con las 33 y el segundo con el subconjunto resultante de la selección secuencial de características) de las vistas utilizadas en cada combinación, a partir de estos datos se obtiene un vector de puntajes de predicción de riesgo utilizando validación cruzada, esta divide los datos en 2 partes: la primera con el 80 % para entrenamiento y la segunda con el 20 % para validación. Por medio del vector de predicción de riesgo se realiza la comparación con los datos reales y se obtiene el valor de AUC con un intervalo de confianza del 95 %.

El interés del trabajo de investigación es trabajar con distintas combinaciones de vistas, bajo la misma base de datos, para analizar el impacto de las vistas en la evaluación de riesgo.

3.2. COMPARACIÓN DE RENDIMIENTOS

Habiendo analizado cada uno de los métodos expuestos anteriormente, se busca hacer una comparación entre los rendimientos de los 4 métodos tanto con la selección de características como sin ella.

En la tabla 2, se observa la recopilación de los rendimientos de cada uno de los métodos.

Tabla 2. Recopilación de los rendimientos por cada método.

Método	Combinaciones	AUC	95 % CI
Misma vista-mismo seno	LCC-LCC	0.4339	[0.3590, 0.5089]
	LMLO-LMLO	0.4888	[0.4130, 0.5645]
	RCC-RCC	0.4501	[0.3747, 0.5254]
	RMLO-RMLO	0.4781	[0.4023, 0.5537]
Misma vista-mismo seno Selección de características	LCC-LCC	0.4020	[0.3280, 0.4760]
	LMLO-LMLO	0.4807	[0.4051, 0.5565]
	RCC-RCC	0.4317	[0.3568, 0.5066]
	RMLO-RMLO	0.4295	[0.3547, 0.5044]
Diferente vista-mismo seno *	LMLO-LCC	0.8026	[0.7448, 0.8603]
	LCC-LMLO	0.7972	[0.7387, 0.8556]
	RMLO-RCC	0.8082	[0.7512, 0.8652]
	RCC-RMLO	0.7488	[0.6850, 0.8127]
Diferente vista-mismo seno Selección de características **	LMLO-LCC	0.8297	[0.7757, 0.8838]
	LCC-LMLO	0.7818	[0.7216, 0.8421]
	RMLO-RCC	0.7854	[0.7256, 0.8453]
	RCC-RMLO	0.7366	[0.6715, 0.8016]
Misma vista-diferente seno	LMLO-RMLO	0.4108	[0.3365, 0.4851]
	RMLO-LMLO	0.6197	[0.5466, 0.6928]
	LCC-RCC	0.5169	[0.4412, 0.5927]
	RCC-LCC	0.5065	[0.4307, 0.5823]
Misma vista-diferente seno Selección de características	LMLO-RMLO	0.4753	[0.3997, 0.5510]
	RMLO-LMLO	0.5509	[0.4756, 0.6262]
	LCC-RCC	0.5210	[0.4453, 0.5967]
	RCC-LCC	0.4712	[0.3956, 0.5468]
Diferente vista-diferente seno *	LCC-RMLO	0.8010	[0.7431, 0.8589]
	RMLO-LCC	0.8402	[0.7878, 0.8927]
	RCC-LMLO	0.7873	[0.7277, 0.8470]
	LMLO-RCC	0.7919	[0.7328, 0.8510]
Diferente vista-diferente seno Selección de características **	LCC-RMLO	0.8166	[0.7607, 0.8725]
	RMLO-LCC	0.8698	[0.8223, 0.9175]
	RCC-LMLO	0.7336	[0.6682, 0.7989]
	LMLO-RCC	0.8007	[0.7427, 0.8587]

Fuente: Elaboración propia

La comparación entre los rendimientos de los métodos se realizó con el fin de saber si las diferencias son estadísticamente significativas.

El propósito de la prueba DeLong es comparar dos curvas ROC correlacionadas,

al realizar los experimentos bajo la misma base de datos se cumple la correlación entre las curvas y se puede realizar la prueba. Este método hace comparaciones entre pares de métodos y proporciona el valor AUC en cada método a un intervalo de confianza del 95 % y el p-valor para la diferencia de los AUC.

Para determinar si las diferencias son estadísticamente significativas se tiene en cuenta el valor de umbral $\alpha = 0,5$ definido anteriormente. Las comparaciones cuyo p-valor no supere el umbral, son las que demuestran una diferencia que se puede considerar estadísticamente significativa.

Una vez realizada la prueba, los resultados demuestran que existen diferencias significativas de dos de los grupos, con respecto a los otros dos grupos de combinaciones. En la tabla 2 se muestran estos grupos identificados con un asterisco (*).

Los métodos con selección de características con un desempeño superior, se pueden identificar en la tabla 2 con doble asterisco (**). Sin embargo, se puede observar que el comportamiento fue el mismo en los estudios basados en las 33 características, como en los estudios con el subconjunto de selección de características, dado que la diferencia en los rendimientos entre los dos estudios una vez realizada la prueba DeLong no es considerada estadísticamente significativa.

3.3. MODELO PROPUESTO A PARTIR DE ESTRATEGIAS MULTI-VISTA

El modelo propuesto se implementó a partir de las combinaciones de los grupos que arrojaron un rendimiento estadísticamente significativo respecto a los demás grupos. Estos grupos son los identificados en la tabla 2 con un asterisco (*).

Cada combinación se tomó como modelo para el entrenamiento (*training*) y para la prueba (*test*) se utilizaron las dos vistas de la combinación pasando la primera vista

como casos y controles, de igual manera la segunda vista como casos y controles. Esto con el objetivo de utilizar el modelo en nuevos pacientes, ya que en estos exámenes no se tiene previo conocimiento de si la paciente es un caso o un control.

A partir de este proceso se obtienen 2 vectores de puntuación (*score*). Para obtener el vector de puntuación del modelo se realiza un postprocesamiento hallando el máximo entre los dos vectores, este vector resultante será el utilizado para estimar el nivel de riesgo de un nuevo examen.

Los rendimientos a partir del vector de puntuación resultante del modelo para cada combinación se muestran en la tabla 3.

Tabla 3. Rendimiento de los modelos propuestos a partir de estrategias multi-vista.

Método	Combinaciones	AUC	95 % CI
Diferente vista-mismo seno	LMLO - LCC	0.5067	[0.4309, 0.5825]
	LCC - LMLO	0.5222	[0.4465, 0.5978]
	RMLO - RCC	0.4779	[0.4022, 0.5536]
	RCC - RMLO	0.4389	[0.3638, 0.5140]
Diferente vista-diferente seno	LCC - RMLO	0.5081	[0.4324, 0.5839]
	RMLO - LCC	0.5226	[0.4469, 0.5983]
	RCC - LMLO	0.5067	[0.4309, 0.5825]
	LMLO - RCC	0.4874	[0.4117, 0.5632]

Fuente: Elaboración propia

Con el propósito de interpretar los resultados correctamente, es necesario aclarar que los métodos cuyos valores $AUC < 0.5$ no poseen fuerza predictiva en la evaluación de riesgo. En la tabla 3 se pueden observar valores $AUC > 0.5$, los cuales aparentemente muestran una mejoría en la evaluación, sin embargo, se decide realizar una prueba estadística para comprobar si la mejora es significativa.

Se realizó la prueba de DeLong con $\alpha = 0,05$ y se encontró que dos de los pares comparados obtienen un $p < \alpha$. La combinación RCC-RMLO perteneciente al grupo diferente vista mismo seno mostró diferencias significativas con respecto a LCC-LMLO y RCC-LMLO con $p= 0.0455$, $p=0.0213$ respectivamente. Sin embargo, pierde relevancia al no poseer diferencias significativas con las demás combinaciones de la tabla 3.

Por lo anterior se demuestra que ninguna combinación es significativamente mejor que las demás.

4. CONCLUSIONES

Se realizó la implementación y posterior comparación entre métodos multi-vista, a fin de determinar la existencia de combinaciones de vistas que generen un mejor desempeño en la evaluación de riesgo de cáncer de seno.

La selección secuencial de características logró aumentar el valor de AUC en 8 de las combinaciones de los métodos, y disminuirla en las 8 combinaciones restantes. A pesar de esto, se considera que las diferencias no son estadísticamente significativas ya que no superan el umbral definido para la prueba DeLong.

Realizando el análisis comparativo bajo la prueba de DeLong, se lograron obtener diferencias significativas entre los métodos. Las combinaciones de los métodos diferente vista - misma seno y diferente vista - diferente seno fueron las que alcanzaron un mejor rendimiento.

Se demostró que ninguna combinación de vistas muestra un mejor rendimiento con respecto a las demás, al realizar el análisis a partir del modelo propuesto usando estrategias multi-vistas, el cual puede ser usado en nuevos exámenes.

BIBLIOGRAFÍA

ARAQUE, Oscar y col. "Selecting the Mammographic-View for the Parenchymal Analysis-Based Breast Cancer Risk Assessment". En: mayo de 2019 (vid. pág. 10).

BALLARD, Dana H. "Generalizing the Hough transform to detect arbitrary shapes". En: *Pattern recognition* 13.2 (1981), págs. 111-122 (vid. pág. 23).

BRANDT, Sami S y col. "An anatomically oriented breast coordinate system for mammogram analysis". En: *IEEE transactions on medical imaging* 30.10 (2011), págs. 1841-1851 (vid. págs. 12, 18).

BYNG, Jeffrey W y col. "Automated analysis of mammographic densities and breast carcinoma risk". En: *Cancer: Interdisciplinary International Journal of the American Cancer Society* 80.1 (1997), págs. 66-74 (vid. pág. 12).

CASTRO, Magdalena. "Bioestadística aplicada en investigación clínica: conceptos básicos". En: *Revista Médica Clínica Las Condes* 30.1 (2019), págs. 50-65 (vid. pág. 30).

CHEN, Xin y col. "Breast cancer risk analysis based on a novel segmentation framework for digital mammograms". En: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2014, págs. 536-543 (vid. pág. 12).

DELONG Elizabeth R y DELONG, David M y CLARKE PEARSON Daniel L. "Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach". En: *Biometrics* (1988), págs. 837-845 (vid. pág. 30).

FERLAY, Jacques y col. "Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012". En: *International journal of cancer* 136.5 (2015), E359-E386 (vid. pág. 9).

FERRI, F J y col. *Comparative Study of Techniques for Large-Scale Feature Selection*. Inf. téc. (vid. pág. 27).

GIERACH, Gretchen L y col. "Relationships between computer-extracted mammographic texture pattern features and BRCA1/2 mutation status: a cross-sectional study". En: *Breast Cancer Research* 16.4 (2014), pág. 424 (vid. pág. 12).

HÄBERLE, Lothar y col. "Characterizing mammographic images by using generic texture features". En: *Breast Cancer Research* 14.2 (2012), R59 (vid. págs. 12, 17).

HUO, Zhimin y col. "Computerized analysis of digitized mammograms of BRCA1 and BRCA2 gene mutation carriers". En: *Radiology* 225.2 (2002), págs. 519-526 (vid. págs. 12, 17, 24).

HUO, Zhimin y col. "Computerized analysis of mammographic parenchymal patterns for breast cancer risk assessment: feature selection". En: *Medical Physics* 27.1 (2000), págs. 4-12 (vid. págs. 12, 17, 24).

KAREMORE, Gopal y col. "A method to determine the mammographic regions that show early changes due to the development of breast cancer". En: *Physics in Medicine & Biology* 59.22 (2014), pág. 6759 (vid. págs. 12, 18).

KELLER, Brad M y col. "Estimation of breast percent density in raw and processed full field digital mammography images via adaptive fuzzy c-means clustering and support vector machine segmentation". En: *Medical physics* 39.8 (2012), págs. 4903-4917 (vid. pág. 23).

LEHMAN, Constance D y col. "Cancer yield of mammography, MR, and US in high-risk women: prospective multi-institution breast cancer screening study". En: *Radiology* 244.2 (2007), págs. 381-388 (vid. pág. 9).

LEHMAN, Constance D y col. "Screening women at high risk for breast cancer with mammography and magnetic resonance imaging". En: *Cancer* 103.9 (2005), págs. 1898-1905 (vid. pág. 9).

LI, Hui y col. "Comparative analysis of image-based phenotypes of mammographic density and parenchymal patterns in distinguishing between BRCA1/2 cases, unilateral cancer cases, and controls". En: *Journal of Medical Imaging* 1.3 (2014), pág. 031009 (vid. pág. 12).

LI, Hui y col. "Computerized analysis of mammographic parenchymal patterns for assessing breast cancer risk: effect of ROI size and location". En: *Medical Physics* 31.3 (2004), págs. 549-555 (vid. págs. 12, 24).

LI, Hui y col. "Computerized analysis of mammographic parenchymal patterns on a large clinical dataset of full-field digital mammograms: robustness study with two high-risk datasets". En: *Journal of digital imaging* 25.5 (2012), págs. 591-598 (vid. págs. 12, 24).

LI, Hui y col. "Computerized Texture Analysis of Mammographic Parenchymal Patterns of Digitized Mammograms¹". En: *Academic Radiology* 12.7 (2005), págs. 863-873 (vid. págs. 12, 17).

LI, Hui y col. "Fractal analysis of mammographic parenchymal patterns in breast cancer risk assessment". En: *Academic radiology* 14.5 (2007), págs. 513-521 (vid. págs. 12, 24).

LI, Hui y col. "Power spectral analysis of mammographic parenchymal patterns for breast cancer risk assessment". En: *Journal of digital imaging* 21.2 (2008), págs. 145-152 (vid. págs. 12, 24).

MANDUCA, Armando y col. "Texture features from mammographic images and risk of breast cancer". En: *Cancer Epidemiology and Prevention Biomarkers* 18.3 (2009), págs. 837-845 (vid. pág. 12).

NIELSEN, Mads y col. "A novel and automatic mammographic texture resemblance marker is an independent risk factor for breast cancer". En: *Cancer Epidemiology* 35.4 (2011), págs. 381-387 (vid. pág. 12).

NIELSEN, Mads y col. "Mammographic texture resemblance generalizes as an independent risk factor for breast cancer". En: *Breast Cancer Research* 16.2 (2014), R37 (vid. págs. 12, 18).

PEREIRA, Augusto. "Selección de características para el reconocimiento de patrones con datos de alta dimensionalidad en fusión nuclear". Tesis doct. UNED. Universidad Nacional de Educación a Distancia (España), 2015 (vid. pág. 26).

PERTUZ, Said y col. "Open Framework for Mammography-based Breast Cancer Risk Assessment". En: *IEEE-EMBS International Conference on Biomedical and Health Informatics* (2019). accepted, págs. 1-4 (vid. págs. 10, 12).

PERTUZ, Said y col. "Open Framework for Mammography-based Breast Cancer Risk Assessment". En: *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*. IEEE. 2019, págs. 1-4 (vid. págs. 14, 22-25).

PERTUZ, Said y SASSI Antti y HOLLI-HELENIUS Kirsi y KAMARAINEN Joni y RINTAKIKKA Irina y LAAPERI Anna-Leena y ARPONEN Otso. "Clinical Evaluation of a

Fully-automated Parenchymal Analysis Software for Breast Cancer Risk Assessment: a Pilot Study in a Finnish Sample”. En: *European Journal of Radiology* (2019), pág. 108710 (vid. págs. 10, 14, 15, 22).

PUDIL, Pavel, Jana Novovičová y Josef Kittler. “Floating search methods in feature selection”. En: *Pattern recognition letters* 15.11 (1994), págs. 1119-1125 (vid. pág. 28).

RÜCKSTIEB Thomas y OSENDORFER, Christian y VAN-Patrick. “Sequential feature selection for classification”. En: *Australasian Joint Conference on Artificial Intelligence*. Springer. 2011, págs. 132-141 (vid. pág. 27).

SMITH Karen Lisa y ISAACS, Claudine. “Management of women at increased risk for hereditary breast cancer”. En: *Breast disease* 27.1 (2007), págs. 51-67 (vid. pág. 9).

SUN Wenqing y TSENG, Tzu Liang y QIAN Wei y ZHANG Jianying y SALTZSTEIN-Edward C y ZHENG Bin y LURE Fleming y YU Hui y ZHOU Shi. “Using multiscale texture and density features for near-term breast cancer risk analysis”. En: *Medical physics* 42.6Part1 (2015), págs. 2853-2862 (vid. págs. 12, 17).

TAN Maxine Pu, JINTAO y BIN Zheng. “Reduction of false-positive recalls using a computerized mammographic image feature analysis scheme”. En: *Physics in Medicine & Biology* 59.15 (2014), pág. 4357 (vid. págs. 19-22).

TAN, Maxine y col. “A new approach to develop computer-aided detection schemes of digital mammograms”. En: *Physics in Medicine & Biology* 60.11 (2015), pág. 4413 (vid. pág. 12).

TAN, Maxine y col. "Assessment of a four-view mammographic image feature based fusion model to predict near-term breast cancer risk". En: *Annals of biomedical engineering* 43.10 (2015), págs. 2416-2428 (vid. pág. 12).

TAN M., PU J. CHENG S. LIU H. ZHENG B. "Assessment of a four-view mammographic image feature based fusion model to predict near-term breast cancer risk.." En: *Annals of biomedical engineering*. 2015 () (vid. págs. 10, 14, 16).

TORRES Germán F y PERTUZ, Said. "Automatic detection of the retroareolar region in x-ray mammography images". En: *VII Latin American Congress on Biomedical Engineering CLAIB 2016, Bucaramanga, Santander, Colombia, October 26th-28th, 2016*. Springer. 2017, págs. 157-160 (vid. págs. 23, 24).

TORRES-Mejía Gabriela y DE STAVOLA, Bianca y col. "Mammographic features and subsequent risk of breast cancer: a comparison of qualitative and quantitative evaluations in the Guernsey prospective studies". En: *Cancer Epidemiology and Prevention Biomarkers* 14.5 (2005), págs. 1052-1059 (vid. pág. 12).

VIOLINI, Maria Lucia. "Selección de características. Su aplicación a texturas". Tesis doct. 2014, págs. 8-25 (vid. págs. 26, 28).

WEI, Jun y col. "Association of computerized mammographic parenchymal pattern measure with breast cancer risk: a pilot case-control study". En: *Radiology* 260.1 (2011), págs. 42-49 (vid. págs. 12, 17, 24).

WILKINSON, Leland. "Tests of significance in stepwise regression." En: *Psychological bulletin* 86.1 (1979), pág. 168 (vid. pág. 29).

WU, Yi-Ta y col. "Comparison of mammographic parenchymal patterns of normal subjects and breast cancer patients". En: *Medical Imaging 2008: Computer-Aided*

Diagnosis. Vol. 6915. International Society for Optics y Photonics. 2008, pág. 691520 (vid. págs. 12, 24).

ZHENG, Yuanjie y col. "Parenchymal texture analysis in digital mammography: a fully automated pipeline for breast cancer risk assessment". En: *Medical physics* 42.7 (2015), págs. 4149-4160 (vid. pág. 12).