



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería Industrial

Desarrollo de un modelo predictivo mediante técnicas de
aprendizaje máquina como soporte al diagnóstico del
estadio del cáncer de mama a partir del análisis de
imágenes de PET/RM

Trabajo Fin de Grado

Grado en Ingeniería Biomédica

AUTOR/A: Iborra Roncales, Ignacio

Tutor/a: Moratal Pérez, David

Cotutor/a externo: ROMERO MARTIN, JUAN ANTONIO

CURSO ACADÉMICO: 2023/2024



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



ESCOLA TÈCNICA
SUPERIOR ENGINYERIA
INDUSTRIAL VALÈNCIA

TRABAJO FINAL DE GRADO EN INGENIERÍA BIOMÉDICA

Desarrollo de un modelo predictivo mediante técnicas de aprendizaje máquina como soporte al diagnóstico del estadio del cáncer de mama a partir del análisis de imágenes de PET/RM

AUTOR: Ignacio Iborra Roncales.

TUTOR/A: David Moratal Pérez, Universitat Politècnica de València

COTUTORES: Juan Antonio Romero Martín, ASCIRES Grupo Biomédico

CURSO ACADÉMICO: 2023/2024

AGRADECIMIENTOS

La realización del presente Trabajo de Fin de Grado no habría sido posible gracias a la ayuda y el apoyo continuo de mi tutor David Moratal Pérez y mi cotutor Juan Antonio Romero Martín. Agradecer en especial a Juan Antonio la facilidad y cercanía que ha mostrado en todo momento para ayudarme e interesarse por mis avances.

Agradecer de corazón a toda mi familia, madre, hermanos, tíos, primos, abuelos y amigos, por el apoyo incondicional durante toda la vida, pero en especial durante este último y difícil año y medio, sin ellos, seguro este Trabajo no habría salido adelante.

Finalmente, hacer una mención especial a mi padre, agradecerle de corazón todo lo que ha hecho para hacer de mí quién hoy soy, estoy seguro de que me ha estado ayudando y apoyando para sacar este proyecto y estará muy orgulloso allá donde esté.

RESUMEN

El tratamiento y pronóstico de una paciente con cáncer de mama dependen esencialmente de la estadificación en la escala TNM de dicho cáncer en particular. A la hora de determinar con seguridad el estadio del cáncer de mama se debe llevar a cabo un proceso complejo y delicado que incluye múltiples pruebas, y que a su vez puede resultar muy tedioso para la paciente. Se requieren desde técnicas invasivas como las biopsias para estudiar la afectación de los ganglios linfáticos, hasta pruebas de imagen tales como la mamografía, ecografía, resonancia magnética o tomografía computarizada para determinar así el tamaño del tumor y la posible presencia de metástasis.

La obtención de imágenes por fusión de resonancia magnética (RM) con imágenes PET, nos permite obtener un estudio metabólico y morfológico que facilita la precisión en el diagnóstico de la lesión. Surge así la posibilidad de disponer de una herramienta que permite estadificar el cáncer de mama de un paciente a partir de una única prueba PET/RM, permitiendo extraer de dichas imágenes características relacionadas con la morfología, la textura y la radiómica del tumor.

Por todo ello, durante el presente Trabajo de Fin de Grado se ha desarrollado un modelo predictivo que concluye que las características de las imágenes de la lesión segmentada obtenidas por PET/RM son significativamente relevantes para determinar la presencia o no de metástasis y facilitar así el diagnóstico del estadio del cáncer de mama en la escala TNM.

Para ello se han exportado de la base de datos de ASCIRES Grupo Biomédico, empresa con la que se ha colaborado en el desarrollo del presente TFG, imágenes extraídas del equipo híbrido PET/RM de pacientes con cáncer de mama. Se ha segmentado en 3D el tumor de los pacientes para obtener de esta forma la región de interés (ROI) de la mama. Posteriormente, se han obtenido las características de las texturas y de radiómica de dichas imágenes y, finalmente, se han desarrollado distintos modelos predictivos en función de las características mencionadas. Se ha evaluado de cada uno de estos la capacidad y la relevancia predictiva en relación con la presencia o no de metástasis (M) en base a métricas de evaluación como la matriz de confusión y el área bajo la curva.

Palabras clave: cáncer de mama, escala TNM, resonancia magnética, PET, texturas, radiómica, estadio, mama.

RESUM

El tractament i pronòstic d'una pacient amb càncer de mama depenen essencialment de l'estadificació en l'escala TNM d'aquest càncer en particular. A l'hora de determinar amb seguretat l'estadi del càncer de mama s'ha de dur a terme un procés complex i delicat que inclou múltiples proves, i que al seu torn pot resultar molt tediós per a la pacient. Es requereixen des de tècniques invasives com les biòpsies per a estudiar l'afectació dels ganglis limfàtics, fins a proves d'imatge com ara la mamografia, ecografia, ressonància magnètica o tomografia computada per a determinar així la grandària del tumor i la possible presència de metàstasi.

L'obtenció d'imatges per fusió de ressonància magnètica (RM) amb imatges PET, ens permet obtenir un estudi metabòlic i morfològic que facilita la precisió en el diagnòstic de la lesió. Sorgeix així la possibilitat de disposar d'una eina que permet estadificar el càncer de mama d'un pacient a partir d'una única prova PET/RM, permetent extraure d'aquestes imatges característiques relacionades amb la morfologia, la textura i la radiòmica del tumor.

Per tot això, durant el present Treball de Fi de Grau s'ha desenvolupat un model predictiu que conclou que les característiques de les imatges de la lesió segmentada obtingudes per PET/RM són significativament rellevants per a determinar la presència o no de metàstasi i facilitar així el diagnòstic de l'estadi del càncer de mama en l'escala TNM.

Per a això s'han exportat de la base de dades de ASCIRES Grup Biomèdic, empresa amb la qual s'ha col·laborat en el desenvolupament del present TFG, imatges extretes de l'equip híbrid PET/RM de pacients amb càncer de mama. S'ha segmentat en 3D el tumor dels pacients per a obtenir d'aquesta manera la regió d'interés (ROI) de la mama. Posteriorment, s'han obtingut les característiques de les textures i de radiòmica d'aquestes imatges i, finalment, s'han desenvolupat diferents models predictius en funció de les característiques esmentades. S'ha avaluat de cadascun d'aquests la capacitat i la rellevància predictiva en relació amb la presència o no de metàstasi (M) sobre la base de mètriques d'avaluació com la matriu de confusió i l'àrea sota la corba.

Paraules clau: càncer de mama, escala TNM, ressonància magnètica, PET, textures, radiòmica, estadi, mama.

ABSTRACT

The treatment and prognosis of a patient with breast cancer depends essentially on the TNM staging of that cancer. To determine the stage of breast cancer with certainty, a complex and delicate process involving multiple tests must be carried out, which can be very tedious for the patient. These range from invasive techniques such as biopsies to study lymph node involvement, to imaging tests such as mammography, ultrasound, MRI or CT scans to determine the size of the tumour and the possible presence of metastases.

The obtaining of images by fusion of magnetic resonance imaging (MRI) with PET images allows us to obtain a metabolic and morphological study that facilitates the precision in the diagnosis of the lesion. This gives us the possibility of having a tool that allows the staging of a patient's breast cancer from a single PET/MRI test, allowing us to extract characteristics related to the morphology, texture and radiomics of the tumour from these images.

For all these reasons, during this Final Degree Project a predictive model has been developed that concludes that the characteristics of the images of the segmented lesion obtained by PET/MRI are significantly relevant to determine the presence or absence of metastasis and thus facilitate the diagnosis of the stage of breast cancer on the TNM scale.

For this purpose, images extracted from the hybrid PET/MR equipment of breast cancer patients have been exported from the database of ASCIRES Grupo Biomédico, a company with which we have collaborated in the development of this TFG. The patients' tumours were segmented in 3D to obtain the region of interest (ROI) of the breast. Subsequently, the texture and radiomic characteristics of these images were obtained and, finally, different predictive models were developed based on the characteristics. The capacity and predictive relevance of each of these models in relation to the presence or absence of metastasis (M) has been evaluated based on evaluation metrics such as the confusion matrix and the area under the curve.

Key words: breast cancer, TNM scale, MRI, PET, textures, radiomics, stage, breast.

DOCUMENTOS CONTENIDOS EN EL TFG

1. MEMORIA

2. PRESUPUESTO

MEMORIA

Desarrollo de un modelo predictivo mediante técnicas de aprendizaje máquina como soporte al diagnóstico del estadio del cáncer de mama a partir del análisis de imágenes de PET/RM

AUTOR: Ignacio Iborra Roncales

TUTOR/A: David Moratal Pérez, Universitat Politècnica de València

COTUTORES: Juan Antonio Romero, ASCIRES Grupo Biomédico

CURSO ACADÉMICO: 2023/2024

ÍNDICE

1.	MOTIVACIÓN.....	12
2.	OBJETIVOS.....	13
3.	INTRODUCCIÓN.....	14
3.1	ANATOMÍA DE LA MAMA.....	14
3.2	CÁNCER DE MAMA.....	15
3.2.1	EPIDEMIOLOGÍA.....	16
3.2.2	ETIOLOGÍA.....	16
3.2.3	DIAGNÓSTICO.....	17
3.2.4	TIPOS DE CÁNCER DE MAMA.....	17
3.2.5	ESTADIFICACIÓN.....	18
3.2.6	TRATAMIENTO.....	19
3.3	IMAGEN DE RESONANCIA MAGNÉTICA.....	20
3.3.1	IMAGEN DE RESONANCIA MAGNÉTICA MAMARIA.....	20
3.4	IMAGEN DE TOMOGRAFÍA POR EMISIÓN DE POSITRONES.....	22
3.5	PET/RM.....	23
3.6	ANÁLISIS DE DESCRIPTORES.....	23
3.6.1	DEFINICIÓN DE LA REGIÓN DE INTERÉS.....	24
3.6.2	EXTRACCIÓN DE CARACTERÍSTICAS DE TEXTURAS.....	24
3.6.3	SELECCIÓN DE CARACTERÍSTICAS.....	26
3.6.4	MÉTODOS DE CLASIFICACIÓN.....	27
3.6.5	VALIDACIÓN DEL MODELO.....	28
3.6.6	MÉTRICAS DE EVALUACIÓN.....	29
4.	MATERIALES.....	31
4.1	IMÁGENES PET/RM.....	31
4.2	SOFTWARES PARA EL PREPROCESADO DE IMÁGENES.....	31
4.3	SOFTWARE DE MATLAB.....	32
4.3.1	RADIOMICS 3D.....	32
4.4	SOFTWARE DE OLEA SPHERE.....	32
5.	MÉTODOS.....	33
5.1	ADQUISICIÓN DE LAS IMÁGENES.....	33
5.2	SEGMENTACIÓN DE LA LESIÓN.....	34
5.3	ANÁLISIS DE LAS CARACTERÍSTICAS.....	34
5.3.1	EXTRACCIÓN DE LAS CARACTERÍSTICAS DE TEXTURAS.....	35

5.3.2 EXTRACCIÓN DE LAS CARACTERÍSTICAS DE OLEA	35
5.4 SELECCIÓN DE CARACTERÍSTICAS.....	36
5.5 CLASIFICACIÓN	36
6. RESULTADOS	38
6.1 CLASIFICACIÓN BINARIA.....	38
6.2 ANÁLISIS DEL ESTUDIO BINARIO	40
6.3 ANÁLISIS DATOS DE PRUEBA.....	42
7. DISCUSIÓN.....	44
8. CONCLUSIONES	46
9. LÍNEAS FUTURAS	47
10. BIBLIOGRAFÍA.....	48
11. ANEXOS.....	50
11.1 ANEXO I. RELACIÓN DEL TRABAJO CON LOS OBJETIVOS DE DESARROLLO SOSTENIBLE DE LA AGENDA 2030.....	50

ÍNDICE DE FIGURAS

Figura 1. Estructura anatómica de la mama [5].	14
Figura 2. Vascularización del tejido mamario [6].	15
Figura 3. Cuadrantes de la mama en el plano coronal [7].	15
Figura 4. Sistema de clasificación TNM [14].	18
Figura 5. Componentes del equipo de RM [15].	20
Figura 6. Posicionamiento RM de mama [17].	20
Figura 7. Diferentes planos de una mama, axial, coronal y sagital respectivamente. Fuente: Imágenes proporcionadas por ASCIRES Grupo Biomédico.	20
Figura 8. Series de resonancia magnética, serie DINÁMICA y ADC respectivamente. Fuente: Imágenes proporcionadas por ASCIRES Grupo Biomédico.	21
Figura 9. Proceso de aniquilación dentro del tomógrafo.	22
Figura 10. Imagen PET de una lesión maligna en la mama. Fuente: Imagen proporcionada por ASCIRES Grupo Biomédico.	22
Figura 11. Cálculo de una GLCM para una dirección horizontal 2 para una matriz 4x5 con 8 niveles de gris.	25
Figura 12. Cálculo de una GLRLM para una matriz 4x4 con 3 niveles de gris [24].	25
Figura 13. Cálculo de una GLSZM para una matriz 4x4 con 3 niveles de gris [24].	25
Figura 14. Cálculo de una NGTDM de una matriz 4x4 píxeles con 3 niveles de gris [24].	26
Figura 15. Matriz de confusión [28].	29
Figura 16. Curva ROC con el valor del AUC correspondiente. Fuente: elaboración propia.	30
Figura 17. Imagen en PET/RM de cáncer de mama correspondiente a la serie "WATER: Ph3/3D Ax VIBRANT DINAMIC". Fuente: Imagen proporcionada por ASCIRES Grupo Biomédico.	33
Figura 18. Segmentación de la lesión en mama derecha realizada con 3D Slicer. Fuente: elaboración propia.	34
Figura 19. Interfaz Olea Sphere para la extracción de características. Fuente: elaboración propia.	36
Figura 20. Matriz de confusión del método de clasificación Fine Tree con selección de características por Wrapper (DiagQuadratic). Fuente: elaboración propia.	41
Figura 21. Matriz de confusión del método de clasificación KNN con selección de características por Wrapper (DiagLinear). Fuente: elaboración propia.	41
Figura 22. Matriz de confusión del método de clasificación Fine Tree con selección de características por Fisher Score de 5 características. Fuente: elaboración propia.	42
Figura 23. Matriz de confusión del método de clasificación KNN con selección de características por Fisher Score de 20 características. Fuente: elaboración propia.	42

ÍNDICE DE TABLAS

Tabla 1. Sistema de clasificación TNM [2] [12] [13].	18
Tabla 2. <i>Estadio en función del sistema TNM</i> [14].	19
Tabla 3. Recuento de las características de texturas extraídas.	26
Tabla 4. Datos del estudio binario correspondiente a la presencia o no de metástasis. Fuente: elaboración propia.	38
Tabla 5. Resultados obtenidos en el estudio del conjunto de descriptores extraídos aplicando el método de selección de características de Wrappers para diferentes métodos de clasificación. Fuente: Elaboración propia.	39
Tabla 6. Resultados obtenidos en el estudio del conjunto de descriptores extraídos aplicando el método de selección de características de Fisher Score para diferentes métodos de clasificación y diferente número de características. Fuente: Elaboración propia.	39
Tabla 7. Resultados obtenidos en el estudio del conjunto de descriptores extraídos aplicando el método de selección de características de ReliefF para diferentes métodos de clasificación y diferente número de características. Fuente: Elaboración propia	40

1. MOTIVACIÓN

El cáncer de mama es la patología maligna más frecuentemente diagnosticada entre las mujeres y con tasas más elevadas en países desarrollados, de forma que se estima que el riesgo de padecer esta enfermedad es de 1 de cada 8 mujeres. Además, la incidencia del cáncer de mama ha aumentado en los últimos años, este suceso se relaciona con el aumento y envejecimiento de la población, la exposición a ciertos factores de riesgo, sedentarismo, entre otros. No obstante, la mortalidad de esta patología desciende gracias al desarrollo de programas que permiten una detección precoz y mejora de tratamientos [1].

Una vez es diagnosticada esta enfermedad, se procede con la estadificación que describe la etapa del cáncer en el cuerpo y que determina en cierta manera tanto la gravedad y pronóstico del paciente como la vía óptima de tratamiento. Los cánceres en etapas más tempranas se identifican como estadio 0, mientras que el resto van desde estadio I al IV, siendo IV el peor pronóstico [2].

La escala de estadificación que se utiliza habitualmente es la del sistema TNM del *American Joint Committee on Cancer* (AJCC). Esencialmente a través de este sistema se determina el estadio en función del tamaño del tumor (T), el número de ganglios linfáticos adyacentes infectados (N) y la presencia de metástasis (M). Este tercer factor es crucial, ya que en caso de haber metástasis directamente se determina, indiferentemente del resto de valores, estadio IV [2]. Para establecer el estadio el paciente se ve sometido a numerosas y tediosas pruebas desde técnicas de imagen como mamografía, ecografía, resonancia magnética o tomografía computarizada, hasta pruebas invasivas como la biopsia [3].

Dada la importancia de una correcta y precoz estadificación, y el interés tanto de los especialistas como de los propios pacientes, es de gran relevancia desarrollar un programa que facilite la estadificación minimizando el número de pruebas a las que el paciente se ve sometido.

En este contexto, la extracción y evaluación de las distintas características constituyentes en la región de interés que se obtienen por imagen de resonancia magnética y PET (PET/RM), son un aspecto relevante para la comprensión de las imágenes médicas y, en consecuencia, para obtener un diagnóstico correcto del paciente en función de las características significativas y evitar así las numerosas pruebas que hacen del proceso de diagnóstico un trámite molesto para el individuo.

2. OBJETIVOS

El principal objetivo que define y motiva este Trabajo de Fin de Grado es concluir si los diferentes descriptores que se pueden obtener de las imágenes adquiridas por PET/RM poseen un valor predictivo suficiente de cara a determinar la presencia o no de metástasis en una paciente con cáncer de mama de forma que fuese posible desarrollar una herramienta de soporte al diagnóstico de la escala TNM.

Para alcanzar este objetivo principal se definen los siguientes objetivos específicos:

- Caracterizar las lesiones a partir de las características de texturas y de *Olea Sphere* extraídas de las imágenes.
- Reducir la dimensionalidad de los descriptores adquiridos de las imágenes.
- Realizar diferentes análisis estadísticos predictivos sobre la presencia de metástasis.
- Evaluar los diferentes métodos de clasificación entrenados.

3. INTRODUCCIÓN

3.1 ANATOMÍA DE LA MAMA

Las mamas representan órganos glandulares de secreción externa, situadas de forma subcutánea en la pared anterior del tórax, sobre el músculo pectoral mayor a ambos lados de la línea media, desde la segunda hasta la sexta costilla (Figura 1). Estos órganos están encargados fundamentalmente de producir y secretar la leche materna durante el periodo de lactancia [4].

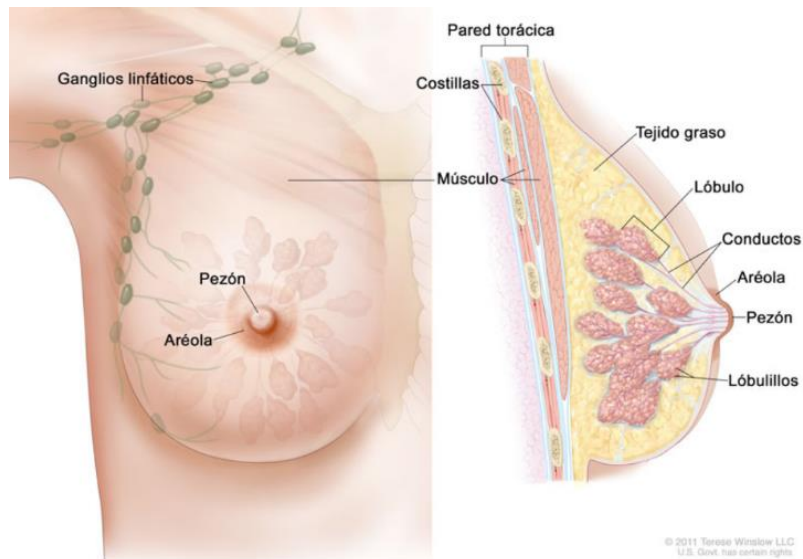


Figura 1. Estructura anatómica de la mama [5].

La mama está compuesta principalmente por tejido adiposo, tejido epitelial y la glándula mamaria, que consiste en aproximadamente 15 a 20 lobulillos glandulares dispuestos en forma circular (Figura 1). Estos lobulillos se conectan a los conductos galactóforos, los cuales desembocan en el seno galactóforo, una estructura que actúa como reservorio de leche y conecta el interior de la mama con el exterior a través del pezón, por donde se libera la leche durante la lactancia.

Además, cada lóbulo principal se descompone en lobulillos más pequeños que finalizan en las células alveolares, las cuales son responsables de la producción de leche. Por lo tanto, se puede observar que la mama tiene una estructura piramidal, con una mayor concentración de vasos y glándulas en la base, cerca de la pared torácica, y en el vértice se encuentra el pezón [5].

Cabe mencionar que la vascularización de la mama se realiza de forma radial desde la periferia del pezón, varias ramas arteriales permiten la irrigación de los distintos sectores mamarios. Esta vascularización está cubierta por vasos situados a ambos lados del esternón, así como por los vasos torácicos (Figura 2).

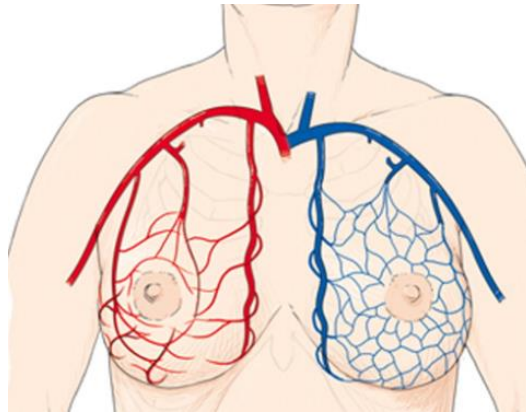


Figura 2. Vascularización del tejido mamario [6].

Es importante conocer y comprender a la hora de hacer referencia a una lesión en la mama, la distribución en cuadrantes de esta. La mama se divide en 4 cuadrantes en el plano coronal (Figura 3), y en tres porciones tanto en el plano sagital como axial.

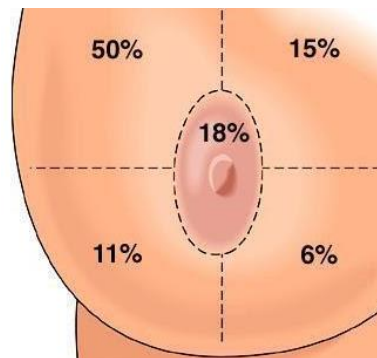


Figura 3. Cuadrantes de la mama en el plano coronal [7].

Los 4 cuadrantes de la mama correspondientes al plano coronal se forman por el entrecruzamiento de una línea vertical y otra horizontal que se cruzan en el pezón. De esta forma se diferencia el cuadrante superior externo (CSE), cuadrante superior interno (CSI), cuadrante inferior externo (CIE) y el cuadrante inferior interno (CII). Estos cuadrantes concurren en el complejo areola-pezón formando la región retroareolar.

Mientras que las 3 regiones asociadas al plano sagital y al plano axial se identifican como porción anterior, medio o posterior, en función de si es en la zona más cercana a la pared torácica (posterior), en la zona media o en la más cercana a la región retroareolar (anterior).

3.2 CÁNCER DE MAMA

El cáncer de mama es una patología en la que las células mamarias alteradas se multiplican de forma descontrolada dando lugar al tumor. Estas células acaban adquiriendo la capacidad para invadir otros tejidos y órganos pudiendo producir metástasis y caracterizando así al tumor como maligno [8].

Representa el cáncer más comúnmente diagnosticado en las mujeres en 157 países. Durante 2022 se diagnosticaron en el mundo, 2,3 millones de casos de cáncer de mama en el mundo, además de registrarse 670.000 defunciones [8].

3.2.1 EPIDEMIOLOGÍA

Como se ha mencionado con anterioridad, el cáncer de mama es el tumor más habitual entre las mujeres, concretamente corren el riesgo de padecerlo a lo largo de la vida del 12%, afectando así a 1 de cada 8 mujeres [1].

En 2022, se diagnosticaron 2,3 millones de casos de cáncer de mama en el mundo. Concretamente en España, se registraron aproximadamente 34.740 nuevos casos en 2022, lo que constituye casi el 29% de los cánceres diagnosticados en mujeres, según el Observatorio de la Asociación Española Contra el Cáncer [1].

La mayoría de los diagnósticos se realizan en mujeres de entre 45 y 65 años, aunque la incidencia aumenta con la edad y se estabiliza a partir de los 60 años. El incremento de las tasas de incidencia en los últimos años puede atribuirse al crecimiento y envejecimiento de la población, cambios en los hábitos reproductivos, mayor exposición a factores de riesgo como la obesidad y el uso de terapias hormonales, el sedentarismo y la globalización del cribado poblacional. A pesar de este aumento, la mortalidad está en descenso gracias a los programas de detección precoz y las mejoras en los tratamientos [1].

3.2.2 ETIOLOGÍA

Existen varios factores que pueden alterar la probabilidad de desarrollar un cáncer de mama, por tanto, se trata de una patología multifactorial. Es importante conocer las causas que pueden conducir a acrecentar el riesgo de padecer la enfermedad y ocasionar las mutaciones celulares que la originan [1].

Comenzando con la edad y el género, son dos factores cruciales en el cáncer de mama puesto que el 99% de los casos se dan en mujeres, y la mayor parte de ellos se diagnostican de forma mayoritaria en torno a los 50 y 70 años [9].

En segundo lugar, se deben considerar los antecedentes tanto a nivel familiar como personal. A nivel familiar, el cáncer de mama puede ser hereditario, es importante conocer que existen diferentes mutaciones genéticas hereditarias asociadas a un aumento del riesgo de desarrollar cáncer de mama, entre otros. Los genes *BRCA1* o *BRCA2* son los genes comúnmente vinculados con el cáncer de mama, así como con otros cánceres [9].

Otra serie de factores asociados a un mayor riesgo de desarrollar la enfermedad podrían ser la densidad de la mama (mujeres con más del 75% de la mama constituida por tejido denso tienen mayor riesgo), menstruación temprana, menopausia tardía, y estilos de vida poco saludables como el sedentarismo, sobrepeso, alcohol, etc. [9].

3.2.3 DIAGNÓSTICO

La detección precoz de la enfermedad se antoja determinante a la hora de obtener un mejor pronóstico del cáncer, puesto que la supervivencia a 5 años varía si la lesión es local, regional o si ya se ha diseminado. En el caso del cáncer de mama, la enfermedad no es especialmente sintomática, sin embargo, puede presentar un bulto en la mama o axila, secreción del pezón, enrojecimiento o descamación en la mama o en el pezón [10].

Generalmente el proceso de evaluación o diagnóstico del cáncer de mama consta de un primer examen clínico en el que en caso de detectarse alguna anomalía desencadena una serie de pruebas de imagen médica que permiten obtener el diagnóstico final [10].

Las técnicas de imagen más habituales son la mamografía y la ecografía en las que si no se obtiene un diagnóstico claro o se sospecha malignidad pueden desencadenar el uso de técnicas como la resonancia magnética [10].

Una vez diagnosticada la malignidad o en caso de que no se pueda determinar con certeza en las pruebas mencionadas, se solicita una biopsia guiada por imágenes. Este método permite obtener tejido mamario sin necesidad de cirugía para analizarlo, con la posibilidad incluso de eliminar el tumor mediante extirpación quirúrgica [10].

3.2.4 TIPOS DE CÁNCER DE MAMA

El cáncer de mama se trata de un adenocarcinoma, es decir, es un cáncer de origen glandular que se da debido a una serie de mutaciones celulares. Podemos clasificar la enfermedad según varias características: la presencia o ausencia de invasión, las características histológicas, los rasgos inmunohistoquímicos y el grado de diferenciación celular [1] [11].

La clasificación en función de la presencia o no de invasión diferencia entre carcinoma in situ que hace referencia a la proliferación celular maligna que ocurre en el interior del conducto o del lobulillo mamario, sin infiltración del tejido que lo rodea, y, carcinoma infiltrante que es aquella que si traspasa la frontera del conducto o del lobulillo [1] [11].

En segundo lugar, la clasificación histológica que diferencia según el tipo de célula que origina la enfermedad. Dentro de esta clase los más comunes son los carcinomas ductales que se forman en las células que recubren los conductos galactóforos, este tipo representa en torno al 80% de los cánceres. Y, por otro lado, también son habituales los carcinomas lobulillares que se dan en las células de los lobulillos mamarios [1] [11].

A continuación, está la clasificación por grado histológico, que clasifica en función de la semejanza de las células tumorales a las células normales, existiendo desde grado 1 (bien diferenciado) hasta grado 3 (pobrementemente diferenciado) donde el riesgo de diseminación es superior [1] [11].

Finalmente, la clasificación más importante y considerada es por fenotipo obtenido a partir de técnicas de inmunohistoquímica. Permite clasificar la enfermedad en cuatro subtipos en base a la presencia de receptores hormonales de estrógeno (ER) y progesterona (PR), de la proteína HER₂ y el índice de proliferación (Ki67) [1] [11].

- Luminal A: alta expresión de los receptores hormonales ER y PR, así como un bajo índice de proliferación y baja expresión de HER₂. Beneficiado por las terapias hormonales.
- Luminal B: receptores hormonales ER y PR positivos, así como un alto índice de proliferación y baja expresión de HER₂.
- HER₂ positivo: HER₂ positivo y receptores hormonales ER y PR positivos o negativos. Tratamiento anti- HER₂ promueve un pronóstico más positivo.
- Triple negativo: HER₂ y receptores hormonales ER y PR negativos. Generalmente es el subtipo caracterizado por una mayor agresividad, pero el de menor incidencia.

3.2.5 ESTADIFICACIÓN.

Cuando se confirma la malignidad del cáncer, comienza el proceso de estadificación, este establece la etapa en la que se encuentra la enfermedad, ayuda a determinar la gravedad y la mejor forma de tratarla. El principal método es el comúnmente conocido como sistema TNM, regulado por Comité Conjunto Estadounidense sobre el Cáncer (AJCC, en inglés), con el objetivo de promover uniformidad entre los médicos y los centros de tratamiento a la hora de describir los diferentes casos y resultados de forma que puedan comprenderse con facilidad [2] [12] [12].

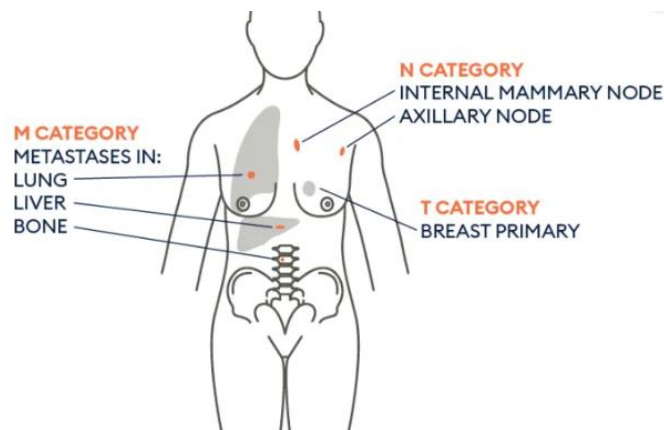


Figura 4. Sistema de clasificación TNM [14].

Las siglas de la escala de clasificación TNM corresponden a las diferentes categorías que se muestran en la Tabla 1.

Tamaño del tumor original (T)	Tis	Carcinoma <i>in situ</i>
	T1	≤ 20 mm
	T2	20 – 50 mm
	T3	> 50 mm
	T4	Afecta a la pared torácica y/o piel
Afectación de los ganglios linfáticos (N)	N0	0 ganglios linfáticos axilares
	N1	1-3 ganglios linfáticos axilares
	N2	4-9 ganglios linfáticos axilares
	N3	> 10 ganglios linfáticos axilares
Presencia o no de metástasis (M)	M0	Ausencia de metástasis
	M1	Presencia de metástasis

Tabla 1. Sistema de clasificación TNM [2] [12] [13].

Una vez se determina cada una de las variables del sistema de clasificación, se asigna el estadio del cáncer en función de la combinación de dichas variables.

TNM	Subestadio	Estadio
Tis N0 M0	0	0
T1 N0 M0	IA	I
T0-1 N1 M0	IB	
T0-1 N1 M0 T2 N0 M0	IIA	II
T2 N1 M0 T3 N0 M0	IIB	
T0-3 N2 M0 T3 N1 M0	IIIA	
T4 N0-2 M0	IIIB	III
T1-4 N3 M0	IIIC	
T1-4 N1-3 M1	IV	
		IV

Tabla 2. Estadio en función del sistema TNM [14].

Asociar un correcto estadio a la enfermedad del paciente se antoja fundamental ya que de este dependen las mejores posibilidades terapéuticas, así como el pronóstico del tratamiento seleccionado.

3.2.6 TRATAMIENTO

El tratamiento que recibirá el paciente depende como ya se ha comentado de distintos factores entre ellos el estadio diagnosticado. Se caracteriza por ser multidisciplinar, es decir, participan distintos especialistas como pueden ser oncólogos, radiólogos, ginecólogos etc. y puede combinar a su vez diferentes tipos de tratamientos.

En cuanto a las principales terapias que se aplican cabe mencionar la cirugía, conocida como mastectomía, que se basa en la extirpación de la mama. La cirugía puede ser más conservadora consistiendo en la extirpación únicamente del tumor con un cierto margen de tejido mamario sano, o por el contrario puede requerir extirpar la totalidad de la mama y ciertos tejidos circundantes.

Por otro lado, la radioterapia consiste en el uso de radiación ionizante con el objetivo de eliminar las células tumorales y, pese a que generalmente se aplica tras la cirugía, también se puede aplicar de forma intraoperatoria. Mientras que la quimioterapia, mediante fármacos inyectados o en pastillas persigue destruir las células tumorales. También se emplea la terapia hormonal, que impide que el tejido tumoral reciba las hormonas para su proliferación. Y finalmente otras alternativas como puede ser la terapia biológica dirigida, la terapia endocrina o la inmunoterapia [1] [13].

3.3 IMAGEN DE RESONANCIA MAGNÉTICA

La resonancia magnética (RM) es una técnica de imagen que permite la evaluación, detección, diagnóstico y seguimiento de diferentes patologías, capaz de obtener imágenes 2D y 3D de estructuras anatómicas con gran nivel de detalle por su sensibilidad y resolución. Se trata de una técnica no invasiva que consiste en la emisión de radiación electromagnética, es decir, radiación por radiofrecuencia no ionizante [15].

Un equipo de resonancia magnética se sirve de tres componentes principales hardware (Figura 5). En primer lugar, el imán, es el responsable de generar un campo magnético de una determinada intensidad y uniforme. A continuación, las bobinas de gradiente son las encargadas de generar variaciones a nivel tridimensional en la intensidad de campo magnético. Y, por último, la bobina de radiofrecuencia, cuyo papel es excitar los tejidos anatómicos a una frecuencia concreta [16].

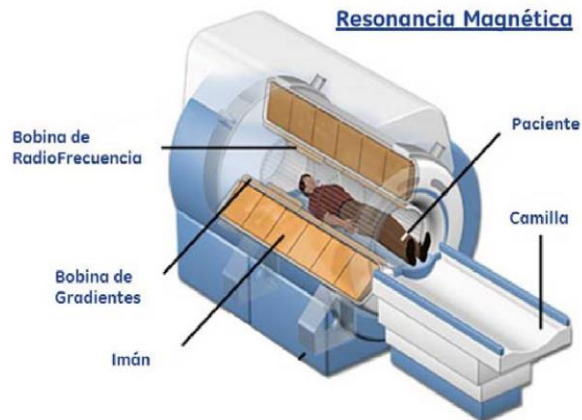


Figura 5. Componentes del equipo de RM [15].

La obtención de imágenes por resonancia magnética consiste en el siguiente fenómeno físico. El imán genera un campo magnético, bajo el cual aquellos átomos con número de protones y/o neutrones impar, que tengan movimiento de espín, rotan sobre sí mismos para alinearse con el campo magnético. Las bobinas de radiofrecuencia emiten ondas de forma que los núcleos se ven estimulados y absorben energía a una determinada frecuencia, la frecuencia de resonancia. Una vez los pulsos de radiofrecuencia se desconectan, los núcleos liberan la energía absorbida mediante la interacción con el medio y vuelven a su estado inicial. Esa energía liberada por los núcleos es percibida por las bobinas receptoras como señal generando una corriente inducida, señal que se traducirá en la imagen de resonancia final [15].

3.3.1 IMAGEN DE RESONANCIA MAGNÉTICA MAMARIA

Más allá de otras aplicaciones, el uso de la resonancia magnética de mamas tiene como objetivo evaluar, diagnosticar o monitorear el cáncer de mama de forma que se obtenga información detallada sobre la enfermedad o se esclarezcan las sospechas de padecerla facilitando su detección [16].

La adquisición de imagen por RM de mama requiere un equipamiento especial con el que no todos los centros hospitalarios cuentan. El equipo debe contar con un dispositivo especial denominado bobina del seno destinada a recrear la imagen de la mama. La obtención de la imagen requiere el posicionamiento del paciente en decúbito prono con las mamas colocadas como se puede observar en la Figura 6.



Figura 6. Posicionamiento RM de mama [17].

Esta técnica de imagen médica permite obtener secuencias en los tres planos principales, axial, sagital y coronal. Para mejorar la visualización de las imágenes de RM para la detección del cáncer de mama, se emplea generalmente un medio de contraste, comúnmente Gadolinio combinado con un quelante por su toxicidad por sí solo [16].

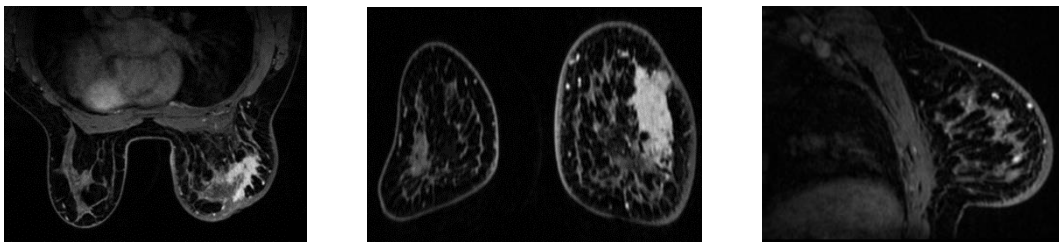


Figura 7. Diferentes planos de una mama, axial, coronal y sagital respectivamente. Fuente: Imágenes proporcionadas por ASCIRES Grupo Biomédico.

Además de la adición de contraste, variando los parámetros de adquisición se pueden potenciar ciertas propiedades de las imágenes, de forma que se pueden obtener distintas modalidades. Durante este trabajo se van a emplear fundamentalmente las expuestas en la figura 8.

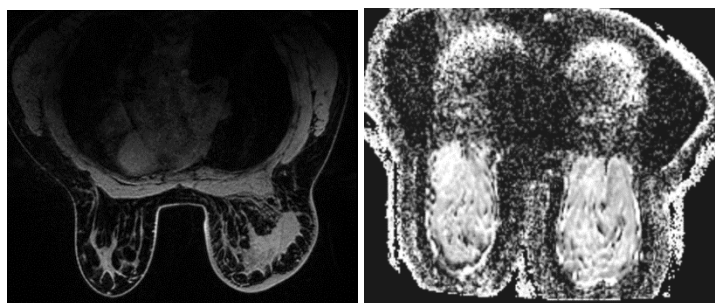


Figura 8. Series de resonancia magnética, serie DINÁMICA y ADC respectivamente. Fuente: Imágenes proporcionadas por ASCIRES Grupo Biomédico.

3.4 IMAGEN DE TOMOGRAFÍA POR EMISIÓN DE POSITRONES

La tomografía por emisión de positrones (PET) es una técnica de imagen médica que permite obtener imágenes funcionales y cuantitativas, a partir de radionucleidos emisores de positrones, aunque la técnica requiere la detección coincidente de dos fotones gamma.

Comienza a partir de la administración de un trazador radioactivo en el paciente que se reparte por el organismo del individuo mediante el sistema circulatorio. Con el transcurso del tiempo, se da la desintegración en forma de emisión de positrones. Cuando estos positrones colisionan se produce el fenómeno de aniquilación que produce dos rayos gamma en direcciones opuestas con una energía de 511 keV. Estos rayos son detectados permitiendo la obtención de la imagen en función de la distancia antes de la aniquilación que depende de la energía del positrón [18].

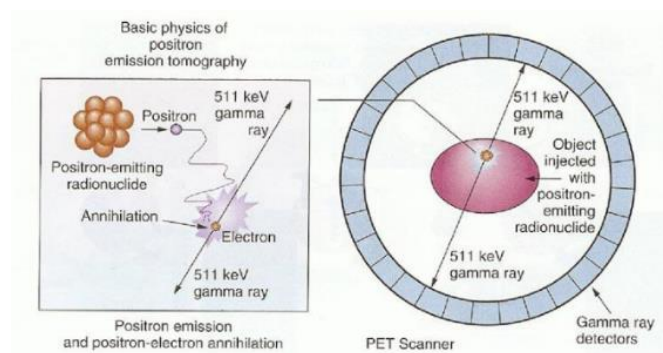


Figura 9. Proceso de aniquilación dentro del tomógrafo.

Uno de los puntos más interesantes que ofrece la técnica de imagen PET, es conocer la actividad metabólica a partir de la acumulación de la cantidad de radiofármaco. El índice que permite cuantificar esta actividad es conocido como *Standardized Uptake Value* máximo (SUV máx.), índice que se añade a las características utilizadas durante el estudio, calculado como la relación entre la captación y el volumen del *vóxel* de máxima captación [19].

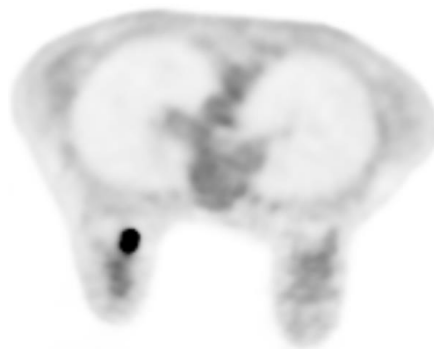


Figura 10. Imagen PET de una lesión maligna en la mama. Fuente: Imagen proporcionada por ASCIRES Grupo Biomédico.

3.5 PET/RM

El PET/RM consiste en un sistema híbrido que combina la tomografía por emisión de positrones y la resonancia magnética, es una técnica ya ampliamente utilizada para el diagnóstico, valoración y estadificación por imagen. La fundamental ventaja que ofrece el PET/RM es que integra de forma simultánea la información morfológica anatómica y funcional que ofrece la RM, con la información metabólica de los procesos bioquímicos y moleculares que se obtiene por PET [20] [21].

Otra de las principales ventajas que presenta esta técnica de adquisición es que reduce de forma significativa la dosis de radiación sobre el paciente. Por un lado, la resonancia magnética como bien se ha indicado no emite radiación ionizante y, por otro lado, el PET al aumentarse en este caso el tiempo de exposición va a necesitar una menor dosis. Todo esto lo hace muy beneficioso para todos los pacientes en general, sobre todo en aquellos casos en los que se requieren numerosas pruebas de seguimiento. Además, la tecnología avanzada con resonancia de 3 Teslas, permite que las imágenes de PET/RM ofrezcan un excelente resolución [22].

No obstante, ni todos los individuos pueden someterse a la técnica PET/RM por su contraindicación en pacientes con marcapasos, prótesis, claustrofobia, etc., ni es accesible para todos los centros médicos debido a su alto coste [22].

3.6 ANÁLISIS DE DESCRIPTORES

Para poder entender e interpretar de forma cualitativa y con mayor precisión una imagen médica es necesario comprender los diferentes descriptores que se pueden extraer de ella y que engloban una información determinada. Existen múltiples tipos de características, pero principalmente se conocen los descriptores de texturas y los de formas, que son englobados por el concepto de radiómica.

En este estudio, en primer lugar, se obtendrán los descriptores de texturas. Estos consisten en un conjunto de métricas que representan la disposición espacial de las distintas intensidades de gris dentro de la región de interés (ROI), definida previamente en el proceso de la segmentación, y se obtendrán a partir de modelos matemáticos evaluando la intensidad y su distribución [23].

En segundo lugar, se utilizarán características cuantitativas que se obtienen a partir de un análisis semiautomático a través del software de Olea Sphere sobre las diferentes series obtenidas por el equipo de PET/RM.

El análisis sobre los descriptores de una imagen puede ser de gran ayuda para recabar información complementaria, y que permita desarrollar herramientas de soporte para el diagnóstico, evaluación, toma de decisiones, entre otras aplicaciones dentro del ámbito de la salud.

3.6.1 DEFINICIÓN DE LA REGIÓN DE INTERÉS

Cuando se tienen las imágenes adquiridas por el equipo de PET/RM, es esencial para el estudio definir la región de interés (ROI). En este caso, la ROI para la extracción de las características de texturas debe delimitar la lesión presente en el tejido correspondiente al tumor de mama. Esto se debe, a que posteriormente, en el proceso de extracción de características, será necesario utilizar esta máscara para la obtención de las características de texturas sobre las diferentes imágenes de interés.

Para la obtención de la ROI, en el caso de los descriptores de texturas, se ha combinado métodos de segmentación semiautomáticos como *grow from seeds* o *smoothing* que implementa el software utilizado, junto a la delineación manual, para tratar de contrarrestar la falta de precisión de dichos métodos.

Por otro lado, en el caso de las características cuantitativas obtenidas en el software de Olea Sphere, no es necesario que la ROI se ajuste con precisión a la lesión mamaria, sino que es suficiente con delimitar de forma manual una ROI independientemente de su tamaño, pero asegurando que todos los píxeles internos pertenecen a la lesión.

3.6.2 EXTRACCIÓN DE CARACTERÍSTICAS DE TEXTURAS

Para la adquisición y evaluación de los descriptores de texturas existen diferentes métodos. Como bien se ha introducido con anterioridad, las características de texturas son aquellas que representan la disposición espacial de las distintas intensidades de gris dentro de la ROI. A continuación, se presentan diferentes alternativas para su obtención.

- Los métodos basados en modelos representan las texturas a partir de modelos matemáticos, ya sean generativos o estocásticos. Esta técnica no se emplea con frecuencia debido al alto coste computacional asociado a la estimación de los parámetros de dichos modelos, un proceso necesario para el análisis subsiguiente de texturas.
- Los métodos estructurales interpretan las texturas estructuras básicas que poseen una composición y un orden específicos. Este método busca definir las reglas que rigen las texturas a partir de la comprensión de su jerarquía.
- Los métodos de transformación muestran la imagen en un nuevo espacio, donde las coordenadas están relacionadas con las texturas de la imagen, utilizando transformadas como Fourier o Wavelet.
- Los métodos estadísticos, son aquellos que representan texturas en función de la distribución espacial de los diferentes niveles de grises presentes en la imagen y las relaciones entre estos.

Estos últimos, los métodos estadísticos, son los que se van a utilizar a lo largo de este estudio para la obtención de las características de texturas. Estos métodos se basan tanto en matrices como en histogramas, y depende a su vez del orden del estadístico.

Se diferencian, por lo tanto, los estadísticos de primer orden basados en el histograma, que, sí consideran los niveles de gris, pero no tienen en cuenta la relación que hay entre los píxeles. Y, en segundo lugar, existen los siguientes estadísticos que además son los que se emplean para obtener las características en el presente trabajo [24].

- El estadístico de primer orden que define la *Gray-Level Co-occurrence Matrix* (GLCM). Este estadístico determina la frecuencia con la que se relacionan distintos niveles de gris, para ello estudia un par de píxeles definiendo una dirección y una distancia ²⁴.

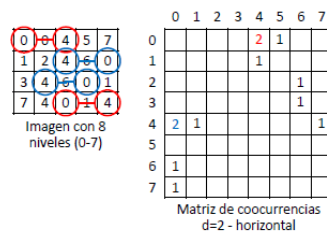


Figura 11. Cálculo de una GLCM para una dirección horizontal 2 para una matriz 4x5 con 8 niveles de gris.

- La *Gray-level Run-length Matrix* (GLRLM) es un estadístico de orden superior que analiza la interacción entre los píxeles de la ROI, para ello representa el número de píxeles consecutivos que contienen el mismo nivel de gris en una determinada dirección ²⁴.

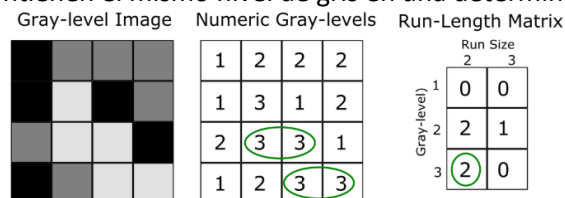


Figura 12. Cálculo de una GLRLM para una matriz 4x4 con 3 niveles de gris [24].

- La *Gray-level Size Zone Matrix* (GLSZM), es otro estadístico de orden superior que de la misma forma que el anterior analiza la interacción entre los píxeles, en este caso representa el número de píxeles consecutivos con el mismo nivel de gris, pero en a lo largo de una región ²⁴.

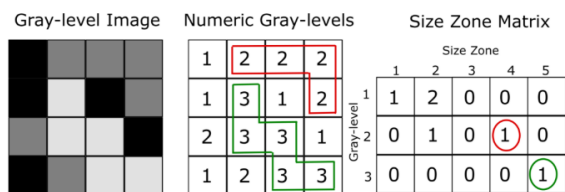


Figura 13. Cálculo de una GLSZM para una matriz 4x4 con 3 niveles de gris [24].

- Finalmente, el último estadístico de orden superior empleado en este estudio es el de la *Neighborhood Gray-tone Difference Matrix* (NGTDM). Este estadístico considera entre vecindarios de píxeles, y calcula la suma diferencia de niveles de gris de todos los pertenecientes a dicho vecindario respecto a un píxel determinado de este [24].

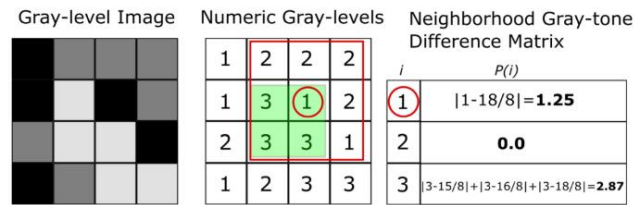


Figura 14. Cálculo de una NGTDM de una matriz 4x4 píxeles con 3 niveles de gris [24].

Todos estos estadísticos detallados nos permiten obtener de forma resumida un total de 40 correspondientes a las texturas para 4 niveles diferentes de gris, lo que conforma un total de 160 características a las que debemos añadir los 3 descriptores globales que también se extraen en este proceso.

CARACTERÍSTICAS DE TEXTURAS				
Global	GLCM	GLRLM	GLSZM	NGTDM
3	9	13	13	5

Tabla 3. Recuento de las características de texturas extraídas.

3.6.3 SELECCIÓN DE CARACTERÍSTICAS

Cuando se dispone del conjunto de características correspondientes a las texturas y a las obtenidas a partir del software de Olea Sphere, se procede con lo que se conoce como reducción de la dimensionalidad a partir de la selección de características. Este proceso consiste en escoger un subconjunto de la totalidad de características en función de su rendimiento predictivo en el modelo final. No es conveniente escoger todas las características ya que en general, en el ámbito sanitario se trabaja con bases de datos muy amplias, lo que su pone un alto coste computacional y, además, las variables no significativas pueden introducir ruido al modelo predictivo [24].

Por tanto, durante esta sección, se van a exponer algunos de los métodos más habituales para escoger un subconjunto de características dependiendo del valor de dichos descriptores para discriminar entre diferentes clases [24].

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

Los métodos *filters* son aquellos que seleccionan las variables que tienen un mejor ajuste respecto a una función que no está centrada en el acierto o error del modelo predictivo. Estos métodos devuelven un ranking de las variables en función de su importancia, siendo las más relevantes las situadas en las primeras posiciones. Algunas de las funciones más utilizadas son, *mutual information*, *Fisher score*, o *Relief* [24].

Por otro lado, los métodos *wrappers* utilizan algoritmos de aprendizaje y seleccionan las variables que alcanzan los mejores resultados predictivos a partir de una función determinada. Existen para este tipo de selección, dos métodos básicos, *forward selection* que es más eficiente a nivel computacional y por otro lado, la que se ha utilizado durante este estudio, eliminación *backward* que tiene un mayor potencia predictiva. En general, este método se caracteriza por obtener mejores resultados sobre el grupo específico de clasificación pero generaliza peor y tiene un ejecución más lenta.

3.6.4 MÉTODOS DE CLASIFICACIÓN

El análisis de características culmina con la clasificación, lo cual puede lograrse mediante métodos estadísticos simples o técnicas de aprendizaje automático. Los métodos de aprendizaje automático implican un proceso automatizado que aprende de los datos suministrados, con el objetivo de asistir en la toma de decisiones médicas. Este aprendizaje puede ser supervisado o no supervisado. En el aprendizaje supervisado, el sistema se entrena con datos específicos y luego intenta predecir los resultados de nuevos datos de prueba. En el aprendizaje no supervisado, el sistema no cuenta con datos de entrenamiento previos, solo con datos de entrada, y por lo tanto se adapta a estos datos para generar predicciones [25].

Este estudio se centrará en métodos de aprendizaje automático supervisado, ya que la división de datos permite disponer de conjuntos de entrenamiento adecuados para un aprendizaje efectivo, donde concretamente se realizará para un estudio binario.

A continuación, se describen algunos de los principales modelos de clasificación habitualmente utilizados, y que se estudiarán durante el Trabajo.

- Modelo ***K-Nearest Neighbors (KNN)***: este modelo se basa en la clasificación de las variables en función de las distancias a un número K de vecinos más cercanos. De esta forma, se le asigna al caso la clase más cercana como resultado de una ponderación de votos entre los K vecinos seleccionados. Para calcular las distancias existen varias fórmulas como la distancia Euclídea, la distancia de Manhattan, entre otras [26].
- Modelo ***Naive Bayes (NB)***: este modelo realiza la clasificación aplicando el Teorema de Bayes por lo que considera la probabilidad y presupone que las variables predictoras son independientes condicionalmente. Se basa en obtener una tabla de frecuencias a raíz de los datos, y en función de estas asignar la clase más probable para cada caso.
- Modelo ***Support Vector Machines (SVM)***: esta técnica de aprendizaje supervisado tiene como principal objetivo encontrar un hiperplano en un espacio de alta dimensión que separe las diferentes clases de datos de la manera más clara posible. Trata de separar las clases, pero también maximizar la distancia de los puntos más cercanos entre dos clases. Para esto hace uso de vectores de soporte y, si es necesario, transformando los datos a un espacio de mayor dimensión para manejar casos no lineales [26].

- Modelo **Bagged Tree**: es una técnica que implica entrenar múltiples árboles de decisión en diferentes subconjuntos del conjunto de datos y combinar sus predicciones para mejorar la precisión y la estabilidad del modelo final [26].
- Modelo **Boosted Tree**: consiste en entrenar secuencialmente varios árboles de decisión simples, donde cada árbol se enfoca en corregir los errores de los árboles anteriores. Al final, las predicciones de todos los árboles se combinan de manera ponderada para mejorar la precisión y robustez del modelo final.
- Modelo **Fine Tree**: consiste en el uso de árboles de decisión de gran profundidad, se basa en definir la estructura en la que cada nodo interno representa una "prueba" o condición sobre una característica, cada rama representa el resultado de la prueba, y cada nodo hoja representa una etiqueta de clase.
- Modelos de **Regresión logística**: modela la probabilidad de un resultado binario utilizando una combinación lineal de variables predictoras transformada por la función sigmoide.

3.6.5 VALIDACIÓN DEL MODELO

Los métodos de validación son estrategias utilizadas para estimar la capacidad predictiva de los modelos al aplicarse a nuevos datos, utilizando únicamente los datos de entrenamiento. La premisa básica de estos métodos es ajustarse usando un subconjunto de las observaciones del conjunto de entrenamiento y se evalúa con las observaciones restantes. Este procedimiento se repite varias veces y los resultados se combinan y promedian. Las repeticiones ayudan a compensar las posibles variaciones que pueden surgir debido a la asignación aleatoria de las observaciones. A continuación, se exponen dos de los principales métodos de validación [27].

- Validación **Hold Out**: es una técnica básica en la que se divide al azar el conjunto de datos en dos partes, un conjunto de entrenamiento y un conjunto de prueba. El modelo se entrena utilizando el conjunto de entrenamiento y se evalúa utilizando el conjunto de prueba. Esta evaluación proporciona una estimación de cómo se desempeñará el modelo en datos no vistos.
- **K-Fold Cross-Validation**: este método de validación divide el conjunto de datos en k subconjuntos de aproximadamente el mismo tamaño. El modelo se entrena k veces, utilizando cada uno de los k subconjuntos como conjunto de prueba una vez, y el resto de los datos como conjunto de entrenamiento. Posteriormente, se promedian los resultados de las k iteraciones para obtener una estimación final del rendimiento del modelo [27].

3.6.6 MÉTRICAS DE EVALUACIÓN

La evaluación objetiva de modelos predictivos es fundamental para garantizar que un modelo no solo se ajuste bien a los datos de entrenamiento, sino que también funcione eficazmente con datos nuevos. Existen diversas métricas y técnicas para evaluar la precisión, robustez y generalización de los modelos, en este estudio se van a implementar el área bajo la curva y la matriz de confusión.

- **Matriz de confusión:** es una tabla que compara las predicciones del modelo con las verdaderas etiquetas de clase y muestra la frecuencia de cada tipo de predicción. Se compone de cuatro componentes, número de casos positivos correctamente clasificados (TP), número de caso negativos clasificados incorrectamente como positivos (FP), número de casos negativos correctamente clasificados (TN) y número de positivos clasificados incorrectamente como negativos (FN). A partir de ella, se pueden establecer valores como, la precisión que es el valor de predicciones correctas total, la precisión que representa los verdaderos positivos entre el total de positivos predichos, la sensibilidad que es la proporción de verdaderos positivos sobre todos los positivos reales y la especificidad, proporción de verdaderos negativos sobre todos los negativos reales.

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Figura 15. Matriz de confusión [28].

- **Área Bajo la Curva (AUC):** es un valor numérico que representa el área bajo la Curva de Característica Operativa del Receptor (ROC). Este valor varía entre 0 y 1, donde un AUC de 1 indica un clasificador perfecto y un AUC de 0.5 sugiere un clasificador que no es mejor que el azar. La curva ROC hace referencia a una representación gráfica que ilustra el rendimiento de un modelo de clasificación y, se construye trazando la Tasa de Verdaderos Positivos (TPR) contra la Tasa de Falsos Positivos (FPR) para diferentes valores del umbral.

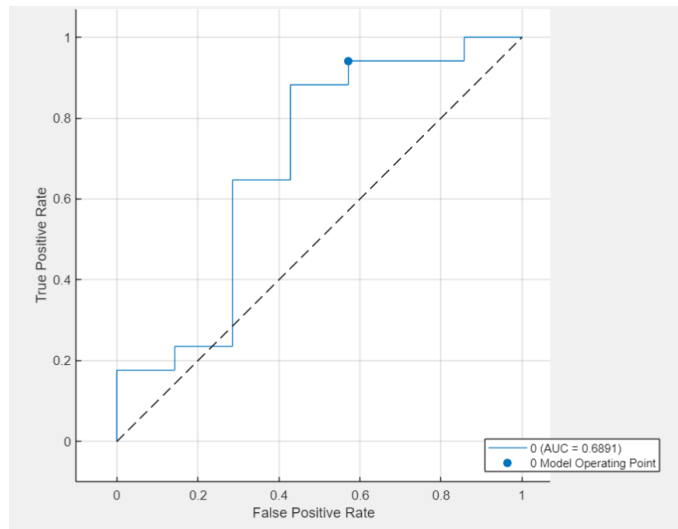


Figura 16. Curva ROC con el valor del AUC correspondiente. Fuente: elaboración propia.

4. MATERIALES

En este apartado se procede a exponer todos los materiales empleados para el desarrollo del presente Trabajo de Fin de Grado, desde las imágenes adquiridas por PET/RM proporcionadas y realizadas por el propio ASCIRES grupo biomédico hasta los diferentes softwares utilizados para el preprocesado de las imágenes y para realizar el modelo de aprendizaje con su correspondiente análisis estadístico.

4.1 IMÁGENES PET/RM

Todas las imágenes empleadas para el presente trabajo han sido obtenidas mediante el equipo híbrido de PET/RM de ASCIRES Grupo Biomédico. Concretamente se han utilizado imágenes de 30 pacientes, 29 mujeres y 1 hombre, que han sido tomadas desde junio de 2020 hasta mayo de 2022, este elenco de pacientes tiene una media de $51,56 \pm 12,92$ años. Dentro de todas las modalidades que se obtienen mediante el PET/RM, se han exportado las siguientes:

- SELECTIVA MAC
- SELECTIVO AXIAL *Cube* T2
- SELECTIVO Ax DWI 0-800
- ADC (mm^2/s)
- WATER: 3D Ax VIBRANT DINAMIC
- WATER: Ph2/3D Ax VIBRANT DINAMIC
- WATER: Ph3/3D Ax VIBRANT DINAMIC

4.2 SOFTWARES PARA EL PREPROCESADO DE IMÁGENES

Para poder trabajar con las imágenes previamente han tenido que ser preprocesadas, para ello, se han utilizado distintos softwares que seguidamente se explicarán.

1. **PACS CareStream** (*Carestream Health*, Rochester, Nueva York, Estados Unidos). Se han observado los diferentes casos en el software de *PACS CareStream*, estos casos posteriormente han sido descargados desde el mismo software en forma DICOM.
2. **RadiAnt DICOM Viewer** (*Medixant*, Poznan, Polonia) El software de *RadiAnt DICOM* permite visualizar las imágenes médicas en formato DICOM descargadas del PACS, este software se ha empleado como soporte para facilitar la segmentación.
3. **3D Slicer** (*Harvard Medical School*, Massachusetts, Estados Unidos). El software de 3D Slicer se ha utilizado para realizar la segmentación de las lesiones mamarias de las imágenes exportadas desde PACS. Este software permite exportar la segmentación en formato *Nifti* lo cual ha sido útil para su posterior uso en el software de *MATLAB*.
4. **MRICron** (*NITRC*, Estados Unidos). El software de MRICron se ha empleado para cambiar el formato de las imágenes DICOM descargadas del PACS a formato *Nifti* para su posterior uso en el software de *MATLAB*.

4.3 SOFTWARE DE MATLAB

El software de MATLAB R2023a (The MathWorks Inc., Natick, MA, USA) es un software matemático de programación propio. Durante el presente Trabajo Final de Grado, el software de MATLAB se ha utilizado para realizar la selección de las características más significativas, así como para entrenar y desarrollar el modelo de aprendizaje en base a las características previamente seleccionadas.

4.3.1 RADIOMICS 3D

Se ha empleado un paquete de herramientas de Martin Vallières con el objetivo de obtener las características correspondientes a las texturas en 3D, extrayendo como resultado un archivo Excel con 40 texturas para cada nivel de gris (16, 32, 64, 126) y las 3 características de texturas globales, es decir, un total de 163 características para cada uno de los 30 pacientes.

4.4 SOFTWARE DE OLEA SPHERE

El software de Olea Sphere (Olea Medical, La Ciotat, Francia) constituye un visor médico y analítico conectado con el software de PACS, lo que permite importar los pacientes para realizar los estudios requeridos. En el presente trabajo se ha utilizado Olea Sphere para obtener datos cuantitativos del estudio dinámico de perfusión T1 y de los mapas de difusión.

5. MÉTODOS

A lo largo de este apartado se busca exponer el proceso que ha permitido desarrollar el presente Trabajo de Fin de Grado con el que se ha obtenido un modelo predictivo mediante técnicas de aprendizaje como soporte al diagnóstico del estadio del cáncer de mama a partir del análisis de imágenes de PET/RM.

Este Trabajo se divide en dos principales bloques, el primero de ellos consta de la segmentación de las imágenes exportadas del PACS, concretamente la segmentación de la serie *WATER: Ph3/3D Ax VIBRANT DINAMIC*. Y, en segundo lugar, el desarrollo del modelo predictivo a partir del uso de los correspondientes filtros para la selección de características y la posterior elección del propio modelo de aprendizaje de clasificación. En este caso, el modelo predictivo que se realiza es un modelo de dos clases, en el que se estudia la presencia o no de metástasis.

5.1 ADQUISICIÓN DE LAS IMÁGENES

Para poder proceder con la obtención y análisis de las características es necesario adquirir previamente las imágenes para el estudio. Estas imágenes se obtienen por el equipo de PET/RM de ASCIRES Grupo Biomédico, como bien se ha explicado anteriormente. Estas imágenes se pueden obtener en múltiples modalidades en función de los parámetros de adquisición.

En el estudio que nos ocupa, se han utilizado las diferentes series obtenidas por el equipo de PET/RM, la serie correspondiente al coeficiente de difusión aparente (ADC) y la serie dinámica correspondiente al 3 minuto de la administración de contraste.

En cuanto a la obtención de la serie dinámica, se trata de un estudio potenciado en contraste. En este caso, se adquieren imágenes antes de la administración del contraste y durante los 6 minutos posteriores de la administración. Como se ha indicado, en este estudio la serie utilizada es la correspondiente al tercer minuto, "*WATER: Ph3/3D Ax VIBRANT DINAMIC*".

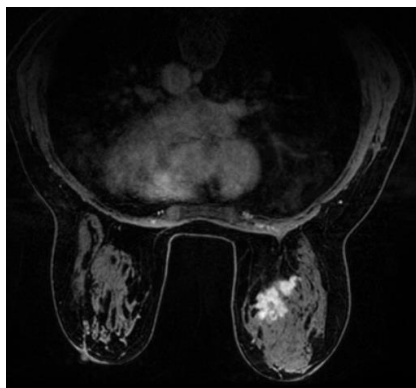


Figura 17. Imagen en PET/RM de cáncer de mama correspondiente a la serie "*WATER: Ph3/3D Ax VIBRANT DINAMIC*". Fuente: Imagen proporcionada por ASCIRES Grupo Biomédico.

5.2 SEGMENTACIÓN DE LA LESIÓN

Con las imágenes descargadas del PACS y correctamente anonimizadas, se debe definir la región de interés (ROI). En nuestro estudio, la ROI debe delimitar la lesión correspondiente al tumor de mama. Es un paso importante ya que a la hora de extraer a partir de los modelos matemáticos los descriptores de las texturas, se utiliza esta máscara tridimensional. Se llevó a cabo la obtención de la máscara de segmentación mediante el software de 3D Slicer. Como se ha comentado anteriormente, la segmentación se realizó sobre la serie *WATER: Ph3/3D Ax VIBRANT DINAMIC*, y de esta se obtiene la máscara en formato *Nifti* para su posterior implementación en MATLAB.

Una vez se carga la imagen en el entorno de 3D Slicer se procede a realizar la segmentación, la segmentación de todas las imágenes ha seguido el mismo procedimiento. Para ella se ha empleado la herramienta de segmentación *grow from seeds* (crecimiento por semillas), *smoothing* (suavizado), así otras herramientas más básicas como la delineación manual, recorte, entre otras. Dado que el proceso de segmentación a partir del crecimiento por semillas no es del todo preciso, se ha tenido que rectificar los errores en la segmentación semiautomática de forma manual. Con todas estas herramientas, se han obtenido las máscaras de segmentación para las lesiones mamarias de cada uno de los pacientes, como se puede observar en el ejemplo de la figura 18.

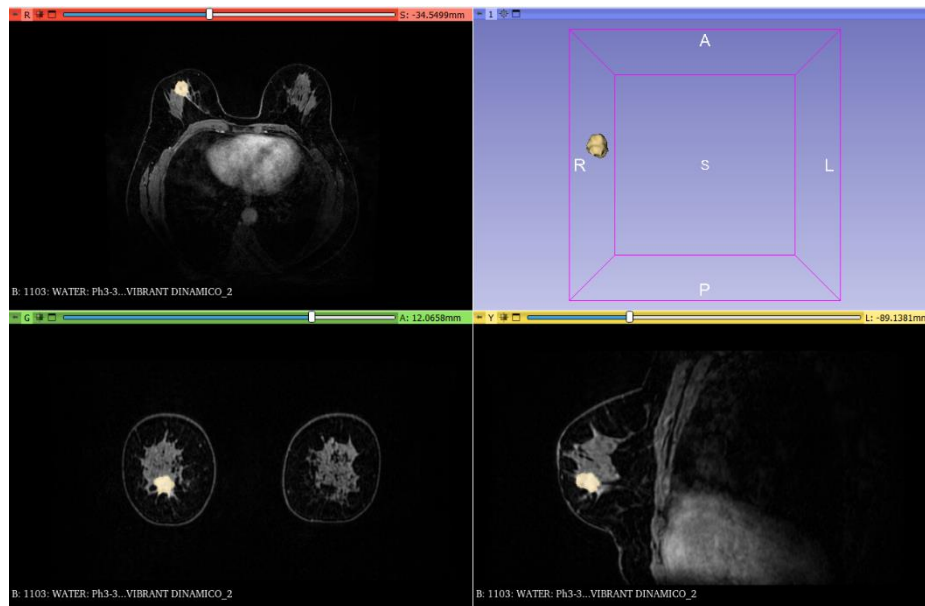


Figura 18. Segmentación de la lesión en mama derecha realizada con 3D Slicer. Fuente: elaboración propia.

5.3 ANÁLISIS DE LAS CARACTERÍSTICAS

Para poder alcanzar el objetivo del presente trabajo, es necesario obtener información interpretable de las imágenes médicas de forma precisa. Esta información se basa en una serie

de descriptores que se obtienen de diferentes formas y que engloban una información determinada.

Dentro de las características, en primer lugar, se obtendrán los descriptores de texturas. Estos consisten en un conjunto de métricas que representan la disposición espacial de las distintas intensidades de gris dentro de la región de interés (ROI), definida previamente en el proceso de la segmentación, y se obtendrán a partir de modelos matemáticos ²⁴.

En segundo lugar, se utilizarán características cuantitativas que se obtienen del estudio dinámico de perfusión T1 y de los mapas de difusión, para ello se emplea el software de *Olea-Sphere* que extrae de forma automática estos valores para una ROI definida de forma manual.

Y como se ha comentado al inicio, también se incluirán durante el estudio los valores de *SUVmax* y *SUVpico* presentes en el informe del radiólogo para cada caso.

5.3.1 EXTRACCIÓN DE LAS CARACTERÍSTICAS DE TEXTURAS

En este estudio, los descriptores de texturas se han obtenido a través de la herramienta de MATLAB, utilizando un paquete de herramientas de Martin Vallières que permite la extracción de esta información. Concretamente, la función empleada del paquete de herramientas es el de *Radiomics* 3D, donde se obtienen las relacionadas con los métodos estadísticos anteriormente mencionados; Global, GLCM, GLRLM, GLSZM y NGTDM.

Una vez ejecutada esta función, se genera un archivo .csv con las características correspondientes a la imagen seleccionada. Los inputs de la función son un archivo .mat que se ha generado a través de la imagen seleccionada y la máscara generada en el proceso de segmentación.

5.3.2 EXTRACCIÓN DE LAS CARACTERÍSTICAS DE OLEA

El segundo conjunto de características son las características que representan datos cuantitativos y que se obtienen en el apartado de *Permeability* sobre el estudio dinámico de perfusión T1 y de los mapas de difusión. De nuevo para poder obtener estos descriptores se debe definir una nueva ROI, en este caso no es necesario ajustarla precisa al contorno de la lesión, sino que simplemente con asegurarnos de que la ROI sólo contiene pixeles correspondientes a la zona de interés se puede considerar como una ROI válida.

Una vez definida la ROI sobre los estudios realizados se devuelven automáticamente los descriptores cuantitativos que posteriormente se exportan en formato .csv. Estos descriptores son los valores *Peak*, *Peak Enhancement*, *Wash In*, *Wash Out*, *Curve Wash Out*, AUC, TME y SER.

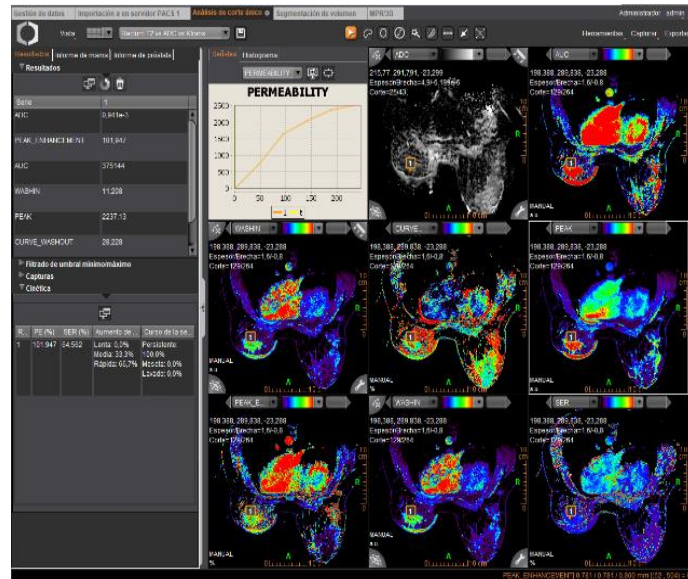


Figura 19. Interfaz Olea Sphere para la extracción de características. Fuente: elaboración propia.

5.4 SELECCIÓN DE CARACTERÍSTICAS

La selección de características es el proceso que nos permite generar un subconjunto dentro del total de descriptores en función del valor predictivo sobre las clases. En este estudio se han utilizado métodos de selección *filter*, concretamente el filtro *Relief* y el *Fisher Score* que devuelven un ranking ordenado en función de los pesos de los descriptores. Así como también se han empleado métodos de selección basados en *wrappers*.

Para aplicar cualquiera de los filtros sobre los descriptores anteriormente extraídos, es importante unificar todas las características de interés, es decir, las características de texturas, las obtenidas con *Olea Sphere*, y ciertos valores del informe del especialista como el SUV máx. y el SUV pico, que también ha sido introducidos en el estudio.

Finalmente, una vez tenemos nuestro archivo .csv con todo el conjunto de características, los métodos *filters* de *Fisher Score* y *Relief* devuelven el ranking de características en función de sus pesos predictivo, y se seleccionan de forma arbitraria un número concreto de dichas características de mayor a menor relevancia. Mientras que el método *Wrapper*, seleccionan las características que se ajustan a una función previamente definida. Es importante considerar que una mayor cantidad de características no conlleva un mejor resultado, ya que ciertas variables pueden incrementar el ruido y además añadir coste computacional, por tanto, se trata de buscar la cantidad óptima de características que forman el subconjunto para la posterior clasificación.

5.5 CLASIFICACIÓN

Para el proceso de clasificación se ha utilizado el apartado de MATLAB, *Classification Learner*, con el que una vez se han dividido los datos en entrenamiento y test, permite entrenar en base a los diferentes métodos de clasificación un modelo de aprendizaje. Como bien se ha indicado anteriormente los métodos empleados han sido *Naive Bayes*, *KNN*, *SVM*, *Bagged Tree*, *Boosted*

Tree, *Fine Tree* y regresión logística. Y para evaluarlos se han considerado las métricas de evaluación, AUC y matriz de confusión de la cual además se extraen los valores de exactitud, precisión, sensibilidad y especificidad.

6. RESULTADOS

Tal y como se ha indicado, el principal objetivo de este Trabajo Fin de Grado es esclarecer si con los descriptores que se pueden obtener a partir de un equipo de PET/RM, es decir, únicamente con una resonancia magnética y un PET simultáneo, se podría llegar a determinar o no la presencia de metástasis en el paciente. Esto permitiría desarrollar un sistema de soporte al diagnóstico del estadio en cáncer de mama. En caso de ser así, esto permitirá evitar al paciente enfrentar las numerosas y tediosas pruebas que se han nombrado anteriormente. Para ello se ha realizado un estudio de dos clases, en el que las etiquetas a predecir han sido la presencia o no de metástasis. Para realizar el presente estudio se han utilizado los datos de la Tabla 4.

PACIENTES SIN METÁSTASIS	PACIENTES CON METÁSTASIS
21	9

Tabla 4. Datos del estudio binario correspondiente a la presencia o no de metástasis. Fuente: elaboración propia.

Los diferentes estudios se han realizado empleando distintos métodos de selección de características y distintos sistemas de clasificación, además de variar el número de características seleccionadas para la clasificación. Posteriormente, todos ellos han sido evaluados con los valores del AUC, la matriz de confusión y su correspondiente sensibilidad y especificidad.

6.1 CLASIFICACIÓN BINARIA

En esta sección se van a presentar los resultados obtenidos para este estudio binario. Se van a diferenciar 3 estudios distintos en función del filtro de características utilizado, *Relief*, *Fisher Score* o *Wrappers*, y a su vez, para cada uno de ellos se analizarán los resultados según el método de clasificación y el número de características seleccionadas.

- Para el método de **selección de características *Wrappers***, se ha diseñado una función en la que se han aplicados tipos de discriminación que implementa MATLAB, la discriminación diagonal cuadrática (*DiagQuadratic*) y la diagonal lineal (*DiagLinear*). Este método no permite determinar el número de características seleccionadas si no que selecciona directamente las que mejor se ajustan a la función definida, en nuestro caso se escogen 5 características para la función discriminadora de diagonal cuadrática, y 8 para la de diagonal lineal. Por ello se obtienen los resultados de la Tabla 5.

Método de Clasificación	Métricas de Evaluación	<i>DiagLinear</i>	<i>DiagQuadratic</i>
<i>Fine Tree</i>	AUC	0,86	0,86
	SENSIBILIDAD	1	1
	ESPECIFICIDAD	0,82	0,82
<i>Naive Bayes</i>	AUC	0,82	0,80
	SENSIBILIDAD	0,71	0,57
	ESPECIFICIDAD	0,82	0,88
<i>SVM Gaussiano</i>	AUC	0,74	0,79
	SENSIBILIDAD	0	0
	ESPECIFICIDAD	1	1
<i>KNN</i>	AUC	0,87	0,77
	SENSIBILIDAD	0,85	0,71
	ESPECIFICIDAD	0,88	0,82
<i>Bagged Trees</i>	AUC	0,91	0,87
	SENSIBILIDAD	0,57	0,57
	ESPECIFICIDAD	0,82	0,88

Tabla 5. Resultados obtenidos en el estudio del conjunto de descriptores extraídos aplicando el método de selección de características de Wrappers para diferentes métodos de clasificación. Fuente: Elaboración propia.

- Para el método de **selección de características Fisher Score**, se ha realizado un estudio en el que de nuevo se varían los sistemas de clasificación y en el que ahora sí, se va a estudiar cómo cambian los resultados en función de las características seleccionadas. En este caso, se van a evaluar los resultados para 5, 10, 15, y 20 características seleccionadas. De esta forma se pueden observar los resultados en la Tabla 6.

Método de Clasificación	Métricas de Evaluación	5 características	10 características	15 características	20 características
<i>Fine Tree</i>	AUC	0,84	0,84	0,84	0,84
	SENSIBILIDAD	1	1	1	1
	ESPECIFICIDAD	0,82	0,82	0,82	0,82
<i>Naive Bayes</i>	AUC	0,84	0,81	0,79	0,82
	SENSIBILIDAD	0,43	0,43	0,43	0,43
	ESPECIFICIDAD	0,82	0,76	0,76	0,76
<i>SVM Gaussiano</i>	AUC	0,75	0,79	0,74	0,76
	SENSIBILIDAD	0,28	0,14	0,14	0,14
	ESPECIFICIDAD	1	1	1	1
<i>KNN</i>	AUC	0,85	0,85	0,85	0,93
	SENSIBILIDAD	0,71	0,71	0,71	0,86
	ESPECIFICIDAD	1	1	1	1
<i>Bagged Trees</i>	AUC	0,88	0,87	0,87	0,89
	SENSIBILIDAD	0,71	0,71	0,71	0,43
	ESPECIFICIDAD	0,88	0,82	0,82	0,88

Tabla 6. Resultados obtenidos en el estudio del conjunto de descriptores extraídos aplicando el método de selección de características de Fisher Score para diferentes métodos de clasificación y diferente número de características. Fuente: Elaboración propia.

- Para el método de **selección de características Relief**, se ha realizado un estudio en el que de nuevo se varían los sistemas de clasificación y en el que se va a estudiar cómo

modifican los resultados en función de las características seleccionadas. En este caso, se van a evaluar los resultados para 5, 10, 15, y 20 características seleccionadas. Así se pueden observar los resultados en la Tabla 7.

Método de Clasificación	Métricas de Evaluación	5 características	10 características	15 características	20 características
<i>Fine Tree</i>	AUC	0,35	0,37	0,49	0,49
	SENSIBILIDAD	0	0	0,14	0,14
	ESPECIFICIDAD	0,76	0,76	0,76	0,76
<i>Naive Bayes</i>	AUC	0,53	0,45	0,66	0,56
	SENSIBILIDAD	0,28	0,28	0,28	0,43
	ESPECIFICIDAD	0,76	0,82	0,82	0,82
<i>SVM Gaussiano</i>	AUC	0,63	0,65	0,57	0,56
	SENSIBILIDAD	0	0	0	0
	ESPECIFICIDAD	1	1	1	1
<i>KNN</i>	AUC	0,55	0,52	0,55	0,48
	SENSIBILIDAD	0,28	0,28	0,28	0,14
	ESPECIFICIDAD	0,82	0,76	0,82	0,82
<i>Bagged Trees</i>	AUC	0,70	0,45	0,51	0,46
	SENSIBILIDAD	0,14	0,14	0,14	0,14
	ESPECIFICIDAD	0,94	0,82	0,82	0,64

Tabla 7. Resultados obtenidos en el estudio del conjunto de descriptores extraídos aplicando el método de selección de características de Relief para diferentes métodos de clasificación y diferente número de características. Fuente: Elaboración propia

6.2 ANÁLISIS DEL ESTUDIO BINARIO

Para finalizar con el análisis de los resultados del proyecto, se exponen a continuación las matrices de confusión de los estudios marcados verdes en las Tablas 5, 6 y 7, estos estudios han sido escogidos por considerar los mejores modelos de clasificación.

En primer lugar, en la Figura 20, podemos observar como en la matriz de confusión obtenida en el modelo de entrenamiento de *Fine Tree* con selección de características por *Wrappers (DiagQuadratic)*, donde únicamente se clasifican erróneamente 3 de los 17 casos negativos mientras que todos los casos positivos se asignan correctamente.

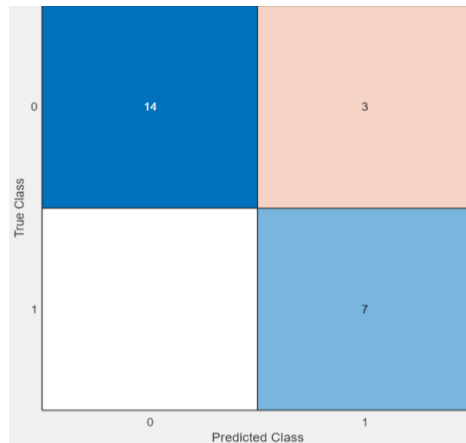


Figura 20. Matriz de confusión del método de clasificación *Fine Tree* con selección de características por *Wrapper (DiagQuadratic)*. Fuente: elaboración propia.

A continuación, se muestra en la Figura 21, la matriz de confusión correspondiente al modelo de clasificación basado en KNN, con selección de características por *Wrappers (DiagLinear)*. En este caso se clasifican de forma incorrecta 2 casos negativos, y uno positivo.

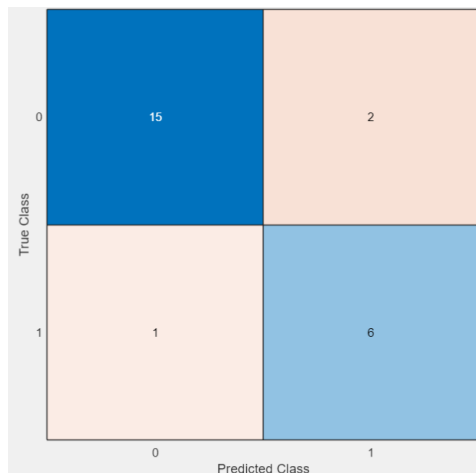


Figura 21. Matriz de confusión del método de clasificación *KNN* con selección de características por *Wrapper (DiagLinear)*. Fuente: elaboración propia.

En este caso, se observa en la Figura 22, la matriz de confusión del modelo de clasificación basado en *Fine Tree*, con selección de características por *Fisher Score* de 5 características seleccionadas. Se puede apreciar que se clasifican de forma incorrecta 3 casos negativos.

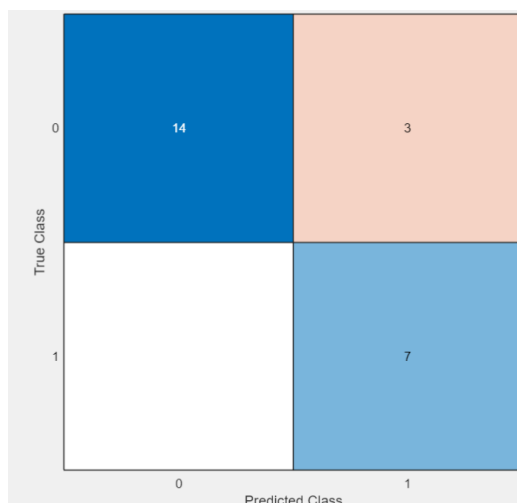


Figura 22. Matriz de confusión del método de clasificación Fine Tree con selección de características por Fisher Score de 5 características. Fuente: elaboración propia.

Por último, se aprecia en la Figura 23, la matriz de confusión del modelo de clasificación basado en KNN, con selección de características por Fisher Score de 20 características seleccionadas. Se puede apreciar que se clasifica de forma incorrecta un único caso positivo, siendo por tanto el modelo que presenta una mayor exactitud.

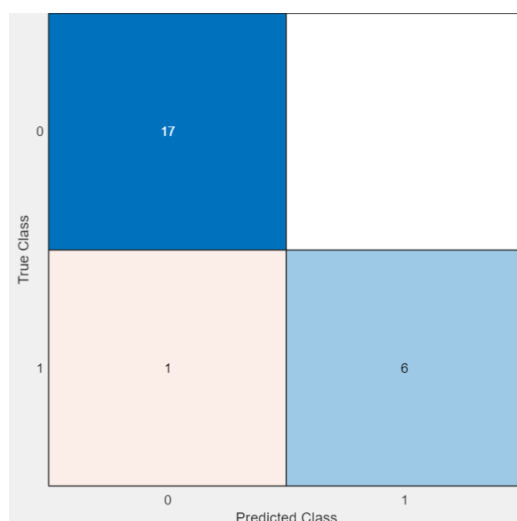


Figura 23. Matriz de confusión del método de clasificación KNN con selección de características por Fisher Score de 20 características. Fuente: elaboración propia.

6.3 ANÁLISIS DATOS DE PRUEBA

En esta sección, una vez se han exportado los modelos de clasificación entrenados en el apartado de *Classification Learner* de MATLAB, se procede a introducir en ellos los datos de prueba. Como se ha indicado, estos datos de prueba se obtuvieron de forma aleatoria mediante la partición de datos de *K Fold*, en este caso, el 5% de los casos se destinan a la prueba lo que son un total de 6 pacientes. A continuación, se indica la exactitud en la predicción de los modelos mostrados en verdes en las Tablas 5, 6 y 7, para los 6 casos de *test*.

En primer lugar, para el modelo de clasificación *Fine Tree* con selección de características por *Wrappers (DiagQuadratic)*, se consigue una exactitud de 66'66% para los de datos de prueba, concretamente se asignan de forma correcta 3 de los 4 casos negativos y de forma incorrecta los 1 de los 2 positivos.

Por otro lado, en cuanto al modelo de clasificación *KNN* con selección de características por *Wrappers (DiagLinear)*, se consigue una exactitud de 83'33% para los de datos de prueba, concretamente se asignan de forma incorrecta 1 de los 4 casos negativos y de forma correcta los 2 casos positivos.

Por último, tanto para el modelo de clasificación *KNN* con selección de características por Fisher Score de 5 y 20 características, se obtiene una exactitud de 66'66% para los de datos de test, concretamente se asignan de forma correcta todos los casos negativos y de forma incorrecta los dos positivos.

7. DISCUSIÓN

La discusión de los resultados obtenidos se centrará en lo obtenido para cada uno de los métodos de selección de características y de los modelos de clasificación empleados. En primer lugar, se comentará lo que se aprecia para cada uno de ellos por separado y a continuación se comparará.

En primer lugar, se puede apreciar que los modelos de clasificación basados en las características seleccionadas por el método *Wrappers* no presentan ninguno de ellos un valor de AUC por debajo del 77%, siendo el valor más elevado 91%.

En general se trata de valores muy elevados lo que hace pensar que efectivamente, tal y como se pretendía demostrar, los descriptores que se pueden extraer de las imágenes obtenidas por PET/RM se pueden vincular a la presencia o no de metástasis en el paciente. Las variaciones de los valores de AUC, sensibilidad y especificidad en los diferentes modelos a lo largo de estudio se puede deber, y es algo para tener en cuenta para futuras investigaciones, a que el número de pacientes con metástasis en el estudio es significativamente menor al de pacientes sin metástasis.

A continuación, en el caso de los modelos basados en las características seleccionadas por el filtro Fisher Score, se puede observar que el valor de AUC no desciende en ningún momento por debajo de 74%, siendo en este caso su valor máximo 93%.

De la misma forma, para la selección de características por el método *RelieF*, se observa una disminución evidente respecto a los dos métodos anteriores. En este caso, los valores de AUC oscilan entre 35% y 70%, esta diferencia respecto los otros dos métodos se puede deber a que la función definida para el método no se ajusta adecuadamente a los valores del estudio binario.

En segundo lugar, si consideramos todos los resultados obtenidos a través de los diferentes métodos de selección y clasificación, vemos que todos ellos tienen en común una especificidad bastante elevada, de hecho, todas están por encima del 75% siendo el máximo 100%. Esto quiere decir que los diferentes métodos presentados, en caso de indicar presencia de metástasis es muy probable que realmente esté presente. No obstante, a nivel general, los datos no son tan positivos en cuanto a la sensibilidad, pese a que hay modelos que tienen una sensibilidad del 100%, también hay modelos que no clasifican ninguno de los pacientes como positivos, obteniendo así una sensibilidad nula.

Por último, en cuanto a la discusión de los resultados obtenidos a partir de los métodos de clasificación exportados a MATLAB a los cuales se le han introducido los datos de prueba, se ratifican las ideas anteriores. Se tratan de test con una elevada especificidad, es decir, si el modelo clasifica como positivo, es muy probable que así sea. En los 4 análisis de prueba, únicamente dos casos negativos han sido clasificados como positivos. Mientras que, por el contrario, durante los 4 análisis, en dos de ellos, KNN con selección de características por Fisher Score de 5 y 20 características, se han catalogado de negativos los dos positivos, y en otro, *Fine Tree* con selección de características por *Wrappers (DiagQuadratic)*, se ha catalogado 1 de los 2 positivos como negativo.

Mientras que el modelo que mejores resultados ha proporcionado en relación con los datos de test es el modelo de clasificación *KNN* con selección de características por *Wrappers*

(DiagLinear). Únicamente clasifica de forma incorrecta uno del total de casos negativos (4) y de forma correcta todos los positivos (2).

Por tanto, para finalizar podemos concretar que aquellos modelos que presentan una elevada especificidad clasifican de forma positiva aquellos casos que con una alta probabilidad realmente sean positivos, como sería el modelo de clasificación basado en *KNN*, con selección de características por *Fisher Score* de 20 características seleccionadas. Mientras que aquellos con una elevada sensibilidad, clasifican como negativos aquellos casos que con elevada probabilidad lo serán realmente, como el método de clasificación *KNN* con selección de características por *Wrappers (DiagLinear)*.

La disparidad entre los diferentes resultados como se ha comentado anteriormente se puede deber a que en el conjunto de casos con el que se ha trabajado en este estudio, la cantidad de pacientes con metástasis es considerablemente menor a la de pacientes sin metástasis, por ello en general los métodos de clasificación tiene una elevada especificidad mientras que la sensibilidad es más oscilante.

8. CONCLUSIONES

Dada la elevada incidencia que presenta el cáncer de mama, resulta de vital importancia continuar trabajando e investigando en técnicas y procesos que permitan agilizar la detección, diagnóstico y tratamiento. Uno de los procesos más molesto, sobre todo de cara al paciente, es a la hora de estadificar la patología concluyendo la posible presencia o no de metástasis. Siendo este un proceso que incluye numerosas pruebas que vuelven la estadificación muy tediosa, teniendo que enfrentarse a las diferentes pruebas de imagen médica, así como a técnicas invasivas como la biopsia.

El presente Trabajo de Fin de Grado ha implementado un análisis binario basándose en los diferentes descriptores, concretamente las texturas y las características por el software de *Olea Sphere*, con el objetivo de poder determinar la presencia o no de metástasis en 30 pacientes con cáncer de mama. Para ello, se han desarrollado los correspondientes métodos de clasificación en base a diferentes características seleccionadas en función de su peso para explicar la clasificación.

Para ello se ha empleado imágenes correspondientes a 30 pacientes donde 9 de ellas presentan metástasis, estas imágenes han sido preprocesadas y segmentadas obteniendo así la región de interés (ROI) que se ha analizado para obtener los descriptores. Una vez se tiene definida la ROI se extraen los descriptores haciendo uso del software de *Olea Sphere* y de MATLAB. Finalmente se han desarrollado en MATLAB tres filtros distintos para la clasificación de características y con ellas, se han entrenado diferentes métodos de clasificación (*Naive Bayes*, *SVM*, *KNN*, *Fine Tree*, *Bagged Tree*, *Boosted Tree* y *Regresión Logística*).

Con todo esto se han obtenido resultados que hacen indicar que los descriptores que se obtienen de las imágenes PET/RM en pacientes con cáncer de mama si están correlacionados con la presencia o no de metástasis. Concretamente, de todos los métodos de clasificación utilizados se ha obtenido un AUC máximo de 93% para la selección de características por *Fisher Score*, mientras que utilizando *Wrappers* se ha obtenido un 91% como máximo, y un 70% utilizando *Relief*.

Finalmente, cuando se han aplicado los mejores métodos de clasificación obtenidos sobre los datos de prueba, se ha visto que, pese a tener un conjunto de pacientes de entrenamiento (24) relativamente pequeño, en las predicciones se han alcanzado hasta un 83% de exactitud.

9. LÍNEAS FUTURAS

Una vez finalizado el estudio, resulta también esencial proponer posibles vías de interés y mejora para continuar con la investigación correspondiente al presente trabajo.

Dados los resultados obtenidos, sería de gran interés aumentar el tamaño de los datos disponibles pudiendo hacer así un estudio con un gran número de pacientes de ambas clases permitiendo así determinar si realmente se puede obtener un gran valor predictivo a través de los modelos de aprendizaje.

Así mismo, viendo que los descriptores obtenidos sí parecen tener valor predictivo en la presencia o no de metástasis y, por lo tanto, valor predictivo para diferencia entre estadio I-III y estadio IV respecto a la escala de estadificación TNM, es de interés investigar si estas características pueden tener valor predictivo para clasificar entre los cuatro estadios diferentes. Poder diseñar una herramienta de soporte al estadio final del cáncer de mama sería un gran avance ya que ahorraría muchas pruebas a los pacientes.

10. BIBLIOGRAFÍA

- [1] *¿Qué es el cáncer de mama?* (2023). <https://www.contraelcancer.es/es/todo-sobre-cancer/tipos-cancer/cancer-mama/que-es-cancer-mama>. (acceso 2024-05-22).
- [2] Martín, M., Herrero, A., & Echavarría, I. (2015). *El cáncer de mama*. *Arbor*, 191(773), a234. <https://doi.org/10.3989/arbor.2015.773n3004>
- [3] *Exámenes de detección del cáncer de seno (mama)*. (2023). Cancer.gov. <https://www.cancer.gov/espanol/tipos/seno/paciente/deteccion-seno-pdq>. (acceso 2024-05-22).
- [4] Amy D., Amorós Oliveros F.J., Teboul M. (2002). *Anatomía mamaria y su correlación ecográfica*. <https://semg.info/mgyf/medicinageneral/mayo2002/369-372.pdf>
- [5] Cantero, R. (2022). *Anatomía de la mama*. Raimundo Cantero. <https://raimundocantero.es/blog/anatomia-de-la-mama/>. (acceso 2024-05-22).
- [6] Bruant-Rodier C., Bodin F. (2011). *Mamoplastias para el tratamiento de hipertrofia y ptosis (I). Principios generales*, EMC - Cirugía Plástica Reparadora y Estética, Volumen 19, Problema 1, 2011, Páginas 1-17. [https://doi.org/10.1016/S1634-2143\(11\)70983-9](https://doi.org/10.1016/S1634-2143(11)70983-9).
- [7] "Frecuencia de cáncer de mama por zonas." <http://alvarezrad.mx/web/index.php/blog/12-frecuencia-de-cancer-de-mama-por-zonas>. (acceso 2024-05-22).
- [8] World Health Organization: WHO & World Health Organization: WHO. (2024). *Cáncer de mama*. <https://www.who.int/es/news-room/fact-sheets/detail/breast-cancer>. (acceso 2024-05-22).
- [9] Osorio BN, Bello HC, Vega BL. (2020). *Factores de riesgo asociados al cáncer de mama*. <https://www.medigraphic.com/pdfs/revcubmedgenint/cmi-2020/cmi202i.pdf>
- [10] Villarreal-Ríos, E., Escorcía-Reyes, V., Martínez-González, L., Vargas-Daza, E. R., Galicia-Rodríguez, L., Cervantes-Becerra, R., & López-Ramos, J. M. (2014). *Historia natural del proceso diagnóstico del cáncer de mama*. *Revista Panamericana de Salud Pública*, 35, 172-178. <https://www.scielosp.org/pdf/rpsp/2014.v35n3/172-178/es>
- [11] Abellán Lucas, M. (2021). *Variación del tratamiento quirúrgico del cáncer de mama tras la introducción de la resonancia magnética en la estadificación preoperatoria*. <http://hdl.handle.net/10803/673204>
- [12] Ramos, J., Benaissa, L., Ahumada, G., & Rivas, F. (2012). *Distribución por estadios del cáncer de mama en el Hospital Universitario de Neiva, 2011*. *RFS Revista Facultad de Salud*, 4(1), 51-59. <https://core.ac.uk/download/pdf/229704768.pdf>
- [13] Arceo-Martínez, M. T., López-Meza, J. E., Ochoa-Zarzosa, A., & Palomera-Sánchez, Z. (2021). *Estado actual del cáncer de mama en México: principales tipos y factores de riesgo*. *Gaceta mexicana de oncología*, 20(3), 101-110. <https://www.scielo.org.mx/pdf/gamo/v20n3/2565-005X-gamo-20-3-101.pdf>
- [14] *UICC and the TNM Classification of Malignant Tumours | UICC*. (2023). UICC. <https://www.uicc.org/who-we-are/about-uicc/uicc-and-tnm>. (acceso 2024-05-22).
- [15] Daimiel Naranjo, I. (2022). *Estrategias de mejora en el diagnóstico del cáncer de mama por imagen de resonancia magnética: Avances en la secuencia potenciada en difusión*. <https://docta.ucm.es/rest/api/core/bitstreams/b6969ecf-d603-4116-9e1d-af1e32634397/content>
- [16] Castellote Huguet, M. P. (2021). *Desarrollo de una metodología para la ayuda al diagnóstico del cáncer de mama mediante el análisis de texturas en imágenes de resonancia magnética*. Universitat Politècnica de València. <https://riunet.upv.es/handle/10251/174018>

- [17] *Información sobre la imagen por resonancia magnética (IRM) de mama.* (2023). Memorial Sloan Kettering Cancer Center. <https://www.mskcc.org/es/cancer-care/patient-education/breast-magnetic-resonance-imaging-mri-breast>
- [18] Zornoza, A., M-Regueira, F., Rotellar, F., R-Spiteri, N., & Zornoza, G. (2001). *Valoración de la tomografía por emisión de positrones (PET-18FDG) en la estadificación del cáncer de mama.* *Rev Senología Pat Mam*, 14, 135-44. https://www.sespm.es/wp-content/uploads/revista/2001_14_4/2.pdf
- [19] Ladrón De Guevara H., D.; Catalán V., P.; Hernández M., C.; Pablo Zhindon P., J.; Ladrón De Guevara H., D.; Catalán V., P.; Hernández M., C.; Pablo Zhindon P., J. (2019). *Valor Del PET/CT Semicuantitativo (SUVmax) En El Estudio de Nódulo Pulmonar Sólido.* *Revista chilena de enfermedades respiratorias.* <https://doi.org/10.4067/S0717-73482019000200116>.
- [20] *PET/RM – Cetir (Grupo ASCIRES) – Expertos en Diagnóstico por la Imagen y Medicina Nuclear.* (2024). <https://www.cetir.com/especialidades/pet-rm/> (acceso 2024-05-22).
- [21] Barrio, P.; López-Melgar, B.; Fidalgo, A.; Romero-Castro, M. J.; Moreno-Arciniegas, A.; Field, C.; Garcerant, M.; Anmad Shihadeh, L.; Díaz-Antón, B.; Ruiz de Aguiar, S.; García Cañamaque, L.; Fernández-Friera, L. (2021). *Valor adicional de la tecnología híbrida de PET/RM frente a la RM y la PET en la enfermedad cardiovascular.* *Rev Esp Cardiol*, 74 (4), 303–311. <https://doi.org/10.1016/j.recesp.2020.06.023>.
- [22] Riola-Parada, C., García-Cañamaque, L., Pérez-Dueñas, V., Garcerant-Tafur, M., & Carreras-Delgado, J. L. (2016). *PET/RM simultánea vs. PET/TC en oncología. Una revisión sistemática.* *Revista Española de Medicina Nuclear e Imagen Molecular*, 35(5), 306-312. https://www.sciencedirect.com/science/article/pii/S2253654X16300440?casa_token=6fgt-pAHdYD8AAAAA:7l-0FjrpWuCKo8Vzdz6ivll9rjiuacVunwNgYDIlljyUAcVAUWxMsZ2-WgumFzqOmw9u2jG9QRBoy
- [23] Mejía Bayas, T. M. (2015). *Segmentación y análisis de imágenes termográficas mediante descriptores de textura simples* (Bachelor's thesis, Universidad Técnica de Ambato. Facultad de Ingeniería en Sistemas, Electrónica e Industrial. Carrera de Ingeniería en Electrónica y Comunicaciones). https://repositorio.uta.edu.ec/bitstream/123456789/10383/1/Tesis_t979ec.pdf
- [24] Santacruz, A. M. L. (2017). *Diagnosis and prognosis of cardiovascular diseases by means of texture analysis in magnetic resonance imaging* (Doctoral dissertation, Tesis doct. Universitat de València. <https://core.ac.uk/download/pdf/93038147.pdf>
- [25] Sáez Silvestre C. (2024). *Sistemas de Información y Telemedicina, Tema 3: Sistemas de Ayuda a la Decisión Basados en Datos Biomédicos: Métodos de Aprendizaje.* https://poliformat.upv.es/access/content/group/GRA_13068_2023/SIT1-Tema%203.5.pdf (acceso 2024-05-21).
- [26] González Vilanova, A. (2019). *Métodos de machine learning en estudios biomédicos* (Doctoral dissertation, Universitat Politècnica de València). <https://riunet.upv.es/bitstream/handle/10251/127574/Gonz%C3%A1lez%20-%20M%C3%A9todos%20de%20machine%20learning%20en%20estudios%20biom%C3%A9dicos.pdf?sequence=4>
- [27] Rodrigo, J. A. (2020). *Validación de modelos predictivos (machine learning): Cross-validation, OneLeaveOut, Bootstrapping.* https://cienciadedatos.net/documentos/30_cross-validation_oneleaveout_bootstrap
- [28] Izco, F. (2018). *Base de datos corporativa de personas.* https://bookdown.org/f_izco/BDC-POC/metricas.html. (acceso 2024-05-22).

11. ANEXOS

11.1 ANEXO I. RELACIÓN DEL TRABAJO CON LOS OBJETIVOS DE DESARROLLO SOSTENIBLE DE LA AGENDA 2030

En el siguiente apartado se propone el grado de relación con los Objetivos de Desarrollo Sostenible de la agenda 2030 con los que está comprometido el presente Trabajo de Fin de Grado.

Objetivos de Desarrollo Sostenibles	Alto	Medio	Bajo	No Procede
ODS 1. Fin de la pobreza.				✗
ODS 2. Hambre cero.				✗
ODS 3. Salud y bienestar.	✗			
ODS 4. Educación de calidad.				✗
ODS 5. Igualdad de género				✗
ODS 6. Agua limpia y saneamiento.				✗
ODS 7. Energía asequible y no contaminante				✗
ODS 8. Trabajo decente y crecimiento económico.				✗
ODS 9. Industria, innovación e infraestructuras.	✗			
ODS 10. Reducción de las desigualdades.				✗
ODS 11. Ciudades y comunidades sostenibles.				✗
ODS 12. Producción y consumo responsables.				✗
ODS 13. Acción por el clima.				✗
ODS 14. Vida submarina.				✗
ODS 15. Vida de ecosistemas terrestres				✗
ODS 16. Paz, justicia e instituciones sólidas.				✗
ODS 17. Alianzas para lograr objetivos				✗

PRESUPUESTO

Desarrollo de un modelo predictivo mediante técnicas de aprendizaje máquina como soporte al diagnóstico del estadio del cáncer de mama a partir del análisis de imágenes de PET/RM

AUTOR: Ignacio Iborra Roncales

TUTOR/A: David Moratal Pérez, Universitat Politècnica de València

COTUTORES: Juan Antonio Romero, ASCIRES Grupo Biomédico

CURSO ACADÉMICO: 2023/2024

ÍNDICE PRESUPUESTO

1. INTRODUCCIÓN.....	53
2. CUADRO DE PRECIOS DE MANO DE OBRA.....	53
3. CUADRO DE PRECIOS DE MATERIALES.....	53
4. CUADRO DE PRECIOS UNITARIOS.....	54
5. CUADRO DE PRECIOS DESCOMPUESTOS.....	55
6. CUADROS DE PRECIOS PARCIALES Y MEDICIONES.....	57
7. PRESUPUESTO DE EJECUCIÓN POR CONTRATA.....	58

1. INTRODUCCIÓN

El objetivo del documento del Presupuesto es obtener una estimación aproximada sobre la inversión económica que es requerida para llevar a cabo el desarrollo y la ejecución de este Trabajo de Fin de Grado. Para obtener este elaborar el documento se ha utilizado el programa de Arquímedes, y, en los siguientes apartados, se describen las distintas secciones que dan forma al Presupuesto, en este caso, destinado al desarrollo de un modelo predictivo mediante técnicas de aprendizaje máquina como soporte al diagnóstico del estadio del cáncer de mama a partir del análisis de imágenes de PET/RM.

2. CUADRO DE PRECIOS DE MANO DE OBRA

En esta sección se describe la mano de obra necesaria para elaborar el proyecto y, a su vez, el coste aproximado de esta. En este caso se ha considerado la mano de obra de un ingeniero biomédico como tutor y otro como cotutor, el estudiante de ingeniería biomédica como autor del proyecto, y el radiólogo.

Para el cálculo estimado de cada mano de obra se han considerado los siguientes salarios base anuales para los distintos integrantes. En los casos tanto del ingeniero tutor como cotutor se han asignado 55.000 euros anuales y en el caso del radiólogo 60.000 euros anuales. Mientras que, para el ingeniero biomédico estudiante, considerado junior, se ha definido un salario de 18.000 euros anuales.

A su vez, se ha considerado una jornada laboral de 8 horas diarias, de forma que las horas laborales efectivas, descontando fines de semana, días hábiles de vacaciones y festivos, acumulan un total de 1792 horas. De forma que el precio de mano de obra se adquiere dividiendo el salario base anual entre el cúmulo de horas trabajadas, obteniendo la siguiente Tabla.

Núm.	Código	Denominación de la mano de obra	Precio	Horas	Total
1	MO.RAD	Radiólogo	33,48	1,00 h	33,48
2	MO.TU	Ingeniero Biomédico Tutor	30,69	99,00 h	3.038,31
3	MO.CO	Ingeniero Biomédico Cotutor	30,69	37,50 h	1.151,01
4	MO.IBJ	Ingeniero Biomédico Junior	10,00	324,00 h	3.240,00
				Total mano de obra:	7.462,80

3. CUADRO DE PRECIOS DE MATERIALES

A continuación, el Cuadro de Precios de Material, refleja de forma aproximada el coste que suponen todos los materiales que han sido utilizados a lo largo de la elaboración del proyecto. Se han incluido los diferentes softwares, *PACS CareStream*, *Olea Sphere*, *3D Slicer*, *Radiant Dicom Viewer*, *Mricron*, *MATLAB*, así como los diferentes dispositivos utilizados que han sido el ordenador proporcionado por ASCIRES Grupo Biomédico y un ordenador de uso personal.

Núm.	Código	Denominación del material	Precio	Cantidad	Total
1	MAT.MATLAB	Licencia MATLAB R2023b	35,00	43,00 h	1.505,00
2	MAT.OLEA	Software Olea Sphere	5,00	50,00 h	250,00
3	MAT.Radiant	Software Radiant Dicom Viewer	5,00	10,00 h	50,00
4	MAT.PCp	Ordenador Portátil Personal	1,90	281,00 h	533,90
5	MAT.PCA	Ordenador Ascires Grupo Biomédico	1,50	92,00 h	138,00
6	MAT.Anom	Software Dicom Anonymizer	1,00	40,00 h	40,00
7	MAT.3DSlic...	Software 3D Slicer	1,00	50,00 h	50,00
8	MAT.Windows	Licencia Windows 10	0,90	223,00 h	200,70
9	MAT.Office	Licencia de Office 365	0,90	180,00 h	162,00
10	MAT.PACS	Software PACS CareStream	0,15	40,00 h	6,00
11	MAT.Mri	Software Mricron	0,10	10,00 h	1,00
				Total materiales:	2.936,60

4. CUADRO DE PRECIOS UNITARIOS

En esta sección se describe cada unidad y su respectivo importe, expresado tanto en cifras como en letras. Los precios unitarios provienen de los precios descompuestos que se presentan en el apartado posterior.

Cuadro de precios nº 1			
Nº	Designación	Importe	
		En cifra (Euros)	En letra (Euros)
1.1	1 Definición del proyecto h Reunión inicial Tutor y Cotutor para definir el TFG	109,05	CIENTO NUEVE EUROS CON CINCO CÉNTIMOS
1.2	h Planificación del TFG	74,24	SETENTA Y CUATRO EUROS CON VEINTICUATRO CÉNTIMOS
1.3	h Revisiones Periódicas del TFG	76,21	SETENTA Y SEIS EUROS CON VEINTIUN CÉNTIMOS
2.1	2 Obtención y preparado de las imágenes h Aprendizaje e instalación de softwares	17,13	DIECISIETE EUROS CON TRECE CÉNTIMOS
2.2	h Descarga y anonimización de las imágenes	17,28	DIECISIETE EUROS CON VEINTIOCHO CÉNTIMOS
2.3	h Segmentación de las imágenes	25,37	VEINTICINCO EUROS CON TREINTA Y SIETE CÉNTIMOS
3.1	3 Desarrollo del modelo predictivo mediante técnicas de aprendizaje máquina h Obtener de los archivos .mat	52,90	CINCUENTA Y DOS EUROS CON NOVENTA CÉNTIMOS
3.2	h Extracción y selección de las características	52,90	CINCUENTA Y DOS EUROS CON NOVENTA CÉNTIMOS
3.3	h Obtención y evaluación del método de clasificación	52,90	CINCUENTA Y DOS EUROS CON NOVENTA CÉNTIMOS
4.1	4 Redacción de la memoria del TFG h Redacción de los documentos	13,31	TRECE EUROS CON TREINTA Y UN CÉNTIMOS
4.2	h Revisión de los documentos y corrección de errores	78,08	SETENTA Y OCHO EUROS CON OCHO CÉNTIMOS
4.3	h Preparación defensa del TFG	46,17	CUARENTA Y SEIS EUROS CON DIECISIETE CÉNTIMOS
Valencia, Mayo 2024 Ingeniero Biomédico			

5. CUADRO DE PRECIOS DESCOMPUESTOS

En este apartado se muestran los distintos capítulos que dan lugar al Trabajo Fin de Grado y se resume el proceso correspondiente a cada uno de ellos. Se detalla además cada uno de los capítulos en función de las unidades de obra, materiales, y tiempos requeridos por cada uno de ellos.

- Capítulo 1: Definición del proyecto
- Capítulo 2: Obtención y preparado de las imágenes
- Capítulo 3: Desarrollo del modelo predictivo mediante técnicas de aprendizaje máquina
- Capítulo 4: Redacción de la memoria del TFG

Cuadro de Precios Descompuestos					
Nº	Código	Ud	Descripción		Total
1 Definición del proyecto					
1.1	01.01	h	Reunión inicial Tutor y Cotutor para definir el TFG		
	MO.TU	1,00 h	Ingeniero Tutor	30,69	30,69
	MO.CO	1,00 h	Ingeniero Cotutor	30,69	30,69
	MO.IBJ	1,00 h	Ingeniero Biomédico Junior	10,00	10,00
	MO.RAD	1,00 h	Radiologo	33,48	33,48
		4,00 %	Costes indirectos	104,86	4,19
			Precio total por h		109,05
					Son ciento nueve Euros con cinco céntimos
1.2	01.02	h	Planificación del TFG		
	MO.CO	1,00 h	Ingeniero Cotutor	30,69	30,69
	MO.TU	1,00 h	Ingeniero Tutor	30,69	30,69
	MO.IBJ	1,00 h	Ingeniero Biomédico Junior	10,00	10,00
		4,00 %	Costes indirectos	71,38	2,86
			Precio total por h		74,24
					Son setenta y cuatro Euros con veinticuatro céntimos
1.3	01.03	h	Revisiones Periódicas del TFG		
	MO.CO	1,00 h	Ingeniero Cotutor	30,69	30,69
	MO.IBJ	1,00 h	Ingeniero Biomédico Junior	10,00	10,00
	MO.TU	1,00 h	Ingeniero Tutor	30,69	30,69
	MAT.PCp	1,00 h	Ordenador Portátil Personal	1,90	1,90
		4,00 %	Costes indirectos	73,28	2,93
			Precio total por h		76,21
					Son setenta y seis Euros con veintiun céntimos

Nº	Código	Ud	Descripción	Total
2 Obtención y preparado de las imágenes				
2.1	02.01	h	Aprendizaje e instalación de softwares	
	MO.IBJ	1,00 h	Ingeniero Biomédico Junior	10,00
	MO.CO	0,10 h	Ingeniero Cotutor	30,69
	MAT.PCp	1,00 h	Ordenador Portátil Personal	1,90
	MAT.PCA	1,00 h	Ordenador Ascires Grupo Biomédico	1,50
		4,00 %	Costes indirectos	16,47

Precio total por h 17,13

Son diecisiete Euros con trece céntimos

2.2	02.02	h	Descarga y anonimización de las imágenes	
	MO.CO	0,10 h	Ingeniero Cotutor	30,69
	MO.IBJ	1,00 h	Ingeniero Biomédico Junior	10,00
	MAT.PCA	1,00 h	Ordenador Ascires Grupo Biomédico	1,50
	MAT.Windows	1,00 h	Licencia Windows 10	0,90
	MAT.PACS	1,00 h	Software PACS CareStream	0,15
	MAT.Anom	1,00 h	Software Dicom Anonymizer	1,00
		4,00 %	Costes indirectos	16,62

Precio total por h 17,28

Son diecisiete Euros con veintiocho céntimos

2.3	02.03	h	Segmentación de las imágenes	
	MO.IBJ	1,00 h	Ingeniero Biomédico Junior	10,00
	MAT.PCp	1,00 h	Ordenador Portátil Personal	1,90
	MAT.Radiant	0,20 h	Software Radiant Dicom Viewer	5,00
	MAT.3DSlicer	1,00 h	Software 3D Slicer	1,00
	MAT.Windows	1,00 h	Licencia Windows 10	0,90
	MAT.Mri	0,20 h	Software Mricron	0,10
	MAT.OLEA	1,00 h	Software Olea Sphere	5,00
	MAT.PCA	1,00 h	Ordenador Ascires Grupo Biomédico	1,50
	MO.CO	0,10 h	Ingeniero Cotutor	30,69
		4,00 %	Costes indirectos	24,39

Precio total por h 25,37

Son veinticinco Euros con treinta y siete céntimos

Nº	Código	Ud	Descripción	Total
----	--------	----	-------------	-------

3 Desarrollo del modelo predictivo mediante técnicas de ...

3.1	03.01	h	Obtener de los archivos .mat	
	MO.IBJ	1,00 h	Ingeniero Biomédico Junior	10,00
	MAT.Windows	1,00 h	Licencia Windows 10	0,90
	MAT.MATLAB	1,00 h	Licencia MATLAB R2023b	35,00
	MAT.PCp	1,00 h	Ordenador Portátil Personal	1,90
	MO.CO	0,10 h	Ingeniero Cotutor	30,69
		4,00 %	Costes indirectos	50,87

Precio total por h 52,90

Son cincuenta y dos Euros con noventa céntimos

3.2	03.02	h	Extraccion y selección de las características	
	MO.IBJ	1,00 h	Ingeniero Biomédico Junior	10,00
	MAT.PCp	1,00 h	Ordenador Portátil Personal	1,90
	MAT.Windows	1,00 h	Licencia Windows 10	0,90
	MAT.MATLAB	1,00 h	Licencia MATLAB R2023b	35,00
	MO.CO	0,10 h	Ingeniero Cotutor	30,69
		4,00 %	Costes indirectos	50,87

Precio total por h 52,90

Son cincuenta y dos Euros con noventa céntimos

3.3	03.03	h	Obtencion y evaluacion del método de clasificación	
	MO.IBJ	1,00 h	Ingeniero Biomédico Junior	10,00
	MAT.PCp	1,00 h	Ordenador Portátil Personal	1,90
	MAT.MATLAB	1,00 h	Licencia MATLAB R2023b	35,00
	MAT.Windows	1,00 h	Licencia Windows 10	0,90
	MO.CO	0,10 h	Ingeniero Cotutor	30,69
		4,00 %	Costes indirectos	50,87

Precio total por h 52,90

Son cincuenta y dos Euros con noventa céntimos

Nº	Código	Ud	Descripción		Total	
4 Redacción de la memoria del TFG						
4.1	04.01	h	Redacción de los documentos			
		MO.IBJ	1,00 h	Ingeniero Biomédico Junior	10,00	10,00
		MAT.Office	1,00 h	Licencia de Office 365	0,90	0,90
		MAT.PCp	1,00 h	Ordenador Portátil Personal	1,90	1,90
			4,00 %	Costes indirectos	12,80	0,51
				Precio total por h		13,31
Son trece Euros con treinta y un céntimos						
4.2	04.02	h	Revisión de los documentos y corrección de errores			
		MO.TU	1,00 h	Ingeniero Tutor	30,69	30,69
		MO.CO	1,00 h	Ingeniero Cotutor	30,69	30,69
		MO.IBJ	1,00 h	Ingeniero Biomédico Junior	10,00	10,00
		MAT.PCp	1,00 h	Ordenador Portátil Personal	1,90	1,90
		MAT.Office	1,00 h	Licencia de Office 365	0,90	0,90
		MAT.Windows	1,00 h	Licencia Windows 10	0,90	0,90
			4,00 %	Costes indirectos	75,08	3,00
				Precio total por h		78,08
Son setenta y ocho Euros con ocho céntimos						
4.3	04.03	h	Preparación defensa del TFG			
		MO.TU	1,00 h	Ingeniero Tutor	30,69	30,69
		MO.IBJ	1,00 h	Ingeniero Biomédico Junior	10,00	10,00
		MAT.Windows	1,00 h	Licencia Windows 10	0,90	0,90
		MAT.Office	1,00 h	Licencia de Office 365	0,90	0,90
		MAT.PCp	1,00 h	Ordenador Portátil Personal	1,90	1,90
			4,00 %	Costes indirectos	44,39	1,78
				Precio total por h		46,17
Son cuarenta y seis Euros con diecisiete céntimos						

6. CUADROS DE PRECIOS PARCIALES Y MEDICIONES

En esta sección se incluyen los Presupuestos Parciales junto a las mediciones. Para ello se separa entre los distintos capítulos y se define cada unidad, su cantidad y además se indica el presupuesto de su ejecución.

Presupuesto TFG				Página 12
PRESUPUESTO PARCIAL Nº 1 Definición del proyecto				
Nº	DESCRIPCION	CANTIDAD	PRECIO	IMPORTE
1.1	H. Reunión inicial Tutor y Cotutor para definir el TFG			
		1,00	109,05	109,05
1.2	H. Planificación del TFG			
		2,00	74,24	148,48
1.3	H. Revisiones Periódicas del TFG			
		6,00	76,21	457,26

PRESUPUESTO PARCIAL N° 2 Obtención y preparado de las imágenes				
Nº	DESCRIPCION	CANTIDAD	PRECIO	IMPORTE
2.1	H. Aprendizaje e instalación de softwares			
		2,00	17,13	34,26
2.2	H. Descarga y anonimización de las imágenes			
		40,00	17,28	691,20
2.3	H. Segmentación de las imágenes			
		50,00	25,37	1.268,50

PRESUPUESTO PARCIAL N° 3 Desarrollo del modelo predictivo mediante técnicas de aprendizaje máquina				
Nº	DESCRIPCION	CANTIDAD	PRECIO	IMPORTE
3.1	H. Obtener de los archivos .mat			
		3,00	52,90	158,70
3.2	H. Extracción y selección de las características			
		20,00	52,90	1.058,00
3.3	H. Obtencion y evaluacion del método de clasificación			
		20,00	52,90	1.058,00

PRESUPUESTO PARCIAL N° 4 Redacción de la memoria del TFG				
Nº	DESCRIPCION	CANTIDAD	PRECIO	IMPORTE
4.1	H. Redacción de los documentos			
		90,00	13,31	1.197,90
4.2	H. Revisión de los documentos y corrección de errores			
		15,00	78,08	1.171,20
4.3	H. Preparación defensa del TFG			
		75,00	46,17	3.462,75

7. PRESUPUESTO DE EJECUCIÓN POR CONTRATA

Para finalizar, en esta sección se ofrece la suma de los Presupuestos Parciales juntos los gastos generales (16% de los Presupuestos Parciales, beneficio industrial (7% de los Presupuestos Parciales + gastos generales), y el 21% de IVA.

Proyecto: Presupuesto TFG

Capítulo	Importe
Capítulo 1 Definición del proyecto	714,79
Capítulo 2 Obtención y preparado de las imágenes	1.993,96
Capítulo 3 Desarrollo del modelo predictivo mediante técnicas de aprendizaje máquina	2.274,70
Capítulo 4 Redacción de la memoria del TFG	5.831,85
Presupuesto de ejecución material	10.815,30
16% de gastos generales	1.730,45
7% de beneficio industrial	757,07
Suma	13.302,82
21% IVA	2.793,59
Presupuesto de ejecución por contrata	16.096,41

Asciende el presupuesto de ejecución por contrata a la expresada cantidad de DIECISEIS MIL NOVENTA Y SEIS EUROS CON CUARENTA Y UN CÉNTIMOS.

Valencia, Mayo 2024
Ingeniero Biomédico

Ignacio Iborra Roncales