



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA



ESCUELA TÉCNICA  
SUPERIOR INGENIEROS  
INDUSTRIALES VALENCIA

**TRABAJO FIN DE GRADO EN INGENIERÍA BIOMÉDICA**

# **ANÁLISIS DE TEXTURAS DE IMÁGENES PET/TC PARA LA CARACTERIZACIÓN, CUANTIFICACIÓN Y AYUDA EN EL DIAGNÓSTICO DEL CÁNCER DE MAMA**

AUTOR: MARÍA MANCEBO GONZÁLEZ

TUTOR: IGNACIO BOSCH ROIG

COTUTOR: ÁNGEL ALBERICH BAYARRÍ

**Curso Académico: 2018-19**



## AGRADECIMIENTOS

“Me gustaría aprovechar la ocasión para agradecer a mi tutores, Nacho y Ángel, la confianza que han depositado en mí y por darme la oportunidad de trabajar en este proyecto junto a un equipo, como es el de QUIBIM. También agradecer a la Dra. Francisca Mulero, por su ayuda y apoyo para llevar a cabo este proyecto.

A mis compañeros de la carrera, por todos los buenos momentos que hemos pasado, y también por los malos, gracias a los cuáles salimos más fuertes tras cuatro años juntos.

Por último, agradecer a mi familia y amigos, por apoyarme siempre en las decisiones que he tomado, por mostrarme su eterna confianza y por animarme a estudiar esta carrera lejos de casa. Soy todo lo que soy gracias a ellos.”



## RESUMEN

Una de las características más comunes que presentan los tumores cancerígenos es la existencia de condiciones de hipoxia en algunas regiones de la lesión, lo que está relacionado con su proliferación y metástasis, además de disminuir su sensibilidad a tratamientos como la radioterapia (radioresistencia). A partir de la evaluación y del desarrollo de la hipoxia en el microentorno del tumor, se podrían obtener parámetros de interés para la caracterización y planificación de terapias personalizadas contra el cáncer.

El principal radiofármaco utilizado en el ámbito de la investigación en imagen para la evaluación y cuantificación de la hipoxia es el  $^{18}\text{F}$ -misonidazol. Esta sustancia radiactiva se une a moléculas grandes de las células cancerosas con hipoxia, de forma que, la radiación que emite se puede detectar mediante PET. Con el objetivo de obtener una mayor cantidad de información, en este estudio se van a utilizar imágenes de una modalidad híbrida como la PET/TC, que proporciona diagnósticos más exactos al disponer de información funcional (PET) localizada espacialmente sobre la anatomía con alta resolución espacial (TC).

Para la caracterización del tumor en cáncer de mama se ha utilizado el análisis de texturas, específicamente diseñado para objetivar la agresividad tumoral que ofrece información cuantitativa de aspectos subjetivos, principalmente la heterogeneidad de tejidos, debida por ejemplo a la existencia de diferentes hábitats en el tumor, como una elevada proliferación vascular, necrosis, fibrosis, entre otros. Considerándose este aspecto esencial en los tumores, sobre todo en los más agresivos, el uso de esta aplicación es fundamental, ya que con una correcta medida de la heterogeneidad se puede ayudar a la caracterización de esta patología y a la determinación del pronóstico de los pacientes.

Este trabajo ha sido propuesto y realizado en la empresa Quibim, la cual ha proporcionado las interfaces gráficas necesarias para la visualización y segmentación de las imágenes PET/TC y el programa de análisis de texturas utilizado en el postprocesado sobre MatLab. Las imágenes corregistradas han sido facilitadas por el Centro Nacional de Investigaciones Oncológicas (CNIO), que, en colaboración con Quibim, tiene el objetivo de valorar el uso del radiofármaco misonidazol en predicción, diagnóstico y tratamiento en pacientes con cáncer de mama.

Este trabajo se ha realizado en el marco de un proyecto de investigación piloto y sin financiación asociada, realizado en ambas instituciones.

**Palabras clave:** Cáncer de mama, MISO, hipoxia, análisis de texturas, segmentación.

# RESUM

Una de les característiques més comunes que presenten els tumors cancerígens és l'existència de condicions de hipoxia en algunes regions de la lesió, la qual cosa està relacionat amb la seua proliferació i metàstasi, a més de disminuir la seua sensibilitat a tractaments com la radioteràpia (radioresistència) . A partir de l'avaluació i del desenrotllament de l'hipoxia en el microentorn del tumor, es podrien obtenir paràmetres d'interès per a la caracterització i planificació de teràpies personalitzades contra el càncer.

El principal radiofàrmaco utilitzat l'àmbit de la investigació en imatge per a l'avaluació i quantificació de l'hipoxia és el  $^{18}\text{F}$ -misonidazol. Esta substància radioactiva s'unix a molècules grans de les cèl·lules canceroses amb hipoxia, de manera que, la radiació que emet es pot detectar per mitjà de tomografia per emissió de positrons (PET) . Amb l'objectiu d'obtenir una major quantitat d'informació, en este estudi es van a utilitzar imatges d'una modalitat híbrida com la PET/TC, que proporciona diagnòstics més exactes al disposar d'informació funcional (PET) localitzada espacialment sobre l'anatomia amb alta resolució espacial (TC) .

Per a la caracterització del tumor en càncer de mama s'ha utilitzat l'anàlisi de textures, específicament dissenyat per a objectivar l'agressivitat tumoral que ofereix informació quantitativa d'aspectes subjectius, principalment l'heterogeneïtat de teixits, deguda per exemple a l'existència de diferents hàbitats en el tumor, com una elevada proliferació vascular, necrosi, fibrosi, entre altres. Considerant-se este aspecte essencial en els tumors, sobretot en els més agressius, l'ús d'esta aplicació és fonamental, ja que amb una correcta mesura de l'heterogeneïtat es pot ajudar a la caracterització d'esta patologia i a la determinació del pronòstic dels pacients.

Este treball ha sigut proposat i realitzat en l'empresa Quibim, la qual ha proporcionat les interfícies gràfiques necessàries per a la visualització i segmentació de les imatges PET/TC i el programa d'anàlisi de textures utilitzat en el postprocesado sobre MatLab. Les imatges corregistradas han sigut facilitades pel Centre Nacional d'Investigacions Oncològiques (CNIO) , que, en col·laboració amb Quibim, té l'objectiu de valorar l'ús del radiofàrmaco misonidazol en predicció, diagnòstic i tractament en pacients amb càncer de mama.

**Paraules clau:** Càncer de mama, MISO, hipoxia, anàlisi de textures, segmentació.





# ABSTRACT

One of the most common features of cancerous tumour is the existence of hypoxia in some regions of the lesion, which directly affects their proliferation and metastases and, moreover, decreases their sensitivity to the effects of medical treatments, such as radiation therapy (*Radiation Resistance*). Based on the evaluation and development of hypoxia in the microenvironment of the tumour, measurement parameters could be obtained for the characterization and planning of personalized therapies against cancer.

The main radiopharmaceutical used in the field of imaging research for the evaluation and quantification of hypoxia is  $^{18}\text{F}$ -Misonidazole. This radioactive drug binds to big molecules in cancer cells with oxygen deficiency, so, the radiation it emits can be detected by Positron Emission Tomography (PET). In order to obtain a greater amount of information, this study will use images of a hybrid modality such as PET/CT, which provide more accurate diagnostics when having functional information (PET) spatially located on anatomy with high spatial resolution (TC).

A texture analysis biomarker has been used to characterize the lesion in breast cancer, which has been designed to measure tumour aggressiveness that provides quantitative information on subjective aspects, mainly the heterogeneity in the images, due to, for example, high vascular proliferation, necrosis, etc. Considering this essential aspect in tumours, especially in the most aggressive ones, the use of this application is fundamental, because a correct measurement of heterogeneity can help in characterizing these tumours and in determining the prognosis of patients.

This paper has been proposed and carried out in the Quibim company, which has provided the necessary graphical interfaces for the visualization and segmentation of PET/CT images and, the texture analysis biomarker used in the postprocessing on MatLab. The coregistered images have been provided by the National Center for Oncology Research (CNIO), which, in collaboration with Quibim, aims for the evaluation of the use of misonidazole in prediction, diagnosis and treatment in breast cancer patients.

**Keywords:** Breast cancer, MISO, hypoxia, texture analysis, segmentation.



# INDICE GENERAL

<b>CAPÍTULO 1. INTRODUCCIÓN.....</b>	<b>1</b>
1. 1 Visión general .....	1
1. 2 Diagnóstico del cáncer de mama.....	2
1. 3 Biomarcadores de imagen .....	2
<b>CAPÍTULO 2. OBJETIVOS Y MOTIVACIÓN .....</b>	<b>4</b>
<b>CAPÍTULO 3. CONTEXTUALIZACIÓN DEL TRABAJO.....</b>	<b>5</b>
3. 1 Patología.....	5
3. 2 Cáncer de mama .....	6
3. 2. 1 Anatomía de la mama.....	6
3. 2. 2 Causas y crecimiento del tumor .....	7
3. 2. 3 Tipos de cáncer de mama.....	8
3. 2. 4 Tipos de tratamiento.....	9
3. 3 Imagen médica .....	11
3. 4 Hipoxia en los tumores.....	13
3. 4. 1 Medida de la hipoxia con radiofármacos.....	14
3. 5 Análisis de texturas .....	15
3. 5. 1 Métodos para la obtención de análisis de texturas .....	16
3. 5. 2 Métodos estadísticos .....	17
3. 5. 3 Métodos basados en modelos .....	19
3. 6. Sistemas de Ayuda a la Decisión Clínica.....	20
<b>CAPÍTULO 4. MÉTODO DE TRABAJO .....</b>	<b>22</b>
4. 1 Base de datos.....	22
4. 2 Procesamiento de las imágenes .....	24
4. 3 Procesamiento y análisis de las imágenes .....	25
4. 3. 1 Segmentación de las lesiones .....	25
4. 3. 2 Análisis de texturas .....	28
4. 3. 3 Reducción de la dimensionalidad de los datos .....	29
4. 3. 4 Clasificación de los datos .....	34
4. 4 Diseño de un clasificador .....	35
<b>CAPÍTULO 5. RESULTADOS Y DISCUSIÓN.....</b>	<b>38</b>

5. 1 Métricas para la evaluación de los resultados .....	38
5. 2 Resultados de los diferentes modelos de clasificación.....	40
5. 2. 1 <i>Support Vector Machines (SVM)</i> .....	41
5. 2. 2 <i>Nearest Neighbour Classifiers (KNN)</i> .....	41
5. 2. 3 <i>Ensemble Classifiers</i> .....	42
5. 2. 4 <i>Árboles de decisión</i> .....	43
5. 3 Validación del análisis de texturas con un conjunto de datos balanceado .....	45
5. 4 Discusión de los resultados obtenidos.....	47
5. 5 Limitaciones y líneas futuras de trabajo.....	48
<b>CAPÍTULO 6. CONCLUSIONES .....</b>	<b>49</b>
<b>ANEXO.....</b>	<b>51</b>
1. Código MatLab de Análisis de Texturas.....	51
2. Código MatLab <i>Data Reduction</i> .....	58
3. Código MatLab clasificación de biomarcadores .....	59
<b>BIBLIOGRAFÍA .....</b>	<b>61</b>
<b>PRESUPUESTO .....</b>	<b>66</b>
1. INTRODUCCIÓN Y OBJETIVOS .....	67
2. COSTES PARCIALES .....	67
2. 1 <i>Coste de Personal</i> .....	67
2. 2 <i>Coste de Herramientas Software</i> .....	68
2.3 <i>Coste de Herramientas Hardware</i> .....	69
3. PRESUPUESTO TOTAL .....	70

## INDICE DE FIGURAS

<b>Figura 1.</b> Tipos de cáncer más frecuentes en mujeres según el número de casos clínicos [aecc,2018] .....	1
<b>Figura 2.</b> Anatomía de la mama en detalle [INC, 2019] .....	7
<b>Figura 3.</b> Ejemplo sencillo de obtención de matriz GLCM con distancia par y dirección 0° .....	18
<b>Figura 4.</b> Ejemplo sencillo de obtención de matriz de largo recorrido GLRL.....	19
<b>Figura 5.</b> Ejemplo de cálculo de la medida D Dimensión Fractal.....	20
<b>Figura 6.</b> Variables de entrada para el desarrollo de un Sistema de Apoyo a la Decisión Clínica (CDSS) [Fuster, 2018] .....	21
<b>Figura 7.</b> Esquema de las etapas que constituyen la metodología del trabajo.....	22
<b>Figura 8.</b> Resumen del procedimiento seguido en la selección de los casos a estudiar .....	24
<b>Figura 9.</b> Segmentación inicial de la lesión en imágenes TC de una paciente en varios cortes axiales. Las imágenes de la izquierda muestran el corte antes de dibujar sobre el tumor. Cada fila de imágenes representa el mismo corte.....	26
<b>Figura 10.</b> Delineación de las segmentaciones en los cortes de la Figura 9 con el apoyo de la radiación del 18F-MISO en PET. ....	27
<b>Figura 11.</b> Ejemplo de tumor no discernible en PET y señalización por la Dra. en la segunda imagen. ....	27
<b>Figura 12.</b> Diagrama de bloques de Texturas.m.....	28
<b>Figura 13.</b> Representación de la matriz de correlación cruzada en un mapa de calor.....	31
<b>Figura 14.</b> Agrupación de los datos según clusterización jerárquica y representación de los coeficientes de correlación .....	32
<b>Figura 15.</b> Mapa de calor tras aplicarle el umbral [-0.8, 0.8]. ....	32
<b>Figura 16.</b> Ejemplo de pasos a seguir en Classification Learner.....	36
<b>Figura 17.</b> Estructura de la matriz de confusión.....	39

<b>Figura 18.</b> Valor diagnóstico de las curvas ROC [Parra, 2019] .....	40
<b>Figura 19.</b> Comparación de matriz de confusión de modelos con exactitud 85,7% y matriz de confusión de modelos con exactitud 81%.....	44
<b>Figura 20.</b> Comparación de matriz de confusión de modelos con exactitud 72,6% y matriz de confusión de modelos con exactitud 78,6%.....	44
<b>Figura 21.</b> Curva ROC para modelo Cubic SVM con exactitud igual a 81% y AUC 0.51	45
<b>Figura 22.</b> Curva ROC y Area Under the Curve del clasificador Subspace Discriminant.	46

## INDICE DE TABLAS

<b>Tabla 1.</b> Grupos de variables relacionados en más de 0,8. (Elevada correlación directa)..	33
<b>Tabla 2.</b> Grupos de variables relacionados en menos de - 0,8. (Elevada correlación indirecta) .....	33
<b>Tabla 3.</b> Desviación estándar de las variables y señalización en amarillo de la varianza máxima en cada grupo.....	35
<b>Tabla 4.</b> Evaluación de modelos – Support Vector Machines y exactitud en la predicción de respuesta .....	41
<b>Tabla 5.</b> Evaluación de modelos – Nearest Neighbour Classifiers y exactitud en la predicción de respuesta.....	42
<b>Tabla 6.</b> . Evaluación de modelos – Ensemble Classifiers y exactitud en la predicción de la respuesta .....	42
<b>Tabla 7.</b> Evaluación de modelos – Decision Tree y exactitud en la predicción de respuesta .....	43
<b>Tabla 8.</b> Índices estadísticos del modelo de clasificación Subspace Discriminant .....	46
<b>Tabla 9.</b> Comparación de las métricas más importantes entre modelos con muestra desbalanceada y muestra balanceada.....	47
<b>Tabla 10.</b> Cuadro de precios Personal .....	67
<b>Tabla 11.</b> Cuadro de precios Herramientas Software .....	68
<b>Tabla 12.</b> Cuadro de precios Herramientas Hardware.....	69
<b>Tabla 13.</b> Cálculo del presupuesto total del trabajo.....	70



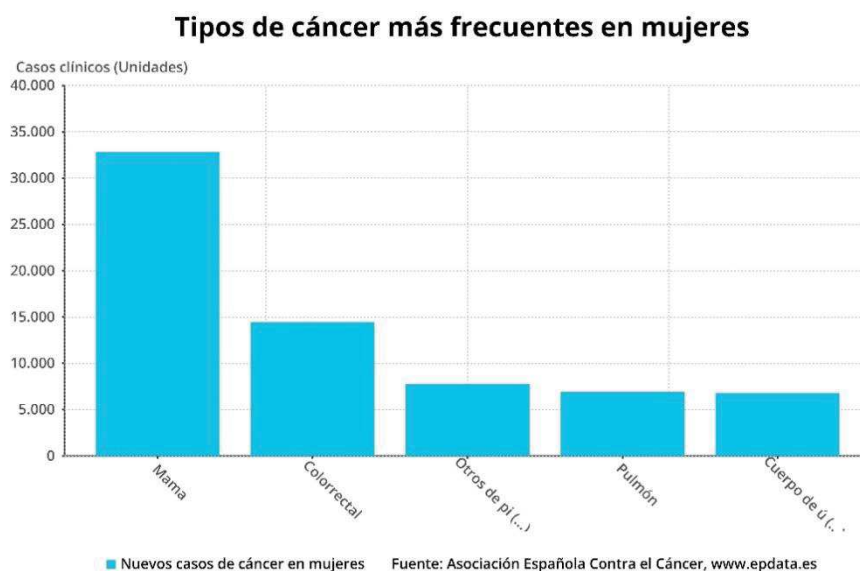


# **MEMORIA**

# CAPÍTULO 1. INTRODUCCIÓN

## 1.1 Visión general

Actualmente, el cáncer de mama es el tumor maligno más frecuente en mujeres y la primera causa de muerte en países desarrollados y en vías de desarrollo, estimándose que, 1 de cada 8 mujeres tendrá cáncer de mama a lo largo de su vida. Según la Asociación Española Contra el Cáncer (AECC), en el año 2018 se diagnosticaron, aproximadamente, 1.380.000 casos nuevos de cáncer de mama en el mundo. En comparación al total, la incidencia en España es baja, afectando en más de 30.000 casos al año (*Figura 1*), lo que representa casi el 30% de los tumores totales diagnosticados a las mujeres [aecc, 2019].



**Figura 1.** Tipos de cáncer más frecuentes en mujeres según el número de casos clínicos [aecc,2018]

Debido al envejecimiento de la población y a un diagnóstico cada vez más precoz, en los últimos años, tanto el número de casos como las tasas de incidencia han aumentado lentamente, sin embargo, gracias a los avances en investigación, la mejora de los tratamientos y del diagnóstico temprano, se ha conseguido estabilizar el número de fallecidos y disminuir la tasa de mortalidad [aecc, 2019].

## **1.2 Diagnóstico del cáncer de mama**

El cáncer de mama es una de las pocas enfermedades cancerosas que no presenta síntomas desde su aparición, por lo que puede ser detectada de forma temprana. Las posibilidades de curación de este tipo de cáncer cuando son detectados en su etapa inicial son prácticamente del 100%, sin embargo, a medida que se produce el crecimiento del tumor, la probabilidad disminuye. Por este motivo, la importancia de realizar un buen diagnóstico inicial aumenta, además de que, elegir un tratamiento óptimo para cada paciente es fundamental para su recuperación.

Hoy en día, el método más efectivo para la detección del cáncer de mama es la mamografía en programas de cribado poblacional [ASCO, 2018]. Esta técnica consiste, de manera general, en una radiografía de las mamas capaz de detectar lesiones en estadios muy incipientes de la enfermedad, siendo posible esta detección hasta dos años antes de que sean palpables y cuando aún no han invadido en profundidad. En estos casos, cuando se halla el tumor en estos estadios es posible aplicar tratamientos de menor agresividad, los cuales dejan menos secuelas físicas y psicológicas en la mujer.

Actualmente, se está investigando la efectividad del uso de otros métodos de imagen en la detección del cáncer de mama, con el objetivo principal de realizar un procesamiento computacional completo de las imágenes obtenidas y ofrecer al personal sanitario una mayor cantidad de información. Por ello, en este trabajo, se han utilizado imágenes de PET/TC, modalidad híbrida formada por la tomografía por emisión de positrones (PET) y la tomografía computarizada (TC), dando como resultado la combinación de información anatómica y funcional de la patología.

## **1.3 Biomarcadores de imagen**

La necesidad de extraer información cuantitativa de las imágenes médicas radiológicas y de medicina nuclear es crucial para su evaluación y ayuda en el diagnóstico y tratamiento posterior. Los parámetros extraídos reciben el nombre de Biomarcadores de Imagen [Alberich-Bayarri, 2018].

Un biomarcador es una característica que se mide y se evalúa como un indicador de procesos normales biológicos, procesos patogénicos o respuestas farmacológicas a una intervención terapéutica [Strimbu y Tavel, 2010]. Son utilizados en diagnóstico, como herramienta para el análisis de diferentes enfermedades, como indicador del estado de una patología y para la predicción y el control de la respuesta clínica a una intervención. La primera característica que debe cumplir para que su utilización sea válida es la precisión en la cuantificación del estado que se esté estudiando, además de que debe ser específico, sensible y reproducible para poder confiar en ese valor.

En este trabajo se ha utilizado el análisis de texturas, herramienta algorítmica para la obtención de información cuantitativa de imágenes médicas y específicamente diseñada para medir la agresividad tumoral (ver 3.5. *Análisis de texturas*). Esta aplicación no invasiva intenta cuantificar cualidades definidas por términos como áspero, liso, sedoso, etc. en función de la variación espacial en las intensidades de píxeles. El aspecto fundamental a registrar, la heterogeneidad, rugosidad o *bumpiness*, hace referencia a estas variaciones en los valores de intensidad, o en los niveles de gris. A partir de esta cuantificación de la lesión, se pretende realizar una caracterización de mayor precisión y, además, determinar tratamientos personalizados para cada paciente con cáncer de mama.

## CAPÍTULO 2. OBJETIVOS Y MOTIVACIÓN

Hoy en día, la imagen médica ofrece al personal sanitario una gran cantidad de información, siendo, sobre todo, la visual la más utilizada en la valoración y en el diagnóstico de diferentes patologías. Con el creciente desarrollo de la tecnología con aplicación médica, se puede realizar una mejor interpretación de la información extraída de forma cuantitativa, lo que supone una mejora en la atención sanitaria. Este trabajo nace con el objetivo final de evaluar la efectividad del uso de análisis de texturas en imágenes PET/TC del cáncer de mama y de su utilidad como apoyo a la decisión clínica.

Con este fin, se ha utilizado el radiofármaco  $^{18}\text{F}$ -Misonidazol (ver *3.4.1 Medida de la hipoxia con radiofármacos*), el cual, gracias a su capacidad de unión a las moléculas del tumor, permite evaluar y cuantificar la hipoxia, característica fundamental en este tipo de lesiones. Por ello, el segundo objetivo que se plantea es valorar el uso del Misonidazol para la cuantificación de la hipoxia y la predicción de respuesta de un tratamiento.

Para interpretar este tipo de imágenes se ha utilizado análisis de texturas para la medida de la heterogeneidad existente entre los diferentes hábitats del tumor. El principal objetivo y sobre el cual gira en torno este trabajo, es evaluar la utilidad de esta aplicación, a través de la cuantificación de la lesión en forma de biomarcadores, relevantes para el diagnóstico y la selección de un tratamiento específico.

Por último, el objetivo final planteado es obtener un Sistema de Apoyo a la Decisión Médica (*Clinical Decision Support System, CDSS*) [Zikos y DeLellis, 2018] a partir de estos biomarcadores y con el diseño de un clasificador, el cual, una vez entrenado, sea capaz de responder a preguntas clínicas determinadas.

## CAPÍTULO 3. CONTEXTUALIZACIÓN DEL TRABAJO

### 3.1 Patología

El término cáncer hace referencia a un conjunto de numerosas enfermedades, todas ellas relacionadas entre sí y con características independientes, como las causas de aparición, su evolución y tratamientos específicos. Sin embargo, el denominador común que presentan todas ellas es la capacidad de multiplicación y diseminación sin control de las células por todo el organismo [aecc, 2018].

Cada una de las células que conforman un organismo tienen una función o misión determinada, que, cuando ésta es cumplida, o cuando las células han sido dañadas, mueren y son reemplazadas por otras nuevas, manteniendo así la integridad y el correcto funcionamiento de los distintos órganos. En condiciones normales, las células humanas se reproducen para formar nuevas células, a medida que son necesitadas en el organismo, de forma regulada, según una serie de mecanismos que indican a la célula cuando comenzar a dividirse y cuando permanecer estable. La alteración de estos mecanismos en una célula inicia una división descontrolada de ésta y sus descendientes, dando lugar, con el tiempo, a un tumor o nódulo. Además, si estas células adquieren la facultad de invadir tejidos y órganos de alrededor, de trasladarse y proliferar en otras partes del organismo se denomina tumor maligno, o cáncer [aecc, 2018].

Por ello, se pueden distinguir dos tipos de tumores, cuyas principales características se describen a continuación:

#### ❖ Tumores benignos

Masa de tejido celular anormal que crece en una única parte del cuerpo, y no tiene la capacidad de invadir a tejidos cercanos ni de poder desplazarse por el cuerpo humano por vía venosa o linfática a tejidos situados a distancia. Aun así, se consideran peligrosos patológicamente, ya que dependiendo de su situación pueden presionar órganos importantes.

Generalmente, los tumores benignos no suelen crecer nuevamente, por lo que el tratamiento usual es extirparlos quirúrgicamente.

## ❖ Tumores malignos

La principal característica que presentan este tipo de tumores es la capacidad de infiltrarse en tejidos cercanos y de propagarse desde su región de origen a otras partes del cuerpo, produciéndose metástasis. Las células cancerosas se separan del tumor primario y pueden viajar a través del sistema linfático o por la sangre. El tumor metastásico está relacionado con el tumor primario, por ejemplo, si el primario es de pulmón y las células cancerosas se propagan al hueso, sigue presentando una estirpe relacionada con el primero [INC, 2019].

## 3.2 Cáncer de mama

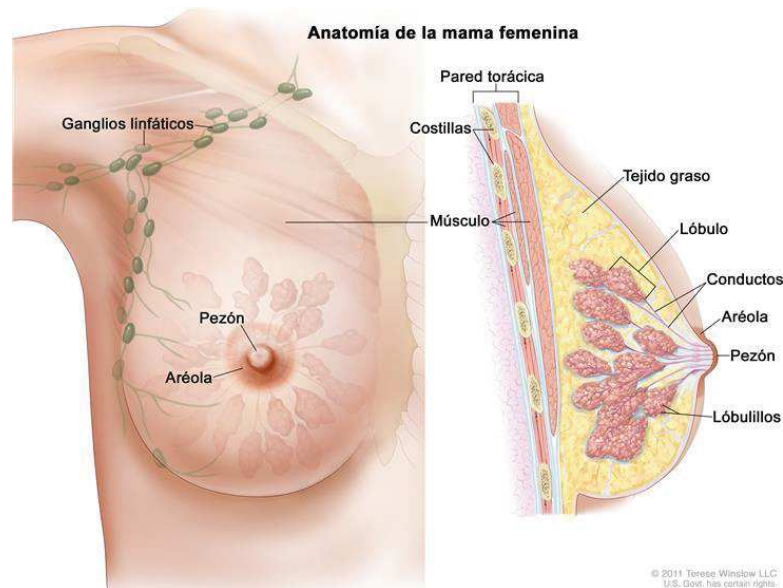
Actualmente, el cáncer de mama es el tumor maligno más frecuente en mujeres y la primera causa de muerte en países desarrollados. El tumor se genera por la proliferación descontrolada de células, lo que produce cambios en el material genético de las mismas, en este caso, de las células correspondientes del tejido de la glándula mamaria.

### 3.2.1 Anatomía de la mama

La glándula mamaria es un parénquima glandular compuesto de alveolos y conductos del estroma de soporte (ver *Figura 2*). Los alveolos son estructuras tubulares huecas recubiertas de células epiteliales y rodeados por células mioepiteliales, que se reúnen en grupos denominados lóbulos, los cuales, cada uno de ellos se conecta con los orificios del pezón a través de conductos lactíferos.

Las mamas se encuentran en la parte anterior del tórax, estando su crecimiento influenciado por características personales, genéticas, maternidad, edad, entre otros. La glándula está rodeada de tejido adiposo que proporciona consistencia y volumen, además de tejido glandular, cuyo tamaño varía según los niveles hormonales, aumentando en período de embarazo y lactancia.

Cada mama tiene además vasos sanguíneos, cuya función es proporcionar sangre a la glándula, y vasos linfáticos, encargados de transportar la linfa entre los ganglios linfáticos, donde se filtra y se almacenan los glóbulos blancos para combatir infecciones y enfermedades.



**Figura 2.** Anatomía de la mama en detalle [INC, 2019]

### 3. 2. 2 Causas y crecimiento del tumor

La mayoría de los casos de cáncer de mama son esporádicos, es decir, se desarrollan a partir del daño a dichos genes, que se produce por casualidad después del nacimiento. Los casos de cáncer de mama hereditarios son menos frecuentes y se deben por las mutaciones en los genes de supresión de tumores (tales con BRCA1 o BRCA2), transmitidas de generación en generación [ASCO, 2018].

Existen múltiples factores que elevan el riesgo de desarrollarlo, por ejemplo, edad, predisposición genética, antecedentes personales o familiares, factores hormonales, factores ambientales y proliferaciones benignas. Aunque todos ellos aumentan esta probabilidad, se puede dar el caso de que una misma persona que posea varios de estos factores nunca llegue a desarrollar cáncer, o sin tener ningún factor de riesgo, pueda llegar a padecerlo.

De esta manera, se pueden causar cambios en los mecanismos encargados de la división celular, provocándose un tumor de proliferación descontrolada, a lo que se le puede sumar



la facultad de invadir tejidos y órganos de alrededor, y de trasladarse a otras partes del organismo, por lo que pasa a denominarse, tumor maligno o metastásico. Este tumor puede crecer de tres maneras distintas [aecc, 2018]:

❖ **Diseminación en el tejido**

El cáncer de mama crece por invasión directa, infiltrando otras estructuras vecinas como la pared torácica y la piel desde su lugar de origen.

❖ **Diseminación linfática**

La red de vasos linfáticos que posee la mama permite que el drenaje de la linfa se efectúe a varios grupos ganglionares. Los más frecuentemente afectados son los ganglios axilares, junto con los situados en la zona central del tórax y los ganglios supraclaviculares.

❖ **Diseminación hematológica**

El cáncer se extiende a través de los vasos sanguíneos, preferentemente hacia los huesos, pulmón, hígado y piel, donde se forma el tumor metastásico.

Una vez que se ha diagnosticado el cáncer de mama, se realizan otro tipo de pruebas con las que se pretende determinar si las células cancerosas se han diseminado. Este proceso se denomina estadificación o determinación del estadio de la enfermedad. La importancia de conocer el estadio de la enfermedad aumenta, ya que, gracias a ello, se pueden planificar tratamientos de mayor precisión y efectividad.

### ***3. 2. 3 Tipos de cáncer de mama***

Como se ha explicado en apartados anteriores, el cáncer de mama puede ser invasivo o no a otros tejidos del organismo. Cuando el tumor no se disemina, no se extiende más allá de los conductos ni de los lobulillos, y se denominan carcinomas ductales o carcinomas lobulares:

### ❖ **Carcinoma ductal**

Tipo más común de cáncer de mama que se origina en las células que recubren internamente los conductos galactóforos, los cuáles conducen la leche desde los lobulillos de la mama hasta el pezón. Este tipo de carcinoma puede ser carcinoma ductal in situ (*CDIS*) o carcinoma ductal invasivo. El *CDIS* se ubica únicamente en el revestimiento del conducto mamario, sin diseminarse a tejidos circundantes u a otras partes vía sanguínea o linfática [ASCO, 2018].

### ❖ **Carcinoma lobular**

Con origen en los lóbulos, al igual que en el carcinoma anterior, existen dos tipos, carcinoma lobular in situ (*CLIS*) y carcinoma lobular invasivo. Aunque el *CLIS* no se considera cáncer, es un factor de riesgo ya que aumenta la probabilidad de presentar cáncer invasivo en cualquiera de las mamas [ASCO, 2018].

## **3. 2. 4 Tipos de tratamiento**

Un plan de tratamiento es un resumen de la patología y del tratamiento planeado contra esta. Algunas de las características más influyentes en este plan son la biología y el comportamiento del cáncer de mama, ya que, por ejemplo, el tamaño y el crecimiento del tumor no siempre son directamente proporcionales y no afectan de la misma manera.

Existen numerosos tipos de tratamientos que se pueden plantear para cada paciente, a la cual se le aplica aquel que corresponda en mayor medida a las características que presenta. Los tratamientos disponibles son los siguientes [ASCO, 2018]:

### ❖ **Cirugía**

Extirpación del cáncer a través de una operación quirúrgica. Existen dos tipos de cirugía, lumpectomía, con la cual se intenta conservar la mayor cantidad de mama posible, se extirpa el tumor y parte del tejido circundante sano; y mastectomía, o extirpación total de la mama.

Además, otro procedimiento que se realiza, con el objetivo de evitar la metastatización, es la biopsia (extracción) del ganglio linfático centinela, el primer ganglio que recibe el drenaje linfático del tumor primario y donde es más probable que se disemine.

#### ❖ **Radioterapia**

Uso de rayos X de alta energía u otros tipos de radiación para impedir o destruir células cancerosas. Su aplicación puede ser de forma externa, en la que el haz se aplica al exterior del cuerpo, e interna, con la cual, mediante el uso de agujas, semillas o catéteres radiactivos se trata el tumor desde dentro. Su utilización pasa por la configuración de la dosis de radiación en las herramientas de planificación. En este ámbito tiene especial interés la utilización de los resultados que se pueden derivar de este trabajo, pues interesa identificar las zonas hipóxicas del tumor, que son a su vez las que presentan mayor resistencia a la radiación, para tratar con una mayor dosis estas regiones.

#### ❖ **Quimioterapia**

Uso de radiofármacos para destruir células cancerosas, poniendo fin a su capacidad de crecer y dividirse. Se puede administrar antes o después de la cirugía, según se pretenda disminuir el tamaño de un tumor grande, facilitar la cirugía y/o disminuir el riesgo de recurrencia.

#### ❖ **Terapia con hormonas**

También conocida como terapia endocrina, es un tratamiento eficaz en aquellos casos de tumores que dan resultado positivo para los receptores de estrógeno o progesterona. Mediante el bloqueo de las hormonas se puede ayudar a prevenir la recurrencia del cáncer ya sea como un tratamiento único o después de la quimioterapia [INC, 2019].

#### ❖ **Terapia dirigida**

Tratamiento dirigido a los genes específicos del cáncer de mama, o a las condiciones del tejido que contribuyen a su crecimiento y supervivencia. Se utilizan

medicamentos y radiofármacos para identificar y atacar células cancerosas específicas sin dañar las células normales.

### **3.3 Imagen médica**

Hoy en día, las imágenes de Radiología y Medicina Nuclear y su correspondiente análisis juegan un papel fundamental en el sector médico, pasando de ser un elemento auxiliar a un elemento esencial en el diagnóstico temprano de diferentes patologías, interviniendo tanto en la medicina preventiva como en el seguimiento de un tratamiento.

Algunas de las principales aplicaciones del estudio por imágenes, y las que más interesan en este trabajo, son la detección del cáncer, conocer su propagación por el tejido y demostrar la efectividad del tratamiento seleccionado [ACS, 2016].

Aunque a menudo los resultados de los análisis son muy útiles, presentan límites a la hora de demostrar con certeza que diversos diagnósticos o cambios en el tejido se deben al cáncer. Se pueden encontrar concentraciones o grandes grupos de células cancerosas, nunca una sola aislada, necesiéndose además millones de células para hacer que un tumor sea visible en el estudio por imágenes. Sin embargo, se mantienen los tratamientos incluso cuando las células cancerosas no son visibles, ya que una sola de ellas puede volver a crecer y formar un nuevo tumor. Por último, no todo lo que parece ser cáncer en las imágenes se corrobora, ya que, con la realización de otras pruebas, por ejemplo, una biopsia, puede indicar que no lo es [ACS, 2016]. Por ello, un estudio completo del cáncer debe incluir, además de los resultados obtenidos de la radiología, la historia clínica del paciente (síntomas y factores de riesgo), biopsias, un examen físico, análisis de sangre y otras pruebas de interés.

Gracias a los avances en tecnología en los últimos años y la experiencia en la interpretación de imagen se ha mejorado considerablemente la detección y caracterización de lesiones mamarias. Cada vez es más posible detectar lesiones de menor tamaño y sutiles, predecir la probabilidad de metastatización e incluso confirmar su naturaleza histológica de forma percutánea [Benito, 2012].

En lo referente a la imagenología en cáncer de mama, actualmente, las tecnologías utilizadas son las siguientes [ASCO, 2018]:

## ❖ **Mamografía**

Método de imagen de búsqueda del cáncer de mama en pacientes sintomáticas o asintomáticas con el objetivo de detectarlo de forma temprana. Tanto en las mamografías de diagnóstico como de detección se aplica una radiación ionizante sobre la mama comprimida para obtener estudios de calidad.

La sensibilidad en el diagnóstico oscila entre el 61 y 95%, siendo menor en pacientes con mamas de mayor densidad, donde se estima que el 10% de ellos no son visibles con este método, aunque sean palpables [Benito, 2012].

La aparición de la mamografía digital, en sustitución de la convencional, se debe a la necesidad de obtener mejores resultados, respecto a la calidad y manipulación de las imágenes, facilidad en el almacenamiento, aumento de rango de sensibilidad en el diagnóstico, posibilidad de envío a otros centros y menor dosis de radiación.

## ❖ **Ecografía**

La ecografía mamaria es un método de imagen mediante ondas de sonido utilizado en el estudio de la patología mamaria, de forma aislada o como complemento de la mamografía. Gracias a las técnicas de los últimos años se ha incrementado la calidad de las imágenes, pudiéndose detectar con el ultrasonido lesiones pequeñas y microcalcificaciones, además de diferenciar una masa sólida de un quiste lleno de líquido (no canceroso).

## ❖ **Mamografía con tomosínteis**

Método que consiste en la fusión de dos métodos radiológicos, ya conocidos, como son la mamografía digital 2D y la tomografía. Es una técnica de imagen tridimensional, en la que, mediante la utilización de un mamógrafo, se realizan 20 tomas desde diferentes ángulos, en un intervalo de tiempo muy corto (8-30 segundos), las cuales son posteriormente reconstruidas. La tomosíntesis tiene la principal ventaja de que complementa la mamografía, no la elimina, de modo que resuelve sus problemas, por ejemplo, la superposición de estructuras normales y patológicas, lo que impide la evaluación de lesiones [Apesteguía, 2019]

### ❖ **Resonancia Magnética (RM)**

La resonancia magnética hace uso de campos magnéticos para producir imágenes detalladas del cuerpo. Se realiza sobre todo una vez se ha detectado el cáncer de mama en la paciente, ya que, aunque actualmente es el método más sensible para el diagnóstico del cáncer de mama, su especificidad no es muy elevada (37-97%), por lo que se utiliza principalmente para controlar el crecimiento o expansión del tumor, previo a una cirugía y comprobar el correcto funcionamiento de la quimioterapia.

La mamografía es el método de imagen más importante utilizado hoy en día en el estudio de la patología mamaria, además de ser el único que ha demostrado disminuir la mortalidad por cáncer de mama, aunque otros métodos están tomando cada vez un papel más importante. A parte de las comentadas, la utilización de diferentes técnicas en este campo, como la tomografía, la mamografía y ecografía de contraste, la espectroscopia o PET, a pesar de que suponen estudios más amplios para definir sus indicaciones y concretar su aportación, ofrecen un futuro muy prometedor en el diagnóstico temprano del cáncer de [Benito, 2012].

### **3.4 Hipoxia en los tumores**

Una característica común de los tumores sólidos, entre los que se incluyen los sarcomas, los carcinomas y los linfomas, es la hipoxia o falta de oxígeno, la cual contribuye de manera local y sistemática a la progresión tumoral, provocando un incremento en la probabilidad de recurrencia de un tumor.

El motivo por el cual los tumores se desarrollan preferentemente en este tipo de regiones con falta de O<sub>2</sub>, es la afección a diversos factores fundamentales en la progresión tumoral, como lo son la infiltración, la metástasis y la angiogénesis. Éste último, responde a las necesidades metabólicas de crecimiento de un tumor, siendo su ausencia una limitación de la progresión y la metástasis [Arvelo y Cotte, 2009].

La sensibilidad de los tumores hipóxicos a la radioterapia disminuye hasta en tres veces, en comparación con aquellos que están bien oxigenados. Por ello, los pacientes con tumores

hipóxicos presentan una peor supervivencia, además de ser más resistentes a otras formas de tratamiento, puesto que son tumores más agresivos y con mayor facilidad de metastatizar. La comprensión y la evaluación de los cambios en el metabolismo y de la proliferación celular debidos al desarrollo de la hipoxia en el microentorno del tumor es de gran importancia ya que proporciona un medio objetivo para la selección y planificación de una terapia personalizada contra el cáncer [Rajendran y Krohn, 2015].

### **3. 4. 1 Medida de la hipoxia con radiofármacos**

Durante la última década, se han producido diversos avances significativos en la investigación de la hipoxia tumoral, con relación a la planificación de un tratamiento y a la individualización del suministro de radiación a un volumen objetivo. La tomografía por emisión de positrones (PET), junto con la utilización de un radiofármaco específico, es el principal método para la evaluación, pronóstico y cuantificación de la hipoxia [Rajendran y Krohn, 2015].

<sup>18</sup>F-MISO, también conocido como <sup>18</sup>F-fluoromisonidazol, es una sustancia radiactiva en estudio, cuyo uso en pruebas de imágenes de diferentes tipos de cáncer se debe a su capacidad de unión a las moléculas de las células cancerosas que presentan una concentración baja de oxígeno [INC, 2019].

La absorción del fármaco MISO en un tumor sólido refleja el grado de hipoxia y, por lo tanto, los cambios que afectan a la biología del cáncer. La identificación no invasiva de los tumores con la evaluación de hipoxia en el microambiente tumoral, junto con otros marcadores moleculares (por ejemplo, el metabolismo de la glucosa, la proliferación celular, etc.) servirá de ayuda en la selección de una terapia adecuada en una fase temprana y con un enfoque personalizado [Rajendran y Krohn, 2015].

De forma general, la imagen médica y las imágenes de hipoxia se están convirtiendo en una herramienta fundamental de imagen *in vivo* para el futuro, debido principalmente, a su complementación con los métodos tradicionales y directos de muestreo del tejido, proporcionándose una instantánea del tumor primario y de la metastatización de la enfermedad.

### 3.5 Análisis de texturas

Hoy en día, el empleo de imágenes digitales en medicina es ampliamente conocido como medio de diagnóstico de diferentes patologías, las cuales son detectadas por la diferenciación de las zonas de interés en estas. Debido a ello, ha crecido la necesidad de utilizar herramientas matemáticas o algoritmos que sirvan de apoyo al análisis de las imágenes en la práctica clínica, y ofrezcan una información mayor en la segmentación de estructuras anatómicas y en la detección de lesiones.

En primer lugar, antes de definir el análisis en cuestión, es necesario entender qué son las texturas en imágenes médicas. Las texturas se definen como atributos o regiones que presentan cualidades específicas visuales, por ejemplo, rugosidad, aspereza, suavidad, etc., representadas como variaciones entre niveles de gris en la imagen [Itzel, 2018].

El análisis de texturas es una aplicación no invasiva de postprocesamiento de imágenes, diseñada para cuantificar una lesión y medir la agresividad tumoral, la cual ofrece información cuantitativa de aspectos subjetivos, principalmente de la heterogeneidad de las imágenes, debido a una elevada proliferación celular, necrosis, etc. El creciente interés en este tipo de análisis deriva de la necesidad de caracterizar esta irregularidad tumoral, debido a que es considerado como factor relevante en el pronóstico tumoral de los pacientes. En este contexto, el análisis de texturas, consistente en una variedad de técnicas matemáticas que pueden describir patrones de los diferentes niveles de gris de una imagen, juega un papel importante en la evaluación de la organización espacial de diferentes tejidos y órganos [Scalco y Rizzo, 2017].

Por lo tanto, el análisis de texturas se está utilizando en gran parte en el ámbito de la radioterapia, sobre todo, para la caracterización del tumor en la fase de planificación y para la predicción de la respuesta al tratamiento. Sin embargo, aunque se trata de una técnica prometedora, no se ha terminado de explotar, ya que existen algunas limitaciones relacionadas con la repetibilidad (cantidad de variabilidad en el sistema de medición causada por el dispositivo de medición) o robustez que necesitan ser mejoradas [Scalco y Rizzo, 2017].



### 3. 5. 1 Métodos para la obtención de análisis de texturas

Con la evaluación de la distribución de intensidad de los niveles de gris entre los píxeles que forman una región de interés en una imagen, se pueden obtener diferentes características de la textura, parámetros matemáticos obtenidos a partir de esta distribución de los píxeles. Las diferentes formas de obtener un análisis de este tipo dependen de la manera en la que se miden las interrelaciones entre los píxeles, siendo en total, tres métodos los existentes [Benet, 2015]:

#### ❖ Métodos basados en modelos

Representación de texturas a partir del uso de modelos matemáticos con los que se estiman parámetros con relación al grado de irregularidad y heterogeneidad de los tejidos, siendo el principal método el análisis fractal (ver 3. 5. 3 *Métodos basados en modelos*).

#### ❖ Métodos basados en transformación

Análisis de las características de textura en un espacio obtenido tras la transformación de *Fourier* o *Wavelet*, en el cual se ajusta más fácilmente el análisis. La transformación más utilizada, hoy en día, es la *Wavelet*, que, frente a la transformada de *Fourier*, la cual maneja una base de funciones localizadas en frecuencia, pero no el tiempo, esta primera presenta una buena localización en ambos.

La transformada *Wavelet* consiste en la división de la imagen en estudio en dos tipos de menor tamaño, la tendencia y las fluctuaciones, siendo una copia de la imagen a menor resolución y el almacenamiento de los cambios realizados a la imagen original, respectivamente. La tendencia y las fluctuaciones más significativas permiten una compresión de la imagen, descartando información irrelevante y ruido [UPV, 2018]. No nos extenderemos más en esta clase, puesto que no se han utilizado en este trabajo.

#### ❖ Métodos estadísticos

Uso de propiedades que hacen referencia tanto a la distribución como a las relaciones entre los valores de niveles de gris, entre las que se destacan el histograma y las matrices de segundo y elevado orden (ver 3. 5. 2 *Métodos estadísticos*).

Además de los métodos estudiados, existe un cuarto método estructural, el cual no se incluye en la clasificación anterior debido a que no serviría para realizar un análisis de texturas, sino para una síntesis de las imágenes, es decir, representan las texturas con el uso de primitivas bien definidas [Epifanio, 2002].

### **3. 5. 2 Métodos estadísticos**

Dentro del análisis de texturas estadístico, se puede realizar una subclasificación diferenciando según el orden estadístico utilizado, obteniendo análisis y características diferentes. El orden hace referencia al número de pasos realizados para extraer información cuantitativa relevante con relación a las variaciones espaciales de intensidad entre los píxeles. Para obtener una buena discriminación de las texturas es necesario utilizar estadísticos de segundo orden.

Según esta clasificación, existen diferentes técnicas estadísticas, entre las que destacan:

#### **❖ Análisis de texturas de primer orden**

Describe la distribución de los valores de los vóxeles individuales sin afectar las relaciones espaciales. Es un método basado en histogramas, por lo que se tienen las ventajas tales como simplicidad e intuitividad, pero no describen las texturas ya que no se consideran las interrelaciones entre píxeles.

En total, se puede obtener hasta doce parámetros significativos: media, mediana, máximo, mínimo, percentil, desviación y varianza de los niveles de gris en la imagen, uniformidad (sobre la distribución de los niveles de gris), desviación media absoluta, entropía, asimetría y kurtosis [García, 2019].

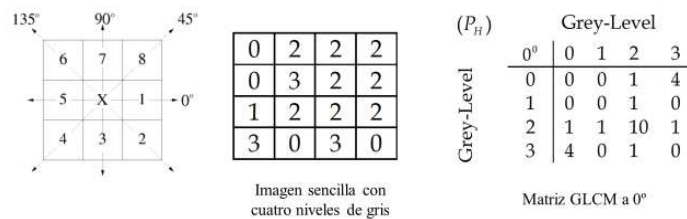
De todas ellas, en este trabajo se han obtenido kurtosis (medida de la forma de la distribución de la intensidad), skewness o asimetría (valor igual a 0 si el histograma es simétrico) y media, mediana, desviación estándar y percentiles 25 y 75 de la intensidad.

❖ **Análisis de texturas de segundo orden con matrices de co-ocurrencia (*Gray-Level Co-occurrence Matrices, GLCM*)**

A diferencia de la técnica anterior, este análisis basado en estadísticos de segundo orden tiene en cuenta la relación espacial entre los píxeles de la imagen, lo que añade cierta dificultad al método.

La matriz de co-ocurrencia permite extraer información estadística gracias al análisis de los valores de niveles de gris para pares de píxeles, es decir, determina cómo de frecuente un píxel con intensidad  $i$  se encuentra relacionado con otro píxel de intensidad  $j$  [Benet, 2015].

Para calcular la matriz (**Figura 3**), es necesario definir una distancia y una dirección, además de los pares de píxeles separados esa distancia. A partir de esto, se pasa a ver la cantidad de pares de píxel separados esa distancia.



**Figura 3.** Ejemplo sencillo de obtención de matriz GLCM con distancia par y dirección 0°

A partir de esta matriz se pueden obtener una serie de parámetros, entre los cuales, contraste (variación local de intensidad en una imagen), correlación (medida de la dependencia lineal de los niveles de gris), autocorrelación (evaluación tanto de la regularidad como de la tosquedad de la textura), *cluster prominence* y *cluster shade* (ofrecen información sobre el grado de simetría), energía (mide la repetición del píxel y expresa la regularidad de la textura), entropía (representa la irregularidad en la distribución de los valores de intensidad y es inversa a la energía), diferencia de entropía, diferencia de la varianza, homogeneidad, información sobre la medida de la correlación, *inverse difference* (con relación a la homogeneidad entre píxeles), probabilidad máxima, suma de la entropía, suma de la media, suma de la varianza y varianza cuadrada [Alobaidli, 2014] [Alberich-Bayarri y Martí-Bonmatí, 2018].

❖ **Análisis de texturas de orden elevado con matrices de largo recorrido (*Gray-Level Run-Length, GLRL*)**

Método de búsqueda en la imagen, según una dirección dada, para carreras de píxeles que tienen el mismo valor de nivel de gris. Una carrera de nivel de gris hace referencia a un conjunto de píxeles alineados consecutivos con el mismo nivel de gris [Benet, 2015].

2	2	2	2	Horizontal Runs	Run Length			
1	1	0	1		2	3	4	
1	1	1	2	Gray Level	0	1	0	0
0	1	0	0		1	1	1	0
					2	0	0	1

**Figura 4.** Ejemplo sencillo de obtención de matriz de largo recorrido GLRL

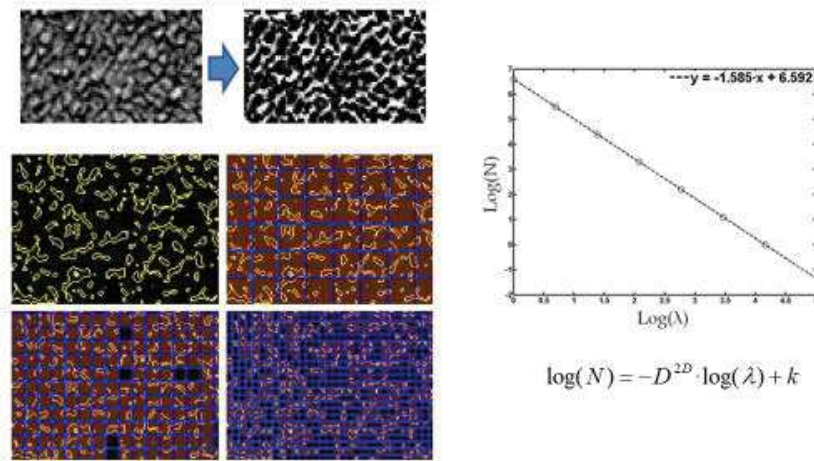
### 3. 5. 3 Métodos basados en modelos

Como se ha mencionado anteriormente, estos métodos se basan en la construcción de un modelo, cuyos parámetros describen las cualidades de la textura. Una de las técnicas más utilizadas es el análisis de texturas basados en fractales, el cual muestra la correlación existente entre la dimensión fractal de una textura y su *coarseness* (tosquedad, aspereza) [Epifanio, 2002].

En primer lugar, se puede definir fractal como un conjunto matemático que presenta autosimilitud a cualquier escala, lo que significa que, aunque se realice cualquier aumento o disminución de tamaño del objeto fractal, este se mantiene exactamente igual al inicial [Epifanio, 2002], es decir, es un objeto cuya estructura se repite a diferentes escalas. Se puede entender los fractales como modelos para explicar la naturaleza o la realidad, por lo que, a partir de la construcción de fractales y compararlos con un repositorio de ellos, se puede conocer las cualidades o propiedades de la región en estudio [Benoit, s.f.]

El análisis de texturas basado en fractales en imágenes médicas se basa en la determinación de la medida numérica adimensional  $D$  o Dimensión Fractal, para cada píxel o conjunto de píxeles, la cual es calculada a partir de los  $N$  elementos resultantes de tamaño  $\lambda$  y representa

la pendiente de la recta de regresión logarítmica (**Figura 5**). Debido a la relación existente entre la dimensión fractal y la cualidad de aspereza o heterogeneidad, el análisis permite cuantificar el grado de irregularidad del tejido calculando como de rápido aumenta el volumen, cuando la unidad de medida es cada vez menor [Navas, 2009].



**Figura 5.** Ejemplo de cálculo de la medida D Dimensión Fractal [Alberich-Bayarri]

### 3. 6. Sistemas de Ayuda a la Decisión Clínica

Los Sistemas de Ayuda a la Decisión Clínica (CDSS) son sistemas computacionales que aportan conocimiento específico para las decisiones médicas que deben tomarse en los planes de tratamiento de pacientes, ya sea en el diagnóstico, pronóstico, tratamiento o administración [Fuster, 2018].

Muchas decisiones hoy en día son de tan complejidad que se ha sobrepasado la capacidad de la mente humana para operar sin ayuda o soporte para la decisión clínica. El mayor desafío al que se enfrenta es tomar la información adecuada (discriminar datos no influyentes clínicamente pero que afectan al sistema) y aplicarla a cada paciente en el momento en que la información sea necesaria.

Los CDSS están altamente relacionados con el concepto de Medicina Basada en la Evidencia (ver **Figura 6**) ya que, recogen el conocimiento médico del problema a través de diferentes protocolos, datos biomédicos y diversos controles de calidad, dan soporte computacional para la implantación en la práctica médica de conocimiento científico y evalúan el

rendimiento de la solución para cada problema específico [Fuster, 2018]. En conclusión, el uso de este sistema pretende potenciar las habilidades de los profesionales de sanidad para manejar múltiples variables biomédicas, además de aportar herramientas de valor añadido a sus tareas y al sistema sanitario en el que desarrollan los procesos asistenciales.



**Figura 6.** Variables de entrada para el desarrollo de un Sistema de Apoyo a la Decisión Clínica (CDSS) [Fuster, 2018]

## CAPÍTULO 4. MÉTODO DE TRABAJO

Antes de comenzar a explicar cuál es la metodología seguida en este trabajo y con el objetivo de aclarar todos los pasos realizado, la **Figura 7** muestra un esquema del procedimiento seguido y un breve resumen de cada etapa.



**Figura 7.** Esquema de las etapas que constituyen la metodología del trabajo

### 4.1 Base de datos

Con el objetivo de valorar la respuesta al tratamiento en pacientes con cáncer de mama y evaluar de forma específica su evolución, se ha trabajado con una base de datos proporcionada por el Centro Nacional de Investigación Oncológica (CNIO).

Se trata de un estudio del cáncer de mama en neoadyuvancia (anterior a la cirugía) en el que, de manera aleatoria, a las pacientes se les ha aplicado dos tratamientos diferentes, quimioterapia estándar o la combinación de quimioterapia y medicación (tratamiento experimental). A este segundo grupo, además del PET en situación basal, se le realizó un segundo a los 15 días, con el objetivo de evaluar los cambios de oxigenación, irrigación y metabolismo debidos al fármaco extra inyectado.

Las respuestas obtenidas de los estudios, con las que se contrastarán los resultados obtenidos en este trabajo, están graduadas entre 0 y 4, según la escala patológica *Symmans* y *Pusztai* [Tresseraa, 2013]. De manera general, hay tres categorías: 0 y 1 son casi equivalentes y vienen a ser una respuesta completa; 2 y 3, también biológicamente parecidos, hacen referencia a una respuesta escasa o ausencia de ella; y, por último, el 4 significa el

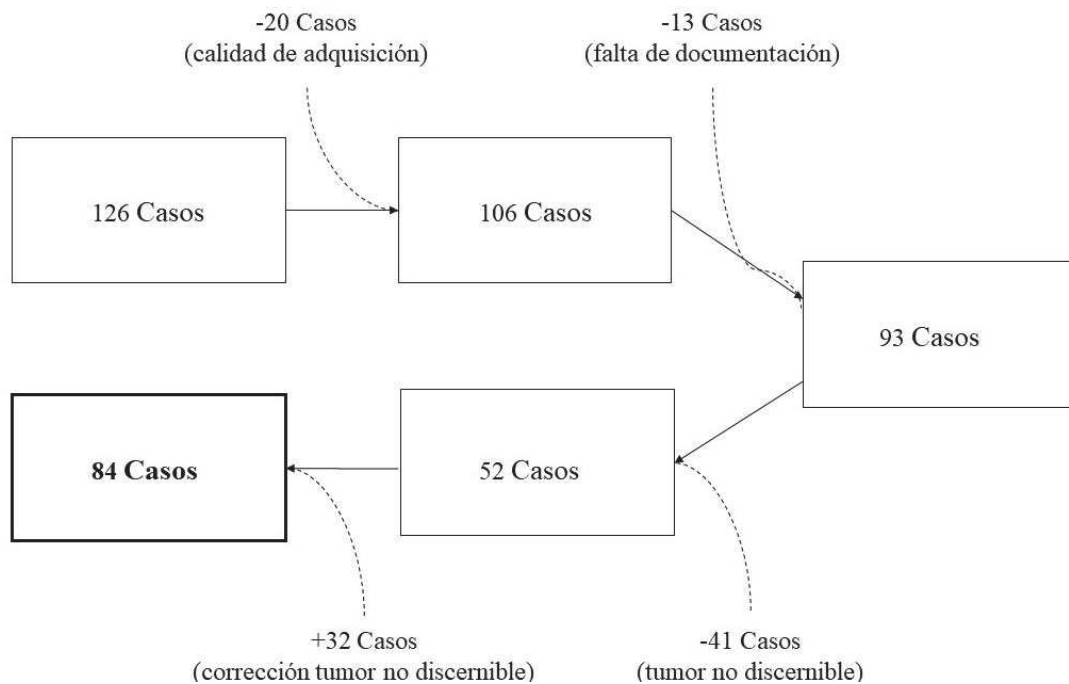
crecimiento del tumor durante el tratamiento, aunque al tratarse de un hecho bastante raro y no darse el caso en esta situación, se ha ignorado. Al ser una serie muy pequeña, no es conveniente hacer subgrupos por lo que se ha decidido dicotomizar la respuesta como 0/1 (respuesta completa o casi completa) y 2/3 (ausencia del beneficio del tratamiento).

La base de datos inicial constaba de 126 estudios de F<sup>18</sup>-MISO PET/TC de pacientes con uno o dos instantes temporales (adquisiciones basales y de seguimiento), dependiendo del grupo de tratamiento perteneciente. De esta totalidad, se excluyeron en primer lugar 20 casos por presentar una calidad de adquisición insuficiente para el análisis. Otros 13 casos fueron también extraídos por no constar de la documentación necesaria para el estudio. Por lo tanto, previo al análisis y basándose únicamente en criterios de la calidad de imagen y criterios de falta de documentación, los estudios restantes de la base de datos y con posibilidad de analizar suman un total de 93.

Previo al procesamiento de las imágenes, se realizó una revisión de la base de datos en función de la visibilidad del tumor en la adquisición de las imágenes o por no presentar márgenes bien delimitados. Por ello, se excluyeron 41 casos más, dejando un total de 52 casos listos para analizar.

Para el análisis específico de este trabajo, la base de datos final se puede considerar insuficiente para la obtención de resultados coherentes, por lo que, antes de continuar con el procesamiento de las imágenes, se ha trabajado en profundidad con los 41 casos correspondientes a imágenes con tumor no discernible, con el objetivo de obtener una base de datos final mayor. Con la ayuda de la Doctora Francisca Mulero, responsable de la Unidad de Imagen Molecular del CNIO, se ha conseguido segmentar las lesiones de estas imágenes y obtener así, un total de 84 casos finales. En la **Figura 8** se muestra un resumen del proceso de exclusión de casos de la base de datos inicial.





**Figura 8.** Resumen del procedimiento seguido en la selección de los casos a estudiar

## 4.2 Procesamiento de las imágenes

Tras la selección de los casos con posibilidad de analizar, estos fueron sometidos a una etapa de corrección en la cual se alinearon las adquisiciones de PET a sus correspondientes adquisiciones de TC. Este proceso es fundamental en el preprocesado de imágenes, ya que, mediante la alineación de los puntos de un objeto con sus correspondientes, se permite solucionar posibles incoherencias en la matriz de adquisición. Cabe destacar que se realizó este proceso a modo de ajuste fino, al tratarse la PET/TC de una modalidad híbrida que ya proporciona imágenes de PET y TC corregistradas, aunque con algunos errores locales de coherencia espacial. Todo este proceso de corrección para la mejora del alineamiento entre la imagen PET y la imagen TC fue realizado utilizando el paquete software *Elastix*.

Para el alineamiento de las adquisiciones de PET y TC se utilizó un corrección rígida de 6 grados de libertad (traslación y rotación), el cual mantiene todas las distancias entre los diferentes puntos de cada imagen, la rectitud de las líneas, la planitud de las superficies y los ángulos. Este tipo de transformación se encuentra definida en la ecuación (1).

Debido a la existencia de adquisiciones basales y de seguimiento en algunos de los casos, una vez alineadas las imágenes, el corregistro entre estas se realizó calculando la transformación necesaria para trasladar las imágenes de seguimiento a las imágenes del estudio basal. En este caso, la transformación utilizada es no rígida y basada en el uso de *b-splines*, cuya ecuación se encuentra definida en la ecuación (2). En ambas transformaciones, se ha utilizado la información mutua como métrica para medir el desempeño tras cada iteración y el método de gradiente descendiente como optimizador.

$$T_{rígida}(x) = R \cdot (x - c) + t + c \quad (1)$$

$$T_{b-splines}(x) = x + \sum_{x_k \in N_x} p_k \beta^3\left(\frac{x - x_k}{\sigma}\right) \quad (2)$$

Donde  $R$  es la matriz de rotación,  $t$  es el vector de traslación,  $c$  es el centro de rotación,  $x_k$  son los puntos de control de las *b-splines*,  $\beta^3(x)$  es el polinomio multidimensional de las *b-splines* cúbicas,  $p_k$  es el vector de coeficientes de las *b-splines* y  $N_x$  es el conjunto de puntos de control.

## 4.3 Procesamiento y análisis de las imágenes

### 4.3.1 Segmentación de las lesiones

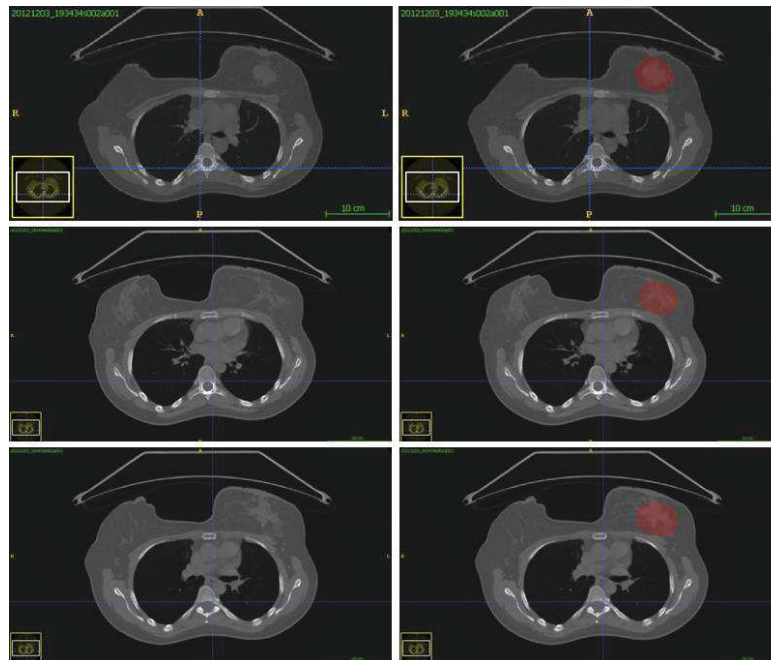
Una vez se han corregistrado las imágenes PET/TC, el siguiente paso es realizar una segmentación lo más precisa posible del tumor con la ayuda del 18F-MISO, para así, posteriormente, analizarlas con el módulo de análisis de texturas.

La segmentación anatómica es una técnica de tratamiento digital con la que se permite extraer información de objetos y puede ser concebida como la consecución de dos tareas, el reconocimiento y la delineación. El reconocimiento se puede definir como la determinación, a grandes rasgos, de la ubicación de la lesión en la imagen estudiada, mientras que la delineación consiste en la determinación precisa de la extensión espacial del objeto en cuestión.

En este procedimiento se han utilizado las aplicaciones software *RadiAnt DICOM Viewer* e *ITK-Snap*. El primero es un visor DICOM-PACS de imágenes médicas, el cual proporciona la visualización de las adquisiciones y su información asociada; mientras que el segundo se ha utilizado para segmentar estructuras en 3D de imágenes médicas en formato *NIfTI*. Ambas aplicaciones disponen de descarga gratuita a través de sus páginas webs oficiales.

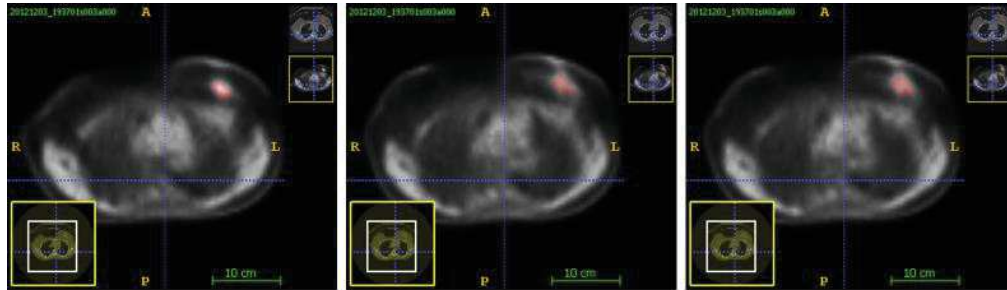
La etapa de reconocimiento se comenzó con la diferenciación anatómica de la lesión, la cual era lo suficientemente visible en la mayoría de los casos en las imágenes de TC, por lo que, en primer lugar, se realizó una segmentación inicial de la zona en la cual se encontraba el tumor, delimitando esta zona asegurando un contenido completo. El proceso que se ha realizado consiste en dibujar sobre la imagen, la zona en cada uno de los cortes en los que se apreciaba el tumor. En la **Figura 9** se muestra un ejemplo del procedimiento.

Aunque se ha realizado un corregistro de las imágenes, se he preferido trabajar en primer lugar con las imágenes de TC, ya que la calidad y visualización es mucho más elevada que en las de PET.



**Figura 9.** Segmentación inicial de la lesión en imágenes TC de una paciente en varios cortes axiales. Las imágenes de la izquierda muestran el corte antes de dibujar sobre el tumor. Cada fila de imágenes representa el mismo corte.

Gracias a la radiación que emite el radiofármaco  $^{18}\text{F}$ -MISO visible en PET, a partir de las aproximaciones realizadas en TC y de la alineación de ambas imágenes, se han corregido las segmentaciones hasta obtener delineaciones lo más precisas posible de las lesiones (*Figura 10*).



**Figura 10.** Delineación de las segmentaciones en los cortes de la Figura 9 con el apoyo de la radiación del  $^{18}\text{F}$ -MISO en PET.

Sin embargo, como se ha comentado en el apartado *Base de datos*, no en todos los casos el tumor era fácilmente discernible en TC o PET, ya que, si los bordes no están claramente diferenciados, la tarea de la delineación se dificulta enormemente (*Figura 11.1*).



**Figura 11.** Ejemplo de tumor no discernible en PET y señalización por la Dra. en la segunda imagen.

Con la ayuda de la Dra. Francisca Mulero se ha podido segmentar un total de 32 casos más, aumentando así la N de trabajo para el posterior análisis. Mediante la señalización (ver *Figura 11.2*) de la lesión en un corte para cada caso, se ha dibujado la región de interés alrededor, para ese corte, y en los cinco cortes anteriores y siguientes, obteniendo así una segmentación completa 3D.

### 4. 3. 2 Análisis de texturas

Tras obtener las segmentaciones para los 84 casos estudiados, el siguiente paso es realizar el análisis de texturas sobre cada uno de ellos y obtener así la cuantificación de la lesión en estudio en forma de parámetros.

La función de análisis implementada es *Texturas.m*, se encuentra especificada en el ANEXO con mayor detalle y en la **Figura 12** se muestra un esquema-resumen del funcionamiento del código. El análisis de texturas se ha realizado a través de *MatLab R2018a*, herramienta software que ofrece un entorno de desarrollo integrado y un lenguaje de programación propio, y con la que se ha llevado a cabo todo el análisis correspondiente al procesamiento de las imágenes.

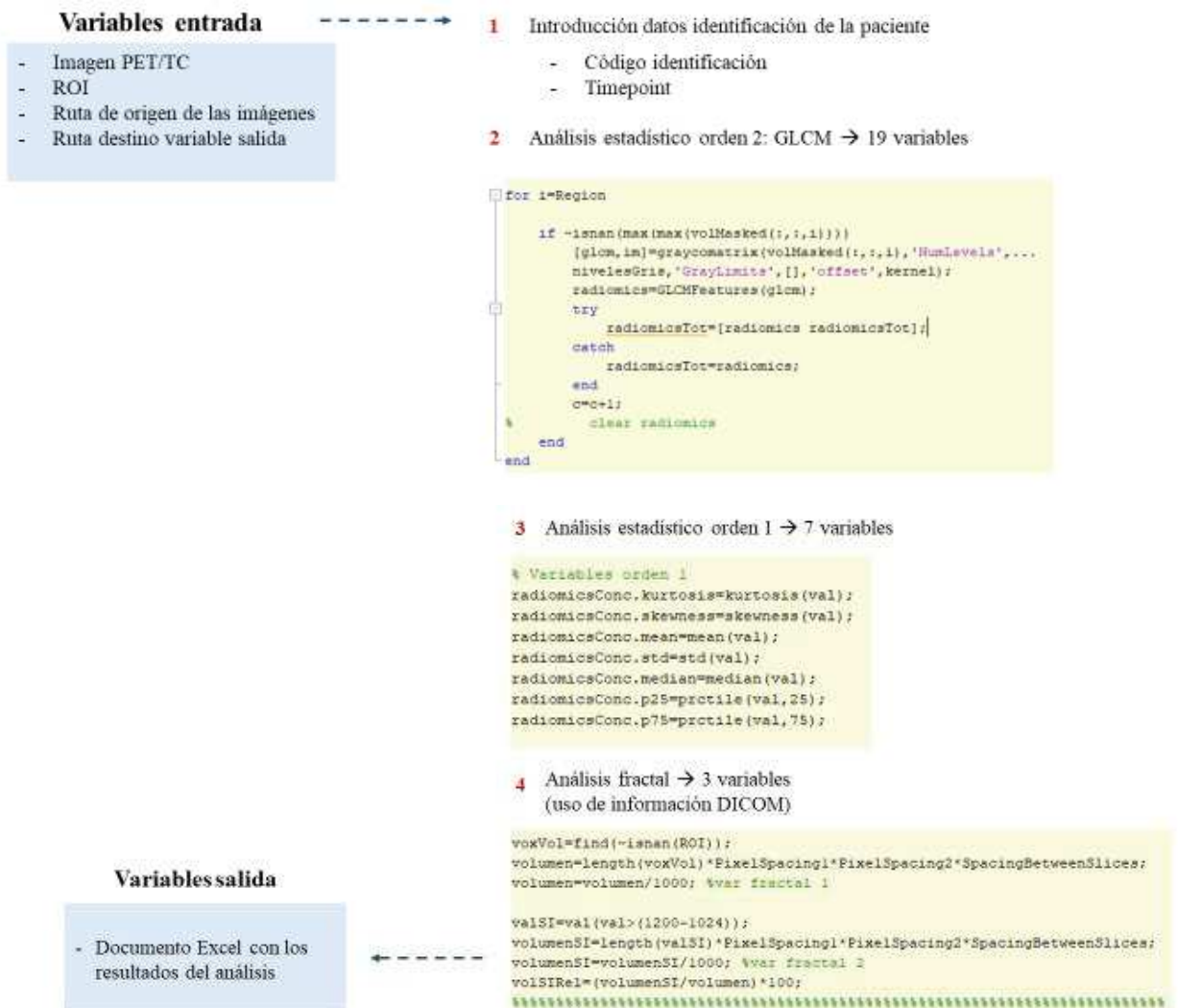


Figura 12. Diagrama de bloques de Texturas.m

Teniendo en cuenta que, de la base de casos seleccionada, las pacientes a las que se les llevó a cabo el tratamiento experimental constan de dos imágenes en diferentes instantes temporales, para responder al reto clínico de poder aportar biomarcadores de imagen de valor pronóstico en la fase diagnóstica, se ha decidido que para el análisis de este trabajo se tenga en cuenta únicamente las primeras imágenes o de instante temporal basal, ya que, al utilizarse tratamientos diferentes, es difícil que una característica basal prediga beneficio de ambos, salvo que sea una característica de beneficio específico de la quimioterapia.

Una vez discriminadas las imágenes de seguimiento, se ha utilizado el código especificado en cada uno de los 84 casos, tarea para la cual se tiene que introducir a la plataforma software los archivos en formato *NIfTI* de la imagen de las mamas y de la segmentación realizada, además de las rutas de origen y de almacenamiento de los resultados.

Como resultado, en la ruta destino seleccionada, se ha creado un archivo Excel *Texturas.xls*, en el cual, en formato tabla, se almacenan todas las variables de salida, y se va actualizando con cada caso analizado. La información que devuelve el programa incluye el *timepoint* seleccionado (0 o 1, en este caso basal 0), código de la paciente, apellido, nombre, identificación, sexo, nacimiento, fecha de toma de la imagen y los parámetros de cuantificación extraídos del análisis. Estas variables texturales suman un total de 28 y corresponden a parámetros de primer orden (kurtosis, asimetría, intensidad media, mediana y desviación estándar de la intensidad, y los percentiles 25 y 75 correspondientes), parámetros de segundo orden (autocorrelación, *cluster prominence*, *cluster shade*, contraste, correlación, entropía, diferencia de entropía, diferencia de varianza, energía, homogeneidad, probabilidad máxima, *dissimilarity*, etc. ver en 3.5.2. *Métodos estadísticos*) y dos parámetros relacionados con el análisis fractal, la dimensión fractal del volumen completo y la de la segmentación o ROI. Debido a la gran cantidad de información extraída, el siguiente paso consiste en la reducción de estos parámetros en aquellos biomarcadores clínicamente relevantes en este caso de estudio.

#### **4. 3. 3 Reducción de la dimensionalidad de los datos**

El número de parámetros descriptivos puede dar una idea de la complejidad de la información obtenida con la caracterización del tumor. No siempre, una mayor cantidad de

datos es sinónimo de información relevante y útil, es más, una elevada complejidad puede suponer un *overfitting* o sobreentrenamiento en el aprendizaje automático posterior. Por ello, es necesario realizar una reducción de la dimensionalidad de los datos a través de la priorización de los biomarcadores de mayor relevancia en la población.

Las técnicas de selección de los biomarcadores, de manera general, pueden agruparse según sean dependientes o independientes. Los métodos dependientes o *wrappers* utilizan un clasificador particular de precisión como medida de utilidad para un subconjunto de parámetros, sin embargo, esto puede suponer un elevado coste computacional y sobreentrenamiento del clasificador. Por otra parte, dentro de las técnicas independientes, existen los filtros, los cuales evalúan la estadística de los datos sin utilizar ningún clasificador particular, por lo que se extraen parámetros genéricos. El principal componente que incluyen los filtros es el índice de medida de relevancia que cuantifica la utilidad de incluir una característica particular en un conjunto [Alberich-Bayarri y Martí-Bonmatí, 2018].

Dentro de esta última técnica, uno de los métodos más simples para conocer la relevancia y relación entre los parámetros extraídos con el análisis de texturas, y el que se ha seguido en este trabajo, es el cálculo de matrices de correlación cruzada, por lo que se obtiene la correlación entre pares de parámetros. A la matriz resultante se le aplica a su vez un umbral, con el objetivo de identificar todos aquellos subconjuntos de características que están altamente relacionadas. Por último, dentro de cada uno de ellos, se elegirá un único parámetro según un máximo de relevancia, pasando a ser un biomarcador para este estudio [Alberich-Bayarri y Martí-Bonmatí, 2018].

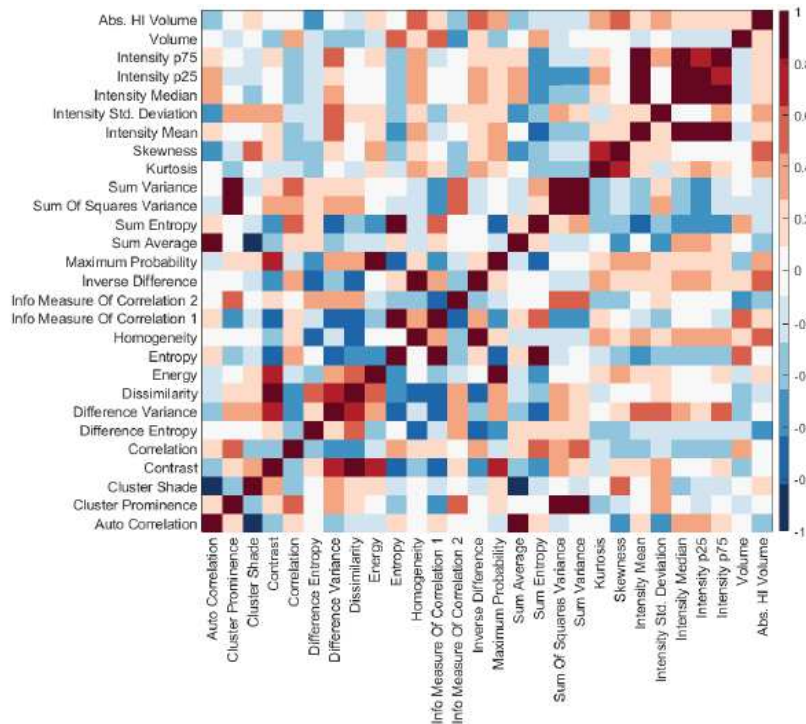
Por lo tanto, una vez se han obtenido las 28 características para cada caso, continuando con la herramienta de software *MatLab*, el primer paso que se ha realizado es la obtención de la matriz de correlación, la cual se ha calculado con una función, propia de la plataforma, *corrcoef.m* y ha devuelto una matriz de 28 x 28 con los coeficientes de correlación correspondientes a cada par de características, en la escala de -1 a 1, siendo -1 una correlación negativa, 0 no representa ninguna correlación y 1 se refiere a correlación positiva.

Para visualizar la matriz de correlación cruzada se han utilizado mapas de calor, en los cuales se puede conocer con mayor facilidad los subconjuntos de parámetros de mayor relación. La función utilizada es *HeatMap.m*, de *MatLab*, con la que se crea este mapa sin realizar ningún



tipo de agrupación de los parámetros, es decir, la representación mantiene el mismo orden que se sigue en la base de datos introducida.

En la **Figura 13** se muestra la relación entre pares de características, sin embargo, no es suficiente para seleccionar subconjuntos de parámetros, por lo que es necesario reagruparlos siguiendo unos criterios establecidos.



**Figura 13.** Representación de la matriz de correlación cruzada en un mapa de calor

La clusterización o agrupación consiste en la agrupación de los datos y es uno de los métodos más comunes de análisis no supervisado. La función *Clustergram.m* realiza un mapa de calor junto con una clusterización simple en forma de dendograma, lo que permite la visualización, no solo de aquellas características con un elevado grado de correlación (positivo y negativo) sino que las agrupa según estén relacionadas. Además, debido a que cada parámetro se mide en diferentes unidades, la propia función los normaliza para la representación. En la **Figura 14** se muestra este resultado, y en la **Figura 15** se hace referencia al mismo mapa, tras aplicarle un umbral de  $[-0.8, 0.8]$ , lo que permite una mejor visualización de las características más relevantes, necesarias para la reducción de dimensionalidad:



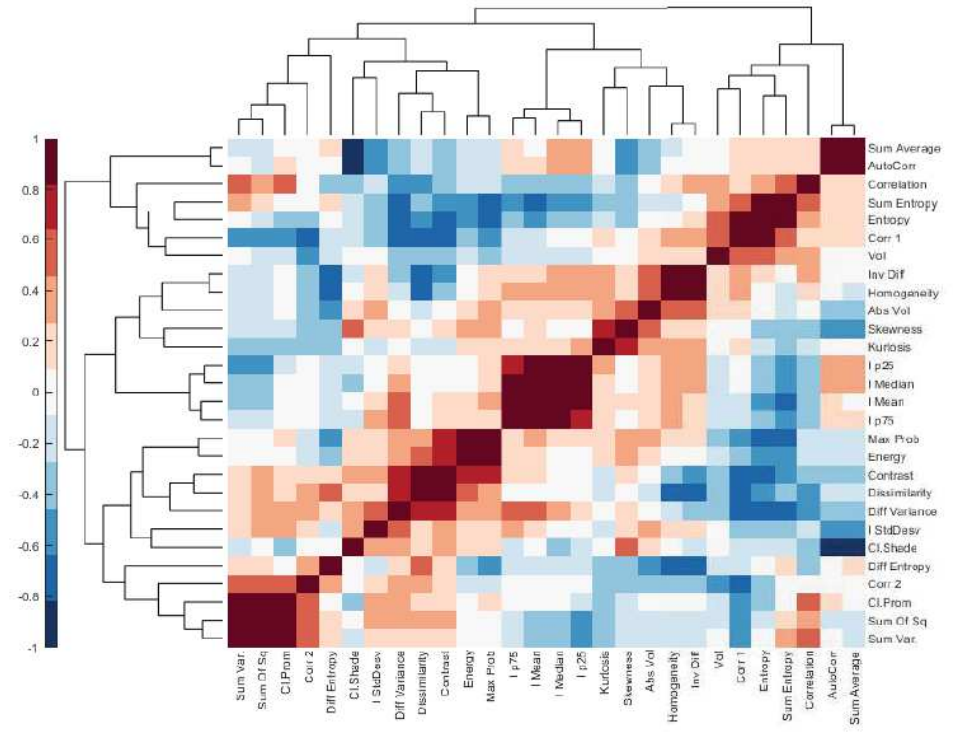


Figura 14. Agrupación de los datos según clusterización jerárquica y representación de los coeficientes de correlación

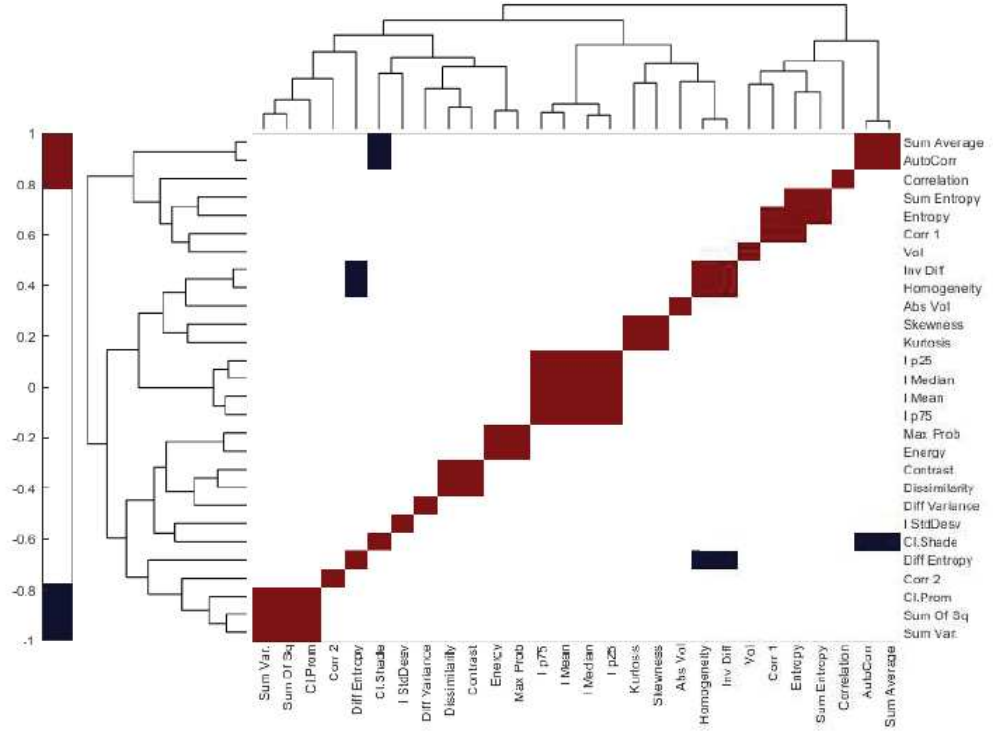


Figura 15. Mapa de calor tras aplicarle el umbral [-0.8, 0.8].

Como se puede apreciar en la anterior figura, tras aplicarle el umbral seleccionado y descartar así las variables que no están relacionadas, se puede visualizar fácilmente cuales son los grupos de mayor relevancia en el análisis de texturas, tanto con correlación directa (*colormap* rojo) como indirecta (*colormap* azul). En este tipo de mapas, la diagonal siempre representa el valor máximo de correlación, ya que se comparan las variables consigo mismas. Por lo tanto, hay que escoger las variables que se reúnen en bloque, es decir, que cuentan con más de un cuadrado unidad:

**Tabla 1.** Grupos de variables relacionados en más de 0,8. (Elevada correlación directa)

<b>Correlación directa &gt; 0.8</b>		
Sum Average Auto Correlation	Skewness Kurtosis	Dissimilarity Contrast
Sum Entropy Entropy	Intensity p25 Intensity Median	Cluster Prominence Sum of Squares Variance
Entropy Info Measure of Correlation1	Intensity Mean Intensity p75	Sum Variance
Inverse Difference Homogeneity	Energy Contrast	Dissimilarity Contrast

**Tabla 2.** Grupos de variables relacionados en menos de - 0,8. (Elevada correlación indirecta)

<b>Correlación indirecta &lt; -0.8</b>	
Cluster Shade	Difference Entropy
Auto Correlation	Homogeneity
Sum Average	Inverse Difference

A continuación, de cada grupo, se escogerá un solo parámetro, el de mayor relevancia para la población, para después utilizarlos como variables en el clasificador junto con otras variables independientes de cada paciente y las respuestas a los tratamientos.

#### 4.3.4 Clasificación de los datos

El criterio escogido para seleccionar una característica de cada grupo es ser el de mayor repercusión y máxima relevancia en la población en estudio, es decir, los que presenten una mayor variabilidad.

Debido a que los parámetros obtenidos se encuentran en diferentes unidades, antes de compararlos, se han normalizado con el *z-score*, obteniéndose así la desviación estándar de cada característica, para toda la población, en el mismo rango de medida. En la ecuación (3) se muestra la fórmula para calcular el valor *z* para cada variable *i* de cada paciente *j*, utilizando *x* como valor de la variable del paciente, y la media y la desviación estándar de cada variable para todo el conjunto de la población.

$$z_{i,j} = \frac{x_{i,j} - \mu_i}{\sigma_i} \quad (3)$$

De los 28 parámetros obtenidos con el análisis de texturas, se ha reducido a 21 (**Tablas 1 y 2**) según la matriz de correlación y el umbral escogido para reducir la dimensionalidad de los resultados. De estas 21 variables, se ha calculado el valor *z-score* de toda la matriz resultante (para cada valor en cada paciente) para poder hacer una comparación entre todas ellas. Una vez se tienen los parámetros normalizados, se ha calculado la variación entre el valor máximo y mínimo de cada variable, con el fin de escoger de cada subconjunto extraído de la **Figura 15**, el biomarcador que presente la variación máxima en la población. Este procedimiento se encuentra detallado en el ANEXO.

En la **Tabla 3**, se puede ver que son once los subconjuntos de parámetros con mayor correlación, tanto positiva como negativa, por lo que, se deberían obtener un máximo de 11 biomarcadores. Sin embargo, algunas características aparecen en varios grupos, por lo que, del total de variables extraídas del análisis de texturas, se han reducido a 8 biomarcadores, disminuyendo así la dimensionalidad de los resultados y la complejidad del modelo de clasificación, desarrollado en el siguiente apartado.

**Tabla 3.** Desviación estándar de las variables y señalización en amarillo de la varianza máxima en cada grupo.

<b>Variables</b>	<b><i>z-score</i></b>	<b>Variables</b>	<b><i>z-score</i></b>	<b>Variables</b>	<b><i>z-score</i></b>
Sum Average	4,4618	Skewness	5,8095	Cluster Prominence	4,8743
Auto Correlation	4,1887	Kurtosis	6,2695	Sum of Squares Variance	4,6024
Sum Entropy	4,1898	Intensity p25	4,0548	Sum Variance	4,8179
Entropy	4,5223	Intensity Median	4,8386	Cluster Shade	5,6920
Entropy	4,5223	Intensity Mean	5,9215	Auto Correlation	4,1887
Info Measure of Correlation1	4,1904	Intensity p75	7,6429	Sum Average	4,4618
Inverse Difference	5,6130	Energy	5,7638	Difference Entropy	5,2584
		Contrast	5,8012	Homogeneity	5,6702
Homogeneity	5,6702	Dissimilarity	5,2135	Inverse Difference	5,6130
		Contrast	5,8012		

#### 4. 4 Diseño de un clasificador

Por último, una vez se han obtenido los biomarcadores de mayor relevancia en la cuantificación de la hipoxia tumoral, el siguiente paso es diseñar un clasificador que, a partir de la información dada, sea capaz de dar respuestas clínicas o prediga las respuestas a tratamientos, con relación al cáncer de mama.

Para ello, se necesitan otras variables independientes de entrada que aporten información sobre la paciente y ayuden en este objetivo. Estos factores pronósticos son:

- ❖ **Medida de estadiaje tumoral (T):** hace referencia al tamaño y extensión del tumor primario. Se mide en una escala del 0 al 4, siendo mayor el tumor y la posibilidad de diseminarse a otros tejidos a mayor número.
- ❖ **Medida de estadiaje ganglionar (N):** se refiere a la extensión de cáncer que se ha diseminado a los ganglios linfáticos cercanos. Se clasifica del 0 al 3, y al igual que en caso anterior, a mayor número, más son los ganglios linfáticos a los que se ha extendido el tumor.
- ❖ **Edad:** existe cierta controversia sobre si este factor influye en la respuesta del tratamiento, pero debido a que sí es un factor de riesgo en el desarrollo del tumor, se ha incluido en el estudio.
- ❖ **Grado de agresividad del tumor:** al contrario que en el resto de los factores, en general, los tumores registrados con el grado más alto (3), han respondido mejor al tratamiento que los de menor grado (1).

- ❖ **Receptores hormonales (RRHH):** se clasifican en positivos (1) y negativos (0), siendo estos últimos los que responden mejor al tratamiento.

Para entrenar el modelo de clasificación, se ha utilizado *Classification Learner App*, aplicación propia de *MatLab*, lo que ha facilitado el proceso al continuar con el mismo lenguaje de programación que en el resto del trabajo.

La principal ventaja que ofrece esta aplicación, basada en aprendizaje supervisado, es la capacidad de entrenar múltiples modelos diferentes a la vez, a partir de las variables de entrada independientes y una variable respuesta o etiquetas. Algunos de los servicios que ofrece son: explorar los datos a través de diversas gráficas lo que permite una mejor interpretación, seleccionar diferentes parámetros, especificar esquemas de validación, etc. Los modelos de clasificación resultantes incluyen árboles de decisión, modelos de regresión logística, clasificadores k-NN (*k-Nearest Neighbor*) y SVM (*Support Vector Machines*), los cuales pueden ser comparados, para elegir así el mejor modelo de descripción con la automatización del entrenamiento [MathWorks, 2019].

A continuación, se explica el procedimiento seguido en *Classification Learner App* para entrenar los modelos de clasificación:

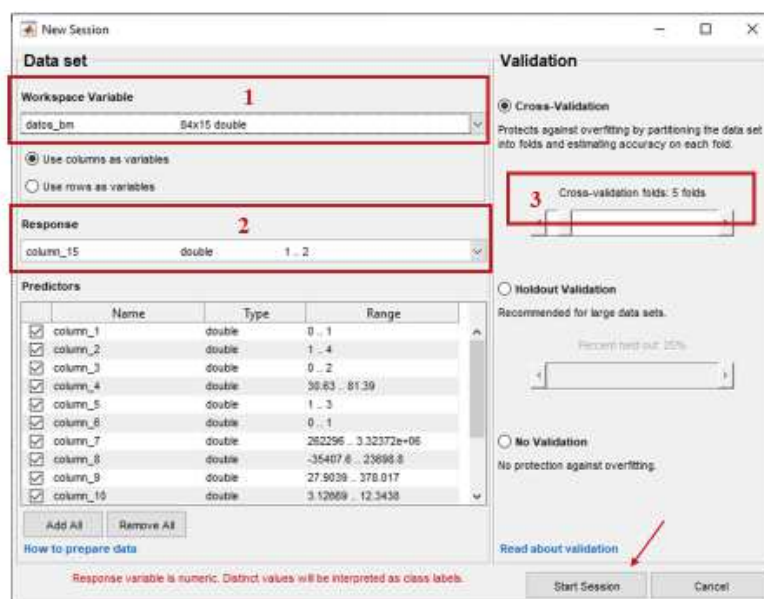


Figura 16. Ejemplo de pasos a seguir en Classification Learner

1. En primer lugar, se crea una nueva sesión en la aplicación y se selecciona en la ventana flotante la matriz con todas las variables con las que se quiere trabajar, siendo estas los cinco factores independientes explicados anteriormente (T, N, edad, grado y RRHH), los ocho biomarcadores escogidos y las etiquetas correspondientes a la respuesta. En total, una matriz de tamaño 84x15 (pacientes x variables).
2. A continuación, de la matriz de datos cargada, se especifica qué columna corresponde a las etiquetas para el aprendizaje supervisado. En este caso, la variable respuesta corresponde a la respuesta a los tratamientos estándar y experimental especificados en el apartado *Base de datos*, que, al ser dicotomizada, se ha simplificado de cuatro tipos de respuestas a dos, en función de si el resultado es positivo (1) o negativo (2).
3. Antes de comenzar la sesión, queda especificar la validación. Se ha escogido *Cross-Validation* con el objetivo de evitar el *overfitting* y garantizar la independencia entre los datos de entrenamiento y los de prueba.  
De esta manera, se ha escogido un valor k estándar igual a 5 particiones y se evalúa el modelo seleccionado variando las agrupaciones, además de calcular al final el error promedio. Este método proporciona una buena estimación de la predicción del modelo entrenado con todos los datos.
4. Por último, se empieza la sesión y se puede elegir que modelos se quiere entrenar, pero, como nos interesa saber cuál nos ofrece mejores resultados, se ha seleccionado el entrenamiento múltiple, lo que ha proporcionado resultados en 23 modelos, lo cuales se han comentado en el apartado en el *Capítulo 5. Resultados*.

# CAPÍTULO 5. RESULTADOS Y DISCUSIÓN

## 5.1 Métricas para la evaluación de los resultados

La principal ventaja que ofrece el uso de *Classification Learner App* frente a otros métodos de clasificación estadísticos simples es el aprendizaje supervisado que realiza sobre los datos. El aprendizaje supervisado es un algoritmo de *Machine Learning* que emplea un conjunto de datos conocidos de entrenamiento, entre los que se incluyen datos de entrada y valores de respuesta, clases o etiquetas de datos, para realizar predicciones.

Generalmente, el proceso de validación del aprendizaje automático puede complicarse, ya que dependiendo de cómo se haga la partición de los datos en los conjuntos de entrenamiento y prueba, se pueden obtener resultados diferentes en la precisión, afectando a la evaluación del rendimiento del modelo en función de esta métrica. Además, teniendo en cuenta que la base de datos sobre la que se está trabajando es de una  $N$  bastante pequeña, este problema puede verse en aumento. La solución aplicada a ello, como se ha comentado en el apartado 5.4 *Diseño de un clasificador*, es utilizar la validación cruzada (CV), en este caso, se ha tomado una  $k = 5$ , garantizando la validación de los resultados.

Esta aplicación ha permitido entrenar múltiples modelos en un tiempo mínimo y ofrece, además, gran cantidad de información sobre su rendimiento y métricas de validación, por ejemplo, la matriz de confusión, la curva ROC y el área debajo de la curva AUC. Los modelos de aprendizaje supervisado testados son *Decision Trees*, *Support Vector Machines (SVM)*, *Ensemble Classifiers* y *Nearest Neighbour Classifiers*. De cada uno de ellos, existen diferentes tipos, los cuales se muestran en las **Tablas 4-7**, junto con el porcentaje de precisión en la predicción tras la clasificación.

Para la evaluación de cada sistema de clasificación se ha estimado su rendimiento a partir de métricas, las cuales son las siguientes:

### ❖ Matriz de confusión

De un problema de clase  $n$ , la matriz de confusión (**Figura 17**) es una matriz  $n \times n$  en la que las filas se refieren a la clasificación real de los datos, y las columnas a la predicción realizada por el modelo. Con ella, de manera explícita y visual, se puede saber cuándo una clase es confundida por otra [Recuero, 2019].

		Predicción	
		Positivos	Negativos
Observación	Positivos	Verdaderos Positivos (VP)	Falsos Negativos (FN)
	Negativos	Falsos Positivos (FP)	Verdaderos Negativos (VN)

**Figura 17.** Estructura de la matriz de confusión

VP y VN se refieren a la cantidad de positivos y negativos, respectivamente, clasificados correctamente por el modelo, mientras que, FN es la cantidad de positivos que fueron clasificados incorrectamente como negativos y FP es la cantidad de negativos que fueron clasificados incorrectamente como positivos [Zelada, 2017]. A partir de estos índices es posible calcular otras métricas, como son:

$$Exactitud = \frac{VP + VN}{Total}$$

$$Tasa\ de\ error = \frac{FP + FN}{Total}$$

$$Sensibilidad = \frac{VP}{Total\ Positivos}$$

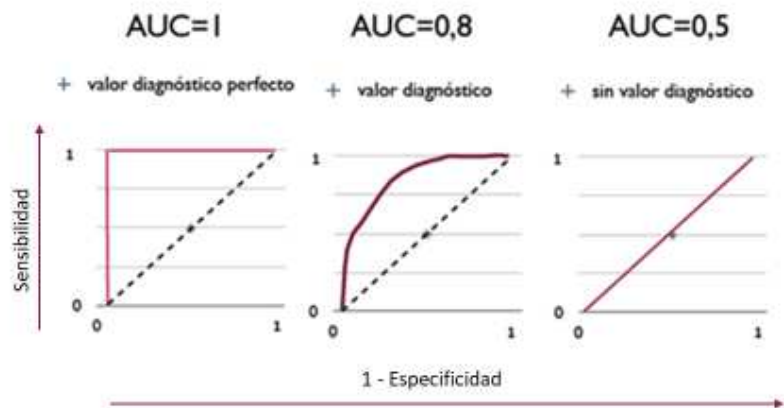
$$Especificidad = \frac{VN}{Total\ Negativos}$$

#### ❖ Curva ROC

La curva ROC es una gráfica, en un sistema de ejes cartesianos, en la que se representa en el eje de abscisas la probabilidad de falso positivo (sensibilidad) frente a la probabilidad de verdadero positivo (1 - especificidad), en el eje de ordenadas. Además, de la curva se puede extraer otro parámetro, *Area Under the Curve* (AUC), el cual representa la probabilidad de que un valor aleatorio sea clasificado correctamente [Parra, 2019].



La curva ROC representa el rendimiento del clasificador, enfrentado la probabilidad de falso positivo y la probabilidad de verdadero positivo. VPR se refiere a la sensibilidad, probabilidad de clasificar correctamente a una paciente cuya clase real coincida con la predicción, y FPR es igual a resta uno de la especificidad, es decir, probabilidad de clasificar a una paciente en una clase establecida como positiva, cuando la respuesta real es negativa.



**Figura 18.** Valor diagnóstico de las curvas ROC [Parra, 2019]

El mejor método posible de predicción situaría la curva en torno a la esquina superior de la gráfica, siendo el punto máximo (0,1), representando un 100% de sensibilidad (ningún falso negativo) y un 100% de especificidad (ningún falso positivo). Cuando la curva ROC recorre la diagonal y el factor AUC es menor a 0,50, se considera que el modelo no clasifica correctamente [Parra, 2019].

## 5.2 Resultados de los diferentes modelos de clasificación

A continuación, se van a presentar cuales son los modelos de clasificación que se han entrenado, junto con la exactitud de predicción de cada uno ellos, lo que va a permitir ver fácilmente cuál puede ser el mejor para la clasificación de estos datos.

### 5. 2. 1 Support Vector Machines (SVM)

SVM es un algoritmo de aprendizaje supervisado que se puede emplear para clasificación binaria o regresión. Se construye un hiperplano óptimo en forma de superficie de decisión, de modo que el margen de separación entre las dos clases en los datos se amplía al máximo. Los vectores de soporte se refieren a un pequeño subconjunto de las observaciones de entrenamiento que se utilizan como soporte para la ubicación óptima de la superficie de decisión.

En la construcción de la superficie de decisión solo hacen falta los vectores de soporte seleccionados de los datos de entrenamiento, para la cual es necesario utilizar el *kernel*, ya que los datos nunca se transforman específicamente al espacio de características. [MathWorks, 2019]. Dependiendo del *kernel* escogido para ello, existen diferentes tipos de SVM (**Tabla 4**), de los cuales, en todos ellos, se han obtenido resultados de exactitud superiores al 80%.

**Tabla 4.** Evaluación de modelos – Support Vector Machines y exactitud en la predicción de respuesta

<b>Support Vector Machines SVM</b>	
Modelos	Exactitud (%)
Linear SVM	<b>85,70%</b>
Quadratic SVM	84,50%
Cubic SVM	81,00%
Fine Gaussian SVM	<b>85,70%</b>
Medium Gaussian SVM	<b>85,70%</b>
Coarse Gaussian SVM	<b>85,70%</b>

### 5. 2. 2 Nearest Neighbour Classifiers (KNN)

KNN es un clasificador de aprendizaje basado en instancias, de forma que el modelo generalizado no se construye a partir del aprendizaje con datos de entrenamiento, sino que sucede en el mismo momento en el que se testan los datos de prueba. Sus principales ventajas son la flexibilidad, buen rendimiento y de funcionamiento sencillo: el algoritmo clasifica cada dato nuevo en el grupo que corresponda, según tenga k datos vecinos más cerca de una clase u otra [Flores, Ruiz, Peñalver y Zabala, 2019].

Al igual que en los modelos SVM, la mayoría de los clasificadores KNN (*Tabla 5*) obtienen una exactitud entorno al 80%:

**Tabla 5.** Evaluación de modelos – Nearest Neighbour Classifiers y exactitud en la predicción de respuesta

<i>Nearest Neighbour Classifiers KNN</i>	
Modelos	Exactitud (%)
Fine KNN	78,60%
Medium KNN	<b>85,70%</b>
Coarse KNN	<b>85,70%</b>
Cosine KNN	<b>85,70%</b>
Cubic KNN	<b>85,70%</b>
Weighted KNN	<b>85,70%</b>

### 5. 2. 3 Ensemble Classifiers

Los clasificadores de tipo *ensemble* son combinaciones de múltiples clasificadores, lo que supone, a menudo, ofrecer mejores resultados que cualquiera de ellos por separado y reduciendo los posibles errores que pudieran dar. Los clasificadores base pueden diferir en el algoritmo, parámetros, representación o conjunto de entrenamiento, pero con el aprendizaje en su conjunto, se busca reducir el sesgo y la varianza de la muestra [Acharya, 2019].

De los modelos que obtienen valores de exactitud del 85,70%, se obtienen los mismos resultados que en los modelos anteriores. Como se puede ver en la *Figura 19*, aunque muestren un rendimiento óptimo, el método de clasificación no es completamente correcto, ya que no ha aprendido a clasificar ambas clases, directamente se decanta por la clase mayoritaria (respuesta igual a 2).

**Tabla 6.** . Evaluación de modelos – Ensemble Classifiers y exactitud en la predicción de la respuesta

<i>Ensemble Classifiers</i>	
Modelos	Exactitud (%)
Boosted Trees	<b>85,70%</b>
Bagged Trees	<b>85,70%</b>
Subspace Discriminant	<b>85,70%</b>
Subspace KNN	81,70%
RUSBoosted Trees	47,60%

#### 5. 2. 4 Árboles de decisión

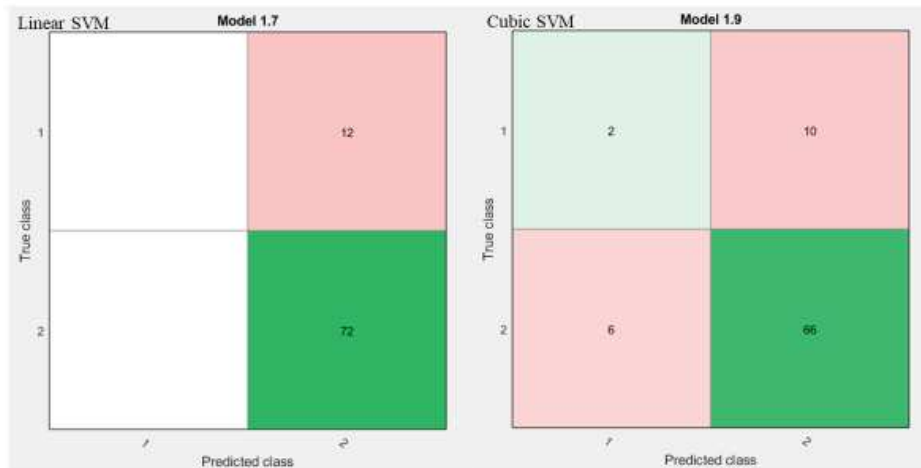
Por último, los árboles de decisión es un algoritmo que identifica la variable más significativa y su valor que proporciona los mejores conjuntos homogéneos de población. Potencian modelos predictivos de alta precisión, estabilidad y facilidad de interpretación [Parra, 2019]. En la **Tabla 7** se recogen los resultados de tres modelos de árboles de decisión que, aunque son menores en comparación con el resto de los modelos, siguen considerándose de elevada exactitud.

**Tabla 7.** Evaluación de modelos – Decision Tree y exactitud en la predicción de respuesta

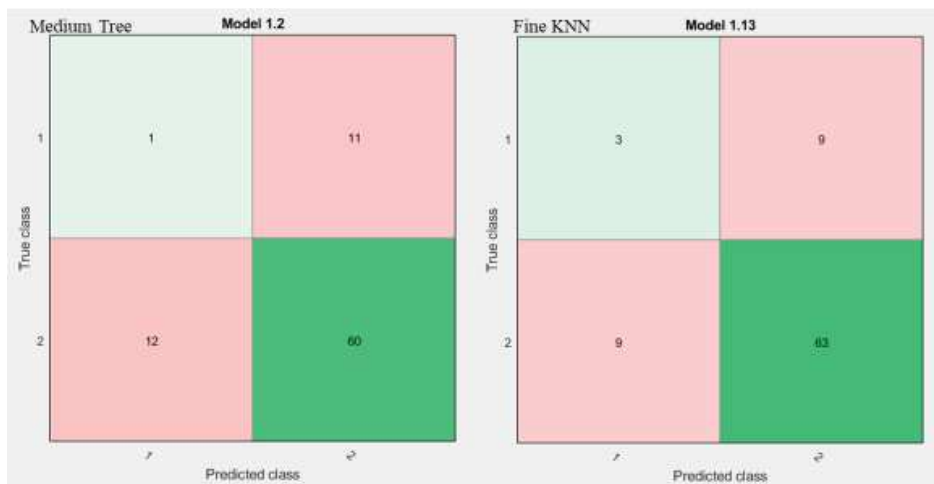
<i>Decision Tree</i>	
Modelos	Exactitud (%)
Fine Tree	72,60%
Medium Tree	72,60%
Coarse Tree	77,40%

En la mayoría de los modelos tipo SVM, KNN y *Ensemble*, el modelo clasifica correctamente las respuestas en un 85,70%, sin embargo, al obtener la matriz de confusión de uno de ellos, **Figura 19**, vemos que, debido a que existen una mayor cantidad de datos clasificados como 2 (no funcionamiento del tratamiento) respecto a 1, el modelo ha aprendido a clasificar todo el conjunto de datos como 2, obteniendo así, mejores resultados de rendimiento.

Por otro lado, si tomamos como ejemplo un modelo con una exactitud menor, como por ejemplo *Cubic SVM* con exactitud de 81%, o los modelos que se muestran en la **Figura 20**, se obtienen mejores resultados de clasificación de los datos, ya que, en estos casos, el modelo aprende a clasificar la clase minoritaria.

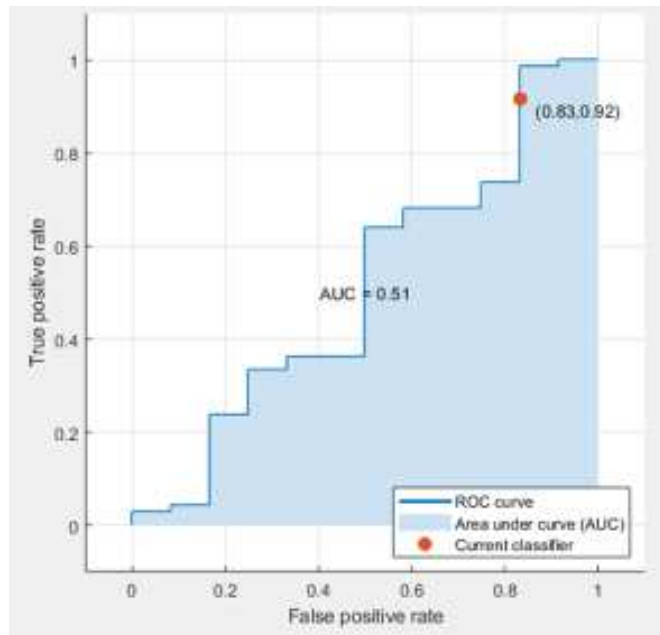


**Figura 19.** Comparación de matriz de confusión de modelos con exactitud 85,7% y matriz de confusión de modelos con exactitud 81%



**Figura 20.** Comparación de matriz de confusión de modelos con exactitud 72,6% y matriz de confusión de modelos con exactitud 78,6%

Debido a las limitaciones explicadas en mayor profundidad en el apartado 5.2 *Limitaciones y líneas futuras de trabajo*, el desbalanceo de los datos a favor de la clase 2, (72 pacientes están clasificadas como 2 y 12 como 1) hace que todos los modelos de aprendizaje supervisado aprendan, de manera general, que casi siempre que se analice un caso se clasifique en el grupo de respuesta 2, independientemente de cuales sean las variables pronostico independientes y los biomarcadores. Por ello, las curvas ROC resultantes no ofrecen buenos resultados y el valor del área bajo la curva (AUC) no es mayor a 0,55. En la **Figura 21** se muestra la curva del modelo *Cubic SVM* en la que, tomando la respuesta 2 como positiva, se obtiene una sensibilidad elevada, pero una especificidad muy baja.



**Figura 21.** Curva ROC para modelo *Cubic SVM* con exactitud igual a 81% y AUC 0.51

### 5. 3 Validación del análisis de texturas con un conjunto de datos balanceado

Con el objetivo de comprobar si el análisis de texturas y la medida de la hipoxia tumoral es un método válido para la caracterización en cáncer de mama, se ha decidido equilibrar la base de datos de trabajo, de manera que existan el mismo número de pacientes con respuesta 1 que con respuesta 2, para poder entrenar de nuevo un modelo de aprendizaje que ofrezca unos resultados válidos.

Para conseguir balancear los datos, se ha seleccionado el máximo de los casos de la respuesta minoritaria, siendo en total 12, y del conjunto de respuesta mayoritaria (72 casos), se ha seleccionado al azar el mismo número de casos, obteniendo así una muestra de datos de 24 pacientes.

```
data = xlsread('Texturas.xls'); % Variables para cada paciente
respuesta = xlsread('Respuesta.xlsx'); % Factores pronosticos
independientes y respuesta

% Buscamos que filas de pacientes corresponden al trat 1 y al trat 2
tr1 = find(respuesta(:,8)==1);
tr2 = find(respuesta(:,8)==2);

% Separamos en dos grupos LOS BIOMARCADORES
```

```

dat_tr1 = data(tr1,:);
dat_tr2 = data(tr2,:);

k = randsample([1:1:72],12) %Tomamos 12 numeros al azar PARA RESPUESTA 2
rand_tr2 = dat_tr2(k,:); % Grupo 2 al azar

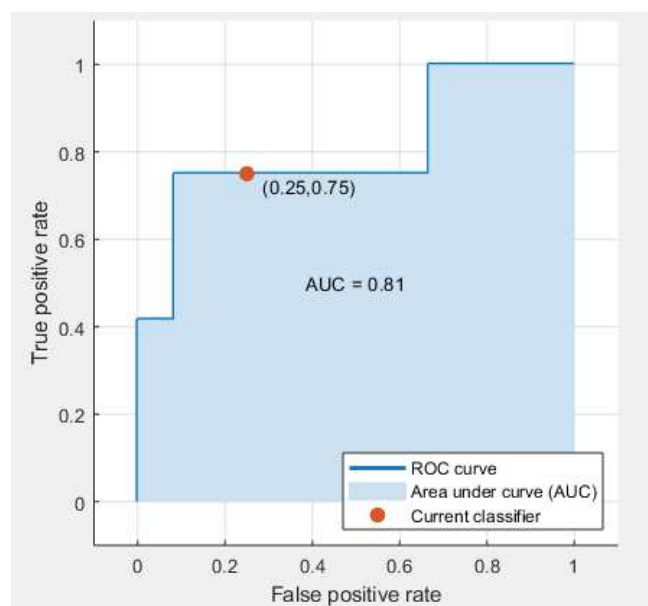
% Separamos en dos grupos LA RESPUESTA
respuesta_1 = respuesta(tr1,:);
respuesta_2 = respuesta(tr2,:);
DATOS = [dat_tr1; rand_tr2]; % 24 X 28

```

Tras repetir el análisis de texturas en este nuevo conjunto de datos, se ha vuelto a utilizar el *Classification Learner App* para entrenar los modelos de aprendizaje supervisado para la clasificación. De todos los modelos utilizados, el que ha proporcionado mejores resultados es *Subspace Discriminant (Ensemble Classifiers)* con una exactitud igual a 75%.

**Tabla 8.** Índices estadísticos del modelo de clasificación Subspace Discriminant

INDICES Y MÉTRICAS			
<i>Subspace Discriminant</i>			
VP	FP	VN	FN
9	3	9	3
Sensibilidad		Especificidad	
75%		75%	
Exactitud		Tasa de error	
75%		25%	



**Figura 22.** Curva ROC y Area Under the Curve del clasificador *Subspace Discriminant*

En la **Figura 22** se muestra la curva ROC resultante del modelo de clasificación. En primer lugar, el índice AUC (*Area Under the Curve*) es igual a 0,81, y por ser una métrica de comparación en la validación, puede considerarse de valor diagnóstico. Este valor se refiere a que existe un 81% de probabilidad de que la respuesta al tratamiento elegido como positivo de un paciente se prediga correctamente, frente a un paciente (de clase negativa) escogido al azar.

## 5.4 Discusión de los resultados obtenidos

A la vista de los resultados obtenidos en los apartados anteriores y del desbalanceo de la base de datos utilizada, se ha elegido que el modelo que mejor clasifica la muestra es *Subspace Discriminant (Ensemble Classifiers)*, porque, aunque en un primer momento, por los valores de exactitud iniciales, se obtengan mejores resultados con los modelos descritos en 5.2 *Resultados de los diferentes modelos de clasificación*, tras obtener el resto de las métricas, como la sensibilidad, la especificidad y el índice AUC, se puede ver una clara diferencia entre ellos (**Tabla 9**).

**Tabla 9.** Comparación de las métricas más importantes entre modelos con muestra desbalanceada y muestra balanceada

COMPARACIÓN DE MÉTRICAS							
<i>Cubic SVM</i>				<i>Subspace Discriminant</i>			
VP	FP	VN	FN	VP	FP	VN	FN
66	10	2	9	9	3	9	3
<b>Sensibilidad</b>		<b>Especificidad</b>		<b>Sensibilidad</b>		<b>Especificidad</b>	
91%		17%		75%		75%	
<b>Exactitud</b>		<b>Tasa de error</b>		<b>Exactitud</b>		<b>Tasa de error</b>	
81%		23%		75%		25%	
<i>Area Under the Curve</i>				<i>Area Under the Curve</i>			
0,51				0,81			

Por ello, aunque la reducción de la base de datos a 24 pacientes sea una reducción bastante grande del conjunto inicial, sirve para dar una idea de los resultados aproximados que se obtendrían con la clasificación de los datos, aunque lo ideal, sería utilizar una base del tamaño inicial que se tiene, pero balanceada lo más equitativamente posible.



En conclusión, se ha decidido elegir los resultados obtenidos con la muestra balanceada y con un tamaño de muestra menor, que con una base de datos mayor desbalanceada, ya que, es mejor elegir unos resultados más bajos, pero con sentido estadístico y equilibrados, que elegir un modelo con algunas métricas mejores *a priori*.

## 5.5 Limitaciones y líneas futuras de trabajo

El principal problema que se ha presentado a la hora de interpretar los resultados de los diferentes modelos de clasificación es que existe un desbalanceo de clases en la base de datos con la se está trabajando. El desbalanceo de clases es una característica del conjunto que se produce cuando una o más clases (clases minoritarias) se encuentra representada en menor medida que otras (clases mayoritarias). Esta cuestión supone un problema en el ámbito de la clasificación de nuevos datos ya que puede derivar en un deterioro notable en la eficiencia de nuestro clasificador [Calvino, 2017].

Por ello, y como se puede comprobar en los resultados y en la curva ROC particularmente, el clasificador diseñado desempeña una buena labor sobre la clase mayoritaria (respuesta igual a 2), ya que detecta una mayor presencia en la muestra, decantando la balanza sobre esta clase a la hora de clasificar.

En el ámbito de la medicina, este problema de desbalanceo se está abordando en la actualidad de forma activa y se están investigando continuamente nuevas técnicas para solventarlo. De esta manera, una posible línea futura de este trabajo sería afrontar el desbalanceo de datos, siguiendo diferentes técnicas de remuestreo, en las que se modificaría la distribución inicial de los datos para balancear las clases, o bien cambiar los algoritmos existentes para mejorar la predicción [Calvino, 2017].

Una vez se hubiese solventado este problema propio de la base de datos, con el análisis de texturas que se ha realizado en este trabajo, se podría plantear resolver otro tipo de preguntas clínicas, es decir, no solo predecir la respuesta a un tratamiento.

## CAPÍTULO 6. CONCLUSIONES

Este trabajo nació con el objetivo de evaluar el uso del análisis de texturas en imágenes de cáncer de mama, a partir de la medida de la hipoxia tumoral en este tipo de patologías. Una vez realizado este análisis se debe concluir qué parámetros realizan una mejor caracterización de este tipo de lesión, para así, posteriormente utilizarlos en el diseño de un clasificador cuyo fin principal sea servir de apoyo en la toma de decisiones del personal sanitario.

A la vista de los resultados obtenidos, los parámetros que mejor describen la heterogeneidad existente entre los tejidos tumorales y no tumorales se han reducido del total de 28 variables a 8 biomarcadores, siendo estos *Sum Average*, *Entropy*, *Homogeneity*, *Kurtosis*, *Intensity p75*, *Contrast*, *Cluster Prominence* y *Cluster Shade*, la mayoría obtenidos con el análisis estadístico de segundo orden, GLCM.

Para la validación de este método, se ha utilizado el aprendizaje supervisado, por el cual, a partir del entrenamiento de múltiples modelos de clasificación, se pueden obtener métricas que describan el alcance y la utilidad del trabajo. Para ello, no solo se han utilizado los biomarcadores extraídos del análisis, sino que, además es necesario utilizar factores independientes propios del paciente, junto con una variable respuesta con la que entrenar el modelo.

Debido al desbalanceo de la base de datos, con el objetivo de realizar una simulación de la clasificación del conjunto de datos y evaluar así el análisis, se ha reducido la muestra hasta conseguir equilibrarla y evitar clases minoritarias y mayoritarias de las respuestas al tratamiento. De ella, se han obtenido resultados válidos para el diagnóstico, superando el 80% en la probabilidad de acierto. Con ello, se cumple el fin inicial del trabajo, aportar información al personal sanitario en su toma de decisiones.

Tras la finalización del trabajo, queda en el aire conseguir balancear la muestra de datos en su totalidad, con lo que se podrían obtener resultados de mayor valor para el diagnóstico. Además, con este tipo de análisis se consigue extraer gran cantidad de información de las imágenes médicas, por lo que, dependiendo de cuál sea la necesidad requerida, se podrían evaluar otro tipo de respuestas clínicas, ya sean, la posible metastatización del tumor, la supervivencia del paciente, o la respuesta a otros tratamientos experimentales.



# ANEXO

## 1. Código MatLab de Análisis de Texturas

```
%% Texturas.m

% Ejemplo de análisis de texturas en el caso 2555:

clear all
clc
folder='D:\CNIO_Maria\SEGMENTACIONES FINALES\2555\ST0\TAC_2555_ST0'; %
Ruta del NIFTI original
folder2='D:\CNIO_Maria\SEGMENTACIONES FINALES'; % Ruta donde se crean y
almacenan resultados
file='20130603_14443941F18MISOs004a001.nii'; % Imagen
hdr=dicominfo(strcat(folder,'\IM0'));

Volumen=load_untouch_nii(strcat(folder,'\ ',file));
Volumen=Volumen.img; %Almacenamiento imagen TAC o PET

ROI=load_untouch_nii(strcat(folder,'\Mascara.nii')); % Segmentación
realizada
ROI=ROI.img; %Almacenamiento imagen ROI

Volumen=rot90(Volumen);
Volumen=(Volumen*hdr.RescaleSlope)+(hdr.RescaleIntercept);

ROI=rot90(ROI);
ROI=double(ROI);

% Introducción de datos de identificación:
Timepoint=input('introduce el timepoint:', 's');
Patient=input('introduce el paciente:', 's');

Tercio=0;

Volumen=double(Volumen);
ROI(ROI==0)=NaN;
volMasked=Volumen.*ROI;

nivelesGris=64;

kernel = [0 1; -1 1; -1 0; -1 -1];

c=0;

if Tercio==0
    Region=1:size(volMasked,3);
end

for i=Region

    if ~isnan(max(max(volMasked(:, :, i))))
```

```

[glcm,im]=graycomatrix(volMasked(:,:,i),'NumLevels',nivelesGris,'GrayLimits',[],'offset',kernel);
radiomics=GLCMFeatures(glcm);
try
    radiomicsTot=[radiomics radiomicsTot];
catch
    radiomicsTot=radiomics;
end
c=c+1;
% clear radiomics
end
end

stats=fieldnames(radiomicsTot);
for i=1:length(stats)

radiomicsConc.(char(stats(i)))=horzcat(radiomicsTot.(char(stats(i))));
valores=radiomicsConc.(char(stats(i)));
voxeles=find(~isnan(valores));
valores=valores(voxeles);
radiomicsConc.(char(stats(i)))=mean(valores);
end

%RADIOMICS DE CADA CORTE
for i=1:length(stats)
    for j=1:length(radiomicsTot)

radiomicsConcFull(j).(char(stats(i)))=mean(radiomicsTot(j).(char(stats(i)))));
    end
end

volMasked1=volMasked(:,:,Region);

vox=find(~isnan(volMasked1));
val=volMasked1(vox);
radiomicsConc.kurtosis=kurtosis(val);
radiomicsConc.skewness=skewness(val);
radiomicsConc.mean=mean(val);
radiomicsConc.std=std(val);
radiomicsConc.median=median(val);
radiomicsConc.p25=prctile(val,25);
radiomicsConc.p75=prctile(val,75);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%

%VOLUMEN/VOLUMEN SUPERINTENSOS
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%

PixelSpacing1=hdr.PixelSpacing(1);
PixelSpacing2=hdr.PixelSpacing(2);
SpacingBetweenSlices=hdr.SliceThickness;

voxVol=find(~isnan(ROI));
volumen=length(voxVol)*PixelSpacing1*PixelSpacing2*SpacingBetweenSlices;
volumen=volumen/1000;

valSI=val(val>(1200-1024));

```

```

volumenSI=length(valSI)*PixelSpacing1*PixelSpacing2*SpacingBetweenSlices;
volumenSI=volumenSI/1000;
volSIRel=(volumenSI/volumen)*100;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%

%ALMACENAMIENTO EN EXCEL
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
radiomics=radiomicsConc;
direccionExcel=folder2;
nombreexcel=strcat(direccionExcel,'\','Texturas.xls');

if exist(nombreexcel)~=2
    %CREAR EXCEL

    titulos={'Timepoint','Patient','Surname','Name','ID','Sex','Birthdate','S
    tudy Date','Auto Correlation','Cluster Prominence','Cluster
    Shade','Contrast','Correlation','Difference Entropy','Difference
    Variance','Dissimilarity','Energy','Entropy','Homogeneity','Information
    Measure Of Correlation 1','Information Measure Of Correlation 2','Inverse
    Difference','Maximum Probability','Sum Average','Sum Entropy','Sum Of
    Squares Variance','Sum Variance','Kurtosis','Skewness','Intensity
    Mean','Intensity Std. Deviation','Intensity Median','Intensity
    p25','Intensity p75','Volume','Abs. HI Volume','Rel. HI Volume'};
        estado=xlswrite(nombreexcel,titulos,'Hojal','A1');
    end

    disp('Almacenando en Excel');
    paramsExcel={Timepoint, Patient,hdr.PatientName.FamilyName,'-
    ',hdr.PatientID,hdr.PatientSex,hdr.PatientBirthDate,hdr.StudyDate,radiomi
    cs.autoCorrelation,radiomics.clusterProminence,radiomics.clusterShade,rad
    iomics.contrast,radiomics.correlation,radiomics.differenceEntropy,radiomi
    cs.differenceVariance,radiomics.dissimilarity,radiomics.energy,radiomics.
    entropy,radiomics.homogeneity,radiomics.informationMeasureOfCorrelation1,
    radiomics.informationMeasureOfCorrelation2,radiomics.inverseDifference,ra
    diomics.maximumProbability,radiomics.sumAverage,radiomics.sumEntropy,radio
    mics.sumOfSquaresVariance,radiomics.sumVariance,radiomics.kurtosis,radio
    mics.skewness,radiomics.mean,radiomics.std,radiomics.median,radiomics.p25
    ,radiomics.p75,volumen,volumenSI,volSIRel};
    [datos,texto]=xlsread(nombreexcel,'Hojal');
    tamDatos=size(datos);
    tamDatosFil=tamDatos(1);

    estado=xlswrite(nombreexcel,paramsExcel,'Hojal',strcat('A',num2str(tamDat
    osFil+2),'AJ',num2str(tamDatosFil+2)));

    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function [out] = GLCMFeatures(glcm)
%
% Features computed
% Autocorrelation: [2]
% Cluster Prominence: [2]
% Cluster Shade: [2]
% Contrast: [1]
% Correlation: [1]
% Difference entropy [1]
% Difference variance [1]

```

```

% Dissimilarity: [2]
% Energy: [1]
% Entropy: [2]
% Homogeneity: (Inverse Difference Moment) [2,1]
% Information measure of correlation1 [1]
% Informaiton measure of correlation2 [1]
% Inverse difference (Homogeneity in matlab): [3]
% Maximum probability: [2]
% Sum average [1]
% Sum entropy [1]
% Sum of sqaures: Variance [1]
% Sum variance [1]
%
% References:
% 1. R. M. Haralick, K. Shanmugam, and I. Dinstein, Textural Features of
% Image Classification, IEEE Transactions on Systems, Man and
Cybernetics,
% vol. SMC-3, no. 6, Nov. 1973
% 2. L. Soh and C. Tsatsoulis, Texture Analysis of SAR Sea Ice Imagery
% Using Gray Level Co-Occurrence Matrices, IEEE Transactions on
Geoscience
% and Remote Sensing, vol. 37, no. 2, March 1999.
% 3. D A. Clausi, An analysis of co-occurrence texture statistics as a
% function of grey level quantization, Can. J. Remote Sensing, vol. 28,
no.
% 1, pp. 45-62, 2002
%
%
% Started from Avinash Uppupuri's code on Matlab file exchange. It has
then
% been vectorized. Three features were not implemented correctly in that
% code, it has since then been changed. The features are:
% * Sum of squares: variance
% * Difference variance
% * Sum Variance

if ((nargin > 1) || (nargin == 0))
    error('Too many or too few input arguments')
else
    if ((size(glcm,1) <= 1) || (size(glcm,2) <= 1))
        error('The GLCM should be a 2-D or 3-D matrix. ');
    elseif ( size(glcm,1) ~= size(glcm,2) )
        error('Each GLCM should be square with NumLevels rows and
NumLevels cols');
    end
end

% Normalize the GLCMs
glcm = bsxfun(@rdivide,glcm,sum(sum(glcm)));

% Get size of GLCM
nGrayLevels = size(glcm,1);
nglcm = size(glcm,3);

% checked
out.autoCorrelation = zeros(1,nglcm); %
Autocorrelation: [2]
out.clusterProminence = zeros(1,nglcm); % Cluster
Prominence: [2]

```

```

out.clusterShade = zeros(1,nglcm); % Cluster
Shade: [2]
out.contrast = zeros(1,nglcm); % Contrast:
matlab/[1,2]
out.correlation = zeros(1,nglcm); % Correlation:
[1,2]
out.differenceEntropy = zeros(1,nglcm); % Difference
entropy [1]
out.differenceVariance = zeros(1,nglcm); % Difference
variance [1]
out.dissimilarity = zeros(1,nglcm); %
Dissimilarity: [2]
out.energy = zeros(1,nglcm); % Energy:
matlab / [1,2]
out.entropy = zeros(1,nglcm); % Entropy: [2]
out.homogeneity = zeros(1,nglcm); % Homogeneity:
[2] (inverse difference moment)
out.informationMeasureOfCorrelation1 = zeros(1,nglcm); % Information
measure of correlation1 [1]
out.informationMeasureOfCorrelation2 = zeros(1,nglcm); % Informaiton
measure of correlation2 [1]
out.inverseDifference = zeros(1,nglcm); % Homogeneity
in matlab
% out.inverseDifferenceMomentNormalized = zeros(1,nglcm); % Normalized
Homogeneity
% out.inverseDifferenceNormalized = zeros(1,nglcm); % Normalized
inverse difference
out.maximumProbability = zeros(1,nglcm); % Maximum
probability: [2]
out.sumAverage = zeros(1,nglcm); % Sum average
[1]
out.sumEntropy = zeros(1,nglcm); % Sum entropy
[1]
out.sumOfSquaresVariance = zeros(1,nglcm); % Sum of
sqaures: Variance [1]
out.sumVariance = zeros(1,nglcm); % Sum variance
[1]

glcmMean = zeros(nglcm,1);
uX = zeros(nglcm,1);
uY = zeros(nglcm,1);
sX = zeros(nglcm,1);
sY = zeros(nglcm,1);

% pX pY pXplusY pXminusY
pX = zeros(nGrayLevels,nglcm); % Ng x #glcms[1]
pY = zeros(nGrayLevels,nglcm); % Ng x #glcms[1]
pXplusY = zeros((nGrayLevels*2 - 1),nglcm); %[1]
pXminusY = zeros((nGrayLevels),nglcm); %[1]
% HXY1 HXY2 HX HY
HXY1 = zeros(nglcm,1);
HX = zeros(nglcm,1);
HY = zeros(nglcm,1);
HXY2 = zeros(nglcm,1);

% Create indices for vectorising code:
sub = 1:nGrayLevels*nGrayLevels;
[I,J] = ind2sub([nGrayLevels,nGrayLevels],sub);

% Loop over all GLCMs
for k = 1:nglcm

```



```

currentGLCM = glcm(:,:,k);
glcmMean(k) = mean2(currentGLCM);

% For symmetric GLCMs, uX = uY
uX(k) = sum(I.*currentGLCM(sub));
uY(k) = sum(J.*currentGLCM(sub));
sX(k) = sum((I-uX(k)).^2.*currentGLCM(sub));
sY(k) = sum((J-uY(k)).^2.*currentGLCM(sub));

out.contrast(k) = sum(abs(I-J).^2.*currentGLCM(sub)); %OK
out.dissimilarity(k) = sum(abs(I - J).*currentGLCM(sub)); %OK
out.energy(k) = sum(currentGLCM(sub).^2); % OK
out.entropy(k) = -
nansum(currentGLCM(sub).*log(currentGLCM(sub))); %OK
out.inverseDifference(k) = sum(currentGLCM(sub)./(1 + abs(I-J)
)); %OK
out.homogeneity(k) = sum(currentGLCM(sub)./(1 + (I -
J).^2)); %OK

% out.inverseDifferenceNormalized(k) = sum(currentGLCM(sub)./(1
+ abs(I-J)/nGrayLevels)); %OK
% out.inverseDifferenceMomentNormalized(k)= sum(currentGLCM(sub)./(1
+ ((I - J)/nGrayLevels).^2)); %OK

out.sumOfSquaresVariance(k) = sum(currentGLCM(sub).*((I -
uX(k)).^2)); %<----- N.B! Wrong implementation previously!!
out.maximumProbability(k) = max(currentGLCM(:));

pX(:,k) = sum(currentGLCM,2); %OK
pY(:,k) = sum(currentGLCM,1)'; %OK

tmp1 = [(I+J)' currentGLCM(sub)'];
tmp2 = [abs((I-J))' currentGLCM(sub)'];
idx1 = 2:2*nGrayLevels;
idx2 = 0:nGrayLevels-1;
for i = idx1
    pXplusY(i-1,k) = sum(tmp1(tmp1(:,1)==i,2));
end

for i = idx2
    pXminusY(i+1,k) = sum(tmp2(tmp2(:,1)==i,2));
end

% These can be evaluated for all GLCMs simultaneously, no k-index
% missing. We need the results further down so I keep it in the loop.
out.sumAverage = sum(bsxfun(@times,idx1',pXplusY));
out.sumEntropy = -nansum(pXplusY.*log(pXplusY)); %OK
out.differenceEntropy = -nansum(pXminusY.*log(pXminusY)); %OK
out.differenceVariance(k) = sum((idx2-
out.dissimilarity(k)).^2'.*pXminusY(idx2+1,k)); %<----- N.B! Wrong
implementation previously!! Dissimilarity is "difference Average"
out.sumVariance(k) = sum((idx1-
out.sumAverage(k)).^2'.*pXplusY(idx1-1,k)); %<----- N.B! Wrong
implementation previously AND in [1]

HXY1(k) = -
nansum(currentGLCM(sub)'.*log(pX(I,k).*pY(J,k))); %OK
HXY2(k) = -
nansum(pX(I,k).*pY(J,k).*log(pX(I,k).*pY(J,k))); %OK
HX(k) = -nansum(pX(:,k).*log(pX(:,k))); %OK

```

```

HY(k) = -nansum(pY(:,k).*log(pY(:,k))); %OK

out.autoCorrelation(k) = sum(I.*J.*currentGLCM(sub));
out.clusterProminence(k) = sum((I+J-uX(k)-
uY(k)).^4.*currentGLCM(sub)); %OK
out.clusterShade(k) = sum((I+J-uX(k)-
uY(k)).^3.*currentGLCM(sub)); %OK
out.correlation(k) = (out.autoCorrelation(k) -
uX(k).*uY(k))./(sqrt(sX(k).*sY(k))); %OK

out.informationMeasureOfCorrelation1(k) = (out.entropy(k)-
HXY1(k))./(max(HX(k),HY(k))); %OK
out.informationMeasureOfCorrelation2(k) = (1 - exp(-2.*(HXY2(k)-
out.entropy(k))) ).^(1/2); %OK

end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

## 2. Código MatLab *Data Reduction*

```
% Obtención de los mapas de color de matriz de correlación
data = xlsread('Texturas.xls');
r = find(data(:,1) == 1); % El timepoint está reflejado en la primera
columna
data(r,:)=[]; % Nos quedamos únicamente con datos de timepoint 0
data=data(:,9:end);

matriz_r = corrcoef(data); % Cálculo matriz de correlación cruzada

parametros = {'Auto Correlation','Cluster Prominence','Cluster
Shade','Contrast','Correlation','Difference Entropy','Difference
Variance','Dissimilarity','Energy','Entropy','Homogeneity','Info Measure
Of Correlation 1','Info Measure Of Correlation 2','Inverse
Difference','Maximum Probability','Sum Average','Sum Entropy','Sum Of
Squares Variance','Sum Variance','Kurtosis','Skewness','Intensity
Mean','Intensity Std. Deviation','Intensity Median','Intensity
p25','Intensity p75','Volume','Abs. HI Volume'};

% Mapa de calor sin agrupar
h1 = HeatMap(matriz_r, 'ColorMap', redbluecmap,
'RowLabels',parametros,'ColumnLabels',parametros)

% Mapa de calor agrupado
cg = clustergram(matriz_r,'Colormap', redbluecmap)
set(cg,'RowLabels',parametros,'ColumnLabels',parametros)

%Umbralización: para mantener la reorganización, la umbralización se
realiza utilizando un colormap propio que me indique unicamente los
valores que quiero.

mymap = [0 0 0.2 ;1 1 1; 1 1 1; 1 1 1; 1 1 1;1 1 1; 1 1 1; 1 1 1; 0.5 0
0];

cgu = clustergram(matriz_r,'Colormap', mymap) % Mapa de calor con umbral
set(cgu,'RowLabels',parametros,'ColumnLabels',parametros)
```

### 3. Código MatLab clasificación de biomarcadores

```
data = [datar(:,1) datar(:,2) datar(:,3) datar(:,4) datar(:,6) datar(:,8)
datar(:,9) datar(:,10) datar(:,11) datar(:,12) datar(:,14) datar(:,16)
datar(:,17) datar(:,18) datar(:,19) datar(:,20) datar(:,21) datar(:,22)
datar(:,24) datar(:,25) datar(:,26)]; % 21 parámetros

nombres = {'Auto Correlation','Cluster Prominence','Cluster
Shade','Contrast','Difference
Entropy','Dissimilarity','Energy','Entropy','Homogeneity','Info Measure
Of Correlation 1','Inverse Difference','Sum Average','Sum Entropy','Sum
Of Squares Variance','Sum Variance','Kurtosis','Skewness','Intensity
Mean','Intensity Median','Intensity p25','Intensity p75'};

[Z3,mu,sigma] = zscore(data); % Normalización de las variables con zscore

% Obtención de la variación máxima en la población

rango = zeros(21,2);

for i=1:21
    rango(i,1)= min(Z3(:,i));
    rango(i,2)=max(Z3(:,i));
end

diff = rango(:,2)- rango(:,1);

% Una vez se obtiene esta diferencia, para cada subconjunto señalado en
la figura 11, se escoge el mayor → en total 8

bm = [datar(:,2) datar(:,3) datar(:,4) datar(:,8) datar(:,9) datar(:,12)
datar(:,16) datar(:,21)];
```



## BIBLIOGRAFÍA

- [1] Asociación Española Contra el Cáncer [aecc] (2018). *Cáncer de mama*. Recuperado 16 abril 2019, de <https://www.aecc.es/es/todo-sobre-cancer/tipos-cancer/cancer-mama/que-es-cancer-mama>
- [2] Asociación Española Contra el Cáncer [aecc] (2018). Pronóstico del cáncer de mama: mortalidad y esperanza de vida. Recuperado 12 abril 2019, de <https://www.aecc.es/es/todo-sobre-cancer/tipos-cancer/cancer-mama/mas-informacion/evolucion-cancer-mama>
- [3] American Society of Oncology [ASO] (2018, octubre). Cáncer de mama: Detección. Recuperado 31 agosto 2019, de <https://www.cancer.net/es/tipos-de-cancer/cancer-de-mama/deteccion>
- [4] Alberich-Bayarrí, A. (2018, 9 abril). Biomarcadores de imagen. Recuperado 5 mayo 2019, de <https://quibim.com/tag/biomarcadores-de-imagen/>
- [5] Strimbu, K., and Tavel, J. A. (2010). What are biomarkers? *Curr. Opin. HIV AIDS*. 5, 463–466. doi: 10.1097/COH.0b013e32833ed177
- [6] Zikos, D., & Delellis, N. (2018). CDSS-RM: A clinical decision support system reference model. *BMC Medical Research Methodology*, 18(1), 1–14. <https://doi.org/10.1186/s12874-018-0587-6>
- [7] Asociación Española Contra el Cáncer [aecc] (2018). Tipos de cáncer. Recuperado 31 agosto 2019, de <https://www.aecc.es/es/todo-sobre-cancer/tipos-cancer>
- [8] Instituto Nacional del Cáncer [INC] (2019). ¿Qué es el cáncer? Recuperado 31 agosto 2019, de <https://www.cancer.gov/espanol/cancer/naturaleza/que-es>
- [9] American Society of Clinical Oncology [ASCO] (2018, octubre). Cáncer de mama: Introducción. Recuperado 12 abril 2019, de <https://www.cancer.net/es/tipos-de-cancer/cancer-de-mama/diagnostico>

- [10] American Society of Clinical Oncology [ASCO] (2018, octubre). Cáncer de mama: Diagnóstico. Recuperado 12 abril 2019, de <https://www.cancer.net/es/tipos-de-cancer/cancer-de-mama/diagnostico>
- [11] American Society of Clinical Oncology [ASCO] (2018, octubre). Cáncer de mama: Factores de riesgo y prevención. Recuperado 12 abril 2019, de <https://www.cancer.net/es/tipos-de-cancer/cancer-de-mama/factores-de-riesgo-y-prevencion>
- [12] American Cancer Society [ACS] (2016, enero). Estudios por imágenes (Radiología) y cáncer. Recuperado 16 mayo 2019, de <https://www.cancer.org/es/tratamiento/como-comprender-su-diagnostico/pruebas/estudios-por-imagenes-y-el-cancer.html>
- [13] Dra, Benito, Á., María, D., & García, C. (2012). Capítulo 1. Introducción al diagnóstico por la imagen de la mama. Sociedad Española de Radiología Médica, 1–21. Recuperado 5 mayo 2019, de [http://www.sedim.es/nueva/wp-content/uploads/2015/01/Capitulo\\_1\\_Introduccion.pdf](http://www.sedim.es/nueva/wp-content/uploads/2015/01/Capitulo_1_Introduccion.pdf)
- [14] Apesteguía, L. (2019). *La Tomosíntesis: Mamografía en 3D*. SESPM. Recuperado 4 septiembre 2019, de <https://www.sespm.es/la-tomosintesis-mamografia-en-3d/>
- [15] Arvelo, F., & Cotte, C. (2009). Hipoxia en la malignidad del cáncer. Revisión. *Investigacion Clinica*, 50(4), 529–546. Recuperado de <https://pdfs.semanticscholar.org/76c2/ce30a8c9c6d4da6d304418724c20530d6c95.pdf>
- [16] Rajendran, J., & Krohn, K. (2008). F18 Fluoromisonidazole for Imaging Tumor Hypoxia: Imaging the Microenvironment for Personalized Cancer Therapy. *Bone*, 23(1), 1–7. <https://doi.org/10.1038/jid.2014.371>
- [17] Instituto Nacional del Cáncer [INC] (2019). Tratamiento del cáncer de seno (mama) – Versión para pacientes. Recuperado 26 abril 2019, de <https://www.cancer.gov/espanol/publicaciones/diccionario/def/18f-miso>
- [18] Scalco, E., & Rizzo, G. (2017). Texture analysis of medical images for radiotherapy applications. *British Journal of Radiology*, 90(1070). <https://doi.org/10.1259/bjr.20160642>

- [19] Itzel, M., Vázquez, J.A., Díaz, J.L. y Hernández, C. (2018, septiembre). Análisis de textura en imágenes de resonancia magnética para el diseño de implantes óseos de cráneo. Research Gate. Recuperado 01 septiembre 2019, de [https://www.researchgate.net/publication/328080496\\_ANALISIS\\_DE\\_TEXTURA\\_EN\\_IMAGENES\\_DE\\_RESONANCIA\\_MAGNETICA\\_PARA\\_EL\\_DISENO\\_DE\\_IMPLANTES\\_OSEOS\\_DE\\_CRANEO](https://www.researchgate.net/publication/328080496_ANALISIS_DE_TEXTURA_EN_IMAGENES_DE_RESONANCIA_MAGNETICA_PARA_EL_DISENO_DE_IMPLANTES_OSEOS_DE_CRANEO)
- [20] Benet, M. (2015). Análisis de texturas de imágenes de resonancia magnética de tumores cerebrales para la caracterización y clasificación de distintas regiones de interés. RiuNet Repositorio UPV, 84. Recuperado de <https://riunet.upv.es/handle/10251/67519>
- [21] Universidad Politécnica de Valencia [UPV]. Tratamiento de imágenes digitales mediante wavelets. Recuperado 31 agosto 2019, de <http://personales.upv.es/aperis/wavelets/imagenes/indice.htm>
- [22] Epifanio, I. (2002). Descripción De Texturas Aplicaciones a Su Comprensión Y Clasificación. 4. Recuperado de [https://www.uv.es/vista/vistavalencia/papers/tesis/tesis\\_irene.pdf](https://www.uv.es/vista/vistavalencia/papers/tesis/tesis_irene.pdf)
- [23] García, S. H., Nicot, M. E., & Álvarez, F. E. C. Análisis de Texturas en Imágenes Médicas. (pp. 1-10). Recuperado mayo 2019 de <http://karin.fq.uh.cu/acc/2015/CNE/122-2015/ACC%20Silena%20Herold%202015%20-%20OK/ACC%20Silena%20Herold%202015%20-%20OK/07%20contribuciones%20a%20eventos/18.pdf>
- [24] Alberich-Bayarrí, A & Martí-Bonmatí, L. (julio 2018). Imaging Biomarkers: Development and Clinical Integration. Edición: Softcover
- [25] Alobaidli, S., McQuaid, S., South, C., Prakash, V., Evans, P., & Nisbet, A. (2014). The role of texture analysis in imaging as an outcome predictor and potential tool in radiotherapy treatment planning. *British Journal of Radiology*, 87(1042), 5–14. <https://doi.org/10.1259/bjr.20140369>



- [26] Benoit, B. Fractales. Recuperado 01 septiembre 2019, de [http://www.ub.edu/matefest\\_infifest2011/triptics/fractal.pdf](http://www.ub.edu/matefest_infifest2011/triptics/fractal.pdf)
- [27] Navas, J., Esteban, F. J., Ruiz, J., & Quesada, J. M. (n.d.). La dimensión fractal como método de diagnóstico de enfermedades neurodegenerativas. Universidad de Jaén. Recuperado 01 agosto 2019, de [http://matema.ujaen.es/jnavas/web\\_modelos/curso14\\_15/presentacion%20dimension%20sevilla%2015.pdf](http://matema.ujaen.es/jnavas/web_modelos/curso14_15/presentacion%20dimension%20sevilla%2015.pdf)
- [28] Fuster, E. (2018). Unidad 2: herramientas para el soporte a la decisión médica. Universidad Politécnica de Valencia (UPV).
- [29] Tresserra, F., Martínez, M. Á., González-Cao, M., Rodríguez, I., Viteri, S., Baulies, S., & Fábregas, R. (2013). Respuesta patológica a la quimioterapia neoadyuvante: Correlación entre 2 sistemas de gradación histológica. *Revista de Senología y Patología Mamaria*, 26(3), 77–84. Recuperado 01 septiembre 2019, de <https://doi.org/10.1016/j.senol.2013.01.004>
- [30] MathWorks. (s.f.). *Train models to classify data using supervised machine learning - MATLAB* - MathWorks España. Recuperado 20 agosto, 2019, de <https://es.mathworks.com/help/stats/classificationlearner-app.html>
- [31] Recuero, P. (2019). Machine Learning a tu alcance: La matriz de confusión – Think Big Empresas. Recuperado 24 agosto, 2019 de <https://empresas.blogthinkbig.com/ml-a-tu-alcance-matriz-confusion/>
- [32] Zelada, C. (2017, mayo). Evaluación de modelos de clasificación - RPubs. Recuperado 20 agosto, 2019, de <https://rpubs.com/chzelada/275494>
- [33] Parra, F. (2019). Estadística y Machine Learning con R. Recuperado 26 agosto 2019, de [https://rstudio-pubs-static.s3.amazonaws.com/293405\\_4029f1f23f834b7195189d5504a436b2.html](https://rstudio-pubs-static.s3.amazonaws.com/293405_4029f1f23f834b7195189d5504a436b2.html)
- [34] Flores, D., Ruiz, S., Ruiz, S., Ruiz, S., Peñalver, A., Zabala, A., & Peñalver, A. (2019). El algoritmo K-NN y su importancia en el modelado de datos. Recuperado 26 agosto 2019, de <https://www.analiticaweb.es/algoritmo-knn-modelado-datos/>

- [35] Acharya, T. (2019). Advanced Ensemble Classifiers - Medium. Recuperado 29 agosto 2019, de <https://towardsdatascience.com/advanced-ensemble-classifiers-8d7372e74e40>
- [36] Calvino, H. (2017). Métodos para la mejora de predicciones en clases desbalanceadas en el estudio de bajas de clientes (CHURN) Español. Eio.Usc.Es, (M). Recuperado 26 agosto, 2019, de [http://eio.usc.es/pub/mte/descargas/ProyectosFinMaster/Proyecto\\_1469.pdf](http://eio.usc.es/pub/mte/descargas/ProyectosFinMaster/Proyecto_1469.pdf)

## **PRESUPUESTO**

## 1. INTRODUCCIÓN Y OBJETIVOS

El objetivo de esta parte del trabajo es realizar una valoración económica de los recursos empleados en el desarrollo de este trabajo. Para ello, en primer lugar, se van a tratar los costes parciales, entre los que se incluyen el coste de personal y el coste material, para después obtener el presupuesto total del proyecto.

## 2. COSTES PARCIALES

### 2.1 Coste de Personal

El coste de personal corresponde a los recursos humanos que participan en la elaboración de este trabajo. Para realizar este cálculo, se ha tenido en cuenta el salario medio de las tres personas participantes, además del tiempo dedicado al proyecto. En la **Tabla 10** se muestra el coste estimado de cada persona y el global del personal:

**Tabla 10.** Cuadro de precios Personal

Nº	Perfil	Uds.	Cantidad	Precio Unitario (€)	Total
1	Tutor del TFG	h	30	40.00 (€/h)	1200.00 €
2	Cotutor del TFG	h	30	40.00 (€/h)	1200.00 €
3	Autor del TFG	h	300	20.00 (€/h)	6000.00 €
				<b>TOTAL</b>	<b>8400.00 €</b>

Para ello, se ha tenido en cuenta la participación de Ignacio Bosch Roig y Ángel Alberich-Bayarrí como tutor y cotutor, respectivamente, y de María Mancebo González como autora del trabajo.

## 2. 2 Coste de Herramientas Software

En este apartado se tratan los recursos software empleados en el trabajo. En este caso, este tipo de recursos hace referencia a todas las plataformas utilizadas para la visualización y procesamiento de las imágenes médicas, así como herramientas de programación y aplicaciones ofimáticas de escritorio. En la **Tabla 11** se recogen los costes de cada una de las licencias junto con el período y el intervalo de amortización, ya que no todas ellas se han obtenido específicamente para este proyecto:

**Tabla 11.** Cuadro de precios Herramientas Software

Nº	Licencia	Uds	Cantidad	Precio unitario (€)	Período de amortización (años)	Intervalo amortizado (meses)	Total (€)
1	Office 365	u	1	149.00 €	1 año	12 meses	149.00 €
2	MatLab 2018b	u	1	250.00 €	1 año	12 meses	250.00 €
3	RadiAnt DICOM Viewer	u	1	0.00 €	1 año	4 meses	0.00 €
4	ITK-Snap	u	1	0.00 €	1 año	4 meses	0.00 €
5	Elastix	u	1	0.00 €	1 año	4 meses	0.00 €
						<b>TOTAL</b>	<b>399.00 €</b>

### 2.3 Coste de Herramientas Hardware

Por último, en la **Tabla 12** se incluye el coste del único recurso Hardware utilizado, un ordenador portátil Lenovo Yoga 520 :

**Tabla 12.** Cuadro de precios Herramientas Hardware

Nº	Licencia	Uds	Cantidad	Precio unitario (€)	Período de amortización (años)	Intervalo amortizado (meses)	Total (€)
1	Ordenador portátil	u	1	799.00 €	5 años	4 meses	266.33 €
						<b>TOTAL</b>	<b>266.33 €</b>

### 3. PRESUPUESTO TOTAL

Por último, en la *Tabla 13* se muestra cuál ha sido el coste total del trabajo, en el que se incluyen los costes parciales desglosado en el apartado anterior, a los que se les ha sumado los porcentajes correspondientes a los gastos generales y el beneficio industrial obtenido. A esta suma, se le añade el impuesto sobre el valor añadido (IVA) igual al 21%.

**Tabla 13.** Cálculo del presupuesto total del trabajo

<b>DESCRIPCIÓN</b>	<b>IMPORTE (€)</b>
<b>Coste de Personal</b>	8400.00 €
<b>Coste de Herramientas Software</b>	399.00 €
<b>Coste de Herramientas Hardware</b>	266.33 €
<b>Total presupuesto recursos</b>	<b>9065.33 €</b>
<b>13% Gastos Generales</b>	1178.49 €
<b>6% Beneficio Industrial</b>	543.92 €
<b>Suma</b>	<b>10787.74 €</b>
<b>21% IVA</b>	2265.42 €
<b>TOTAL PRESUPUESTO</b>	<b>13053.16 €</b>

El presupuesto total del presente Trabajo de Fin de Grado asciende a la cantidad de **TRECE MIL CINCUENTA Y TRES EUROS CON DIECISEIS CÉNTIMOS**.