



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



ESCUOLA TÉCNICA
SUPERIOR INGENIERÍA
INDUSTRIAL VALENCIA

Curso Académico:

AGRADECIMIENTOS

La realización de este Trabajo Fin de Grado se ha hecho posible gracias a mi tutor David Moratal Pérez y a mis cotutores Cristian Galán Auge, Silvia Ruiz España y José Manuel Santabárbara Gómez. Agradecer sobre todo a mis cotutores Cristian y Silvia por haberme ayudado en todo momento e interesado en cómo iban mis avances en el proyecto.

Agradecer a mis compañeros de ASCIRES por la ayuda y los continuos conocimientos que me han aportado durante el convenio de prácticas que he podido experimentar gracias a este Trabajo, además del buen ambiente que ha habido todo este tiempo.

Gracias a toda mi familia, mis padres, mis hermanos, mi abuela y Alejandro por los ánimos y el apoyo que he recibido durante todos estos años de universidad e incluso agradecer la confianza que han depositado en mí. Gracias también a todos mis compañeros y amigos que la universidad me ha brindado la oportunidad de conocer.

RESUMEN

La resonancia magnética de mama con contraste representa una herramienta con gran potencial para la detección, evaluación y diagnóstico del cáncer de mama. Debido a su elevada sensibilidad se pueden llegar a resolver contratiempos y limitaciones manifestadas en otras modalidades de diagnóstico por imagen como serían la mamografía o la ecografía.

La extracción y evaluación de las características de textura y de forma en las distintas regiones de interés de la mama conforman un aspecto muy determinante para la exhaustiva comprensión de las imágenes médicas y, en consecuencia, la realización de un adecuado diagnóstico de la paciente en función de aquellas características más significativas.

Por todo ello, se ha pretendido desarrollar una metodología para la ayuda al diagnóstico del cáncer de mama mediante el análisis de texturas y de formas en imágenes de resonancia magnética (RM). De manera que se ha desarrollado un análisis de texturas en 2D y de formas en 3D de la región de interés del tumor de la mama para categorizar la malignidad del mismo en función de la localización de la tumoración en el plano coronal dividido en cuadrantes y el tipo de tumor según su información genética.

Para ello, se seleccionaron de la base de datos de *ASCIREs Grupo Biomédico*, empresa con la cual se ha trabajado en colaboración, las imágenes de RM de 43 pacientes con cáncer de mama, cuya lesión, en este caso, tumor, se segmentó tanto en 2D como en 3D obteniendo así la región de interés (ROI) de la mama.

Una vez preprocesadas las imágenes, se realizaron tres análisis, de texturas, de formas y de ambas características conjuntamente, características que se obtuvieron de la toolbox de *Radiomics* de MATLAB y de *PyRadiomics* de Python, respectivamente. Posteriormente, en MATLAB se hizo uso de dichas características para desarrollar modelos predictivos en función de la localización por cuadrantes del tumor en el plano coronal, estudio multiclase de 4 clases (CIE, CII, CSE y CSI) y del tipo de tumor según su información genética, estudio binario de 2 clases (HER2 positivo y Luminal B con HER2 negativo). Por tanto, se realizaron 6 estudios en total, es decir, los 3 tipos de análisis (de textura, de forma y de ambas) tanto para el estudio multiclase como para el binario.

Como resultados más relevantes se puede comentar que para el estudio multiclase de la localización por cuadrantes, el uso de características de textura y de forma conjuntamente permite obtener el mayor AUC medio del 67'99%, mientras que, para el estudio binario del tipo de tumor, únicamente con el uso de parámetros de textura ya es suficiente para obtener un adecuado valor de AUC medio del 74'04%, pero en ambos casos el mejor modelo de clasificación para nuestros datos es el modelo SVM con Kernel Lineal. Además, en los estudios multiclase, el mejor modelo de predicción se obtiene con el uso de un número muy reducido de características, en cambio, en los estudios binarios, son necesarias prácticamente todas las características iniciales, lo que implica que dichos modelos son más complejos y, por tanto, podrían ser menos generalizables que, en este caso, los presentes estudios multiclase.

Palabras clave: mama, cáncer de mama, resonancia magnética, parámetros de textura, malignidad, localización.

RESUM

La ressonància magnètica de mama amb contrast representa una eina amb gran potencial per a la detecció, avaluació i diagnòstic del càncer de mama. A causa de la seua elevada sensibilitat es poden arribar a resoldre contratemps i limitacions manifestades en altres modalitats de diagnòstic per imatge com serien la mamografia o l'ecografia.

L'extracció i avaluació de les característiques de textura i de forma en les diferents regions d'interés de la mama conformen un aspecte molt determinant per a l'exhaustiva comprensió de les imatges mèdiques i, en conseqüència, la realització d'un adequat diagnòstic de la pacient en funció d'aquelles característiques més significatives.

Per tot això, s'ha pretés desenvolupar una metodologia per a l'ajuda al diagnòstic del càncer de mama mitjançant l'anàlisi de textures i de formes en imatges de ressonància magnètica (RM). De manera que s'ha desenvolupat una anàlisi de textures en 2D i de formes en 3D de la regió d'interés del tumor de la mama per a categoritzar la malignitat del mateix en funció de la localització de la tumoració en el pla coronal dividit en quadrants i el tipus de tumor segons la seua informació genètica.

Per a això, es van seleccionar de la base de dades de *ASCIREs Grupo Biomédico*, empresa amb la qual s'ha treballat en col·laboració, les imatges de RM de 43 pacients amb càncer de mama, la lesió de la qual, en aquest cas, tumor, es va segmentar tant en 2D com en 3D obtenint així la regió d'interés (ROI) de la mama.

Una vegada preprocessades les imatges, es van realitzar tres anàlisis, de textures, de formes i de totes dues característiques conjuntament, característiques que es van obtindre de la toolbox de *Radiomics* de MATLAB i de *PyRadiomics* de Python, respectivament. Posteriorment, en MATLAB es va fer ús d'aquestes característiques per a desenvolupar models predictius en funció de la localització per quadrants del tumor en el pla coronal, estudi multiclasse de 4 classes (CIE, CII, CSE i CSI) i de la mena de tumor segons la seua informació genètica, estudi binari de 2 classes (HER2 positiu i Luminal B amb HER2 negatiu). Per tant, es van realitzar 6 estudis en total, és a dir, els 3 tipus d'anàlisis (de textura, de forma i de totes dues) tant per a l'estudi multiclasse com per al binari.

Com a resultats més rellevants es pot comentar que per a l'estudi multiclasse de la localització per quadrants, l'ús de característiques de textura i de forma conjuntament permet obtindre el major AUC mitjà del 67'99%, mentre que, per a l'estudi binari de la mena de tumor, únicament amb l'ús de paràmetres de textura ja és suficient per a obtindre un adequat valor de AUC mitjà de 74'04%, però en tots dos casos el millor mètode de classificació per a les nostres dades és el model SVM amb Kernel Lineal. A més, en els estudis multiclasse, el millor model de predicció s'obté amb l'ús d'un número molt reduït de característiques, en canvi, en els estudis binaris, són necessàries pràcticament totes les característiques inicials, la qual cosa implica que aquests models són més complexos i, per tant, són menys generalitzables que, en aquest cas, els presents estudis multiclasse.

Paraules clau: mama, càncer de mama, ressonància magnètica, paràmetres de textura, malignitat, localització.

ABSTRACT

Contrast-enhanced breast magnetic resonance represents a tool with great potential for the detection, evaluation and diagnosis of breast cancer. Due to its high sensitivity, it can solve setbacks and limitations of other imaging modalities such as mammography or ultrasound.

Extraction and evaluation of the texture and shape characteristics in the different regions of interest of the breast are a very important aspect for the exhaustive understanding of the medical images and, consequently, the realization of an adequate diagnosis of the patient according to those most significant characteristics.

For all these reasons, we have tried to develop a methodology to aid in the diagnosis of breast cancer through the analysis of textures and shapes in magnetic resonance images (MRI). Thus, a 2D texture analysis and 3D shape analysis of the region of interest of the breast tumour has been developed to categorize the malignancy of the tumour according to the location of the tumour in the coronal plane divided into quadrants and the type of tumour according to its genetic information.

For this purpose, the MR images of 43 patients with breast cancer were selected from the database of *ASCIREs Grupo Biomédico*, a company with which we have worked in collaboration, whose lesion, in this case, tumour, was segmented both in 2D and 3D, thus obtaining the region of interest (ROI) of the breast.

Once the images were preprocessed, three analyses were performed, of textures, shapes and both features together, features obtained from MATLAB's *Radiomics* toolbox and Python's *PyRadiomics*, respectively. Subsequently, these features were used in MATLAB to develop predictive models based on the quadrant location of the tumour in the coronal plane, a multiclass study of 4 classes (CIE, CII, CSE and CSI) and the type of tumour according to its genetic information, a binary study of 2 classes (HER2 positive and Luminal B with HER2 negative). Therefore, a total of 6 studies were carried out, that is to say, the 3 types of analysis (texture, shape and both) for both the multiclass and binary studies.

As the most relevant results we can comment that for the multiclass study of quadrant localization, the use of texture and shape features together allows us to obtain the highest mean AUC of 67.99%, while for the binary study of tumour type, the use of texture parameters alone is sufficient to obtain an adequate mean AUC value of 74.04%, but in both cases the best classification method for our data is the SVM model with Linear Kernel. Moreover, in multiclass studies, the best prediction model is obtained with the use of a very small number of features, whereas in binary studies, practically all the initial features are necessary, which implies that these models are more complex and, therefore, less generalizable than, in this case, the present multiclass studies.

Keywords: breast, MRI, breast cancer, texture parameters, malignancy, localisation.

DOCUMENTOS CONTENIDOS EN EL TFG

1. MEMORIA

2. PRESUPUESTO

MEMORIA

DESARROLLO DE UNA METODOLOGÍA PARA LA AYUDA AL DIAGNÓSTICO DEL CÁNCER DE MAMA MEDIANTE EL ANÁLISIS DE TEXTURAS EN IMÁGENES DE RESONANCIA MAGNÉTICA

AUTORA: PILAR CASTELLOTE HUGUET

TUTOR: DAVID MORATAL PÉREZ

COTUTORES: SILVIA RUIZ ESPAÑA

CRISTIAN GALÁN AUGE

JOSÉ MANUEL SANTABÁRBARA GÓMEZ

Curso Académico: 2020-2021

ÍNDICE MEMORIA

1.	MOTIVACIÓN Y OBJETIVOS	1
2.	INTRODUCCIÓN	2
2.1.	ANATOMÍA DE LA MAMA.....	2
2.2.	CÁNCER DE MAMA	3
2.2.1.	EPIDEMIOLOGÍA.....	3
2.2.2.	ETIOLOGÍA.....	4
2.2.3.	DIAGNÓSTICO	4
2.2.4.	TIPOS DE CÁNCER DE MAMA	5
2.2.5.	TRATAMIENTO	5
2.3.	IMAGEN DE RESONANCIA MAGNÉTICA	6
2.3.1.	Imagen mamaria de resonancia magnética.....	7
2.4.	ANÁLISIS DE TEXTURAS Y DE FORMAS	9
2.4.1.	Adquisición de IRM.....	9
2.4.2.	Definición de la región de interés	10
2.4.3.	Preprocesado de la región de interés	10
2.4.4.	Extracción de características de textura y forma	10
2.4.4.1.	Extracción de características de texturas	11
2.4.4.2.	Extracción de características de formas.....	13
2.4.5.	Selección de características.....	15
2.4.6.	Métodos de clasificación de características	15
2.4.7.	Validación de sistemas	18
2.4.8.	Métodos de evaluación de resultados.....	19
3.	MATERIALES.....	22
3.1.	IMÁGENES DE RM	22
3.2.	SOFTWARES EMPLEADOS PARA EL PRE-PROCESADO DE LAS IMÁGENES	22
3.3.	SOFTWARE MATLAB	23
3.3.1.	Radiomics 2D	23
3.4.	SOFTWARE PYTHON	23
3.4.1.	PyRadiomics.....	23
4.	METODOLOGÍA.....	24
4.1.	SEGMENTACIÓN DE LA LESIÓN	24
4.2.	ANÁLISIS DE TEXTURAS Y FORMAS Y CLASIFICACIÓN	25
4.2.1.	Extracción de características de textura y forma	25

4.2.1.1.	Extracción de texturas	25
4.2.1.2.	Extracción de formas	26
4.2.2.	Selección de características	26
4.2.3.	Clasificación	27
5.	RESULTADOS	28
5.1.	ANÁLISIS DE TEXTURA	29
5.1.1.	Clasificación multiclase.....	29
5.1.2.	Clasificación binario.....	30
5.2.	ANÁLISIS DE FORMA.....	31
5.2.1.	Clasificación multiclase.....	31
5.2.2.	Clasificación binario.....	32
5.3.	ANÁLISIS DE CARACTERÍSTICAS DE TEXTURA Y FORMA	33
5.3.1.	Clasificación multiclase.....	33
5.3.2.	Clasificación binario.....	34
5.4.	COMPARATIVA DE LOS TRES ANÁLISIS	35
6.	DISCUSIÓN	38
7.	CONCLUSIONES.....	41
8.	LÍNEAS FUTURAS.....	42
9.	BIBLIOGRAFÍA	43

ÍNDICE FIGURAS

Figura 1. Corte sagital de la mama incluyendo sus partes más relevantes	2
Figura 2. Vascularización del tejido mamario de la mama derecha	2
Figura 3. Distintas divisiones de la mama en los planos coronal (a), sagital (b) y axial (c)	3
Figura 4. Componentes de un equipo de resonancia magnética.....	6
Figura 5. Principio básico de funcionamiento de la RM.....	6
Figura 6. Modo de funcionamiento de RM, indicando los TE y TR	7
Figura 7. Posición del paciente en una RM de mama	7
Figura 8. Los distintos planos de una mama, a) Plano Axial, b) Plano Sagital, c) Plano Coronal ..	8
Figura 9. Las distintas series que se obtienen en la resonancia magnética, a) Serie T1, b) Serie T2, c) Serie DINÁMICA, todas ellas en plano axial	8
Figura 10. Proceso de análisis de características para la clasificación de IRM	9
Figura 11. Cálculo de una Matriz de Co-ocurrencias de una imagen 4x4 con 3 niveles de gris..	12
Figura 12. Cálculo de una Matriz de Largo Recorrido de una imagen 4x4 con 3 niveles de gris	12
Figura 13. Cálculo de una Matriz de Zona Tamaño de una imagen 4x4 con 3 niveles de gris....	12
Figura 14. Cálculo de una Matriz de Diferencia de Tonos Grises de Vecindad de una imagen 4x4 con 3 niveles de gris	12
Figura 15. Las 15 posibles configuraciones de cubo obtenidas mediante el Algoritmo de Cubos de Marcha	14
Figura 16. Dos posibles aproximaciones de la representación de la superficie de la ROI	14
Figura 17. Los 3 tipos de métricas más utilizadas en el modelo KNN	16
Figura 18. Fronteras de decisión para el modelo KNN con valores de K=1 y K=5.....	17
Figura 19. Modelo SVM ilustrando el hiperplano con el margen máximo y los vectores de soporte de cada clase	17
Figura 20. Clasificación de un Modelo Bagged Tree	18
Figura 21. Estrategia de validación cruzada o K-fold con K iteraciones.....	18
Figura 22. Estrategia de validación simple o Hold-Out	19
Figura 23. Curva ROC, VPR vs. FPR en diferentes umbrales de clasificación	20
Figura 24. Curva ROC indicando el área bajo la curva (AUC)	20
Figura 25. Segmentación de un tumor en la mama derecha mediante el software ITK-SNAP...	25
Figura 26. Ranking de las 15 texturas más relevantes para el nivel de gris 16	27
Figura 27. Valores de AUC del mejor modelo de clasificación para cada nivel de gris en el estudio multiclase de textura.....	29
Figura 28. Top 10 de texturas más utilizadas en el estudio multiclase.....	30

Figura 29. Valores de AUC del mejor modelo de clasificación para cada nivel de gris en el estudio binario de textura.....	31
Figura 30. Valores de AUC del mejor modelo en el estudio multiclase de forma	32
Figura 31. Top 7 de formas más utilizadas en el estudio multiclase.....	32
Figura 32. Valores de AUC del mejor modelo de clasificación en el estudio binario.....	33
Figura 33. Valores de AUC del mejor modelo de clasificación para cada nivel de gris en el estudio multiclase de textura y de forma	34
Figura 34. Top 15 de texturas y formas más utilizadas en el estudio multiclase.....	34
Figura 35. Valores de AUC del mejor modelo de clasificación para cada nivel de gris en el estudio binario de textura y de forma	35
Figura 36. Comparativa de los valores AUCs, tanto en el estudio multiclase como en el binario, de cada análisis indicando el AUC máximo de cada grupo.	36
Figura 37. Comparativa del número de características, tanto en el estudio multiclase como en el binario, de cada análisis indicando el estudio con el AUC máximo correspondiente de cada grupo.	36
Figura 38. a) Matriz de confusión del estudio binario para el Modelo SVM Lineal en el análisis de textura b) Matriz de confusión de la multiclase para el Modelo SVM Lineal en el análisis de textura y forma.....	37

ÍNDICE TABLAS

Tabla 1. Características de textura obtenidas a partir de la extracción de características con Radiomics de MATLAB.	13
Tabla 2. Características de forma obtenidas a partir de la extracción de características con PyRadiomics de Python.....	14
Tabla 3. Matriz de confusión 2x2 para un estudio binario de 2 clases	20
Tabla 4. Matriz de confusión 4x4 para un estudio multiclase de 4 clases	21
Tabla 5. Datos de los estudios realizados en el presente trabajo.....	28
Tabla 6. Estructura organizativa de los resultados	28
Tabla 7. Resultados obtenidos en el estudio multiclase del análisis de textura, indicando los valores de AUC y el número de características usadas para los distintos modelos y niveles de gris, así como el mejor modelo predictivo y sus resultados.....	29
Tabla 8. Resultados obtenidos en el estudio binario del análisis de textura, indicando los valores de AUC y el número de características usadas para los distintos modelos y niveles de gris, así como el mejor modelo predictivo y sus resultados	30
Tabla 9. Resultados obtenidos en el estudio multiclase del análisis de forma, indicando los valores de AUC y el número de características usadas para los distintos modelos y niveles de gris, así como el mejor modelo predictivo y sus resultados.....	31
Tabla 10. Resultados obtenidos en el estudio binario del análisis de forma, indicando los valores de AUC y el número de características usadas para los distintos modelos, así como el mejor modelo predictivo y sus resultados	32
Tabla 11. Resultados obtenidos en el estudio multiclase del análisis de textura y de forma, indicando los valores de AUC y el número de características usadas para los distintos modelos y niveles de gris, así como el mejor modelo predictivo y sus resultados	33
Tabla 12. Resultados obtenidos en el estudio binario del análisis de textura y de forma indicando los valores de AUC y el número de características usadas para los distintos modelos y niveles de gris, así como el mejor modelo predictivo y sus resultados	35
Tabla 13. Comparativa de los tres análisis realizados para ambos estudios, multiclase y binario, con el mejor modelo de cada uno.....	36

1. MOTIVACIÓN Y OBJETIVOS

El cáncer de mama es la patología maligna más frecuente diagnosticada en mujeres y con tasas más elevadas en países desarrollados, pudiendo decir que 1 de cada 8 mujeres padecerán dicha enfermedad. Además, es la principal causa de muerte por cáncer en el sexo femenino, con un 17% de muertes de todas las mujeres diagnosticadas en España. No obstante, cabe destacar que las tasas de incidencia aumentan muy lentamente en el mundo, sobre todo, en España, debido al envejecimiento de la población, a los resultados positivos en los tratamientos y a los programas de detección precoz existentes [1].

El hecho de padecer este tipo de cáncer viene determinado por una serie de factores de riesgo, entre todos ellos destacan el rango de valores de edad entre 35 y 80 años, el género femenino, los antecedentes familiares por el componente genético y la posibilidad de heredar las mutaciones autosómicas dominantes en los genes BRCA1 y BRCA2, la densidad del tejido mamario, el envejecimiento, la menopausia tardía, estilos de vida poco saludables, obesidad, alcohol y tabaco, radiaciones ionizantes, etc. [2].

Por todo ello, y considerando tanto los factores de riesgo como las estadísticas conocidas es de gran relevancia poder desarrollar programas de detección precoz con la finalidad de poder ofrecer el mejor tratamiento y, por tanto, un buen diagnóstico puede implicar de manera indirecta una buena evolución de la lesión. En ocasiones, para diagnosticar de la mejor manera posible es necesario recurrir a imágenes de resonancia magnética debido a su elevada sensibilidad ya que obtiene una detección de tumores malignos del 94-99% frente al 50% en una mamografía [3].

La extracción y evaluación de características de textura y de forma constituyentes en la región de interés de la imagen de resonancia magnética conforman un aspecto muy determinante para la exhaustiva comprensión de estas imágenes médicas y, en consecuencia, la realización de un adecuado diagnóstico de la paciente, en función de aquellas características más significativas, con la finalidad de intentar evitar métodos invasivos, como es en este caso la biopsia.

Cabe destacar que la malignidad de un tumor se suele asociar a un tamaño grande del mismo, a un margen mal definido, a unos contornos irregulares, etc. No obstante, dentro de esta clasificación de malignidad se debería remarcar la importancia del tipo de tumor según su información genética, pudiéndose clasificar en luminal A, luminal B, HER2 y triple negativo, además de remarcar también las posibles localizaciones que puede presentar el tumor en la mama, debido a que, dependiendo de dicha estratificación y localización, el tumor presentará un mejor o peor pronóstico.

El objetivo de este trabajo consiste en desarrollar una metodología para la ayuda al diagnóstico del cáncer de mama mediante el análisis, tanto de texturas como de formas, en imágenes de resonancia magnética, de manera que se desarrolle una posible solución con la cual categorizar la malignidad del tumor en función de tres covariables: la localización de la tumoración en el plano coronal dividido en cuadrantes, la localización en el plano axial y sagital dividido en porciones y el tipo de tumor según su información genética. Para ello, se desarrollará un análisis de texturas en 2D y de formas en 3D de la región central de interés de la tumoración de la mama, incorporando las tres covariables mencionadas previamente, como parámetros iniciales.

2. INTRODUCCIÓN

2.1. ANATOMÍA DE LA MAMA

Las mamas son órganos glandulares de secreción externa cuya función es la producción y secreción de la leche materna durante los periodos de lactancia. Se encuentran ubicadas en la parte frontal de la pared torácica entre la segunda y sexta costilla (Figura 1), medialmente a ambos lados del esternón, lateralmente hasta la línea media axilar y apoyándose en el músculo pectoral mayor y músculos intercostales.

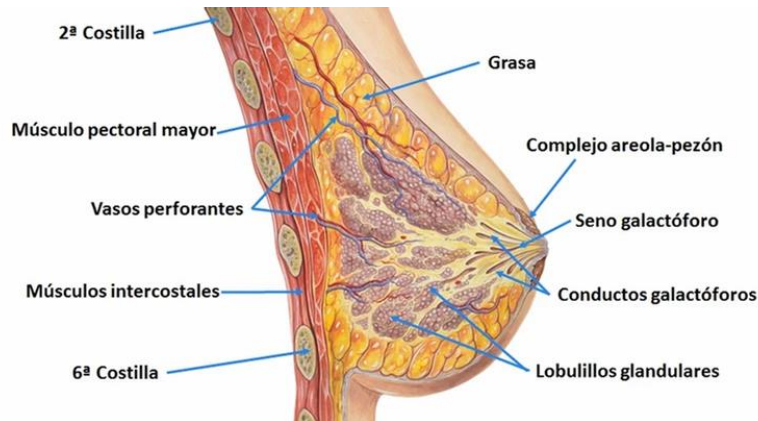


Figura 1. Corte sagital de la mama incluyendo sus partes más relevantes [4].

Una mama está constituida principalmente por tejido adiposo (grasa), tejido epitelial (piel) y por la propia glándula mamaria, la cual está formada por unos 15 o 20 lobulillos glandulares, dispuestos en forma circular, éstos se dirigen a los conductos galactóforos acabando en el seno galactóforo (Figura 1), estructura que actúa de reservorio de leche y conecta el interior de la mama con el exterior, el pezón, por donde se expulsará dicha leche en épocas de lactancia.

Además, cada lóbulo principal se subdivide en lobulillos más pequeños terminando en las células alveolares, que son las que llevan a cabo la producción de la leche. Por tanto, se puede observar que la mama forma una estructura piramidal con mayor concentración de vasos y glándulas en la base, la pared torácica, y en el vértice nos encontramos con el pezón [4].

Cabe destacar la vascularización que presenta el tejido mamario (Figura 2), dicha circulación sanguínea es otorgada por las arterias y las venas mamarias, situadas a ambos lados del esternón, así como los vasos torácicos laterales, además de otras arterias que también aportan vascularización al tejido pero en menor medida [4].

Además, para hacer referencia a cualquier lesión proveniente de la mama, primeramente se habla de dónde se encuentra localizada y, es por ello, que la mama se divide en 4 cuadrantes, vistos desde el plano coronal, y en 3 porciones, vistas tanto desde el plano axial como sagital.

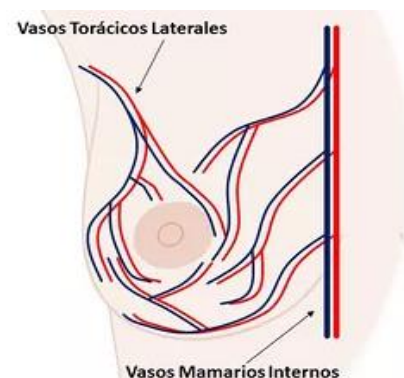


Figura 2. Vascularización del tejido mamario de la mama derecha [4].

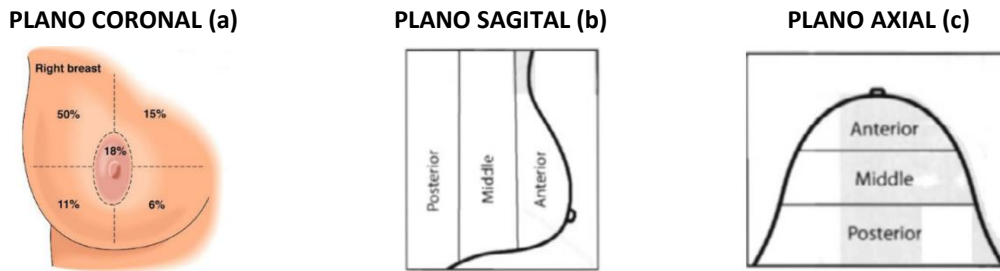


Figura 3. Distintas divisiones de la mama en los planos coronal (a), sagital (b) y axial (c) [5], [6].

Para comprender el concepto de los **4 cuadrantes** (Figura 3a) se trazan dos líneas, una vertical y otra horizontal entrecruzadas en el pezón, obteniendo así un cuadrante superior externo (CSE), uno superior interno (CSI), otro inferior externo (CIE) y otro inferior interno (CII), todos ellos concluyendo en el complejo areola-pezón, conocido como región retroareolar. Se conoce que la mayor cantidad de conductos se encuentran en el CSE y, es por ello, que es donde tienen lugar la mayoría de los tumores malignos de la mama, en un 50%, seguido de la región retroareolar con un 18%, CSI un 15%, CIE un 11% y finalmente CII un 6% [6].

Por otro lado, la mama también se divide en **3 porciones** (Figura 3b y 3c) considerando la porción posterior, la más cercana a la pared torácica, la porción media, la que se encuentra en la zona media de la mama, y la porción anterior, la más cercana al complejo areola-pezón.

Estos dos tipos de localizaciones de una lesión en la mama deben ser consideradas en el presente estudio debido a los posteriores análisis que se desarrollarán.

2.2. CÁNCER DE MAMA

El cáncer de mama es la patología maligna más frecuentemente diagnosticada en mujeres. Es un proceso oncológico donde las células sanas del tejido mamario se degeneran y se convierten en células tumorales, las cuales proliferan sin ningún tipo de control hasta convertirse en tumores [7].

Además, puede decirse que el concepto cáncer de mama no se trata de una sola enfermedad ya que es el término común para los distintos subtipos de tumores de mama que existen, ya sea de orígenes moleculares, celulares o clínicos.

2.2.1. EPIDEMIOLOGÍA

Como bien se ha mencionado, el cáncer de mama es la **patología maligna más frecuente en mujeres**, siendo las tasas de este tipo de cáncer más altas en países desarrollados pese a que están aumentando también en casi todas las regiones del mundo. Además, es la principal causa de muerte por cáncer en el sexo femenino [1].

Las estadísticas internacionales a cerca del cáncer de mama estiman que en el año 2018 se diagnosticaron más de 2 millones de casos nuevos. En concreto en España, la incidencia de nuevos casos es menor que en muchos otros países y, según el Observatorio del Cáncer AECC, se diagnosticaron 33.307 nuevos casos en el año 2019 y murieron 6.579. Entre el total de diagnosticados, el rango de valores de edad más habitual es de 45-65 años, aunque la mayoría de muertes por cáncer de mama ocurren en mayores de 75 años, prácticamente en un 50% [1].

No obstante, cabe destacar que las tasas de incidencia aumentan muy lentamente en el mundo debido al envejecimiento de la población y a los programas de detección precoz existentes [1].

2.2.2. ETIOLOGÍA

La etiología del cáncer de mama es multifactorial, pero básicamente su aparición es la etapa final de una serie de mutaciones celulares. Es por ello que es determinante conocer las causas y factores que contribuyen a un mayor riesgo de desarrollar este tipo de cáncer.

En primer lugar, la **edad** y el **género** son factores fundamentales ya que el 99% de cánceres de mama lo constituyen mujeres, y que la mayoría de casos se suelen diagnosticar en el rango de valores de edad de 35-80 años, habiendo un pico a los 50 años y otro a los 70. Además, destacar que en el periodo de menopausia hay un mayor riesgo debido al aumento de tiempo de exposición del tejido mamario a los estrógenos ováricos [2], [8].

Otro factor a tener en cuenta son los **antecedentes familiares**, factor de riesgo por el componente genético, normalmente el riesgo es hereditario ya que se debe a factores genéticos primarios, que en la mayoría de casos son mutaciones autosómicas dominantes en BRCA 1 y BRCA 2. Aunque el riesgo también puede deberse a factores genéticos desconocidos [2], [8].

Cabe destacar que la **densidad de tejido mamario** también es un posible factor, ya que las mujeres con más del 75% de la mama constituida por tejido denso tienen 4-6 veces más riesgo, además de tener un diagnóstico más difícil por mamografía [2], [8].

Otros factores que también pueden influir en aumentar el riesgo de padecer cáncer podrían ser el envejecimiento, la menopausia tardía, estilos de vida poco saludables, obesidad, alcohol y tabaco, cambios en los patrones reproductivos, radiaciones ionizantes [2], [8]...

2.2.3. DIAGNÓSTICO

El cáncer de mama es una enfermedad poco sintomática, no obstante, puede presentarse simplemente como una masa indolora, un bulto palpable, es por ello que la aparición de una tumoración detectada por la paciente es la expresión inicial en un 75% de los casos, seguida por la secreción del pezón en un 10%.

Normalmente, la evaluación o diagnóstico del cáncer de mama consta de tres pasos, en primer lugar un **examen clínico**, en el caso de detectar alguna anomalía se realiza el diagnóstico mediante imágenes médicas, las técnicas más habituales para ello son la **mamografía** y la **ecografía**, no obstante, si dichas pruebas no determinan ningún tipo de diagnóstico o si son sospechosas de malignidad se debe recurrir a otro tipo de pruebas como la **resonancia magnética** de mama, prueba mucho más precisa y con mejor resolución [8].

Una vez diagnosticado el cáncer y conocida su malignidad, se decide realizar una **biopsia** guiada por imágenes, método que permite obtener tejido mamario sin cirugía y con la posibilidad de poder eliminar el tumor extirpándolo quirúrgicamente [8].

Además, cabría destacar que existen **programas de detección precoz** aumentando así el diagnóstico de lesiones no palpables o asintomáticas, aunque todavía son costosos y difíciles de organizar ya que necesitan rigurosos controles de calidad, así como una gran logística [8].

2.2.4. TIPOS DE CÁNCER DE MAMA

Actualmente, se ha demostrado que todo tipo de cáncer surge debido a una serie de alteraciones moleculares a nivel celular, originando células de carácter maligno y de crecimiento descontrolado. Es por ello que en el cáncer de mama existen distintos tipos de tumores si se considera el análisis genómico del propio cáncer.

Los subtipos de cáncer de mama principales son 4 y se basan en la existencia o no de receptores hormonales, el receptor de estrógeno (ER) y el receptor de progesterona (PR), y del receptor 2 del factor de crecimiento epidérmico humano (HER2) [2], [8].

- **Luminal A.** Presentan receptores hormonales (ER y PR) positivos y HER2 negativo con baja proliferación celular. Su respuesta al tratamiento hormonal es mucho más positiva que un tratamiento de quimioterapia. Pertenecen al 30-40% de las neoplasias de mama.
- **Luminal B.** Presentan ER y PR positivos y HER2 negativo con alta proliferación celular. Su respuesta al tratamiento de quimioterapia es mucho más positiva que un tratamiento hormonal. Pertenecen al 20-30%.
- **HER2 positivo.** Presentan HER2 positivo y ER y PR positivos o negativos. Se aplica un tratamiento anti-HER2. Pertenecen al 10-20% de las neoplasias de mama.
- **Triple Negativo.** Presentan ER, PR y HER2 negativos. Responden a tratamientos de quimioterapia y corresponden al subtipo con mayor mortalidad. Pertenecen al 10-20%.

Estos 4 subtipos de tumores de mama deben ser considerados en el presente estudio debido al posterior análisis que se desarrollará más adelante haciendo referencia a estos conceptos, aunque únicamente se hablará de Luminal B con HER2 negativo y de HER2 positivo.

La realización de este análisis es interesante ya que existen diferencias visuales importantes en imágenes médicas entre estos subtipos. Por ejemplo, la arquitectura del tejido mamario en Luminal B suele ser de bordes irregulares y espiculados mientras que en HER2 positivo existe una masa con bordes mal definidos y calcificaciones [8].

2.2.5. TRATAMIENTO

En general, actualmente en la medicina cada vez los tratamientos son más personalizados ya que dependiendo del subtipo biológico, el estadio tumoral y el grado de diseminación que la paciente presente su pronóstico será mejor o peor.

Los tratamientos que se realizan son básicamente la cirugía, la quimioterapia, la terapia hormonal, la terapia biológica y la radioterapia. Esta amplia ventana de opciones hacen que las decisiones se deban tomar por equipos multidisciplinares compuestos por cirujanos, oncólogos, oncólogos radiólogos, anatomopatólogos, etc. [8], [9].

La **cirugía** más conocida es la mastectomía, que consiste en la extirpación de toda la mama, además también existe la lumpectomía, donde únicamente se extirpa el tumor y sus tejidos circundantes. La **quimioterapia** hace uso de medicamentos en pastilla o inyectados en vena con el fin de acabar con las células tumorales. La **terapia hormonal** evita que la zona tumoral reciba las hormonas necesarias para su proliferación, mientras que la **terapia biológica** se implica con el sistema inmune para combatir con el tumor. Finalmente, la **radioterapia** hace uso de radiación ionizante con intención de eliminar dichas células [9].

2.3. IMAGEN DE RESONANCIA MAGNÉTICA

La resonancia magnética (RM) es una herramienta con gran potencial para la detección, evaluación y diagnóstico de patologías. Esta técnica proporciona imágenes médicas de alta calidad en 2D y 3D, desde la perspectiva que el especialista desee, además, la elevada sensibilidad y alta resolución que ofrece le permite obtener una visualización excelente tanto a nivel anatómico como funcional de los tejidos localizados en distintas partes del cuerpo.

Es una técnica no invasiva que permite la adquisición de imágenes médicas mediante radiaciones electromagnéticas, es decir, una forma de radiación por radiofrecuencia no ionizante. Un equipo de RM consta de tres elementos hardware básicos, además de otros componentes que lo conforman, como se puede observar en la Figura 4.

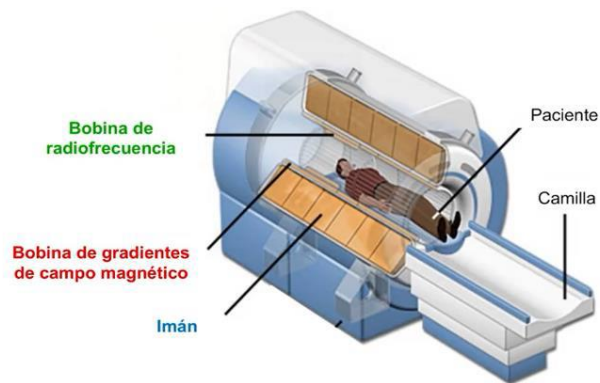


Figura 4. Componentes de un equipo de resonancia magnética [10].

En primer lugar, cabe destacar el **imán**, que es quien crea una intensidad de campo magnético intensa y uniforme. Además, están presentes las **bobinas de gradientes** de campo magnético, que son las que crean una variación de campo magnético en una determinada dirección (x, y, z). Como tercer elemento hardware, se puede considerar la **bobina de radiofrecuencia**, que tiene la función de excitar los tejidos a una determinada frecuencia [10].

El **modo de funcionamiento** de esta técnica, como se puede observar en la Figura 5, se basa en lanzar distintas secuencias de pulsos de radiofrecuencia, de manera que se captan una serie de señales y, en consecuencia, se consigue adquirir la imagen.

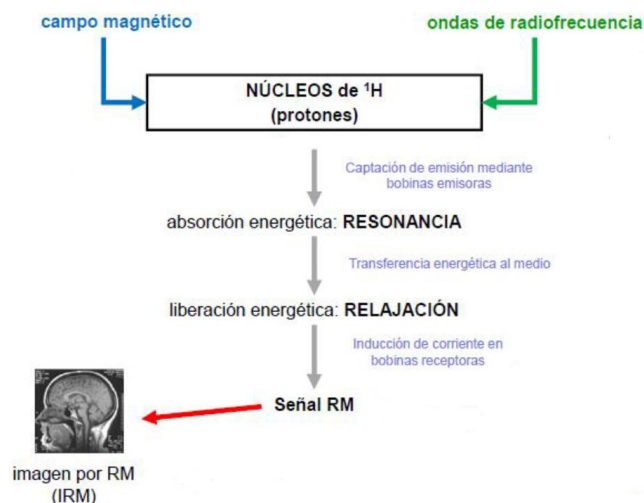


Figura 5. Principio básico de funcionamiento de la RM [10].

DESARROLLO DE UNA METODOLOGÍA PARA LA AYUDA AL DIAGNÓSTICO DEL CÁNCER DE MAMA MEDIANTE EL ANÁLISIS DE TEXTURAS EN IMÁGENES DE RESONANCIA MAGNÉTICA

El imán genera un campo magnético bajo el cual se encuentran los átomos de los tejidos con un número impar de núcleos de protones, que presentan un movimiento de espín, giran sobre sí mismos para alinearse con el campo magnético. En el momento en que las bobinas de radiofrecuencia lanzan ondas de radiofrecuencia, los núcleos, al captarlas, se estimulan y se mueven hacia direcciones diferentes, absorbiendo la energía asociada a dichas ondas a una determinada frecuencia, denominada frecuencia de resonancia [10], [11].

Una vez se desconectan los pulsos de radiofrecuencia, los núcleos excitados interactúan con el medio y liberan la energía que habían captado volviendo a su estado original, realineándose con el campo magnético de nuevo, proceso de relajación. Durante este proceso de liberación de energía por parte de los núcleos, intervienen las bobinas receptoras cerca de ese tejido, y se induce una corriente en dichas bobinas, que es la señal de RM. Finalmente, de dicha señal se obtiene la imagen por resonancia magnética (IRM) [10], [11].

Dentro de todo el funcionamiento de un equipo de RM, existen unos parámetros de adquisición que pueden ser modificados para obtener distintos tipos de imagen y, en consecuencia, poder resaltar tejidos específicos. Estos parámetros son el **tiempo de eco (TE)** y el **tiempo de recuperación (TR)**, como se puede observar en la Figura 6, tiempo que se tarda en captar la señal y tiempo que se tarda en volver a lanzar el siguiente pulso de radiofrecuencia [12].

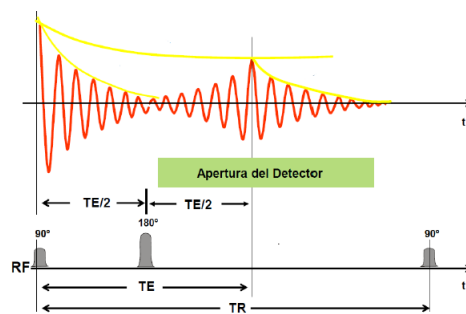


Figura 6. Modo de funcionamiento de RM, indicando los tiempos de eco y de recuperación, TE y TR [12].

Para obtener una imagen potenciada en T2, es necesario un TR corto y un TE largo, de esta manera se obtiene una imagen donde la grasa aparece hipointensa y los líquidos en tonos más brillantes. Mientras que para obtener una imagen potenciada en T1, es necesario tanto el TE como el TR cortos, de manera que aparecerán los líquidos hipointensos y la grasa hiperintensa.

2.3.1. Imagen mamaria de resonancia magnética

Generalmente, una resonancia magnética de mamas se realiza por dos motivos, o bien porque el resultado de la biopsia ha sido positivo de cáncer y, por ello, se desea obtener información más detallada de la enfermedad, o bien en personas con alto riesgo de padecer cáncer de mama y así facilitar la detección.

La adquisición de una IRM de mama requiere de un equipo especial de RM con bobinas específicas de mama. Para ello, es necesario posicionar al paciente sobre la camilla en decúbito prono con los brazos direccionados hacia arriba (Figura 7), de manera que esta posición permite que las mamas queden de la forma más libre posible dentro de la abertura del soporte de dichas bobinas [13].

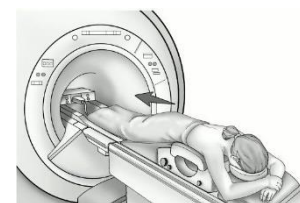


Figura 7. Posición del paciente en una RM de mama [14].

DESARROLLO DE UNA METODOLOGÍA PARA LA AYUDA AL DIAGNÓSTICO DEL CÁNCER DE MAMA MEDIANTE EL ANÁLISIS DE TEXTURAS EN IMÁGENES DE RESONANCIA MAGNÉTICA

Se conoce que en imágenes médicas, se puede hablar de tres planos principales, coronal, axial y sagital. En el caso de imágenes mamarias, para la elección del plano adecuado cabe destacar las ventajas de cada uno de ellos, el **plano sagital** (Figura 8b) requiere un campo de visión (*field of view*, FOV) pequeño y por ello la resolución espacial aumenta, el **plano coronal** (Figura 8c) permite un FOV rectangular al 50-60% y por ello se reduce el tiempo de adquisición y el **plano axial** (Figura 8a) requiere de un menor número de cortes para poder visualizar todas las mamas en su plenitud. No obstante, la mayoría de los protocolos utilizan los cortes en el plano axial debido a dicho número de cortes ya que en el plano sagital y coronal es necesario un mayor número para cubrir las dos mamas completas [14].



Figura 8. Los distintos planos que presenta una mama, a) Plano Axial, b) Plano Sagital, c) Plano Coronal. Fuente: elaboración propia.

Además, para una mejor visualización de la imagen se suele hacer uso de un medio de contraste, de esta manera, el diagnóstico, por ejemplo, de pequeños tumores en mamas densas, se hace más sencillo. El medio de contraste más empleado es el Gadolinio (Gd) combinado con un quelante ya que por sí solo es tóxico [15].

Para mejorar el contraste de la imagen, también es necesario modificar los parámetros de adquisición, de manera que podemos obtener **distintas modalidades de un mismo tejido** mamario, en la Figura 9 se pueden observar algunas de las mismas. No obstante, en este trabajo nos vamos a centrar más en las imágenes potenciadas en T1, apoyándonos también en T2 y DINÁMICA.

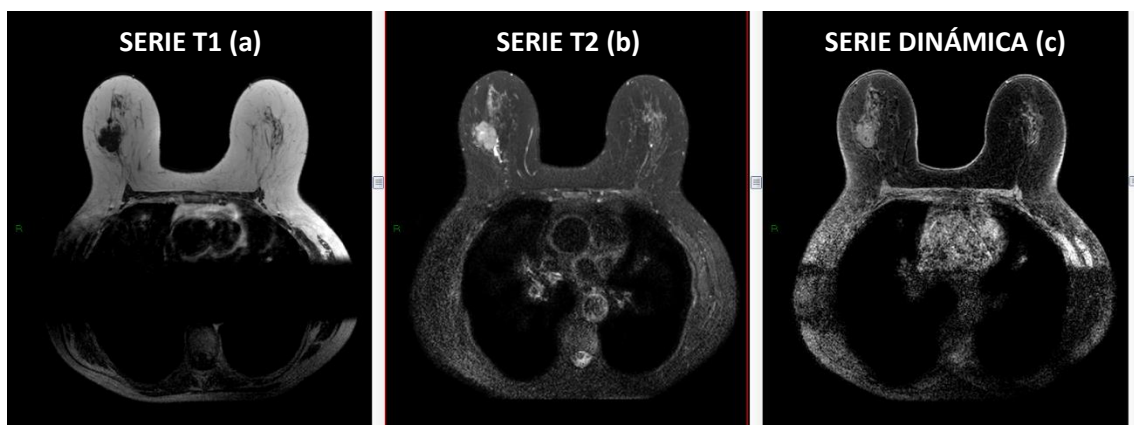


Figura 9. Las distintas series que se obtienen en la resonancia magnética, a) Serie T1, b) Serie T2, c) Serie DINÁMICA, todas ellas en plano axial. Fuente: elaboración propia.

2.4. ANÁLISIS DE TEXTURAS Y DE FORMAS

Para poder interpretar una imagen médica con mayor precisión se deben tener en cuenta los distintos descriptores que pueden ser analizados en la misma. Los descriptores son características (de color, de forma, de textura, de gradiente, etc.), que proporcionan la información básica del contenido de una imagen.

En este caso, se va a proceder al análisis, por un lado de los descriptores de texturas y, por otro lado, de los de formas. Los **descriptores de texturas** de una imagen se definen como las distribuciones espaciales de las diferentes intensidades, es decir, de los distintos niveles de gris existentes entre los píxeles de la región de interés (ROI) de la imagen [16].

Mientras que los **descriptores de formas** se definen como representaciones estructurales o estadísticas de la ROI de la imagen, es decir, se refieren tanto a la geometría como a la estructura del objeto, de manera que el conjunto de este tipo de descriptores permite reconstruir el objeto deseado en su plenitud [17].

Es por ello que para el análisis de este tipo de características se aplicarán métodos matemáticos para poder evaluar tanto la intensidad como la posición de los píxeles en la ROI de una imagen.

Este análisis podría ser una herramienta de gran utilidad para la obtención de información adicional sobre imágenes médicas y para el desarrollo de sistemas cuya finalidad es la ayuda al diagnóstico y la toma de decisiones de los especialistas. De manera que, dependiendo de lo que se pretenda clasificar posteriormente y de los valores obtenidos en cada tipo de análisis, se podrá concluir cuáles son los descriptores que van a aportar mayor información en la pertinente clasificación [17].

Las aplicaciones del análisis de textura y de forma en imágenes de resonancia magnética (IRM) implican un proceso que consta de seis pasos, tal y como se muestra en la Figura 11 [18].



Figura 10. Proceso de análisis de características para la clasificación de IRM [18].

2.4.1. Adquisición de IRM

La realización del análisis de características precisa, en primer lugar, de la adquisición de las imágenes de resonancia magnética, en este caso, de mama. Como se ha comentado anteriormente, dependiendo de los parámetros de adquisición que se indiquen, se pueden obtener distintas modalidades de un mismo tejido, consiguiendo así la posibilidad de resaltar aquellos tejidos que se deseen observar con mayor contraste.

Por este motivo, los **protocolos de adquisición** de este tipo de imágenes condicionan en gran medida los resultados del análisis, sobre todo, del análisis de textura. Por ello, cabe destacar que dichos protocolos sean los adecuados y además, sea el mismo para todas las imágenes que conforman este estudio.

2.4.2. Definición de la región de interés

Una vez obtenidas las imágenes de RM de mama necesarias para dicho estudio, se debe definir la **región de interés (ROI)**, en este caso, la ROI es la tumoración localizada en una región de la mama. En el caso del análisis de texturas, los parámetros obtenidos son a partir de una región de interés bidimensional (ROI 2D). Mientras que en el caso del análisis de formas, los parámetros obtenidos son a partir de una región de interés tridimensional (ROI 3D).

Para obtener la ROI se ha optado por la realización de una segmentación manual de la lesión y pese a que la delineación manual de dicha lesión es un proceso tedioso, es el método más utilizado para ello debido a que es un proceso difícil de automatizar, y además se consigue tener incluida la lesión de la mama definida al completo. Cabe destacar también que el tamaño de la región de interés debe ser lo suficientemente grande, como es en este caso con las lesiones de mama, para poder capturar la información de la textura con una significación estadística [18].

2.4.3. Preprocesado de la región de interés

La región de interés es extraída de las IRM de mama, las cuales se obtienen a partir de un determinado protocolo de adquisición. Es por ello que es necesario el preprocesamiento de dicha región de interés para minimizar los efectos no deseados producidos en dichos protocolos de adquisición, consiguiendo así que todas las ROIs estén bajo las mismas condiciones.

Para ello, es necesaria la aplicación de diversas técnicas de preprocesado, entre las cuales destaca la interpolación espacial, la normalización y la cuantización de los niveles de gris.

En primer lugar, el **método de la interpolación espacial** permite mejorar las imágenes con baja resolución espacial, debido a que una imagen con alta resolución espacial implica la mejora del análisis de textura. Cabe destacar que esta técnica es utilizada en casos de baja resolución o en casos de un análisis tridimensional para hacer que el vóxel sea isotrópico, es decir, que tenga la misma resolución en las tres direcciones (x, y, z).

La **normalización** es la técnica más utilizada para el preprocesado ya que permite que las intensidades de las imágenes sean uniformes y que presenten los mismos niveles de gris.

Finalmente, en el análisis de textura se debe tener en cuenta la **cuantización de niveles de gris**, realizando así una reducción de los niveles de gris que representa una imagen de resonancia magnética. Normalmente, este tipo de imágenes están representadas por 1024 o 4096 niveles de gris, debido a que el número de bits por píxel corresponde a 10 o 12. Esta reducción se debe a que a mayor número de niveles de gris, mayor cantidad de ruido en la imagen, de manera que al reducir los niveles de gris existe una mejora en la relación señal-ruido de la imagen. Por todo ello, el análisis de textura se realiza para 16, 32, 64 y 128 niveles de gris.

2.4.4. Extracción de características de textura y forma

La extracción de características es el paso principal del proceso de análisis de texturas y formas y se basa en el cálculo de las características propias a partir de las ROIs definidas previamente.

2.4.4.1. Extracción de características de texturas

En este apartado, se exponen algunos métodos para evaluar la extracción de características de texturas existentes. Como se ha comentado anteriormente, las texturas se definen como las distribuciones espaciales de los distintos niveles de gris entre los píxeles de la ROI, es decir, son parámetros obtenidos en función de la distribución de los píxeles. Las relaciones entre los distintos píxeles de la ROI pueden ser calculados de distintos modos y, es por ello, que dependiendo de cómo se calculen, se estará haciendo referencia a los distintos métodos para extraer texturas. Seguidamente, se van a tratar aquellos métodos más habituales.

- Los **métodos estructurales** representan las texturas como una serie de estructuras básicas que siguen una composición y un orden determinado. El objetivo está en comprender la jerarquía de la textura, definiendo así las reglas que siguen estas estructuras primitivas.
- Los **métodos basados en modelos** representan las texturas utilizando modelos matemáticos como generativos o estocásticos. Esta técnica no es muy utilizada debido al elevado coste computacional que conlleva la estimación de los parámetros de dichos modelos, proceso necesario para el consecuente análisis de texturas.
- Los **métodos de transformación** representan la imagen en un nuevo espacio, cuyas coordenadas se relacionan con las propias texturas de la imagen, mediante la aplicación de transformadas como Fourier o Wavelet. La más utilizada es la transformada Wavelet ya que permite la representación de las texturas en la escala más adecuada en función de la resolución espacial y el ajuste en un amplio rango para aplicaciones específicas.
- Los **métodos estadísticos** representan las texturas teniendo en cuenta las distribuciones espaciales de los distintos niveles de grises de la imagen y las relaciones entre los mismos.

Este trabajo se va a centrar en los métodos estadísticos para la extracción de características de texturas y es por ello que, dependiendo del orden del estadístico, se hará uso de distintos tipos de métodos estadísticos, tanto basados en el histograma como basados en matrices.

- Los estadísticos de primer orden son los **basados en el histograma**. Se caracterizan por construir dicho histograma considerando el número de píxeles que comparte un cierto nivel de gris. No obstante, al no tener en cuenta las relaciones entre los píxeles, no consiguen describir la textura completamente y, es por ello que suelen utilizarse en combinación con otros métodos.
- Los estadísticos de segundo orden son los basados en la *Gray-Level Co-occurrence Matrix* (GLCM).
- Los estadísticos de orden superior son los basados en la *Gray-Level Run-Length Matrix* (GLRLM), *Gray-Level Size Zone Matrix* (GLSZM) y *Neighborhood Gray-Tone Difference Matrix* (NGTDM).

La **Matriz de Co-ocurrencia o Co-occurrence Matrix (GLCM)** determina la frecuencia con la que se relacionan ciertos niveles de gris. Para analizar cada una de esas distribuciones de niveles de gris, se debe estudiar cada par de píxeles definiendo una dirección y una distancia, en este caso, como se realiza un análisis bidimensional se debe calcular la matriz definiendo 4 direcciones concretas, la horizontal, la vertical, a 45º y a 135º, para posteriormente realizar un promedio de

DESARROLLO DE UNA METODOLOGÍA PARA LA AYUDA AL DIAGNÓSTICO DEL CÁNCER DE MAMA MEDIANTE EL ANÁLISIS DE TEXTURAS EN IMÁGENES DE RESONANCIA MAGNÉTICA

todas ellas y obtener la matriz final. De esta manera, se obtendrá una matriz rotacionalmente invariante evitando así resultados distintos al rotar una misma imagen.

Por ejemplo, en la Figura 11 se observa que la primera entrada de la matriz corresponde a los píxeles con nivel de gris 1 separados por una distancia de 1 y en la dirección horizontal.

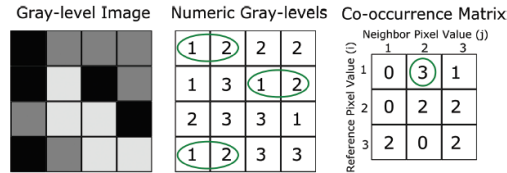


Figura 11. Cálculo de una Matriz de Co-ocurrencias de una imagen de 4x4 píxeles y 3 niveles de gris [18].

La **Matriz de Largo Recorrido o Run Length Matrix (GLRLM)** también determina las interacciones entre píxeles de la ROI. En concreto, esta matriz es un estadístico de órdenes superiores que representa el número de píxeles consecutivos que contienen el mismo nivel de gris en una determinada dirección, como se puede observar en la Figura 12. Por tanto, esta matriz, al igual que en el caso de GLCM, también debe ser calculada para las 4 direcciones existentes en 2D y, posteriormente, realizar un promedio obteniendo así la matriz final.

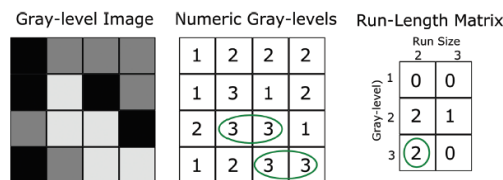


Figura 12. Cálculo de una Matriz de Largo Recorrido de una imagen de 4x4 píxeles y 3 niveles de gris [18].

La **Matriz de Zona de Tamaño o Size Zone Matrix (GLSZM)** es similar a la Matriz de Largo Recorrido, pero considerando áreas o regiones de niveles de gris repetidos (Figura 13). Por tanto, al tener en cuenta dichas regiones de niveles de gris no requiere ser calculada para distintas direcciones, como es en el caso de GLCM y GLRLM y, por tanto, se puede decir que esta matriz es invariable a la rotación de las imágenes.

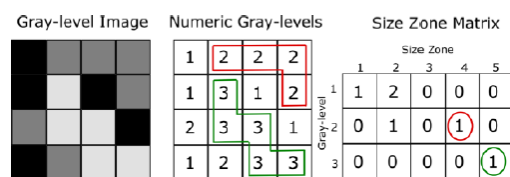


Figura 13. Cálculo de una Matriz de Zona de Tamaño de una imagen de 4x4 píxeles y 3 niveles de gris [18].

La **Matriz de Diferencia de Tonos Grises de Vecindad o Neighborhood Gray-Tone Difference Matrix (NGTDM)**, en este caso, tiene en cuenta vecindarios de píxeles y no regiones. Esta matriz representa la suma de las diferencias de niveles de gris entre todos los píxeles con un determinado nivel de gris de un mismo vecindario (Figura 14).

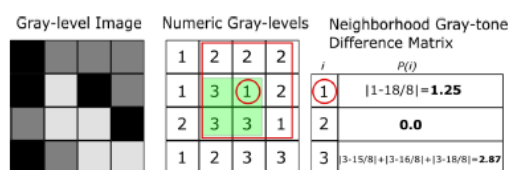


Figura 14. Cálculo de una Matriz de Diferencia de Tonos Grises de Vecindad de una imagen de 4x4 píxeles y 3 niveles de gris [18].

DESARROLLO DE UNA METODOLOGÍA PARA LA AYUDA AL DIAGNÓSTICO DEL CÁNCER DE MAMA
 MEDIANTE EL ANÁLISIS DE TEXTURAS EN IMÁGENES DE RESONANCIA MAGNÉTICA

Como bien se ha comentado previamente, en este trabajo se han utilizado los métodos estadísticos consiguiendo así la extracción de 43 características de textura (Tabla 1).

Tabla 1. Características de textura obtenidas a partir de la extracción de características con Radiomics de MATLAB.

CARACTERÍSTICAS DE TEXTURA				
Global	GLCM	GLRLM	GLSZM	NGTDM
Variance	Energy	Short Run Emphasis (SRE)	Small Zone Emphasis (SZE)	Coarseness
Skewness	Contrast	Long Run Emphasis (LRE)	Large Zone Emphasis (LZE)	Contrast
Kurtosis	Entropy	Gray-Level Nonuniformity (GLN)	Gray-Level Nonuniformity (GLN)	Busyness
	Homogeneity	Run-Lenght Nonuniformity (RLN)	Zone-Site Nonuniformity (ZSN)	Complexity
	Correlation	Run Percentage (RP)	Zone Percentage (ZP)	Strength
	SumAverage	Low Gray-Level Run Emphasis (LGRE)	Low Gray-Level Zone Emphasis (LGZE)	
	Variance	High Gray-Level Run Emphasis (HGRE)	High Gray-Level Zone Emphasis (HGZE)	
	Dissimilarity	Short Run Low Gray-Level Emphasis (SRLGE)	Small Zone Low GrayLevel Emphasis (SZLGE)	
AutoCorrelation		Short Run High Gray-Level Emphasis (SRHGE)	Small Zone High Gray-Level Emphasis (SZHGE)	
		Long Run Low Gray-Level Emphasis (LRLGE)	Large Zone Low Gray-Level Emphasis (LZLGE)	
		Low Run High Gray-Level Emphasis (LRHGE)	Large Zone High Gray-Level Emphasis (LZHGE)	
		GrayLevel Variance (GLV)	Gray-Level Variance (GLV)	
		Run-Lenght Variance (RLV)	Zone-Size Variance (ZSV)	

2.4.4.2. Extracción de características de formas

En este apartado, se trata del proceso realizado para la extracción de características de formas. Como se ha comentado anteriormente, las formas se definen como representaciones estructurales o estadísticas de la ROI de la imagen, es decir, se refieren tanto a la geometría como a la estructura del objeto [17]. Pues bien, estas características de forma se pueden extraer directamente con PyRadiomics, de manera que este paquete de Python brinda la oportunidad de extraer los descriptores del tamaño y de la forma tridimensionales de la ROI predefinida [19].

Para poder extraer este tipo de características de forma, es necesario la construcción de una **malla triangular**, donde los vértices de la misma serán definidos como puntos entre un vóxel del borde de la ROI y su contiguo fuera de la misma.

La construcción de esta malla triangular en 3D se realiza mediante el conocido **algoritmo de cubos de marcha o marching cubes algorithm**. El modo de funcionamiento de este algoritmo es ir desplazándose por el espacio de la imagen el cual se va iterando con cubos 2x2 de 8 vértices. Si la totalidad de vértices son segmentados con un binario 1 o bien no segmentados con un 0, el cubo está completamente dentro de la ROI o bien fuera de ella, mientras que si los vértices presentan tanto 1s como 0s, el vóxel debe contribuir con alguna esquina de la ROI [20], [21].

De esta manera, con dichos vóxeles se generan algunos triángulos y vértices, que se pueden corresponder con algunas de las 15 configuraciones posibles de vértices (Figura 15).

DESARROLLO DE UNA METODOLOGÍA PARA LA AYUDA AL DIAGNÓSTICO DEL CÁNCER DE MAMA
 MEDIANTE EL ANÁLISIS DE TEXTURAS EN IMÁGENES DE RESONANCIA MAGNÉTICA

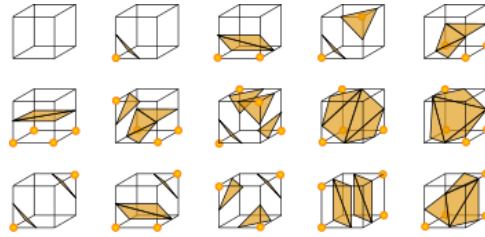


Figura 15. Las 15 posibles configuraciones de cubo obtenidas mediante el Algoritmo de Cubos de Marcha [20].

Una vez recorrida toda la imagen, se obtienen aquellos cubos que dividen la imagen entre la ROI y regiones externas. De manera que, al unir dichos vóxeles, se crea una representación de la superficie de la ROI (Figura 16), que coincidirá en mayor o menor medida según el tamaño de los cubos y la cual, posteriormente, se analizará para extraer las características de forma [20].

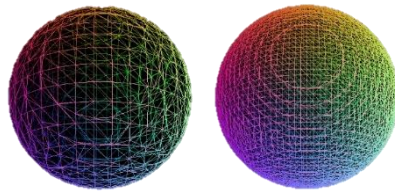


Figura 16. Dos posibles aproximaciones de la representación de la superficie de la ROI [20].

Una vez obtenida la malla de la ROI, ya se pueden extraer las características de forma a partir de dicha aproximación mediante *PyRadiomics*, proceso que se puede realizar de distintas maneras.

En este Trabajo, para dicha extracción se ha hecho uso de un modo denominado Modo de Configuración, en el cual el usuario selecciona aquellas características de interés que desee, de manera que, en este caso, se han extraído únicamente las características de forma, las cuales son 17, como se puede observar en la Tabla 2.

Tabla 2. Características de forma obtenidas a partir de la extracción de características con *PyRadiomics* de Python

CARACTERÍSTICAS DE FORMA
Mesh Volume
Voxel Volume
Surface Area
Surface Volume Ratio
Sphericity
Compactness 1
Compactness 2
Spherical Disproportion
Maximum 3D Diameter
Maximum 2D Diameter Slice
Maximum 2D Diameter Column
Maximum 2D Diameter Row
Major Axis Length
Minor Axis Length
Least Axis Length
Elongation
Flatness

2.4.5. Selección de características

Una vez se han obtenido las características tanto de texturas como de formas mediante los métodos de extracción de características, se debe hacer una reducción de las mismas mediante los métodos de selección de características. Estos métodos se utilizan para definir aquellas características que son más significativas para posteriormente realizar una adecuada clasificación, es decir, permite que el orden de relevancia de todas las características se vea reflejado en un ranking, el cual será utilizado en dicho proceso de creación de modelos de clasificación, donde para cada uno de ellos se seleccionarán el número de texturas óptimas de ese ranking.

Este proceso ofrece ciertas ventajas como descartar variables irrelevantes y redundantes, facilitar la comprensión y visualización de los datos, reducir los requerimientos de almacenamiento y los costes computacionales de los modelos estimados y ofrecer un mejor rendimiento a los modelos predictivos elaborados posteriormente [22].

Los métodos de selección de características se dividen en dos grandes grupos, la **aproximación indirecta o *filters***, y la **aproximación directa o *wrappers***.

Los ***filters*** permiten la selección de aquellas variables con mejor ajuste respecto a una función objetivo sin tener en cuenta la interacción con el clasificador, es decir, centran en la capacidad predictiva individual de cada característica realizando un ranking de todas ellas en función de su relevancia [18], [22].

Mientras que los ***wrappers*** permiten la selección de aquellas variables con mejores resultados para el modelo predictivo, es decir, se basan en un modelo predictivo para seleccionar aquellas características con un rendimiento óptimo [18], [22].

Este Trabajo Fin de Grado se va a centrar en los métodos ***filters***, en concreto, en el ***ReliefF***, un método de filtro que realiza una selección aleatoria de n instancias y busca para cada n su vecino más cercano que pertenezca tanto a la misma clase como a la clase contraria, de manera que proporciona una mayor relevancia a aquellas variables que mejor separan las instancias de distinta clase [22].

Además, se ha decidido la realización de la selección de características mediante el método ***ReliefF*** ya que presentan la ventaja de ser más simples, de ejecución más rápida y de ser robustos frente al sobreajuste o ***overfitting***. A pesar de que este método no considera la capacidad predictiva de un subconjunto de características [22].

2.4.6. Métodos de clasificación de características

El proceso del análisis de textura finaliza en la clasificación de las características, para ello es posible el uso de métodos estadísticos simples o de métodos de aprendizaje automático o ***machine learning***.

Los **métodos de aprendizaje automático** consisten en un proceso automático que aprende de los datos proporcionados y tiene como objetivo la ayuda a la decisión médica. Este método puede ser supervisado o no supervisado, en el **aprendizaje supervisado**, el sistema aprende de los datos de entrenamiento e intenta predecir la salida de los datos de test, mientras que en el

aprendizaje no supervisado, el sistema no dispone de datos de entrenamiento, solo de datos de entrada y, es por ello que se va adaptando a los propios datos para predecir la salida [23].

Este trabajo se va a centrar en los métodos de aprendizaje automático supervisado debido a que la partición de los datos nos va a permitir disponer de datos de entrenamiento para un adecuado aprendizaje. Para llevar a cabo esta clasificación, se desarrollan 4 modelos predictivos válidos tanto para estudios binarios como multiclase.

1. Modelo Naive Bayes (NB).

El modelo de Naive Bayes se basa en una técnica de clasificación estadística denominada Teorema de Bayes. Esta técnica se caracteriza por trabajar con probabilidad y por presuponer que los predictores sean condicionalmente independientes.

El modo de funcionamiento de dicho método es convertir los datos en una tabla de frecuencias, a partir de la cual se podrán calcular las probabilidades que ocurran diversos eventos mediante el Teorema de Bayes (Ecuación 1). Una vez obtenidas todas las probabilidades para cada clase, aquella clase con la probabilidad más alta será el resultado de la predicción [24].

$$P(A|R) = \frac{P(R|A)P(A)}{P(R)} \quad \left\{ \begin{array}{l} P(A): \text{Probabilidad de A} \\ P(R|A): \text{Probabilidad de que se de R dado A} \\ P(R): \text{Probabilidad de R} \\ P(A|R): \text{Probabilidad posterior de que se de A dado R} \end{array} \right. \quad \text{(Ecuación 1)}$$

2. Modelo K-Nearest Neighbour (KNN).

El modelo K-vecinos más próximo es un modelo basado en la memoria que permite la clasificación de los datos haciendo uso del vecindario de una muestra para predecir la clase a la que corresponde. Además, se trata de un modelo no paramétrico basado en distancias, donde las métricas más utilizadas son la distancia del ajedrez, de Manhattan o la distancia euclídea, pudiéndose observar su representación en la Figura 17 [25].

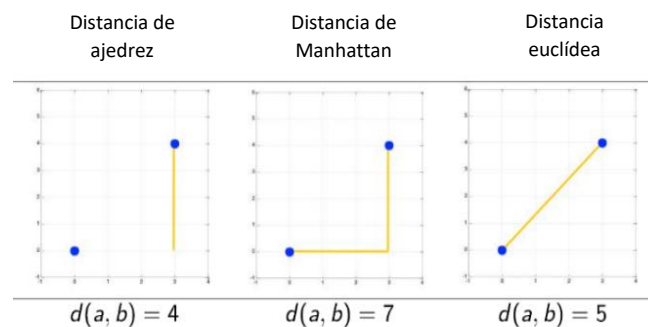


Figura 17. Los 3 tipos de métricas más utilizadas en el modelo KNN [25].

Por tanto, el vecino más próximo busca la clase del vecino más cercano en función de la métrica que se esté utilizando. Cabe destacar que se debe elegir un número de vecinos (K) adecuado para cada conjunto de datos (Figura 18), es decir, cuanto más pequeño sea el valor de K, más pequeñas regiones de cada clase se crearán. Por tanto, valores altos de K implicarán menos regiones de mayor tamaño, lo que reducirá el posible ruido en la clasificación [26].

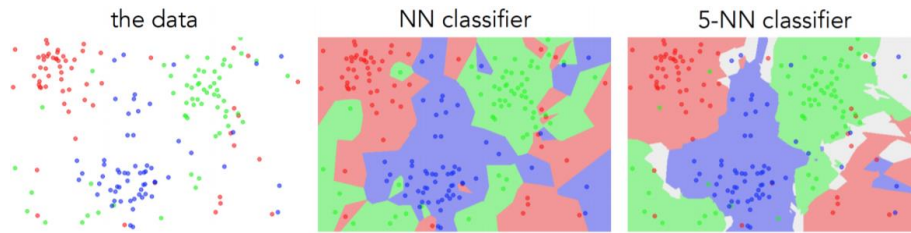


Figura 18. Fronteras de decisión para el modelo KNN con valores de $K=1$ y $K=5$ [27].

3. Modelo Support Vector Machines (SVM). Con Kernel Lineal y Kernel Gaussiano.

El modelo de Máquinas de Vectores de Soporte se basa en el hiperplano óptimo, una extensión de plano capaz de separar los datos en dos clases, obteniendo una distancia o margen máximo con las muestras más cercanas al mismo, denominadas Vectores de Soporte de cada clase.

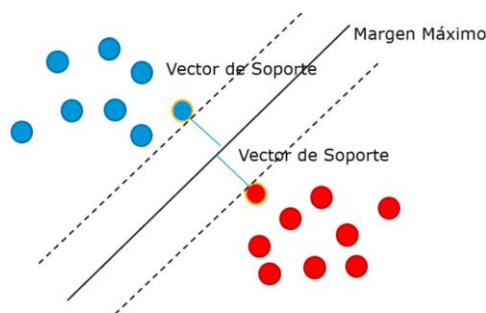


Figura 19. Modelo SVM ilustrando el hiperplano con el margen máximo y los vectores de soporte de cada clase [28].

Para obtener el hiperplano, el sistema debe ser entrenado con unos datos de entrenamiento que pertenecen a la zona donde están presentes los Vectores de Soporte, de tal manera que los datos etiquetados por una clase estarán a un lado del hiperplano y los datos etiquetados por la otra clase estarán al otro lado, como se puede observar en la Figura 19 [29].

Además, cabe destacar los estudios de multiclase para este tipo de modelos ya que el concepto de hiperplano no sería el correcto, sino que serían necesarios los conceptos *one vs. one* o *one vs. all* [29].

En el caso de *one vs. one*, la clasificación se realiza mediante $\frac{N*(N-1)}{2}$ clasificadores binarios, donde N es el número de clases. Cada clasificador binario recibe las muestras de un par de clases de los datos de entrenamiento y aprende a distinguir entre dichas clases. De manera que para clasificar entre todas las clases presentes, todos los clasificadores binarios intentan clasificar cada nuevo dato de test, que pertenecerá a la clase que obtuvo mayor número de predicciones.

Mientras que en el caso de *one vs. all*, la clasificación se realiza con un clasificador para cada clase, de manera que el sistema aprende a distinguir cada clase de las demás y ante un nuevo dato de test, el modelo es capaz de determinar a qué clase pertenece.

4. Modelo Bagged Tree (BT).

El modelo Bagged Tree se basa en algoritmos de aprendizaje con el fin de obtener una mejor rendimiento predictivo. Para ello, se generan aleatoriamente distintos subconjuntos de datos de entrenamiento, de manera que cada subconjunto de datos se utiliza para entrenar distintos árboles de decisión, obteniendo así un modelo de predicción para cada uno. Para obtener el

modelo predictivo Bagged Tree definitivo se realiza un promedio de todas las predicciones obtenidas previamente, como se puede observar en la Figura 20 [30].

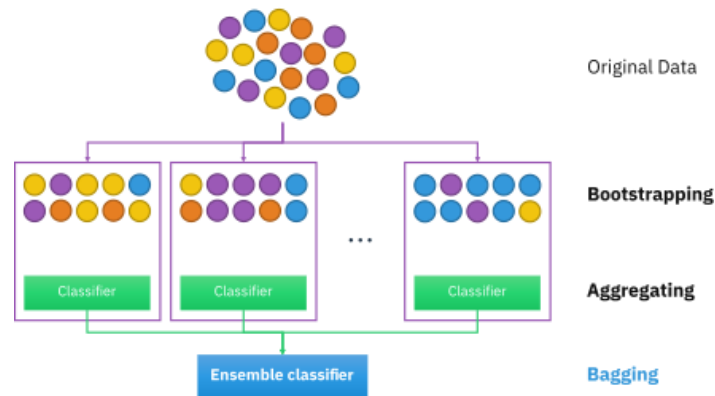


Figura 20. Clasificación de un Modelo Bagged Tree [30].

2.4.7. Validación de sistemas

Para poder realizar una comprobación del correcto funcionamiento predictivo del sistema es necesario llevar a cabo alguna de las existentes estrategias de validación de los modelos, entre las cuales destacan la validación cruzada o K-fold y la validación simple o Hold-out. Básicamente, este proceso consiste en realizar una división de los datos iniciales, de manera que se obtienen dos grupos, el de los datos de entrenamiento para poder entrenar a los distintos modelos predictivos y el de los datos de test para comprobar la eficacia de cada uno de ellos.

- **Validación cruzada o K-fold.**

El fundamento de esta estrategia recae en realizar k divisiones aleatorias de dos subconjuntos de datos, datos de entrenamiento y de test, de manera que para cada iteración se dispondrá de dos subgrupos, siendo el subgrupo de test totalmente independientes y distintos entre ellos. Finalmente, no se obtiene k grupos de entrenamiento y k grupos de test, sino que se realiza una media aritmética de cada iteración y se obtiene un único resultado [31].

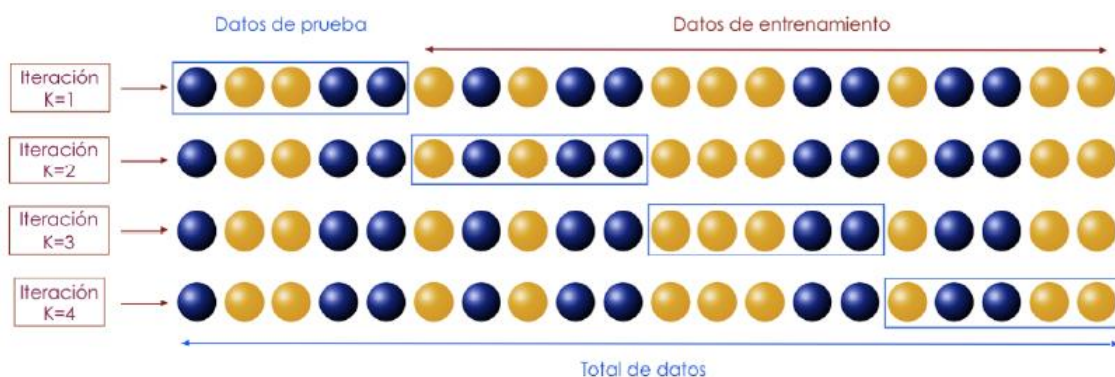


Figura 21. Estrategia de validación cruzada o K-fold con K iteraciones [31].

Cabe destacar que este método es utilizado cuando el conjunto de datos no es muy grande, de manera que para que ambos grupos de datos, entrenamiento y test, sean representativos se hace uso de las K iteraciones.

- **Validación simple o Hold-Out.**

Por otro lado, el fundamento de esta estrategia recae en realizar una única división aleatoria de los datos, obteniendo así unos datos de entrenamiento y otros de test, siendo los de entrenamiento un subgrupo de mayor tamaño. De esta manera, es posible entrenar al sistema con los datos de entrenamiento y posteriormente validarlo con los datos de test [31].



Figura 22. Estrategia de validación simple o Hold-Out [31].

Cabe destacar que este método es utilizado cuando se dispone de un conjunto de datos lo suficientemente grande como para poder separarlo en dos y que sean representativos.

Ambos métodos tienen sus ventajas, la validación cruzada o K-Fold tiene una alta precisión ya que evalúa K subgrupos de datos distintos y no uno solo como en el Hold-Out. No obstante, la validación simple o Hold-Out tiene la ventaja de ser computacionalmente menos costosa debido a que no tiene que realizar todas las iteraciones que en el K-Fold deben hacerse, siendo éste un proceso más lento.

En este Trabajo, se ha decidido trabajar con la validación simple o Hold-Out debido a su fácil implementación y rapidez en la computación. Cabe destacar que el posible riesgo de *overfitting* o sobreajuste del modelo y la menor precisión que presenta dicho método de validación pretenden ser reducidas con la realización de 100 iteraciones de dicho proceso.

2.4.8. Métodos de evaluación de resultados

Finalmente, para evaluar de manera objetiva el rendimiento de los modelos predictivos se hace uso de determinadas métricas de evaluación. Esta evaluación de los resultados obtenidos permite determinar la precisión de cada clasificador y, con ello, el modelo de clasificación que mejor se ajusta a nuestros datos y que mejores resultados nos proporciona.

Seguidamente, se expondrán algunas de las métricas de evaluación de modelos de clasificación existentes, que son las que se utilizarán posteriormente en este Trabajo Fin de Grado.

- **Curva ROC. Curva de Característica Operativa del Receptor**

La curva ROC (Figura 23) es una gráfica que representa la tasa de verdaderos positivos (Ecuación 2) en función de la tasa de falsos positivos (Ecuación 3), es decir, Tasa de VP vs. Tasa de FP. Permite la evaluación del rendimiento de los modelos de clasificación considerando las probabilidades de cada clase [32].

DESARROLLO DE UNA METODOLOGÍA PARA LA AYUDA AL DIAGNÓSTICO DEL CÁNCER DE MAMA MEDIANTE EL ANÁLISIS DE TEXTURAS EN IMÁGENES DE RESONANCIA MAGNÉTICA

$$\text{Sensibilidad} = \text{TVP} = \frac{VP}{VP+FN} \quad (\text{Ecuación 2})$$

$$\text{Especificidad} = \text{TFP} = \frac{FP}{FP+VN} \quad (\text{Ecuación 3})$$

Cabe destacar que la curva ROC representa la sensibilidad (Ecuación 2) en función de 1 menos la especificidad (Ecuación 3), parámetros que se obtienen a partir de la matriz de confusión.

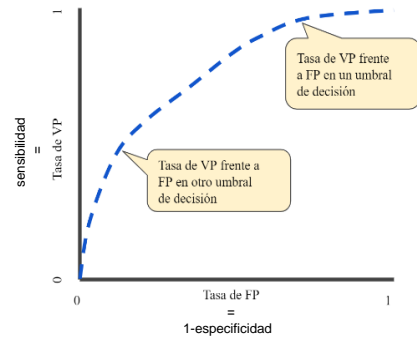


Figura 23. Curva ROC, VPR vs. FPR en diferentes umbrales de clasificación [32].

▪ **AUC. Área Bajo la Curva**

El área bajo la curva o AUC (Figura 24) se obtiene a partir de la curva ROC y mide el área bidimensional por debajo de dicha curva. Representa la probabilidad de que un nuevo dato sea clasificado correctamente, es decir, representa el valor de la predicción de cada modelo cuyo rango de valores es [0,1], siendo 0 un valor de predicción nulo y 1 implica que todas las predicciones que realiza el modelo son correctas, es decir, un modelo de predicción ideal. De este modo, se puede considerar como un modelo poco efectivo aquel que tenga un valor de AUC inferior a 0'5 [32].

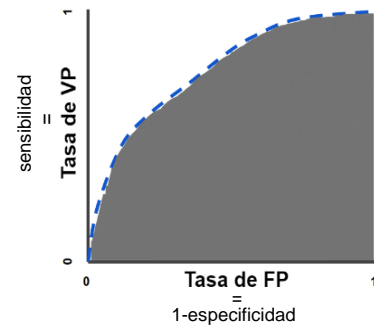


Figura 24. Curva ROC indicando el área bajo la curva (AUC) [32].

Cabe destacar que la obtención de la curva ROC y, con ello, el AUC se puede extender a estudios multiclase mediante la técnica Uno vs Todos (one vs. all), comentada anteriormente, calculando la curva ROC y el AUC considerando una clase y todas las demás agrupadas como una sola.

▪ **Matriz de Confusión**

Dado un número de clases N, la matriz de confusión es una matriz NxN que representa el éxito de las predicciones frente a los datos reales, es decir, la validez de la correlación entre los datos y las clasificaciones que el modelo realiza. Es una matriz donde las columnas representan el valor predicho de cada clase y las filas el valor real de las mismas. De manera que, la diagonal principal hace referencia a aquellas predicciones que han sido acertadas, los denominados Verdaderos Positivos (VP) y los Verdaderos Negativos (VN), mientras que la diagonal inversa de la matriz son errores que el modelo ha cometido, los Falsos Positivos (FP) y los Falsos Negativos (FN) [32].

Para analizar la matriz de confusión, en primer lugar, es necesario hacer referencia a un estudio binario, de dos clases, donde una de las clases es positiva y la otra negativa. Es por ello que la matriz de confusión es de 2x2 (Tabla 3) y es donde se observan los conceptos VP, VN, FP y FN.

Tabla 3. Matriz de confusión 2x2 para un estudio binario de 2 clases. Fuente: elaboración propia.

		VALOR PREDICHO	
		Clase A. Positiva	Clase B. Negativa
VALOR REAL	Clase A. Positiva	Verdaderos Positivos (VP)	Falsos Negativos (FN)
	Clase B. Negativa	Falsos Positivos (FP)	Verdaderos Negativos (VN)

DESARROLLO DE UNA METODOLOGÍA PARA LA AYUDA AL DIAGNÓSTICO DEL CÁNCER DE MAMA MEDIANTE EL ANÁLISIS DE TEXTURAS EN IMÁGENES DE RESONANCIA MAGNÉTICA

En segundo lugar, cabría destacar las matrices de confusión para estudios multiclase, donde hay distintas clases y, por tanto, no hay una definida como positiva y otra como negativa. De esta manera, se obtiene una matriz de confusión de 4x4 (Tabla 4) donde cada columna corresponde a cada clase predicha, cada fila a cada clase real, la diagonal principal son los valores predichos correctamente y el resto de elementos de la matriz son valores predichos incorrectamente, por ejemplo, el elemento (2,1) es Clase B pero se ha predicho como Clase A [33].

Tabla 4. Matriz de confusión 4x4 para un estudio multiclase de 4 clases. Fuente: elaboración propia.

		VALOR PREDICHO			
		Clase A	Clase B	Clase C	Clase D
VALOR REAL	Clase A	Verdaderos Positivos A (VP _A)	Falsos AB (F _{AB})	Falsos AC (F _{AC})	Falsos AD (F _{AD})
	Clase B	Falsos BA (F _{BA})	Verdaderos Positivos B (VP _B)	Falsos BC (F _{BC})	Falsos BD (F _{BD})
	Clase C	Falsos CA (F _{CA})	Falsos CB (F _{CB})	Verdaderos Positivos C (VP _C)	Falsos CD (F _{CD})
	Clase D	Falsos DA (F _{DA})	Falsos DB (F _{DB})	Falsos DC (F _{DC})	Verdaderos Positivos D (VP _D)

En este Trabajo se ha tratado tanto con estudios binarios, para estudiar el tipo de tumoración (Luminal B con HER2 negativo y HER2 positivo), como con estudios multiclase, para analizar la localización del tumor en los distintos cuadrantes de la mama (CSE, CSI, CIE, CII).

A partir de estas matrices se puede obtener dos parámetros adicionales que permiten evaluar de manera distinta los resultados, la sensibilidad y la especificidad. La sensibilidad equivale a TVP (Ecuación 2), representa la proporción de Positivos (Clase A) bien predichos, mientras que la especificidad equivale a TVN (Ecuación 3) y, por tanto, representa la proporción de Negativos (Clase B) correctamente predichos. Estos conceptos también se pueden extrapolar al caso multiclase, de manera que se obtendría una sensibilidad y una especificidad para cada clase.

3. MATERIALES

En esta sección, se expondrán aquellos materiales que han sido utilizados para el desarrollo de este Trabajo Fin de Grado, así como una breve explicación de cada uno de ellos. Se hablará de dónde y cómo se han adquirido las imágenes de RM que se han utilizado, de los distintos software empleados para el preprocesado de las mismas, así como de las herramientas para el desarrollo del propio análisis de texturas y de formas.

3.1. IMÁGENES DE RM

Las Imágenes de Resonancia Magnética que se han sido utilizadas para el desarrollo del presente estudio han sido proporcionadas por ASCIRES Grupo Biomédico. Se disponen de un total de 43 imágenes tomadas entre septiembre de 2017 y marzo de 2020 correspondientes a pacientes con una edad media de $58'84 \pm 10'03$ años. Para poder realizar el estudio se ha tenido que hacer una selección de aquellas pacientes que estaban recientemente diagnosticadas con algún tipo de lesión en las mamas, en concreto, con una tumoración.

Todas las imágenes obtenidas se realizaron en el mismo centro ASCIRES Grupo Biomédico y en la misma máquina de resonancia magnética, con finalidad de disminuir la variabilidad de la toma de imágenes y reducir así grandes diferencias que pudieran influir en los posteriores análisis. Además, cabría destacar que estas imágenes se adquirieron en el equipo de RM de Philips Achieva de 3T (Koninklijke Philips N.V., Ámsterdam, Países Bajos) y con el mismo protocolo, de tal forma que se obtuvieron unas series, de las cuales se han empleado 3 en este estudio: *T1 TSE AXIAL*, *AX T2W_SPAIR* y *DINAMICA AXIAL THRIVE*.

- *T1 TSE AXIAL*. Obtenida con un espesor de corte de 3 mm, matriz de reconstrucción 576x576, tiempo de eco (TE) de 10 ms, tiempo de recuperación (TR) de 360 ms y una resolución en x e y de 0.6250 x 0.6250 mm².
- *AX T2W_SPAIR*. Obtenida con un espesor de corte de 3 mm, matriz de reconstrucción 432x432, un TE de 60 ms, un TR de 4500 ms y con supresión de grasa SPAIR.
- *DINAMICA AXIAL THRIVE*. Obtenida con una matriz de reconstrucción 352x352, el TE más corto, el TR más corto y con supresión de grasa SPAIR. Además, consta de 7 secuencias dinámicas con tiempos de secuencia definidos por el usuario.

Finalmente, para evaluar de igual forma las distintas lesiones de la mama, un mismo radiólogo ha redactado los informes diagnósticos de cada paciente y subido al software *PACS CareStream*.

3.2. SOFTWARES EMPLEADOS PARA EL PRE-PROCESADO DE LAS IMÁGENES

Para poder trabajar con las imágenes previamente han tenido que ser pre-procesadas, para ello, se han utilizado distintos software que seguidamente se explicarán.

1. *PACS CareStream* (Carestream Health, Rochester, Nueva York, Estados Unidos). Se ha realizado la búsqueda de los casos en el software *PACS CareStream* y posteriormente, estas imágenes adquiridas se han exportado desde el propio software en formato DICOM.
2. *Medexprim Suite* (Medexprim, Labège, Francia). Antes de empezar a tratar con dichas imágenes, han tenido que ser anonimizadas debido a la confidencialidad de los datos de

cada paciente y es por ello que se ha usado este programa ya que permite la extracción de cohortes de pacientes extraídos del PACS de forma anonimizada.

3. *ITK-SNAP* (PICSL, Pennsylvania, Estados Unidos y SCI, Utah, Estados Unidos) y *3DSlicer* (National Institutes of Health (NIH), Estados Unidos). Una vez tenemos las imágenes de todos los cortes axiales de ambas mamas, se debe realizar la segmentación de la lesión de cada paciente. Para ello, se ha utilizado el software *ITK-SNAP* como *3DSlicer* y se ha segmentado la lesión en la serie *T1 TSE AXIAL* y las otras *AX T2W_SPAIR* y *DINAMICA AXIAL THRIVE* se han utilizado como apoyo para reconocer con mayor precisión donde se encuentra exactamente la tumoración. Una vez realizada la segmentación en todos los cortes donde había lesión, se extraen dichas imágenes en formato NIFTI.
4. *QDoc* (Agfa HealthCare N. V., Mortsel, Bélgica). Un sistema de información radiológica donde se recogen todos los informes redactados por el radiológica, de los cuales se han hecho uso para facilitar la segmentación.
5. *RadiAnt DICOM Viewer* (Medixant., Poznan, Polonia). Este software es un visor en formato DICOM de imágenes médicas, el cual ha sido utilizado para visualizar las series correspondientes a cada paciente y facilitar así la segmentación.
6. *dcm2niix* (MRICroGL, Universidad de Carolina del Sur, Estados Unidos). Como se ha comentado anteriormente, las imágenes que se han extraído desde el *PACS Carestream* están en formato DICOM. No obstante, para, posteriormente, poder trabajar más fácilmente se ha realizado una conversión del mismo formato DICOM a formato NIFTI debido a que es más sencillo para MATLAB.

3.3. SOFTWARE MATLAB

MATLAB (The MathWorks Inc., Natick, MA, USA) es un software matemático que permite el lenguaje de programación propio y, además, ofrece infinidad de posibilidades. En este caso, se ha utilizado esta herramienta para implementar un código con las funciones de los modelos de clasificación con la finalidad de realizar tanto el análisis de textura como de formas.

3.3.1. Radiomics 2D

Es un paquete de herramientas de programación creado por Martin Vallières que permite la extracción de texturas en 2D ya que gracias a esta función de MATLAB se realiza el preproceso de las imágenes proporcionadas, obteniendo finalmente un archivo Excel con las 43 texturas extraídas para cada nivel de gris presente (16, 32, 64 y 126) y para los 43 pacientes dados.

3.4. SOFTWARE PYTHON

3.4.1. PyRadiomics

Python (Instituto de Investigación en Ámsterdam, Países Bajos) es un lenguaje de programación que soporta orientación a objetos, programación imperativa y funcional. Además, presenta un paquete de herramientas muy útil denominado PyRadiomics ya que permite la extracción automatizada de características, en este caso, de formas. De esta manera, se obtiene un archivo Excel con las 17 formas extraídas para los 43 pacientes dados, para posteriormente poder realizar el análisis de formas en 3D.

4. METODOLOGÍA

En esta sección se va a explicar el procedimiento que se ha seguido en el transcurso de este Trabajo Fin de Grado para llegar a obtener un sistema de ayuda al diagnóstico del cáncer de mama mediante el análisis de texturas y de formas en imágenes de resonancia magnética de pacientes con este tipo de cáncer.

Para ello, se puede observar que el estudio se divide en dos grandes partes, la segmentación de la tumoración en la mama correspondiente a partir de las imágenes potenciadas en T1 y, seguidamente, la obtención y elección del modelo de clasificación, de entre todos los obtenidos, que tenga alta capacidad de predecir.

Para la segunda parte, se ha decidido la realización de tres estudios, uno binario y otros dos multiclase, en uno de los estudios multiclase se analiza la localización por cuadrantes de la mama del tumor y es de 4 clases (CIE, CII, CSE y CSI), mientras que en el otro se analiza la localización por porciones y es de 3 clases (porción anterior, porción media y porción posterior) y en el estudio binario se analiza el tipo de tumoración y es 2 clases (HER2 positivo y Luminal B con HER2 negativo).

4.1. SEGMENTACIÓN DE LA LESIÓN

Una vez obtenidas las imágenes adquiridas anonimizadas, el proceso de la segmentación de la lesión puede ser realizado mediante el software ITK-SNAP, como se ha dicho anteriormente. Cabe destacar que es necesario disponer de la serie T1 TSE AXIAL ya que es sobre la que va a realizar la segmentación, de esta manera se podrá realizar dicho proceso en todos los cortes axiales donde esté presente la tumoración.

El software ITK-SNAP permite la realización de una segmentación semiautomática utilizando métodos de contorno activos, delineación manual, entre otras herramientas de ayuda para la propia segmentación.

Una vez cargada la imagen T1 de una determinada paciente en el software ITK-SNAP, ya se podría empezar a segmentar. Para ello, primeramente se marca la zona del tumor y se realiza una segmentación semiautomática con contorno activos, obteniendo una ROI en cada corte de la lesión.

No obstante, este proceso de semi-automatización no es el más adecuado debido a que el ajuste y la delineación de la lesión no es del todo correcta. Es por ello que se recurre al método de la delineación manual, para ajustar correctamente la lesión con la ayuda de la visualización de las otras series, T2 y DINÁMICA, para comprobar en qué zonas el tejido coge contraste y en cuáles no y así poder realizar con todo detalle la segmentación. De esta forma, y ajustando la segmentación de la lesión en cada corte obtenemos la ROI del tumor en 3D, como se puede observar en la Figura 25.

DESARROLLO DE UNA METODOLOGÍA PARA LA AYUDA AL DIAGNÓSTICO DEL CÁNCER DE MAMA MEDIANTE EL ANÁLISIS DE TEXTURAS EN IMÁGENES DE RESONANCIA MAGNÉTICA

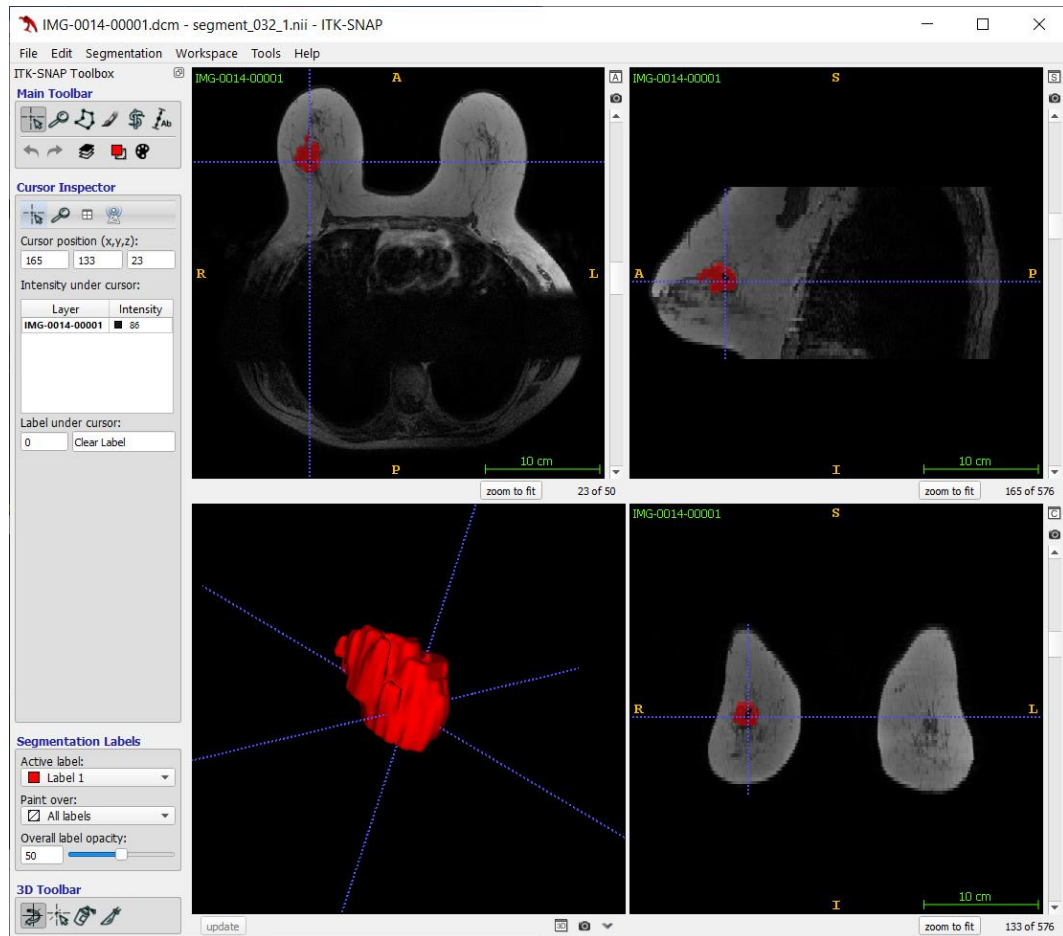


Figura 25. Segmentación de una tumoración en la mama derecha mediante el software ITK-SNAP. Fuente: elaboración propia.

4.2. ANÁLISIS DE TEXTURAS Y FORMAS Y CLASIFICACIÓN

En el momento en que ya se dispone de las imágenes preprocesadas y segmentadas, se podría comenzar el proceso del análisis de características para cada volumen segmentado, es decir, para cada una de las regiones de interés obtenida que, en este caso, es la tumoración de cada una de las pacientes.

Como se ha comentado anteriormente, los tres grandes pasos a seguir y que se van a analizar con mayor detalle son: análisis de texturas y formas, selección de características y clasificación.

4.2.1. Extracción de características de textura y forma

4.2.1.1. Extracción de texturas

La extracción de características de textura es posible gracias a la herramienta de MATLAB *Radiomics 2D*, que permite la extracción de información cuantitativa presente en las imágenes médica proporcionadas. Para llegar a obtener el archivo Excel de las 43 texturas para cada nivel de gris impuesto por el usuario y para las 43 pacientes de partida, se requiere de la realización de ciertos procesos.

Esta función de MATLAB, primeramente, tiene implementado en su código el preprocesado de las imágenes, es decir, lleva a cabo la interpolación, normalización y cuantización del número de niveles de gris quedándonos únicamente con aquellos que se desea. No obstante, en este Trabajo se ha decidido no realizar la interpolación debido a que no es necesario mejorar la resolución ya que la resolución original es suficiente para el análisis que se va a realizar. Una vez se ha preprocesado la imagen, empieza el proceso de la extracción de texturas mediante métodos estadísticos.

Radiomics 2D permite la extracción de un total de 43 texturas a partir de los siguientes métodos estadísticos: Global, GLCM, GLRLM, GLSZM y NGTDM, explicados anteriormente. No obstante, este proceso se repite en 4 ocasiones, obteniendo en cada repetición 43 texturas, pero para niveles de gris distintos, en este caso, para 16, 32, 64 y 128. De esta manera, se determinará qué número de niveles de gris proporciona mejores resultados de clasificación, además se podrá comprobar si a medida que aumenta el número de niveles de gris, disminuye la relación señal ruido, obteniendo así una mejora en el valor predictivo de dichos modelos de clasificación en función de las texturas.

Con todo ello, se obtienen 2 archivos (formato *.csv* y formato *.mat*), donde aparecen las texturas extraídas para los diferentes niveles de gris de todas las pacientes del estudio. Finalmente, para obtener un archivo para cada nivel de gris se debe realizar la separación de los datos, de manera que se obtienen 4 archivos para cada nivel de gris de 43x44 donde las filas corresponden a cada paciente y las columnas a la etiqueta de cada clase seguida de los resultados de las texturas.

4.2.1.2. Extracción de formas

La extracción de características de forma es posible gracias a la herramienta de Python *PyRadiomics*, que permite la extracción automática de información cuantitativa presente en las imágenes médica proporcionadas. Para llegar a obtener el archivo Excel de las 17 formas para las 43 pacientes de partida, se requiere de la realización de una malla triangular que recorra toda la ROI de la imagen, como se ha explicado anteriormente.

Una vez obtenida la malla de la ROI, como no es necesario un preproceso de la imagen, como en el caso anterior, ya se pueden extraer las características de forma a partir de dicha aproximación *PyRadiomics*. De manera que se obtiene un archivo en formato *.csv* de 43x17 donde las filas corresponden a cada paciente y las columnas a los resultados de las formas.

4.2.2. Selección de características

La selección de características es un proceso a destacar debido a la mayor o menor importancia que tienen cada una de las características. Para ello, en este caso, se ha utilizado el método de la aproximación indirecta o *filter*, en concreto, se ha utilizado el *reliefF*, el cual proporciona una mayor relevancia aquellas variables que mejor separan las instancias de distinta clase.

Para la implementación de la función *reliefF* es necesario el archivo *.csv* extraído anteriormente ya que las variables de entrada con los datos y las etiquetas iniciales son extraídas de dicho archivo *.csv*. De estos datos se obtienen los datos de entrenamiento, los cuales son sometidos a 100 iteraciones, obteniendo así 100 matrices, de 1x43 en el caso de las texturas y de 1x17 en el caso de las formas, que corresponden al ranking de dichas características. Finalmente, para obtener el ranking definitivo se obtiene la posición media de cada característica de manera que

DESARROLLO DE UNA METODOLOGÍA PARA LA AYUDA AL DIAGNÓSTICO DEL CÁNCER DE MAMA MEDIANTE EL ANÁLISIS DE TEXTURAS EN IMÁGENES DE RESONANCIA MAGNÉTICA

se obtiene de mayor a menor las características más utilizadas considerando las 100 iteraciones realizadas. Por ejemplo, en la Figura 26 se puede observar el ranking de las 15 características de texturas más relevantes para el nivel de gris 16.

Características de Texturas	Texturas ordenadas por relevancia	Posición media ranking
{'SumAverage_GLCM' }	9	4.7
{'LZHGE_GLSZM' }	36	6.09
{'Kurtosis_GT' }	3	6.37
{'Skewness_GT' }	2	9.15
{'LZE_GLSZM' }	27	10.94
{'GLV_GLRIM' }	24	11.21
{'SRLGE_GLRIM' }	20	12.17
{'Variance_GLCM' }	10	12.23
{'Busyness_NGTM' }	41	14.7
{'LGRE_GLRIM' }	18	15.04
{'LGZE_GLSZM' }	31	17.25
{'SZLGE_GLSZM' }	33	17.34
{'RLV_GLRIM' }	25	17.46
{'LRHGE_GLRIM' }	23	17.9
{'LZLGE_GLSZM' }	35	18.1

Figura 26. Ranking de las 15 texturas más relevantes para el nivel de gris 16. Fuente: elaboración propia.

4.2.3. Clasificación

Se han desarrollado distintos modelos de clasificación cuyos resultados predictivos serán evaluados mediante distintas métricas de evaluación para demostrar cuál es el modelo más válido para los datos aportados.

Los modelos de clasificación que se han llegado a desarrollar en el código han sido los 4 que se han explicado anteriormente, Naive Bayes, KNN, SVM con Kernel Lineal y con Kernel Gaussiano y Bagged Tree, mientras que las métricas de evaluación propuestas han sido los valores de AUC y las matrices de confusión junto con la sensibilidad y la especificidad. Cabe destacar que tanto para el estudio binario como para el multiclase, se han desarrollado los mismos modelos y las mismas métricas de evaluación pero con ligeras modificaciones en el propio código, por ejemplo, la adición del concepto Uno vs. Todos en el modelo SVM y en el AUC o el cálculo de la sensibilidad y la especificidad al obtener una matriz de confusión de mayor tamaño.

Además, para la validación de los modelos implementados se ha realizado la validación simple o Hold-Out con 100 iteraciones, es decir, 100 particiones de los datos en dos grupos, un 75% son datos de entrenamiento para poder entrenar a los distintos modelos predictivos y el 25% restante son los datos de test para comprobar la eficacia de cada uno de ellos.

Además, como bien se ha comentado, una vez obtenido el modelo predictivo validado, se evalúa con determinadas métricas de evaluación, obteniendo el AUC promedio con su correspondiente desviación típica y la matriz de confusión de cada modelo.

5. RESULTADOS

Este Trabajo Fin de Grado tiene como objetivo obtener un sistema de ayuda al diagnóstico del cáncer de mama mediante el análisis de texturas y de formas en imágenes de resonancia magnética de pacientes con este tipo de cáncer. Para ello, como bien se ha comentado anteriormente, se han realizado tres estudios, uno binario y otros dos multiclase (Tabla 5). En primer lugar, en uno de los estudios multiclase se analiza la localización del tumor por cuadrantes de la mama y es de 4 clases (CIE, CII, CSE y CSI), en el otro estudio multiclase se analiza la localización por porciones y es de 3 clases (Porción anterior, porción media y porción posterior) y finalmente, en el estudio binario se analiza el tipo de tumoración y es 2 clases (HER2 positivo y Luminal B con HER2 negativo).

Tabla 5. Datos de los estudios realizados en el presente trabajo. Fuente: elaboración propia.

ESTUDIO	DESCRIPCIÓN	CLASES (Número – Nombre – Tamaño)		
Multiclase	Localización del tumor por cuadrantes de la mama	4	CIE. Cuadrante inferoexterno	15
			CII. Cuadrante inferointerno	6
			CSE. Cuadrante superoexterno	10
			CSI. Cuadrante superointerno	12
Multiclase	Localización del tumor por porciones de la mama	3	Porción anterior	3
			Porción media	32
			Porción posterior	8
Binario	Tipo de tumoración	2	HER2 positivo	23
			Luminal B con HER2 negativo	20

Como se puede observar en la Tabla 5, es de gran relevancia destacar el tamaño de muestras de cada clase ya que, en ocasiones, es muy variable, como es en el caso del estudio multiclase de la localización del tumor por porciones mamarias. Es por ello, que se ha visto dificultado su desarrollo y no se ha podido llegar a implementar debido a la escasez de muestras en la clase ‘Porción anterior’, además de la gran diferencia con respecto a las otras clases, sobre todo, a la clase ‘Porción media’.

Por tanto, la obtención de los resultados se han evaluado los modelos obtenidos del análisis de textura, de formas y, finalmente, del análisis de las características de textura y de formas conjuntamente, todos ellos tanto para el estudio multiclase como para el binario (Tabla 6).

Tabla 6. Estructura organizativa de los resultados. Fuente: elaboración propia.

1. ANÁLISIS DE TEXTURAS	2. ANÁLISIS DE FORMAS	3. ANÁLISIS DE TEXTURAS Y FORMAS
1.1. Estudio multiclase. Localización del tumor por cuadrantes de la mama.	2.1. Estudio multiclase. Localización del tumor por cuadrantes de la mama.	3.1. Estudio multiclase. Localización del tumor por cuadrantes de la mama.
1.2. Estudio binario. Tipo de tumoración.	2.2. Estudio binario. Tipo de tumoración.	3.2. Estudio binario. Tipo de tumoración.

De todos los estudios que se han realizado, finalmente, se ha obtenido el AUC y la matriz de confusión para cada modelo presente en cada estudio, así como las características utilizadas, bien sea de textura, de forma o de ambas, la sensibilidad y especificidad del modelo con mejor poder predictivo de cada estudio.

5.1. ANÁLISIS DE TEXTURA

5.1.1. Clasificación multiclase

Los resultados obtenidos en el estudio multiclase realizado mediante un análisis de textura son los mostrados en la Tabla 7, donde se puede observar para cada nivel de gris el mejor modelo predictivo resultado en un color distinto. Cabe destacar que el valor máximo de AUC es 66,6% y se alcanza para el modelo Naive Bayes con 16 niveles de gris y utilizando 2 características únicamente.

Tabla 7. Resultados obtenidos en el estudio multiclase del análisis de textura, indicando los valores de AUC y el número de características usadas para los distintos modelos y niveles de gris, así como el mejor modelo predictivo y sus resultados. Fuente: elaboración propia.

MODELOS		Nivel de gris 16	Nivel de gris 32	Nivel de gris 64	Nivel de gris 128
Naive Bayes	AUC	0,666 ± 0,095	0,6331 ± 0,106	0,5623 ± 0,115	0,597 ± 0,098
	Características	2	10	43	30
KNN con k impar 5	AUC	0,6357 ± 0,092	0,576 ± 0,115	0,5616 ± 0,119	0,5166 ± 0,102
	Características	1	1	42	17
SVM Lineal	AUC	0,6516 ± 0,092	0,5774 ± 0,118	0,5322 ± 0,118	0,5541 ± 0,121
	Características	2	5	10	17
SVM Gaussiano	AUC	0,661 ± 0,099	0,5975 ± 0,113	0,4793 ± 0,115	0,5196 ± 0,106
	Características	1	3	1	35
Bagged Tree	AUC	0,499 ± 0,038	0,499 ± 0,038	0,499 ± 0,038	0,499 ± 0,038
	Características	1	1	1	1

MODELO DE CLASIFICACIÓN	AUC	Características	Sensibilidad	Especificidad
Naive Bayes (NG 16)	0,666 ± 0,095	2	0,3893	0,7898

En la Figura 27 se pueden observar los distintos valores de AUC medio para el modelo elegido de cada nivel de gris. De manera que, en el caso de los 16 niveles de gris, el AUC máximo se obtiene con un reducido número de características.

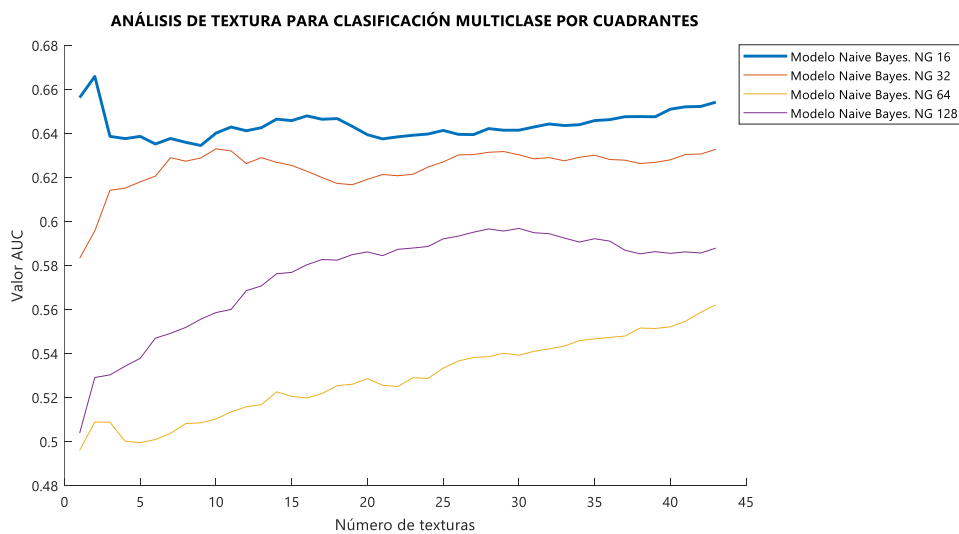


Figura 27. Valores de AUC del mejor modelo de clasificación para cada nivel de gris en el estudio multiclase de textura. Fuente: elaboración propia.

DESARROLLO DE UNA METODOLOGÍA PARA LA AYUDA AL DIAGNÓSTICO DEL CÁNCER DE MAMA MEDIANTE EL ANÁLISIS DE TEXTURAS EN IMÁGENES DE RESONANCIA MAGNÉTICA

En este caso, se utilizan únicamente 2 de las 43 texturas disponibles, es por ello que, seguidamente, en la Figura 28 se muestran las 10 características de textura que, según el ranking generado, son las más utilizadas.

Nombre texturas por relevancia	Texturas más utilizadas en el Modelo Naive Bayes NG16
{ 'SZLGE_GLSZM' }	33
{ 'LGZE_GLSZM' }	31
{ 'SRHGE_GLRLM' }	21
{ 'HGZE_GLSZM' }	32
{ 'Correlation_GLCM' }	8
{ 'Complexity_NGTDm' }	42
{ 'Contrast_NGTDm' }	40
{ 'Contrast_GLCM' }	5
{ 'IGRE_GLRLM' }	18
{ 'HGRE_GLRLM' }	19

Figura 28. Top 10 de texturas más utilizadas en el estudio multiclasa. Fuente: elaboración propia.

5.1.2. Clasificación binario

Los resultados obtenidos en el estudio binario mediante un análisis de textura se muestran en la Tabla 8, donde se puede observar para cada nivel de gris el mejor modelo predictivo resultado en un color distinto. Además, cabe destacar que el valor máximo de AUC es 74,04% y se alcanza para el modelo SVM con Kernel Lineal con 32 niveles de gris y utilizando 43 características.

Tabla 8. Resultados obtenidos en el estudio binario del análisis de textura, indicando los valores de AUC y el número de características usadas para los distintos modelos y niveles de gris, así como el mejor modelo predictivo y sus resultados. Fuente: elaboración propia.

MODELOS		Nivel de gris 16	Nivel de gris 32	Nivel de gris 64	Nivel de gris 128
Naive Bayes	AUC	0,506 ± 0,197	0,464 ± 0,186	0,3944 ± 0,151	0,596 ± 0,166
	Características	17	11	1	3
KNN con k impar 5	AUC	0,45 ± 0,185	0,417 ± 0,172	0,483 ± 0,191	0,6172 ± 0,154
	Características	1	1	19	3
SVM Lineal	AUC	0,6352 ± 0,206	0,7404 ± 0,153	0,6116 ± 0,191	0,5172 ± 0,199
	Características	33	43	41	42
SVM Gaussiano	AUC	0,5874 ± 0,176	0,6246 ± 0,149	0,6952 ± 0,149	0,66 ± 0,166
	Características	43	26	43	3
Bagged Tree	AUC	0,462 ± 0,202	0,4348 ± 0,209	0,4132 ± 0,193	0,6466 ± 0,181
	Características	12	1	1	3

MODELO DE CLASIFICACIÓN MEJOR	AUC	Características	Sensibilidad	Especificidad
SVM con Kernel Lineal (NG 32)	0,7404 ± 0,095	43	0,802	0,498

En la Figura 29 se pueden observar los distintos valores de AUC medio para el modelo elegido de cada nivel de gris. De manera que, en el caso de los 32 niveles de gris, el AUC máximo se obtiene con 43 características, es decir, la totalidad de las texturas disponibles.

DESARROLLO DE UNA METODOLOGÍA PARA LA AYUDA AL DIAGNÓSTICO DEL CÁNCER DE MAMA MEDIANTE EL ANÁLISIS DE TEXTURAS EN IMÁGENES DE RESONANCIA MAGNÉTICA

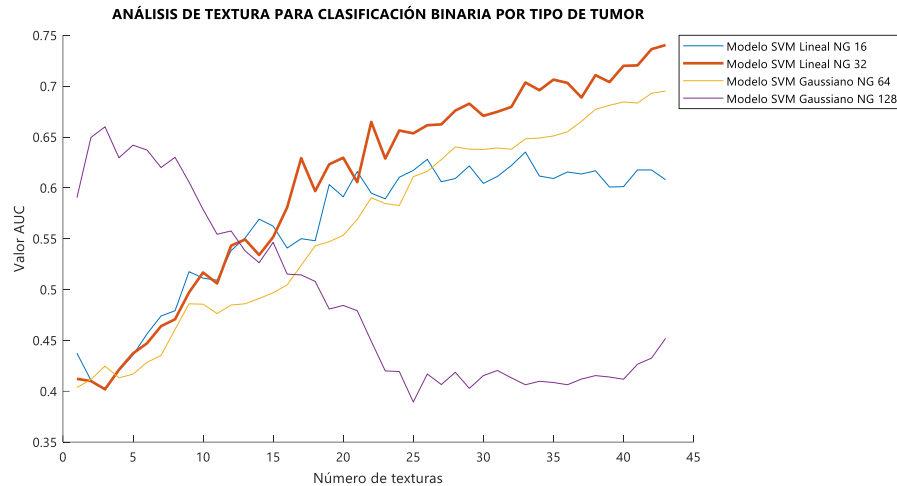


Figura 29. Valores de AUC del mejor modelo de clasificación para cada nivel de gris en el estudio binario de textura. Fuente: elaboración propia.

En este caso, como para el mejor modelo se utiliza la totalidad de las 43 texturas, el método de selección de características de texturas no es relevante ya que todas ellas presentan la misma importancia.

5.2. ANÁLISIS DE FORMA

5.2.1. Clasificación multiclase

Los resultados obtenidos en el estudio multiclase realizado mediante un análisis de forma son los mostrados en la Tabla 9, donde se puede observar el mejor modelo predictivo resultado en un color distinto. Además, cabe destacar que el valor máximo de AUC es 60,74% y se alcanza para el modelo SVM con Kernel Lineal utilizando 1 única característica de forma.

Tabla 9. Resultados obtenidos en el estudio multiclase del análisis de forma, indicando los valores de AUC y el número de características usadas para los distintos modelos y niveles de gris, así como el mejor modelo predictivo y sus resultados. Fuente: elaboración propia.

MODELOS		
Naïve Bayes	AUC	0,4943 ± 0,176
	Características	4
KNN con k impar 5	AUC	0,5625 ± 0,123
	Características	1
SVM Lineal	AUC	0,6074 ± 0,111
	Características	1
SVM Gaussiano	AUC	0,5369 ± 0,105
	Características	1
Bagged Tree	AUC	0,499 ± 0,038
	Características	1

MODELO DE CLASIFICACIÓN	AUC	Características	Sensibilidad	Especificidad
SVM con Kernel Lineal	0,6074 ± 0,111	1	0,3316	0,7785

En la Figura 30 se pueden observar los distintos valores de AUC medio para el modelo elegido. De manera que el AUC máximo se obtiene con un reducido número de características.

DESARROLLO DE UNA METODOLOGÍA PARA LA AYUDA AL DIAGNÓSTICO DEL CÁNCER DE MAMA MEDIANTE EL ANÁLISIS DE TEXTURAS EN IMÁGENES DE RESONANCIA MAGNÉTICA

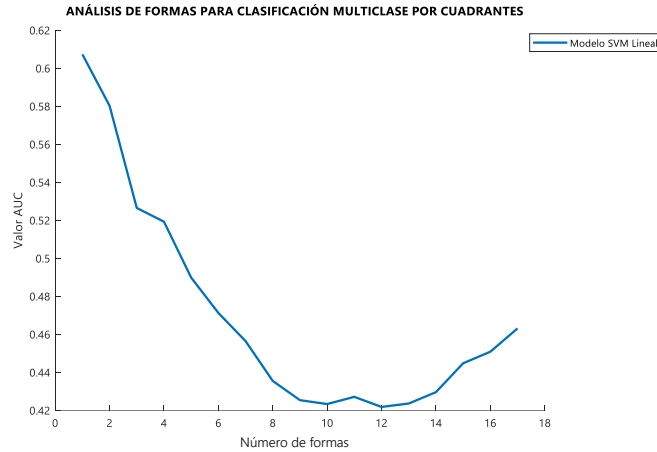


Figura 30. Valores de AUC del mejor modelo en el estudio multiclase de forma. Fuente: elaboración propia.

En este caso, se utilizan únicamente 1 de las 17 formas disponibles, por lo que en la Figura 31 se muestran las 7 características de forma que, según el ranking generado, son las más utilizadas.

Nombre formas por relevancia	Formas más utilizadas en el Modelo SVM Lineal
{'MinorAxisLength' }	12
{'SurfaceArea' }	15
{'SphericalDisproportion' }	13
{'SurfaceVolumeRatio' }	16
{'VoxelVolume' }	17
{'LeastAxisLength' }	5
{'MajorAxisLength' }	6

Figura 31. Top 7 de formas más utilizadas en el estudio multiclase. Fuente: elaboración propia.

5.2.2. Clasificación binario

Los resultados obtenidos en el estudio binario realizado mediante un análisis de forma son los mostrados en la Tabla 10, donde se puede observar el mejor modelo predictivo resultado en un color distinto. Además, cabe destacar que el valor máximo de AUC es 61,52% y se alcanza para el modelo SVM con Kernel Lineal utilizando 17 características.

Tabla 10. Resultados obtenidos en el estudio binario del análisis de forma, indicando los valores de AUC y el número de características usadas para los distintos modelos, así como el mejor modelo predictivo y sus resultados. Fuente: elaboración propia.

MODELOS		
Naive Bayes	AUC	0,4608 ± 0,185
	Características	1
KNN con k impar 5	AUC	0,4776 ± 0,167
	Características	5
SVM Lineal	AUC	0,6152 ± 0,196
	Características	17
SVM Gaussiano	AUC	0,593 ± 0,158
	Características	16
Bagged Tree	AUC	0,4452 ± 0,1686
	Características	1

MODELO DE CLASIFICACIÓN	AUC	Características	Sensibilidad	Especificidad
SVM con Kernel Lineal	0,6152 ± 0,196	17	0,76	0,362

DESARROLLO DE UNA METODOLOGÍA PARA LA AYUDA AL DIAGNÓSTICO DEL CÁNCER DE MAMA MEDIANTE EL ANÁLISIS DE TEXTURAS EN IMÁGENES DE RESONANCIA MAGNÉTICA

En la Figura 32 se pueden observar los distintos valores de AUC medio para el modelo elegido, obtenido el AUC máximo con 17 características, es decir, la totalidad de las formas disponibles.

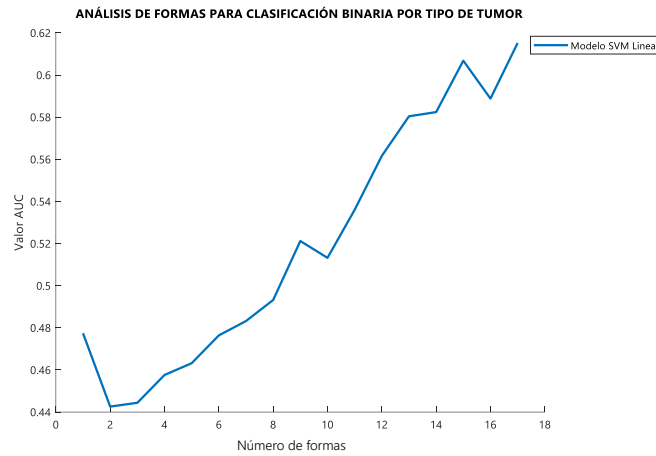


Figura 32. Valores de AUC del mejor modelo de clasificación en el estudio binario. Fuente: elaboración propia.

En este caso, de igual manera que en el análisis binario de textura, como para el mejor modelo se utiliza la totalidad de las 17 formas, el método de selección de características de formas no es relevante ya que todas ellas presentan la misma importancia.

5.3. ANÁLISIS DE CARACTERÍSTICAS DE TEXTURA Y FORMA

5.3.1. Clasificación multiclase

Los resultados obtenidos en el estudio multiclase realizado mediante un análisis de textura y de forma conjuntamente son los mostrados en la Tabla 11, donde se puede observar para cada nivel de gris el mejor modelo predictivo resultado en un color distinto. Además, cabe destacar que el valor máximo de AUC es 67,99% y se alcanza para el modelo SVM con Kernel Lineal con 16 niveles de gris y utilizando 3 características únicamente.

Tabla 11. Resultados obtenidos en el estudio multiclase del análisis de textura y de forma, indicando los valores de AUC y el número de características usadas para los distintos modelos y niveles de gris, así como el mejor modelo predictivo y sus resultados. Fuente: elaboración propia.

MODELOS		Nivel de gris 16	Nivel de gris 32	Nivel de gris 64	Nivel de gris 128
Naive Bayes	AUC	0,6435 ± 0,119	0,5897 ± 0,099	0,5054 ± 0,097	0,5748 ± 0,106
	Características	3	33	30	34
KNN con k impar 5	AUC	0,6147 ± 0,113	0,5534 ± 0,118	0,5273 ± 0,123	0,5357 ± 0,119
	Características	3	2	43	1
SVM Lineal	AUC	0,6799 ± 0,103	0,6034 ± 0,126	0,5708 ± 0,12	0,5714 ± 0,114
	Características	3	2	1	1
SVM Gaussiano	AUC	0,5808 ± 0,109	0,565 ± 0,121	0,5125 ± 0,105	0,5439 ± 0,109
	Características	2	2	1	1
Bagged Tree	AUC	0,499 ± 0,038	0,499 ± 0,038	0,499 ± 0,038	0,499 ± 0,038
	Características	1	1	1	1

MODELO DE CLASIFICACIÓN	AUC	Características	Sensibilidad	Especificidad
SVM con Kernel Lineal (NG 32)	0,6799 ± 0,103	3	0,3787	0,7928

DESARROLLO DE UNA METODOLOGÍA PARA LA AYUDA AL DIAGNÓSTICO DEL CÁNCER DE MAMA MEDIANTE EL ANÁLISIS DE TEXTURAS EN IMÁGENES DE RESONANCIA MAGNÉTICA

En la Figura 33 se pueden observar los distintos valores de AUC medio para el modelo elegido de cada nivel de gris. De manera que, en el caso de los 16 niveles de gris, el AUC máximo se obtiene con un reducido número de características.

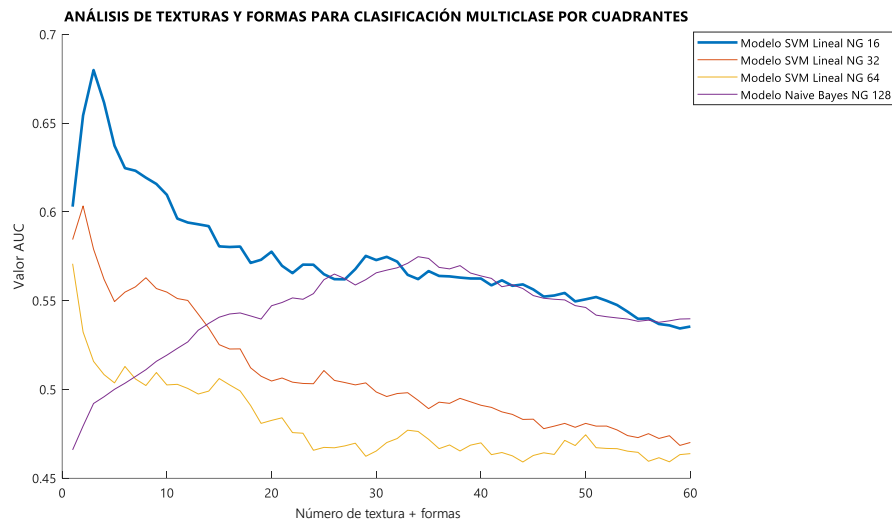


Figura 33. Valores de AUC del mejor modelo de clasificación para cada nivel de gris en el estudio multiclase de textura y de forma. Fuente: elaboración propia.

En este caso, se utilizan únicamente 3 de las 60 texturas y formas disponibles, es por ello que, seguidamente, en la Figura 34 se muestran las 15 características de textura y de forma que, según el ranking generado, son las más utilizadas.

Nombre características por relevancia	Características más utilizadas en el Modelo Naive Bayes NG16
{'SZLGE_GLSZM' }	33
{'LGZE_GLSZM' }	31
{'MinorAxisLength' }	55
{'SRHGE_GLRM' }	21
{'HGZE_GLSZM' }	32
{'SurfaceArea' }	58
{'SphericalDisproportion' }	56
{'Correlation_GLCM' }	8
{'Complexity_NGTDm' }	42
{'Contrast_NGTDm' }	40
{'Contrast_GLCM' }	5
{'LGRE_GLRM' }	18
{'SurfaceVolumeRatio' }	59
{'HGRE_GLRM' }	19
{'Dissimilarity_GLCM' }	11

Figura 34. Top 15 de texturas y formas más utilizadas en el estudio multiclase. Fuente: elaboración propia.

5.3.2. Clasificación binario

Los resultados obtenidos en el estudio binario realizado mediante un análisis de textura y de forma conjuntamente son los mostrados en la Tabla 12, donde se puede observar para cada nivel de gris el mejor modelo predictivo resultado en un color distinto. Además, cabe destacar que el valor máximo de AUC es 71,92% y se alcanza para el modelo SVM con Kernel Lineal con 32 niveles de gris y utilizando 59 características.

DESARROLLO DE UNA METODOLOGÍA PARA LA AYUDA AL DIAGNÓSTICO DEL CÁNCER DE MAMA MEDIANTE EL ANÁLISIS DE TEXTURAS EN IMÁGENES DE RESONANCIA MAGNÉTICA

Tabla 12. Resultados obtenidos en el estudio binario del análisis de textura y de forma indicando los valores de AUC y el número de características usadas para los distintos modelos y niveles de gris, así como el mejor modelo predictivo y sus resultados. Fuente: elaboración propia.

MODELOS		Nivel de gris 16	Nivel de gris 32	Nivel de gris 64	Nivel de gris 128
Naive Bayes	AUC	0,464 ± 0,174	0,4214 ± 0,176	0,3726 ± 0,173	0,5516 ± 0,185
	Características	18	36	26	2
KNN con k impar 5	AUC	0,4276 ± 0,188	0,4388 ± 0,165	0,5562 ± 0,167	0,614 ± 0,176
	Características	20	60	59	42
SVM Lineal	AUC	0,6232 ± 0,2	0,7192 ± 0,163	0,7044 ± 0,175	0,6348 ± 0,188
	Características	60	59	59	60
SVM Gaussiano	AUC	0,5476 ± 0,183	0,5672 ± 0,185	0,5422 ± 0,175	0,5908 ± 0,169
	Características	21	22	19	3
Bagged Tree	AUC	0,4018 ± 0,179	0,36 ± 0,191	0,4228 ± 0,2	0,5828 ± 0,195
	Características	34	1	1	9

MODELO DE CLASIFICACIÓN	AUC	Características	Sensibilidad	Especificidad
SVM con Kernel Lineal (NG 32)	0,7192 ± 0,163	59	0,75	0,538

En la Figura 35 se pueden observar los distintos valores de AUC medio para el modelo elegido de cada nivel de gris. De manera que, en el caso de los 32 niveles de gris, el AUC máximo se obtiene con 59 características, es decir, prácticamente la totalidad de tanto las texturas como las formas disponibles.

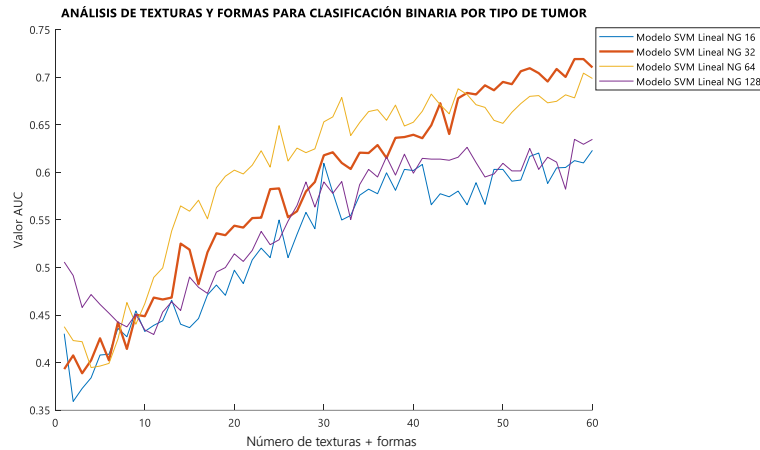


Figura 35. Valores de AUC del mejor modelo de clasificación para cada nivel de gris en el estudio binario de textura y de forma. Fuente: elaboración propia.

5.4. COMPARATIVA DE LOS TRES ANÁLISIS

Finalmente, para acabar de exponer los resultados obtenidos, en la Tabla 13, se muestra un resumen comparativo de los tres análisis realizados y cada uno con sus estudios multiclase y binario correspondientes. De esta manera, como también se puede observar en la Figura 36, el mejor modelo de clasificación para nuestro estudio binario es el modelo SVM con Kernel Lineal con 16 niveles de gris y mediante el análisis de textura con un AUC medio de 74,04%, mientras

DESARROLLO DE UNA METODOLOGÍA PARA LA AYUDA AL DIAGNÓSTICO DEL CÁNCER DE MAMA MEDIANTE EL ANÁLISIS DE TEXTURAS EN IMÁGENES DE RESONANCIA MAGNÉTICA

que para nuestro estudio multiclase es también el modelo SVM con Kernel Lineal pero con 32 niveles de gris y mediante el análisis de textura y de forma.

Tabla 13. Comparativa de los tres análisis realizados para ambos estudios, multiclase y binario, con el mejor modelo de cada uno. Fuente: elaboración propia.

		MODELO DE CLASIFICACIÓN CON NIVEL DE GRIS	AUC	CARACTERÍSTICAS	SENSIBILIDAD ESPECIFICIDAD
ANÁLISIS DE TEXTURA	MULTICLASE	Naive Bayes NG 16	0,666 ± 0,095	2	0,3893 0,7898
	BINARIO	SVM con Kernel Lineal NG 32	0,7404 ± 0,095	43	0,802 0,498
ANÁLISIS DE FORMA	MULTICLASE	SVM con Kernel Lineal	0,6074 ± 0,111	1	0,3316 0,7785
	BINARIO	SVM con Kernel Lineal	0,6152 ± 0,196	17	0,76 0,362
ANÁLISIS DE TEXTURA Y FORMA	MULTICLASE	SVM con Kernel Lineal NG 16	0,6799 ± 0,103	3	0,3787 0,7928
	BINARIO	SVM con Kernel Lineal NG 32	0,7192 ± 0,163	59	0,75 0,538

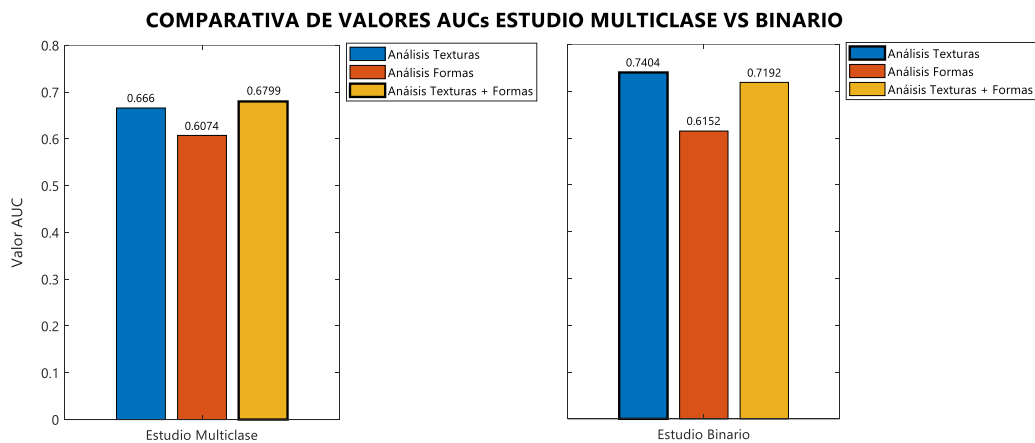


Figura 36. Comparativa de los valores AUCs, tanto en el estudio multiclase como en el binario, de cada análisis indicando el AUC máximo de cada grupo.

Además, cabe destacar que para el estudio multiclase el número de características a utilizar (Figura 37) en los tres análisis es muy reducido, mientras que para el estudio binario es muy elevado, utilizando prácticamente en los tres análisis la totalidad de características correspondientemente.

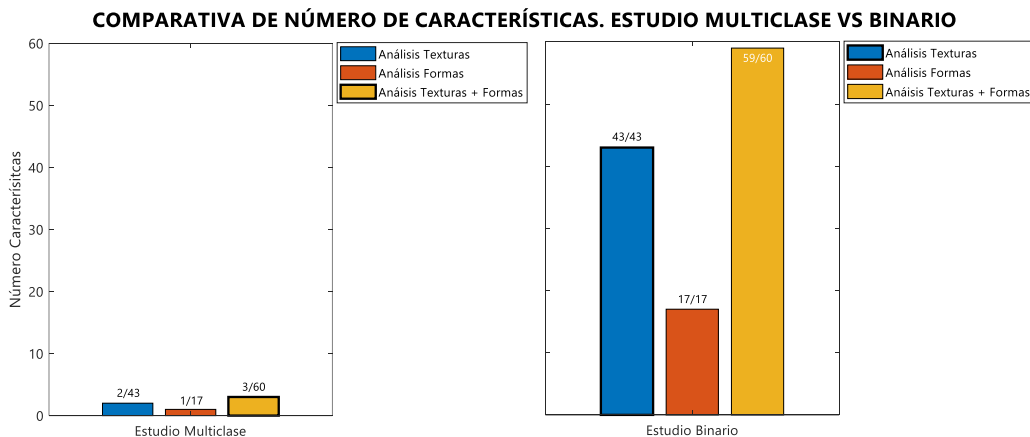


Figura 37. Comparativa del número de características, tanto en el estudio multiclase como en el binario, de cada análisis indicando el estudio con el AUC máximo correspondiente de cada grupo.

DESARROLLO DE UNA METODOLOGÍA PARA LA AYUDA AL DIAGNÓSTICO DEL CÁNCER DE MAMA MEDIANTE EL ANÁLISIS DE TEXTURAS EN IMÁGENES DE RESONANCIA MAGNÉTICA

Finalmente, en la Figura 38 se muestran las matrices de confusión para los mejores modelos de clasificación del estudio binario y del multiclase, además cabe destacar que se muestra una aproximación de las matrices medias de las 100 iteraciones realizadas en el código en ambos casos.

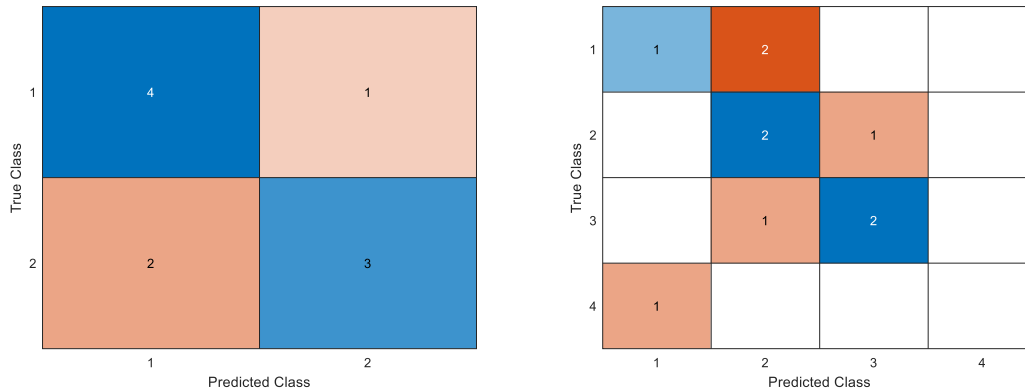


Figura 38. a) Matriz de confusión del estudio binario para el Modelo SVM Lineal en el análisis de textura b) Matriz de confusión de la multiclase para el Modelo SVM Lineal en el análisis de textura y forma. Fuente: elaboración propia.

En la matriz de confusión binaria se observa que 4 pacientes con tumor HER2 positivo y 3 pacientes con tumor Luminal B con HER2 negativo se clasifican correctamente, mientras que 1 paciente con HER2 positivo se clasifican erróneamente como Luminal B con HER2 negativo y de forma inversa ocurre en 2 pacientes.

Por otro lado, en la matriz de confusión multiclase se observa que 1 paciente con localización CSE, 2 pacientes con CSI y 2 con CIE se clasifican correctamente, mientras que 2 pacientes con CSE se clasifican erróneamente como CSI, 1 con CSE como CIE, 1 con CIE como CSI y 1 con CII como CSE.

6. DISCUSIÓN

La discusión de los resultados se basará en los tres análisis expuestos anteriormente, cada uno con sus respectivos estudios. En primer lugar, se comentarán los resultados obtenidos del análisis de textura, posteriormente, los resultados del análisis de formas y, finalmente, se evaluará el análisis de las características de textura y de formas conjuntamente. Para cada uno de estos tres análisis se comentarán también los dos estudios realizados, el estudio multiclase para la clasificación de la localización según el cuadrante de la mama y el estudio binario para la clasificación del tipo de tumoración mamaria.

Tras haber obtenido los **resultados de ambos análisis de textura**, tanto para el estudio multiclase como el binario, podría decirse que dichos resultados de predicción son relativamente significativos. No obstante, cabe destacar que, para el estudio multiclase, los valores de AUC no superan el 66'6%, mientras que en el análisis binario llegan al 74'04%.

Estas diferencias de valores máximos de AUC medios podrían deberse al hecho de que el estudio multiclase, al presentar 4 clases, el número de datos con los que partimos en cada una de ellas se ve reducido respecto al estudio binario ya que, al tener únicamente 2 clases, los datos para cada clase serán mayores y, por tanto, habrá más datos con los que entrenar al modelo y más para testarlo, de manera que la predicción será mayor y, en consecuencia, habrá un mayor AUC medio.

Además, cabría destacar que, respecto al estudio multiclase, de entre todos los modelos de clasificación analizados, el que mejor predicción presenta y mejores resultados se obtienen es el método Naive Bayes para todos los niveles de gris, mientras que, respecto al estudio binario, es el método Máquinas de Vectores de Soporte (SVM) con Kernel Lineal mediante el cual se obtienen el valor máximo de AUC para el nivel de gris 32.

Respecto a los niveles de gris analizados, en este caso, 4 niveles de gris, que son 16, 32, 64 y 128, no existe un nivel de gris determinado para el cual se obtengan mejores resultados de clasificación. En cuanto al estudio multiclase, el nivel de gris que presenta el AUC máximo es el de 16 niveles y, en cuanto al estudio binario, es el de 32 niveles, como bien se ha comentado.

Por tanto, recapitulando, para el análisis multiclase de la localización obtenemos la mejor predicción para el nivel de gris 16 y para el modelo de clasificación Naive Bayes, con un AUC máximo de 0'666, una sensibilidad de 0'3893 y una especificidad de 0'7898. Para el análisis binario del tipo de tumor obtenemos los mejores resultados de predicción para el nivel de gris 32 y para el modelo SVM con Kernel Lineal, con un AUC máximo de 0'7404, una sensibilidad de 0'802 y una especificidad de 0'498.

En el análisis multiclase, la mejor predicción se obtuvo utilizando dicho modelo de clasificación con un número reducido de texturas, en concreto 2. Mientras que, en el análisis binario, los mejores resultados se obtuvieron haciendo uso de todas las texturas disponibles, 43.

Por otro lado, y tras haber obtenido también los **resultados de ambos análisis de formas**, podría decirse que dichos resultados presentan un valor de predicción menor respecto al análisis de las texturas debido a que los valores de AUC no superan el 60'73% para el estudio multiclase, y el 61'52% para el análisis binario.

En ambos casos, de entre todos los modelos de clasificación analizados, el que mejores resultados desarrolla es el método Máquinas de Vectores de Soporte (SVM) con Kernel Lineal, obteniendo el valor máximo de AUC en ambos estudios, multiclase y binario.

Es decir, para el análisis multiclase de la localización obtenemos la mejor predicción para el modelo de clasificación SVM con Kernel Lineal, con un AUC máximo de 0'6073, una sensibilidad de 0'3316 y una especificidad de 0'7785. Para el análisis binario del tipo de tumor obtenemos los mejores resultados utilizando el mismo modelo de clasificación, con un AUC máximo de 0'6152, una sensibilidad de 0'76 y una especificidad de 0'362.

Además, cabría destacar que en el caso del estudio multiclase la mejor predicción se obtiene utilizando un número reducido de formas, 1. Mientras que, en el caso del estudio binario, los mejores resultados se obtienen haciendo uso de todas las formas de las que disponíamos, en concreto 17, de igual manera que en el análisis binario de texturas.

Finalmente, se decidió realizar un **análisis utilizando tanto las características de texturas como de formas**, para comprobar si realmente el uso conjunto de ambos tipos de características se podría traducir en una mejora sustancial del valor de AUC medio.

Para ello, se seleccionó tanto las características de las texturas para cada nivel de gris como las formas, de manera que se obtuvieron resultados dispares en cuanto al análisis multiclase y binario. Primeramente, cabe destacar que el gris de nivel con mayor AUC se corresponde al mismo nivel de gris que habíamos determinado en el análisis de textura, en el caso del análisis multiclase el nivel de gris 16 y en el binario el nivel de gris 32.

Por tanto, centrando la vista únicamente a aquel nivel de gris que mayor AUC proporciona, se puede decir que, para el análisis multiclase, se obtuvo un AUC máximo de 67'99%, que, comparado con el AUC máximo del análisis de texturas (66'6%) y del análisis de formas (60'7%), se ve que hay una mejora significativa respecto a las formas, pero respecto a las texturas la mejora es muy sutil. En cambio, para el análisis binario, se obtuvo un AUC máximo de 71'92%, que, comparado con el AUC máximo del análisis de texturas (74'04%) y del análisis de formas (61'52%), se ve que hay una mejora significativa respecto a las formas, pero respecto a las texturas hay una disminución del valor predictivo.

En ambos casos, de entre todos los modelos de clasificación analizados, el que mejores resultados desarrolla es el método Máquinas de Vectores de Soporte (SVM) con Kernel Lineal.

Por tanto, para el análisis multiclase obtenemos la mejor predicción para el modelo de clasificación SVM con Kernel Lineal, con un AUC máximo de 0'6799, una sensibilidad de 0'3787 y una especificidad de 0'7928. Para el análisis binario, obtenemos los mejores resultados utilizando el mismo modelo de clasificación, con un AUC máximo de 0'7192, una sensibilidad de 0'75 y una especificidad de 0'538.

Además, en el análisis multiclase, la mejor predicción se obtuvo con un número reducido de características, 3, en concreto. Mientras que, para el análisis binario, los mejores resultados se obtienen haciendo uso de prácticamente todas las características, en concreto 59.

Concluyendo, para un estudio multiclase, el uso de características de textura y de forma implica un reducido aumento de la respuesta del valor predictivo, mientras que, para un estudio binario,

únicamente con el uso de parámetros de textura ya es suficiente para obtener un adecuado valor de AUC medio.

Además, cabría destacar que, para los estudios multiclase, el mejor modelo de predicción se obtiene con el uso de un número muy reducido de características, en cambio, en los estudios binarios, son necesarias prácticamente todas las características iniciales, lo que implica dichos modelos son más complejos y, por tanto, son menos generalizables que aquellos que usan menos texturas, como es en el caso de los presentes estudios multiclase. Finalmente, comentar que, en general, el mejor método de clasificación para nuestros datos es el modelo SVM con Kernel Lineal.

7. CONCLUSIONES

Como bien sabemos, el cáncer de mama es la patología maligna más frecuente diagnosticada en mujeres y la principal causa de muerte por cáncer en las mismas. Además, considerando los factores de riesgo que se han comentado a lo largo del estudio, cabe destacar que un buen diagnóstico puede contribuir de manera indirecta a una buena evolución de la lesión y es por ello que las líneas de investigación en el campo de la mama están destinadas a desarrollar metodologías para la ayuda al diagnóstico de esta enfermedad.

En el marco de este Trabajo Fin de Grado se han implementado tres tipos de análisis, en concreto, de texturas, de formas y de ambas características conjuntamente, en IRM sobre un total de 43 sujetos con la finalidad de poder categorizar la malignidad del tumor. Para ello, se han considerado distintas covariables. En primer lugar, la localización por cuadrantes del tumor en el plano coronal, 1º estudio multiclase de 4 clases (CIE, CII, CSE y CSI), seguidamente, la localización por porciones del tumor en el plano sagital, 2º estudio multiclase de 3 clases (porción anterior, porción media y porción posterior), y en tercer lugar, el tipo de tumor según su información genética, estudio binario (HER2 positivo y Luminal B con HER2 negativo).

No obstante, debido al reducido tamaño de muestras de determinadas clases en la covariable de la localización por porciones, se ha visto dificultado su desarrollo y no se ha podido llegar a implementar. Es por ello que, finalmente, se han realizado 6 estudios en total y, consecuentemente, se han obtenido resultados del análisis de textura, de formas y de las características de textura y de formas conjuntamente tanto para el 1º estudio multiclase como para el binario.

Una vez preprocesadas las imágenes, segmentadas la ROIs de las mismas y obtenidos los datos de las características de texturas y de formas, se ha procedido a la realización del proceso de aprendizaje máquina para cada uno de los análisis mencionados previamente, haciendo uso de la herramienta MATLAB. De manera que se han desarrollado los distintos modelos de clasificación (Naive Bayes, KNN, SVM con Kernel Lineal y Gaussiano y Bagged Tree), los cuales han sido entrenados, validados y evaluados, obteniendo así para cada estudio aquel que presenta mayor valor predictivo para la ayuda en el diagnóstico de dicha enfermedad.

Con todo ello y observando los resultados obtenidos (Figura 36), se puede decir que, para el estudio multiclase de la localización por cuadrante, el uso de características de textura y de forma conjuntamente permite obtener un AUC medio máximo del 67'99%, que, comparado con el del análisis de texturas (66'6%) y del análisis de formas (60'7%), implica un reducido aumento de la respuesta del valor predictivo. Mientras que, para el estudio binario del tipo de tumor, únicamente con el uso de parámetros de textura ya es suficiente para obtener un adecuado valor de AUC medio ya que para el análisis de texturas se obtiene un AUC medio máximo de 74'04%, comparado con el de texturas y formas (71'92%) y el de formas (61'52%).

Además, cabe destacar que en los estudios multiclase, el mejor modelo de predicción se obtiene con el uso de un número muy reducido de características, en cambio, en los estudios binarios, son necesarias prácticamente todas las características iniciales, lo que implica que dichos modelos son más complejos y, por tanto, podrían ser menos generalizables que aquellos que usan menos texturas, como es en el caso de los presentes estudios multiclase. Finalmente, cabe comentar que, en general, el mejor método de clasificación para nuestros datos es el modelo SVM con Kernel Lineal.

8. LÍNEAS FUTURAS

Habiendo concluido el presente estudio, sería de gran interés proponer posibles líneas de mejora y ampliación del proyecto, algunas de ellas se muestran seguidamente.

Una de las limitaciones en el transcurso del trabajo ha sido el reducido tamaño de los datos de los cuales disponíamos. Por ello, una mejora sería aumentar dicho tamaño de muestras para poder obtener un mejor valor predictivo en los modelos y así aumentar la robustez del sistema.

Además, en este trabajo se ha realizado la segmentación de las lesiones en la mama de manera semiautomática y también manual. De manera que se podría proponer como mejora una segmentación que consiga ser mucho más precisa y así obtener una ROI que se ajuste perfectamente a los límites de la lesión y que no tenga la posibilidad de incluir tejido circundante que falsee los resultados.

Adentrándonos en el tema del análisis de texturas, en este estudio únicamente se han analizado las texturas en 2D y, es por ello, que se podría plantear realizarlo con toda la lesión, es decir, un estudio tridimensional, en 3D, de igual manera que se ha realizado en el análisis de las formas.

Finalmente, respecto a la metodología a seguir en el análisis de texturas y de formas, podría ser el desarrollo de un algoritmo más complejo, en el cual se incluyan más métodos de selección de características como los *wrappers*, más métodos de clasificación incorporando distintos modelos de clasificación, más métodos de validación como el *K-Fold* ya que todo ello podría implicar una mayor robustez en el código y una mejora sustancial en los resultados.

9. BIBLIOGRAFÍA

- [1] “Pronóstico Cáncer de Mama: Supervivencia y Mortalidad | AECC.” [En línea]. Disponible en: <https://www.aecc.es/es/todo-sobre-cancer/tipos-cancer/cancer-mama/mas-informacion/evolucion-cancer-mama>. [Consultado: 23-Jun-2021].
- [2] P. Chalasani, “Breast Cancer - Medscape,” 2021. [En línea]. Disponible en: <https://emedicine.medscape.com/article/1947145-overview#a1>. [Consultado: 28-May-2021].
- [3] F. Sardanelli *et al.*, “Sensitivity of MRI versus mammography for detecting foci of multifocal, multicentric breast cancer in fatty and dense breasts using the whole-breast pathologic examination as a gold standard,” *Am. J. Roentgenol.*, vol. 183, no. 4, pp. 1149–1157, 2004, doi: 10.2214/ajr.183.4.1831149.
- [4] “Anatomía de la mama | Glándula, complejo areola-pezones, Cooper, lóbulo.” [En línea]. Disponible en: <https://www.cirugiasdelamama.com/anatomia-de-la-mama>. [Consultado: 23-Jun-2021].
- [5] Dra. Marcia Oliva, “Resonancia magnética mamaRia: bi-rads (Breast imaging Report and data system),” *Revista Argentina de Mastología*, pp. 53–74, 2016.
- [6] “Frecuencia de cáncer de mama por zonas.” [En línea]. Disponible en: <http://alvarezrad.mx/web/index.php/blog/12-frecuencia-de-cancer-de-mama-por-zonas>. [Consultado: 23-Jun-2021].
- [7] “Cáncer de Mama: ¿Qué es? | Roche Pacientes.” [En línea]. Disponible en: <https://rochepacientes.es/cancer/mama.html>. [Consultado: 23-Jun-2021].
- [8] M. Abellán Lucas, “Variación del tratamiento quirúrgico del cáncer de mama tras la introducción de la resonancia magnética en la estadificación preoperatoria,” Universitat Autònoma de Barcelona, 2021.
- [9] “¿Cómo se trata el cáncer de mama? | CDC.” [En línea]. Disponible en: https://www.cdc.gov/spanish/cancer/breast/basic_info/treatment.htm. [Consultado: 23-Jun-2021].
- [10] D. Moratal, “Tema 9.1. Introducción a la imagen por RM,” 2020.
- [11] Xavier Pardell, “Resonancia Magnética - Apuntes de Electromedicina,” 2021. [En línea]. Disponible en: <https://www.pardell.es/resonancia-magnetica.html>. [Consultado: 23-Jun-2021].
- [12] D. Moratal, “Tema 9.6. Las secuencias clásicas de pulsos,” 2020.
- [13] I. Pérez Zúñiga, Y. Villaseñor Navarro, M. P. Pérez Badillo, R. Cruz Morales, C. Pavón Hernández, and L. Aguilar Cortázar, “Resonancia magnética de mama y sus aplicaciones,” México D.F., 2012.
- [14] J. Camps Herrero, “Manejo conservador de las lesiones esplénicas experiencia en 136 pacientes con traumatismo esplénico cerrado,” vol. 53, no. 1, pp. 27–38, 2010, doi: 10.1016/j.rx.2010.08.009.
- [15] “Contrastes utilizados en resonancia magnética. TER - Ocronos - Editorial Científico-Técnica.” [En línea]. Disponible en: <https://revistamedica.com/contrastes-resonancia-magnetica-ter/>. [Consultado: 23-Jun-2021].
- [16] G. Castellano, L. Bonilha, L. M. Li, and F. Cendes, “Texture analysis of medical images,” 2004.

- [17] A. B. Pérez, “Desarrollo de algoritmos de procesamiento de imagen avanzado para interpretación de imágenes médicas. Aplicación a segmentación de hígado sobre imágenes de Resonancia Magnética multiseuencia.,” *Tecnalia Res. Innov.*, p. 230, 2016.
- [18] A. Larroza, “Diagnosis and prognosis of cardiovascular diseases by means of texture analysis in magnetic resonance imaging,” no. September, 2017.
- [19] “Radiomic Features — pyradiomics.” [En línea]. Disponible en: <https://pyradiomics.readthedocs.io/en/latest/features.html>. [Consultado: 23-Jun-2021].
- [20] “Matt’s Webcorner - Cubos de marcha.” [En línea]. Disponible en: <https://graphics.stanford.edu/~mdfisher/MarchingCubes.html>. [Consultado: 23-Jun-2021].
- [21] “Polygonising a scalar field (Marching Cubes).” [En línea]. Disponible en: <http://paulbourke.net/geometry/polygonise/>. [Consultado: 23-Jun-2021].
- [22] C. Sáez Silvestre, “Sistemas de Información y Telemedicina I. Unidad 3.5: Reducción de la dimensionalidad,” 2020.
- [23] “Aprendizaje Supervisado y No Supervisado - Fernando Sancho Caparrini.” [En línea]. Disponible en: <http://www.cs.us.es/~fsancho/?e=77>. [Consultado: 23-Jun-2021].
- [24] “Algoritmos Naive Bayes: Fundamentos e Implementación.” [En línea]. Disponible en: <https://medium.com/datos-y-ciencia/algoritmos-naive-bayes-fudamentos-e-implementación-4bcb24b307f>. [Consultado: 23-Jun-2021].
- [25] C. Sáez Silvestre, “Unidad 3.6: Métodos de Aprendizaje Automático,” pp. 1–60, 2017.
- [26] “El algoritmo K-NN y su importancia en el modelado de datos | Blog | España | Merkle.” [En línea]. Disponible en: <https://www.merkleinc.com/es/es/blog/algoritmo-knn-modelado-datos>. [Consultado: 23-Jun-2021].
- [27] “Computer Vision.” [En línea]. Disponible en: http://www.cs.cornell.edu/courses/cs5670/2018sp/lectures/lec20_image_classification.pdf. [Consultado: 23-Jun-2021].
- [28] “Máquinas de Vectores de Soporte (SVM) - IArtificial.net.” [En línea]. Disponible en: <https://www.iartificial.net/maquinas-de-vectores-de-soporte-svm/>. [Consultado: 23-Jun-2021].
- [29] “Máquinas de Vector Soporte (Support Vector Machines, SVMs).” [En línea]. Disponible en: https://www.cienciadedatos.net/documentos/34_maquinas_de_vector_soporte_support_vector_machines#Support_Vector_Machines_para_más_de_dos_clases. [Consultado: 23-Jun-2021].
- [30] “Bootstrap aggregating - Wikipedia.” [En línea]. Disponible en: https://en.wikipedia.org/wiki/Bootstrap_aggregating. [Consultado: 23-Jun-2021].
- [31] C. Gimenez, D. Moratal, and R. Ortiz, “Estudio de parámetros de textura mediante técnicas de aprendizaje máquina para la detección temprana de la enfermedad de alzheimer en imágenes de resonancia magnética,” 2018.
- [32] “Clasificación: ROC y AUC | Curso intensivo de aprendizaje automático.” [En línea]. Disponible en: <https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc?hl=es-419>. [Consultado: 23-Jun-2021].
- [33] “Evaluating Classifiers: Confusion Matrix for Multiple Classes - YouTube.” [En línea]. Disponible en: <https://www.youtube.com/watch?v=FAr2GmWNbT0>. [Consultado: 23-Jun-2021].

PRESUPUESTO

DESARROLLO DE UNA METODOLOGÍA PARA LA AYUDA AL DIAGNÓSTICO DEL CÁNCER DE MAMA MEDIANTE EL ANÁLISIS DE TEXTURAS EN IMÁGENES DE RESONANCIA MAGNÉTICA

AUTORA: PILAR CASTELLOTE HUGUET

TUTOR: DAVID MORATAL PÉREZ

COTUTORES: SILVIA RUIZ ESPAÑA

CRISTIAN GALÁN AUGE

JOSÉ MANUEL SANTABÁRBARA GÓMEZ

Curso Académico: 2020-2021

ÍNDICE PRESUPUESTO

1.	INTRODUCCIÓN	1
2.	CUADRO DE PRECIOS DE MANO DE OBRA	1
3.	CUADRO DE PRECIOS DE MATERIALES.....	2
4.	CUADRO DE PRECIOS UNITARIOS	2
5.	CUADRO DE PRECIOS DESCOMPUESTOS	4
6.	CUADRO DE MEDICIONES.....	7
7.	CUADRO DE PRECIOS PARCIALES	8
8.	PRESUPUESTO DE EJECUCIÓN POR CONTRATA.....	9

1. INTRODUCCIÓN

El Presupuesto es un documento del proyecto cuya finalidad es la estimación aproximada de la inversión económica, en unidades monetarias, que es necesaria para el desarrollo y ejecución del presente Trabajo Fin de Grado, es decir, para la implantación de la metodología desarrollada para la ayuda al diagnóstico del cáncer de mama.

Para ello y haciendo uso del programa Arquímedes, en los siguientes apartados se desarrollarán de forma desglosada las distintas secciones que deben conformar cualquier tipo de Presupuesto.

2. CUADRO DE PRECIOS DE MANO DE OBRA

Esta Sección del Presupuesto considera la mano de obra requerida para la realización del proyecto y, con ello, el coste aproximado que comporta dicha mano de obra. Por tanto, se debería tener en cuenta la mano de obra realizada por el ingeniero biomédico como tutor, los ingenieros biomédicos como cotutores, el estudiante de Ingeniería Biomédica como autor del proyecto y el radiólogo.

Para el cálculo del coste de cada mano de obra necesaria se han estimado los siguientes salarios base anuales. El salario base del ingeniero tutor es de 50.000 euros anuales, el del ingeniero cotutor es de 20.000 euros anuales y el del radiólogo de 65.000 euros anuales. Además, el salario del estudiante considerado ingeniero biomédico junior, autor del Proyecto, al presentar un convenio de prácticas de la UPV con ASCIRES Grupo Biomédico, su salario anual corresponde a 15.000 euros anuales.

Todo ello considerando que la jornada laboral media es de 8 horas diarias y los días efectivos de trabajo anual son un total de 224 días, procedentes de restar los 104 días de fines de semana, 25 días hábiles de vacaciones y 12 días festivos. Por ello, las horas de trabajo anuales serían un total de 1792 horas.

Por tanto, el precio por hora de cada mano de obra se obtendrá realizando la división del salario base anual entre las 1792 horas anuales trabajadas. De manera que se han obtenido los precios por hora mostrados en la siguiente Tabla.

Cuadro de mano de obra					Página 1
Num.	Código	Denominación de la mano de obra	Precio	Horas	Total
1	MO.RAD	Mano de obra de tipo Radiólogo	36,27	2,550 h	92,34
2	MO.TUT	Mano de obra de tipo Ingeniero Tutor	27,90	19,500 h	544,05
3	MO.COT2	Mano de obra de tipo Ingeniero Cotutor 2	16,74	33,700 h	563,86
4	MO.COT1	Mano de obra de tipo Ingeniero Cotutor 1	16,74	36,200 h	605,37
5	MO.AUT	Mano de obra de tipo Ingeniero Junior en Prácticas	8,37	390,000 h	3.264,30
Total mano de obra:					5.069,92

3. CUADRO DE PRECIOS DE MATERIALES

El Cuadro de Precios de Materiales se han considerado todos aquellos materiales que han sido utilizados en el transcurso del presente trabajo. Los softwares que se han empleado han sido *PACS CareStream, Medexprim Suite, ITK-SNAP, QDoc, RadiAnt DICOM Viewer y dcm2niix* para la obtención y preparación de las imágenes, además de MATLAB y Python para el análisis de texturas y de formas, así como del Microsoft Office 2016 para la redacción del trabajo. También cabría destacar los dispositivos utilizados, que han sido el ordenador proporcionado por ASCIRES Grupo Biomédico y un ordenador personal.

Cuadro de maquinaria Página 1

Num.	Código	Denominación de la maquinaria	Precio	Cantidad	Total
1	MAT.Python	Licencia Python	26,00	10,000 h	260,00
2	MAT.MATLAB	Licencia MATLAB R2019b	24,00	141,000 h	3.384,00
3	MAT.DICOM	Software RadiAnt DICOM Viewer	5,00	67,000 h	335,00
4	MAT.MED	Software Medexprim Suite	5,00	30,000 h	150,00
5	MAT.PCp	Ordenador portátil personal	1,55	351,000 h	544,05
6	MAT.PCm	Ordenador local de ASCIRES Grupo Biomédico	1,15	85,000 h	97,75
7	MAT.OFF	Licencia Microsoft 365 Office 2016	0,90	20,000 h	18,00
8	MAT.PACS	Software PACS CareStream	0,10	85,000 h	8,50
9	MAT.NII	Software dcm2niix	0,10	67,000 h	6,70
10	MAT.ITK	Software ITK-SNAP	0,10	67,000 h	6,70
11	MAT.QDoc	Software QDoc	0,10	67,000 h	6,70
12	MAT.W10	Licencia Windows 10	0,10	381,000 h	38,10
				Total maquinaria:	4.855,50

4. CUADRO DE PRECIOS UNITARIOS

Esta sección consiste en la descripción de cada unidad y su precio, expresado tanto en letras como en cifras. Cabe destacar que los precios unitarios se obtienen a partir de los precios descompuestos, no obstante, se mencionan previamente estos primeros.

DESARROLLO DE UNA METODOLOGÍA PARA LA AYUDA AL DIAGNÓSTICO DEL CÁNCER DE MAMA
 MEDIANTE EL ANÁLISIS DE TEXTURAS EN IMÁGENES DE RESONANCIA MAGNÉTICA

Cuadro de precios nº 1

Nº	Designación	Importe	
		En cifra (Euros)	En letra (Euros)
	1 PLANIFICACIÓN DEL PROYECTO Y REVISIÓN BIBLIOGRÁFICA DEL ESTADO DEL ARTE		
1.1	h Reunión inicial con el Tutor, Cotutores y Radiólogo del TFG	110,26	CIENTO DIEZ EUROS CON VEINTISEIS CÉNTIMOS
1.2	h Revisión bibliográfica del estado del arte	11,36	ONCE EUROS CON TREINTA Y SEIS CÉNTIMOS
1.3	h Reuniones periódicas para el seguimiento del proyecto	65,57	SESENTA Y CINCO EUROS CON CINCUENTA Y SIETE CÉNTIMOS
	2 OBTENCIÓN Y PREPARACIÓN DE LAS IMÁGENES		
2.1	h Formación sobre el uso de los distintos softwares y posterior instalación	27,79	VEINTISIETE EUROS CON SETENTA Y NUEVE CÉNTIMOS
2.2	h Descarga de las imágenes y anonimización de los datos	17,05	DIECISIETE EUROS CON CINCO CÉNTIMOS
2.3	h Segmentación de las ROIs de las imágenes	19,34	DIECINUEVE EUROS CON TREINTA Y CUATRO CÉNTIMOS
2.4	h Preparación de los archivos .mat	35,38	TREINTA Y CINCO EUROS CON TREINTA Y OCHO CÉNTIMOS
	3 DESARROLLO DE LA NUEVA METODOLOGÍA, JUNTO CON EL ANÁLISIS DE TEXTURAS Y FORMAS		
3.1	h Extracción de características	72,86	SETENTA Y DOS EUROS CON OCHENTA Y SEIS CÉNTIMOS
3.2	h Programación de la herramienta de clasificación	38,86	TREINTA Y OCHO EUROS CON OCHENTA Y SEIS CÉNTIMOS
3.3	h Obtención y evaluación de los resultados	41,76	CUARENTA Y UN EUROS CON SETENTA Y SEIS CÉNTIMOS
	4 REDACCIÓN Y DEFENSA DEL TFG		
4.1	h Redacción de la memoria del proyecto	10,42	DIEZ EUROS CON CUARENTA Y DOS CÉNTIMOS
4.2	h Revisión y corrección del proyecto	40,01	CUARENTA EUROS CON UN CÉNTIMO
4.3	h Preparación de la defensa del proyecto	10,42	DIEZ EUROS CON CUARENTA Y DOS CÉNTIMOS
	Valencia, Junio 2020 Ingeniera Biomédica		
	María Pilar Castellote Huguet		

5. CUADRO DE PRECIOS DESCOMPUESTOS

En el Cuadro de Precios Descompuestos se muestran los distintos apartados que conforman la realización del Trabajo Fin de Grado y donde se resume el procedimiento a seguir en cada uno de los distintos capítulos.

- Capítulo 1. Planificación del proyecto y revisión bibliográfica del estado del arte
- Capítulo 2. Obtención y preparación de las imágenes
- Capítulo 3. Desarrollo de la nueva metodología, junto con el análisis de texturas y formas
- Capítulo 4. Redacción y defensa del TFG

Seguidamente, se muestra el cuadro de precios descompuestos en el que se explican detalladamente cada uno de los capítulos mencionados, mostrando para cada unidad de obra, las cantidades necesarias de los materiales y las cantidades, en tiempos, de mano de obra.

Anejo de justificación de precios

Nº	Código	Ud	Descripción		Total
1 PLANIFICACIÓN DEL PROYECTO Y REVISIÓN BIBLIO...					
1.1	01.01	h	Reunión inicial con el Tutor, Cotutores y Radiólogo del TFG		
	MO.TUT	1,000 h	Mano de obra de tipo Ingeniero Tutor	27,90	27,90
	MO.COT1	1,000 h	Mano de obra de tipo Ingeniero Cotutor 1	16,74	16,74
	MO.COT2	1,000 h	Mano de obra de tipo Ingeniero Cotutor 2	16,74	16,74
	MO.RAD	1,000 h	Mano de obra de tipo Radiólogo	36,27	36,27
	MO.AUT	1,000 h	Mano de obra de tipo Ingeniero Junior en Prácticas	8,37	8,37
		4,000 %	Costes indirectos	106,02	4,24
			Precio total por h		110,26
1.2	01.02	h	Revisión bibliográfica del estado del arte		
	MO.AUT	1,000 h	Mano de obra de tipo Ingeniero Junior en Prácticas	8,37	8,37
	MAT.W10	1,000 h	Licencia Windows 10	0,10	0,10
	MAT.OFF	1,000 h	Licencia Microsoft 365 Office 2016	0,90	0,90
	MAT.PCp	1,000 h	Ordenador portátil personal	1,55	1,55
		4,000 %	Costes indirectos	10,92	0,44
			Precio total por h		11,36
1.3	01.03	h	Reuniones periódicas para el seguimiento del proyecto		
	MO.TUT	1,000 h	Mano de obra de tipo Ingeniero Tutor	27,90	27,90
	MO.COT1	0,800 h	Mano de obra de tipo Ingeniero Cotutor 1	16,74	13,39
	MO.COT2	0,800 h	Mano de obra de tipo Ingeniero Cotutor 2	16,74	13,39
	MO.AUT	1,000 h	Mano de obra de tipo Ingeniero Junior en Prácticas	8,37	8,37
		4,000 %	Costes indirectos	63,05	2,52
			Precio total por h		65,57

DESARROLLO DE UNA METODOLOGÍA PARA LA AYUDA AL DIAGNÓSTICO DEL CÁNCER DE MAMA
MEDIANTE EL ANÁLISIS DE TEXTURAS EN IMÁGENES DE RESONANCIA MAGNÉTICA

Nº	Código	Ud	Descripción		Total
2 OBTENCIÓN Y PREPARACIÓN DE LAS IMÁGENES					
2.1	02.01	h	Formación sobre el uso de los distintos softwares y posterior instalación		
	MO.COT1	0,400 h	Mano de obra de tipo Ingeniero Cotutor 1	16,74	6,70
	MO.AUT	1,000 h	Mano de obra de tipo Ingeniero Junior en Prácticas	8,37	8,37
	MAT.W10	1,000 h	Licencia Windows 10	0,10	0,10
	MAT.PCm	1,000 h	Ordenador local de ASCIRES Grupo Biomédico	1,15	1,15
	MAT.PACS	1,000 h	Software PACS CareStream	0,10	0,10
	MAT.MED	1,000 h	Software Medexprim Suite	5,00	5,00
	MAT.ITK	1,000 h	Software ITK-SNAP	0,10	0,10
	MAT.QDoc	1,000 h	Software QDoc	0,10	0,10
	MAT.DICOM	1,000 h	Software RadiAnt DICOM Viewer	5,00	5,00
	MAT.NII	1,000 h	Software dcm2niix	0,10	0,10
		4,000 %	Costes indirectos	26,72	1,07
			Precio total por h		27,79
2.2	02.02	h	Descarga de las imágenes y anonimización de los datos		
	MO.COT1	0,100 h	Mano de obra de tipo Ingeniero Cotutor 1	16,74	1,67
	MO.AUT	1,000 h	Mano de obra de tipo Ingeniero Junior en Prácticas	8,37	8,37
	MAT.PCm	1,000 h	Ordenador local de ASCIRES Grupo Biomédico	1,15	1,15
	MAT.W10	1,000 h	Licencia Windows 10	0,10	0,10
	MAT.PACS	1,000 h	Software PACS CareStream	0,10	0,10
	MAT.MED	1,000 h	Software Medexprim Suite	5,00	5,00
		4,000 %	Costes indirectos	16,39	0,66
			Precio total por h		17,05
2.3	02.03	h	Segmentación de las ROIs de las imágenes		
	MO.COT1	0,100 h	Mano de obra de tipo Ingeniero Cotutor 1	16,74	1,67
	MO.RAD	0,010 h	Mano de obra de tipo Radiólogo	36,27	0,36
	MO.AUT	1,000 h	Mano de obra de tipo Ingeniero Junior en Prácticas	8,37	8,37
	MAT.PCm	1,000 h	Ordenador local de ASCIRES Grupo Biomédico	1,15	1,15
	MAT.PCp	1,000 h	Ordenador portátil personal	1,55	1,55
	MAT.W10	1,000 h	Licencia Windows 10	0,10	0,10
	MAT.PACS	1,000 h	Software PACS CareStream	0,10	0,10
	MAT.ITK	1,000 h	Software ITK-SNAP	0,10	0,10
	MAT.QDoc	1,000 h	Software QDoc	0,10	0,10
	MAT.DICOM	1,000 h	Software RadiAnt DICOM Viewer	5,00	5,00
	MAT.NII	1,000 h	Software dcm2niix	0,10	0,10
		4,000 %	Costes indirectos	18,60	0,74
			Precio total por h		19,34
2.4	02.04	h	Preparación de los archivos .mat		
	MO.AUT	1,000 h	Mano de obra de tipo Ingeniero Junior en Prácticas	8,37	8,37
	MAT.PCp	1,000 h	Ordenador portátil personal	1,55	1,55
	MAT.W10	1,000 h	Licencia Windows 10	0,10	0,10
	MAT.MATLAB	1,000 h	Licencia MATLAB R2019a	24,00	24,00
		4,000 %	Costes indirectos	34,02	1,36
			Precio total por h		35,38

**DESARROLLO DE UNA METODOLOGÍA PARA LA AYUDA AL DIAGNÓSTICO DEL CÁNCER DE MAMA
MEDIANTE EL ANÁLISIS DE TEXTURAS EN IMÁGENES DE RESONANCIA MAGNÉTICA**

Nº	Código	Ud	Descripción		Total
3 DESARROLLO DE LA NUEVA METODOLOGÍA, JUNTO ...					
3.1	03.01	h	Extracción de características		
	MO.COT1	0,300 h	Mano de obra de tipo Ingeniero Cotutor 1	16,74	5,02
	MO.COT2	0,300 h	Mano de obra de tipo Ingeniero Cotutor 2	16,74	5,02
	MO.AUT	1,000 h	Mano de obra de tipo Ingeniero Junior en Prácticas	8,37	8,37
	MAT.PCp	1,000 h	Ordenador portátil personal	1,55	1,55
	MAT.W10	1,000 h	Licencia Windows 10	0,10	0,10
	MAT.MATLAB	1,000 h	Licencia MATLAB R2019a	24,00	24,00
	MAT.Python	1,000 h	Licencia Python	26,00	26,00
		4,000 %	Costes indirectos	70,06	2,80
			Precio total por h		72,86
3.2	03.02	h	Programación de la herramienta de clasificación		
	MO.COT2	0,200 h	Mano de obra de tipo Ingeniero Cotutor 2	16,74	3,35
	MO.AUT	1,000 h	Mano de obra de tipo Ingeniero Junior en Prácticas	8,37	8,37
	MAT.PCp	1,000 h	Ordenador portátil personal	1,55	1,55
	MAT.W10	1,000 h	Licencia Windows 10	0,10	0,10
	MAT.MATLAB	1,000 h	Licencia MATLAB R2019a	24,00	24,00
		4,000 %	Costes indirectos	37,37	1,49
			Precio total por h		38,86
3.3	03.03	h	Obtención y evaluación de los resultados		
	MO.TUT	0,100 h	Mano de obra de tipo Ingeniero Tutor	27,90	2,79
	MO.COT1	0,100 h	Mano de obra de tipo Ingeniero Cotutor 1	16,74	1,67
	MO.COT2	0,100 h	Mano de obra de tipo Ingeniero Cotutor 2	16,74	1,67
	MO.AUT	1,000 h	Mano de obra de tipo Ingeniero Junior en Prácticas	8,37	8,37
	MAT.PCp	1,000 h	Ordenador portátil personal	1,55	1,55
	MAT.W10	1,000 h	Licencia Windows 10	0,10	0,10
	MAT.MATLAB	1,000 h	Licencia MATLAB R2019a	24,00	24,00
		4,000 %	Costes indirectos	40,15	1,61
			Precio total por h		41,76
4 REDACCIÓN Y DEFENSA DEL TFG					
4.1	04.01	h	Redacción de la memoria del proyecto		
	MO.AUT	1,000 h	Mano de obra de tipo Ingeniero Junior en Prácticas	8,37	8,37
	MAT.PCp	1,000 h	Ordenador portátil personal	1,55	1,55
	MAT.W10	1,000 h	Licencia Windows 10	0,10	0,10
		4,000 %	Costes indirectos	10,02	0,40
			Precio total por h		10,42
4.2	04.02	h	Revisión y corrección del proyecto		
	MO.TUT	0,300 h	Mano de obra de tipo Ingeniero Tutor	27,90	8,37
	MO.COT1	0,600 h	Mano de obra de tipo Ingeniero Cotutor 1	16,74	10,04
	MO.COT2	0,600 h	Mano de obra de tipo Ingeniero Cotutor 2	16,74	10,04
	MO.AUT	1,000 h	Mano de obra de tipo Ingeniero Junior en Prácticas	8,37	8,37
	MAT.PCp	1,000 h	Ordenador portátil personal	1,55	1,55
	MAT.W10	1,000 h	Licencia Windows 10	0,10	0,10
		4,000 %	Costes indirectos	38,47	1,54
			Precio total por h		40,01
4.3	04.03	h	Preparación de la defensa del proyecto		
	MO.AUT	1,000 h	Mano de obra de tipo Ingeniero Junior en Prácticas	8,37	8,37
	MAT.PCp	1,000 h	Ordenador portátil personal	1,55	1,55
	MAT.W10	1,000 h	Licencia Windows 10	0,10	0,10
		4,000 %	Costes indirectos	10,02	0,40
			Precio total por h		10,42

6. CUADRO DE MEDICIONES

Este apartado es muy similar al Cuadro de Precios Descompuestos ya que está organizado en los mismos capítulos mencionados anteriormente. No obstante, en esta lista se muestra la descripción de cada unidad, así como su cantidad.

01 PLANIFICACIÓN DEL PROYECTO Y REVISIÓN BIBLIOGRÁFICA DEL ESTADO DEL A...

Nº	Ud	Descripción	Medición
01.01	H	Reunión inicial con el Tutor, Cotutores y Radiólogo del TFG	
			Total h : 2,00
01.02	H	Revisión bibliográfica del estado del arte	
			Total h : 20,00
01.03	H	Reuniones periódicas para el seguimiento del proyecto	
			Total h : 7,00

02 OBTENCIÓN Y PREPARACIÓN DE LAS IMÁGENES

Nº	Ud	Descripción	Medición
02.01	H	Formación sobre el uso de los distintos softwares y posterior instalación	
			Total h : 12,00
02.02	H	Descarga de las imágenes y anonimización de los datos	
			Total h : 18,00
02.03	H	Segmentación de las ROIs de las imágenes	
			Total h : 55,00
02.04	H	Preparación de los archivos .mat	
			Total h : 8,00

03 DESARROLLO DE LA NUEVA METODOLOGÍA, JUNTO CON EL ANÁLISIS DE TEXTUR...

Nº	Ud	Descripción	Medición
03.01	H	Extracción de características	
			Total h : 10,00
03.02	H	Programación de la herramienta de clasificación	
			Total h : 48,00
03.03	H	Obtención y evaluación de los resultados	
			Total h : 75,00

04 REDACCIÓN Y DEFENSA DEL TFG

Nº	Ud	Descripción	Medición
04.01	H	Redacción de la memoria del proyecto	
			Total h : 92,00
04.02	H	Revisión y corrección del proyecto	
			Total h : 10,00
04.03	H	Preparación de la defensa del proyecto	
			Total h : 33,00

7. CUADRO DE PRECIOS PARCIALES

El Cuadro de Precios Parciales es una lista donde se muestra todo el contenido descrito en las mediciones junto con el presupuesto de ejecución material.

Presupuesto parcial nº 01 PLANIFICACIÓN DEL PROYECTO Y REVISIÓN BIBLIOGRÁF...

Nº	Ud	Descripción	Medición	Precio	Importe
01.01	H	Reunión inicial con el Tutor, Cotutores y Radiólogo del TFG			
		Total h :	2,00	110,26	220,52
01.02	H	Revisión bibliográfica del estado del arte			
		Total h :	20,00	11,36	227,20
01.03	H	Reuniones periódicas para el seguimiento del proyecto			
		Total h :	7,00	65,57	458,99
Total Presupuesto parcial nº 01 PLANIFICACIÓN DEL PROYECTO Y REVISIÓN BI...					906,71

Presupuesto parcial nº 02 OBTENCIÓN Y PREPARACIÓN DE LAS IMÁGENES

Nº	Ud	Descripción	Medición	Precio	Importe
02.01	H	Formación sobre el uso de los distintos softwares y posterior instalación			
		Total h :	12,00	27,79	333,48
02.02	H	Descarga de las imágenes y anonimización de los datos			
		Total h :	18,00	17,05	306,90
02.03	H	Segmentación de las ROIs de las imágenes			
		Total h :	55,00	19,34	1.063,70
02.04	H	Preparación de los archivos .mat			
		Total h :	8,00	35,38	283,04
Total Presupuesto parcial nº 02 OBTENCIÓN Y PREPARACIÓN DE LAS IMÁGENE...					1.987,12

Presupuesto parcial nº 03 DESARROLLO DE LA NUEVA METODOLOGÍA, JUNTO CO...

Nº	Ud	Descripción	Medición	Precio	Importe
03.01	H	Extracción de características			
		Total h :	10,00	72,86	728,60
03.02	H	Programación de la herramienta de clasificación			
		Total h :	48,00	38,86	1.865,28
03.03	H	Obtención y evaluación de los resultados			
		Total h :	75,00	41,76	3.132,00
Total Presupuesto parcial nº 03 DESARROLLO DE LA NUEVA METODOLOGÍA, JU...					5.725,88

DESARROLLO DE UNA METODOLOGÍA PARA LA AYUDA AL DIAGNÓSTICO DEL CÁNCER DE MAMA
MEDIANTE EL ANÁLISIS DE TEXTURAS EN IMÁGENES DE RESONANCIA MAGNÉTICA

Presupuesto parcial nº 04 REDACCIÓN Y DEFENSA DEL TFG

Nº	Ud	Descripción	Medición	Precio	Importe
04.01	H	Redacción de la memoria del proyecto			
			Total h :	92,00	10,42
					958,64
04.02	H	Revisión y corrección del proyecto			
			Total h :	10,00	40,01
					400,10
04.03	H	Preparación de la defensa del proyecto			
			Total h :	33,00	10,42
					343,86
Total Presupuesto parcial nº 04 REDACCIÓN Y DEFENSA DEL TFG :					1.702,60

8. PRESUPUESTO DE EJECUCIÓN POR CONTRATA

Esta sección del Presupuesto se corresponde con los Presupuestos Parciales sumados a los gastos generales (14% de los Presupuestos Parciales), el beneficio industrial (6% de los Presupuestos Parciales + Gastos Generales) y el 21% de IVA.

Proyecto: PRESUPUESTO TFG

Capítulo	Importe
1 PLANIFICACIÓN DEL PROYECTO Y REVISIÓN BIBLIOGRÁFICA DEL ESTADO DEL ARTE	906,71
2 OBTENCIÓN Y PREPARACIÓN DE LAS IMÁGENES	1.987,12
3 DESARROLLO DE LA NUEVA METODOLOGÍA, JUNTO CON EL ANÁLISIS DE TEXTURAS Y F...	5.725,88
4 REDACCIÓN Y DEFENSA DEL TFG	1.702,60
Presupuesto de ejecución material	10.322,31
15% de gastos generales	1.548,35
6% de beneficio industrial	619,34
Suma	12.490,00
21% IVA	2.622,90
Presupuesto de ejecución por contrata	15.112,90

Asciende el presupuesto de ejecución por contrata a la expresada cantidad de QUINCE MIL CIENTO DOCE EUROS CON NOVENTA CÉNTIMOS.

Valencia, Junio 2020
Ingeniera Biomédica

María Pilar Castellote Huguet