# UNIVERSIDAD COMPLUTENSE DE MADRID
## FACULTAD DE CIENCIAS FÍSICAS


# TESIS DOCTORAL

# Heterogeneidad tumoral en imágenes PET-CT

# Tumor heterogeneity in PET-CT images

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

**Ober Van Gómez López**

DIRECTORES

**José Manuel Udías Moinelo**
**Joaquín López Herraiz**

Madrid

**© Ober Van Gómez López, 2021**

# UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS FÍSICAS
DEPARTAMENTO DE ESTRUCTURA DE LA MATERIA,
FÍSICA TÉRMICA Y ELECTRÓNICA



# TESIS DOCTORAL

HETEROGENEIDAD TUMORAL EN IMÁGENES PET-CT

TUMOR HETEROGENEITY IN PET/CT IMAGES

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR


**Ober Van Gómez López**

DIRECTORES

Dr. José Manuel Udías Moinelo
Dr. Joaquín López Herraiz


**Madrid, 2020**

*A mis amados padres, esposa e hijas,*
*por su apoyo incondicional.*

# AGRADECIMIENTOS

Mi gratitud, principalmente está dirigida a mis Directores de Tesis:

El Prof. Dr. ***José Manuel Udías Moinelo*** y el Prof. Dr. ***Joaquín López Herraiz***, quienes me apoyaron y señalaron siempre el camino a seguir. Gracias por su confianza y por estar siempre que los necesité.

Quiero nuevamente extender mis más profundos agradecimientos a mi director de tesis el *Prof. Dr. José Manuel Udías Moinelo*. Él es el único responsable de que esta tesis vea la luz. Siempre confió en que lo podíamos hacer, y trasladó esa confianza en apoyo constante y efectivo, para que pudiera dedicarme a la elaboración de esta tesis. Así mismo, quiero agradecer a mi director de tesis *el Prof. Dr. Joaquín López Herraiz*, quien siempre me proveyó con valiosas y muy asertivas recomendaciones procedimentales.

Muchas gracias a la Facultad de Ciencias Físicas de la Universidad Complutense de Madrid, especialmente al Departamento de Estructura de la Materia, Física Térmica y Electrónica, por haberme dado la oportunidad de ingresar y cumplir este gran sueño.

Agradezco inmensamente el apoyo y tutoría permanente del *Dr. Ángel Soriano Castrejón*, jefe del servicio de Medicina Nuclear del Hospital Universitario de Ciudad Real y la *Dra. Ana María García* Vicente adjunta adscrita al mismo servicio, asicomo al *Dr. Antonio Francisco Honguero Martínez*, del servicio del Servicio de Cirugía Torácica del Complejo Hospitalario Universitario de Albacete.

Igualmente, quiero agradecer a la División de Medicina Nuclear, Departamento de Imagen Biomédica y Terapia Guiada por Imagen de la Universidad Médica de Viena. Especialmente al *Univ.-Prof. Dr. Marcus Hacker* y al *Assoc. -Prof. Univ. -Doz. Dr. Alexander Haug* por haberme permitido realizar una estancia formativa ("Fellow") en sus instalaciones y permitirme trabajar con algunas de sus bases de datos.

Finalmente, a todos aquellos que directa o indirectamente han colocado un granito de arena para el logro de esta Tesis Doctoral, agradezco de forma sincera su valiosa colaboración.

*"In memoriam"* a mi abuela ***Flor***

# Contents

# List of Figures

## Chapter 2. PET/CT image biomarkers and Radiomics



## Chapter 3. Machine learning in Medical Images and model construction

# List of Tables

# Summary

## TUMOR HETEROGENEITY IN PET/CT IMAGES

## Introduction and objectives

Cancer is a leading cause of morbidity and mortality [1]. The most frequent cancers worldwide are non–small cell lung carcinoma (NSCLC) and breast cancer [2], being their management a challenging task [3]. Tumor diagnosis is usually made through biopsy [4]. However, medical imaging also plays an important role in diagnosis, staging, response to treatment, and recurrence assessment [5]. Tumor heterogeneity is recognized to be involved in cancer treatment failure, with worse clinical outcomes for highly heterogeneous tumors [6,7]. This leads to the existence of tumor sub-regions with different biological behavior (some more aggressive and treatment-resistant than others) [8-10]. Which are characterized by a different pattern of vascularization, vessel permeability, metabolism, cell proliferation, cell death, and other features, that can be measured by modern medical imaging techniques, including positron emission tomography/computed tomography (PET/CT) [10-12]. Thus, the assessment of tumor heterogeneity through medical images could allow the prediction of therapy response and long-term outcomes of patients with cancer [13].

PET/CT has become essential in oncology [14,15] and is usually evaluated through semiquantitative metabolic parameters, such as maximum/mean standard uptake value (SUVmax, SUVmean) or metabolic tumor volume (MTV), which are valuables as prognostic image-based biomarkers in several tumors [16-17], but these do not assess tumor heterogeneity. Likewise, fluorodeoxyglucose ($^{18}$F-FDG) PET/CT is important to differentiate malignant from benign solitary pulmonary nodules (SPN), reducing so the number of patients who undergo unnecessary surgical biopsies. Several publications have shown that some quantitative image features, extracted from medical images, are suitable for diagnosis, tumor staging, the prognosis of treatment response, and long-term evolution of cancer patients [18-20]. The process of extracting and relating image features with clinical or biological variables is called “Radiomics” [9,20-24]. Radiomic parameters, such as textural features have been related directly to tumor heterogeneity [25].

This thesis investigated the relationships of the tumor heterogeneity, assessed by $^{18}$F-FDG-PET/CT texture analysis, with metabolic parameters and pathologic staging in patients with NSCLC, and explored the diagnostic performance of different metabolic, morphologic, and clinical criteria for classifying (malignant or not) of solitary pulmonary nodules (SPN). Furthermore, $^{18}$F-FDG-PET/CT radiomic features of patients with recurrent/metastatic breast cancer were used for constructing predictive models of response to the chemotherapy, based on an optimal combination of several feature selection and machine learning (ML) methods.

## Materials and methods

Patient medical records were reviewed, and variables such as age, gender, histopathologic features [26,27], tumor stage according to American Joint Committee on Cancer (AJCC) staging system [28-32], and treatment response were registered. All had one or more $^{18}$F-FDG-PET/CT images, from which, target lesions were segmented through 3D-Slicer or Hermes Hybrid 3D software [33], getting so several volumes-of- interest (VOI) to extract the metabolic and textural information. For NSCLC patients, tumor size, SUVmax, SUVmean, metabolic tumor volume (MTV), total lesion glycolysis (TLG); were obtained. Besides, several textural features were extracted with MaZda software [34]. Statistical tests were performed with SPSS software [35] to establish correlations between clinical, metabolic, and textural features. For the SPN study, simple visual inspection ($^{18}$F-FDG-uptake or not) and several heuristic combinations of nodule size and SUVmax were used to classify it as malignant or not [36-38]. Alike, a logistic predictive model of malignity, based on PET/CT and clinical variables was constructed [39]. The performance of each approach was evaluated through receiver operating characteristic (ROC) curve analysis [40]. For breast cancer patients, tumor VOIs before and after chemotherapeutical treatment were segmented, of which SUVmax, SUVmean, SUVpeak[1], and SULpeak[2] were obtained [41]. Patients were classified as responder or non-responder to the treatment, according to the PET response criteria in solid tumors (PERCIST) [42]. Also, from pre-treatment VOIs were extracted several radiomic features using MATLAB R2019b [43], which along with the clinical and pathological information were used to construct the prediction models, by using several cross-combinations of feature selection and ML classifiers [44-46]. The tumor lesions were separated into two groups with a ratio of 80:20, the bigger was used to construct the model and for cross-validation [47]; while the smaller dedicated only for validation. Seven feature selection methods [48,49]: ANOVA with F-score, mutual information (MI), least absolute shrinkage and selection operator (LASSO), Wilcoxon test, hierarchical clustering (HC), principal component analysis (PCA),

and independent component analysis (IPA), were cross-combined with seven ML classifier: support vector machines (SVM), random forest (RF), gaussian naive Bayes (GNB), logistic regression (LR), k-nearest neighborhood (KNN), adaptative boosting (AdaBoost) and neural network (NN). Model predictive performances were compared via ROC curve analysis [40].

## Results and conclusions

Tumor heterogeneity in NSCLC assessed by texture analysis of $^{18}$F-FDG-PET/CT images is correlated with metabolic parameters, and both are associated with macroscopic tumor diameter and AJCC staging system (important prognostic factor). However, some textural features have no linear relationship with volume-based metabolic parameters, making them more sensitive to tumor volume definition. However, by an appropriated selection, textural parameters have the potential to be used in clinical practice but are required additional work to further validate their importance.

The assessment of SPN by semiquantitative methods did not improve the sensibility of visual analysis (Se = 95%). There was a limited specificity for all heuristic methods. However, a predictive logistic model combining SUVmax and age had the best global diagnostic performance (*Se* = 87.5% and *Sp* = 46.7%). This model, unlike other models, used the metabolic variable SUVmax, which is shown to be an independent variable of malignancy.

Radiomic models based on 18F-FDG-PET/CT features and ML classifiers can predict the treatment response in patients with recurrent or metastatic breast cancer. LASSO + RF had the highest performance in the validation cohort (0.91±0.05). Although, other combinations also showed significant diagnostic performance. This comparative investigation may be an important reference in identifying reliable and effective machine-learning methods for radiomics-based prognostication in these kinds of patients and shows the great potential of PET/CT radiomics.

---

**1**SUVpeak: calculated as the mean SUV in a VOI of 1.2 *cm* of diameter (volume of 1 *ml*), centered at the most active portion of the tumor [154].

**2**SULpeak: the SUVpeak corrected for lean body mass.

# General overview and motivation of this thesis

Cancer is a leading cause of morbidity and mortality around the world [1, 2]. It can affect any part of the body and consists of malignant cells growing faster than normal cells. Cancerous cells can produce local destruction by the invasion of adjacent tissue, or in an advanced state; they can spread to other organs, which is the main cause of death. The most frequent cancers worldwide are non–small cell lung carcinoma (NSCLC) and breast cancer [1, 2].

Despite upgrades in medical technologies and treatment over the past two decades, the prevention, diagnosis, and treatment of cancer is still a challenging task [3]. In daily clinical practice, it is continuously observed that some patients, with the same tumor and stage of the disease, respond differently to the therapies received, having consequently a different prognosis [50]. This fact has been related to tumor heterogeneity [6, 7], where higher tumor heterogeneity implies a higher chance of developing metastasis, and shorter progression-free and overall survival [51]. Tumor heterogeneity refers to differences between tumor cells regarding cellular morphology, gene expression, metabolism, motility, proliferation, and metastatic potential [10, 52]. This phenomenon occurs both between tumors (inter-tumor heterogeneity) and within tumors (intra-tumor heterogeneity). Tumor heterogeneity is amply recognized to play a major role in cancer treatment failure, with worse clinical outcomes in patients with highly heterogeneous tumors [7, 8]. Therefore, an accurate assessment of tumor heterogeneity is essential for the development of effective therapies [7, 50]. It is believed that it could allow the prediction of therapy response and long-term outcomes, with a better stratification of patients, which could be used to select the treatment on the individual patient characteristics, and improving so the tumor control, avoiding unnecessary toxicity in case of ineffectiveness.

Intratumoral heterogeneity leads finally to the existence of tumor sub-regions, each one with different biological behavior (some more aggressive and treatment-resistant than others) [49, 51]. These sub-regions are characterized at the macroscopic level by a different pattern of vascularization, vessel permeability, metabolism, cell proliferation, cell death, and other features, which can be measured by modern medical imaging techniques, including positron emission tomography/computed tomography (PET/CT) [8], that is routinely used for diagnosis, treatment planning and assessment of tumor therapy response in oncology. Medical imaging methods have the great advantage of evaluating the tumor as a whole, usually non-invasively, being able to appreciate the spatial heterogeneity of the tumor, as well as its temporal evolution, if new images are obtained after a therapeutic intervention

or not [8]. This advantage contrast with pathological studies, where biopsies (samples) of the tumor are necessaries, being these obtained from specific regions of the tumor, which may induce the loss of any tumor spatial heterogeneity assessment.

## Quantification of intratumoral heterogeneity

In recent years, there has been a considerable effort in the medical imaging community to obtain correlations between image features and tumor heterogeneity [15-18]. An approximation to this issue is the texture analysis because image texture gives us information about the spatial arrangement of voxel intensities (i.e. the spatial distribution of radiotracer) in an image or a selected region.

In this thesis, we have focused on tumor images obtained by fluorodeoxyglucose ($^{18}$F-FDG) PET/CT, which allows to accurately localize metabolic abnormalities in the human body after the injection of radiopharmaceutical tracer, enabling the identification of regions with abnormally increased glucose uptake (i.e. high metabolic activity), which is a central characteristic of tumor cells, because of their accelerated growth [53, 54]. Although some very basic quantitative or semi-quantitative parameters have been introduced for the assessment of PET/CT [16, 41, 42], it has historically been a qualitative process, based mainly on the expertise of the medical specialist. Semiquantitative metabolic parameters such as SUVmax, SUVpeak, metabolic tumor volume (MTV), and total lesion glycolysis (TLG), have shown to be valuables as prognostic image-based biomarkers in several cancers. However, they are not designed to assess tumor heterogeneity [16, 17]. On the other hand, in last years, several other advanced PET/CT image features, using texture analysis and other image features, have shown to be more accurate and robust to predict progression-free survival (PFS), overall survival (OS), and treatment response [22, 27]. Further, some textural image features assess the image heterogeneity of the radiotracer. As a multimodal imaging modality, PET/CT allows for noninvasive exploration of intratumor heterogeneity at the macroscopic scale in both the anatomical and functional dimensions [55]. The PET component shows the radiotracer uptake spatial distribution, which may reflect, depending on the radiotracer used, underlying biological processes such as metabolism, hypoxia, cellular proliferation, vascularization, and necrosis [56, 57]. On the other hand, the CT component (usually a low-dose CT without contrast enhancement) shows the tissue density variability, which may result from spatially varying vascularization, necrosis, or cellularity, as well as the proportions of fat, air, and water [58]. For example, intratumoral variations in $^{18}$F-FDG uptake in two patients with non-small cell lung cancer, but different histologic types, can be observed on

PET/CT images in Figure 1. Although not apparent, variations in tumor density on CT are also present.

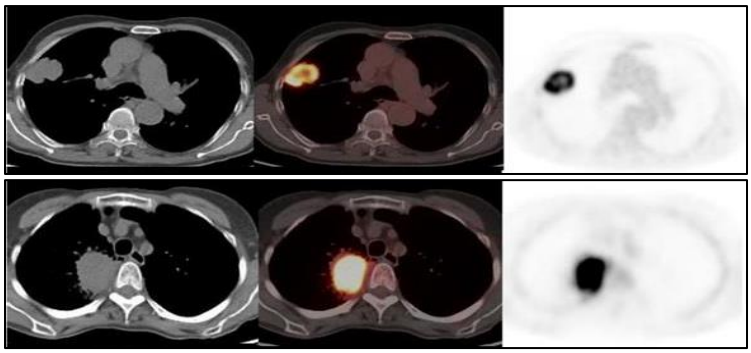


***Figure 1***. 18F-FDG PET/CT images in lung cancer. Upper figures: a patient with squamous cell carcinoma of the right lung. Lower figures: a patient with an adenocarcinoma of the right lung. Adapted from [59].

Thus, the true biological intratumoral variations are translated into different spatial intensity patterns in medical images, which can be quantified by different image processing and analysis methods, including texture analysis (TA) [57], fractal analysis [58], shape models [60-63], intensity histogram analysis [13] and filtering combined with statistical and frequency-based methods [64]. The different methodologies produce several image features, which could act as a surrogate (image-based biomarker) of the true biological intratumoral heterogeneity, providing possibly an appraisal of the tumor aggressiveness, response to the treatment, or metastases capability. These imaging biomarkers, extracted for instance, from pre-treatment medical images, could discriminate between patients responding well to treatment from those who do not, information that could assist physicians in tailoring therapy choices for each patient (personalized medicine).

The process of extracting and relating image features with clinical or biological variables is called "Radiomics", a term introduced by Lambin *et al*. in 2012 [65]. The great potential of radiomic consists in its ability to capture and quantify properties of a tissue, organ, or tumor, which cannot be detected by a visual inspection or simple metrics [25, 66]. Radiomics is used to develop decision support tools by combining radiomic data (image features) with other patients' characteristics. The

general process for building a PET/CT radiomic signature with prognostic/predictive value is independent of the underlying disease and consists of several steps [66, 67]: image acquisition, tumor segmentation, and preprocessing (image resampling and discretization), image features extraction, feature selection, and finally model construction. More detail is given in section 2.3 of this thesis. The textural analysis is only a step in the radiomic process, being included among the different methods for image feature extraction.

Some of the image characteristics that can be extracted are: morphological or shape-based (e.g., volume, eccentricity, compactness, etc.), the histogram of voxel intensity-based (e.g., variance, skewness, kurtosis, etc.), and texture-based. TA is probably the most used method for the characterization of intratumoral heterogeneity, as it involves the quantification of the relationship between intensity and spatial distribution of voxel grey levels within a given ROI. The metrics most commonly used in TA by the medical imaging community are the Gray-Level Co-occurrence Matrix (GLCM) features [68], the Gray-Level Run- Length Matrix (GLRLM) features [69-71], the Gray-Level Size Zone Matrix (GLSZM) features [72] and the Neighborhood Gray-Tone Difference Matrix (NGTDM) features [73]. The methodology used to extract these textural metrics is presented in section 2.3.3 (Appendix C provides the complete description of imaging features used in this thesis).

Several studies have applied TA and other methods in PET/CT images for intratumor heterogeneity characterization, on different tumor types such as lung, breast, esophagus, and lymphomas [22, 74, 75]. Applications of TA in cancer imaging include diagnosis, staging, determining tissue histopathological features, as well as treatment response and clinical outcome prediction. In this thesis, we focus on TA in lung and breast cancer. Although there is a large amount of literature applying TA or radiomics in various tumor entities, here we will restrict to those relevant to this thesis.

## PET/CT texture analysis in lung cancer

Lung cancer usually presents as a solitary pulmonary nodule (SPN) on diagnostic imaging during the early stages of the disease. The early diagnosis of lung cancer is fundamental for a better prognosis of the patients, therefore the accurate diagnosis of SPNs has great clinical importance. The histopathologic study is the mainstay for the classification of lesions into benign and malignant categories because the accurate classification of a lung lesion is a prerequisite for adequate management. However, non-invasive techniques such as CT can estimate the probability of cancer

in solitary pulmonary nodules (SPNs) based on nodule size, growth, and semantic features such as presence and distribution of calcification, fat, or ground-glass opacification (76-78). Despite its high sensitivity of up to 98%, CT has limited specificity (58%) in diagnosing cancer in SPNs. Metabolic activity measured with $^{18}$F-FDG PET/CT (e.g. SUVmax), is more discriminative of benign from malignant nodules (sensitivity 97%, specificity 85%), but there is a need for further analytical tools to improve on the accuracy of existing imaging modalities [38, 79-83] because the surgical resection is still needed to differentiate lung cancer from benign lesions in a significant number of cases [37].

Chen *et al*. [84] used NGTDM features on dual time point imaging (DTPI) PET/CT to differentiate malignant from benign FDG-avid solitary pulmonary nodules. They found that Compared to SUVmax or visual interpretation, NGTDM texture features derived from DTPI PET/CT images could be used as good predictors of SPN malignancy. The AUCs of delayed busyness, delayed coarseness, early busyness, and early SUVmax were 0.87, 0.85, 0.75, and 0.75, respectively. Recently Palumbo *et al.* [85] investigated the role of shape and texture features from $^{18}$F-FDG PET/CT to discriminate between benign and malignant solitary pulmonary nodules. Eighteen three-dimensional imaging features, including conventional, texture, and shape features from PET and CT were tested for significant differences (Wilcoxon-Mann-Whitney-Test) between the benign and malignant groups. Prediction models based on different feature sets and three classification strategies (Classification tree, k-nearest neighbors, and naive Bayes) were also evaluated to assess the potential benefit of shape and texture features compared with conventional imaging features alone. They found that shape and texture features from $^{18}$F-FDG PET/CT could provide benefits in the discrimination between benign and malignant lung nodules compared with conventional imaging features alone. Zhang *et al.* [86] evaluate the diagnostic value of a support vector machine (SVM) model built with texture features based on standard $^{18}$F-FDG-PET in patients with solitary pulmonary nodules (SPNs) at a volume larger than 5 ml. Compared with the SUVmax and MTV models, the texture-based SVM model provided an improvement of approximately 20% in diagnostic accuracy, positive predictive value, negative predictive value, and the area under the operating characteristic curve.

Although the use of textural analysis is a valid approach to classify SPNs, the inclusion of pulmonary nodules with small volumes can distort the quantification of spatial heterogeneity [87]. Likewise, a visual PET/CT analysis has shown great sensitivity to detect malignancy, so the construction of predictive models based on SUV and other clinical variables could achieve a good SPN classification

performance simply. Thus, one proposal of this thesis was to determine an optimum semiquantitative criterion that allows discriminating between malignant and benign nodules, comparing with the visual assessment, and derivate a model to estimate the pretest probability of malignancy of a patient with SPN based on clinical and metabolic $^{18}$F-FDG PET/CT variables.

Likewise, PET/CT texture analysis in patients with NSCLC has been described in the literature. Cook *et al*. [74] evaluated PET textural features in NSCLC and their relationship with response and survival after chemoradiotherapy. They found that in baseline $^{18}$F-FDG PET scan texture as measured by coarseness, contrast, and busyness is associated with nonresponse to chemoradiotherapy and with poorer prognosis. Van Gómez *et al*. [88] assessed the correlation between the texture features like energy, entropy, contrast, correlation, and homogeneity of FDG-PET images, with metabolic parameters such as SUVmax, SUVmean, MTV, and TLG, and pathologic staging in 38 NSCLC patients. A more recent review article by Sollini *et al*. [89] provides a comprehensive review of the literature describing the state of the art of $^{18}$F-FDG-PET/CT texture analysis in NSCLC. In total 85 studies were referenced. They summarize those publications reporting methodological investigations on texture analysis in NSCLC patients, which included the effect of segmentation method on tumor volume estimation, comparison of different discretization methods for textural features, the sensitivity of texture features to tumor motion, the variability of PET textural features using different reconstruction methods, iteration numbers, and voxel size. Likewise, the clinically relevant results of the publications for assessing the diagnostic, prognostic, and predictive role of $^{18}$F-FDG-PET/CT texture analysis were summarized, some of which are as follow:

- Compared with non-malignant lesions, malignant lung nodules are characterized by higher SUVmax and lower morphological and density fractal dimensions

- Large lesions are characterized by high heterogeneity

- Each subtype of NSCLC tumor has different metabolic heterogeneity characteristics. Compared with adenocarcinoma, squamous cell carcinoma is characterized by higher SUVmax and other features that indicate more heterogeneity

- Tumor heterogeneity can predict recurrence in patients with adenocarcinoma tumors without nodal affectation, who have undergone curative surgery but not in squamous cell carcinoma patients (high heterogeneity is associated with a shorter disease-free survival)
- Tumor heterogeneity identifies patients with inoperable stage III NSCLC with poor progression-free survival (for tumor staging, see section 1.3)

- High SUVmax, large MTV, and high heterogeneity are associated with poorer overall survival and progression-free survival in stage I–III NSCLC

- Tumor heterogeneity is associated with response to radiation therapy in NSCLC (disease-specific survival is lower for patients with high entropy)

- Lesions in responders (complete or partial response) to chemoradiotherapy are characterized by lower coarseness, contrast, and busyness than non-responders (stable or progressive disease)

- High coarseness values are associated with an increased risk of progression (increased risk of death), whereas high contrast and busyness values are associated with a lower risk of progression
- Large primary tumors more heterogeneous have a poor prognosis following chemoradiotherapy

- Lesions in responders to erlotinib (a type of targeted cancer drug) are characterized by lower heterogeneity than those in non-responders

Some important pitfalls must be considered by textural analysis. A large number of features can be derived from a single PET/CT scan, so that problems regarding multiple testing and high false discovery rates can be an issue to face [90, 91]. Furthermore, the stability and reproducibility of textural image features can be an issue when used as prognostic markers, as was tested by Leijenaar *et al.* [92]. They used eleven patients for a test-retest comparison and 23 patients for an investigation of interobserver variability, with most of the tested features showing high test-retest (71%) and interobserver (91%) stability, based on the intraclass correlation coefficient. Another important point to address is the correlation that is often found between image features representing tumor

heterogeneity and MTV because larger tumors typically show higher levels of heterogeneity. Hatt *et al.* [93] investigated whether MTV and features of tumor heterogeneity provided complementary prognostic information, or simply described the same part of the variance in patient outcome. Heterogeneity and MTV were both independent prognostic factors for overall survival ($p = 0.009$ and $p = 0.005$, respectively), but more so for larger tumors. For small tumors, there was a considerable correlation between MTV and heterogeneity, and the investigators suggest that there may be complementary prognostic information for tumors greater than 10 $cm^3$.

Texture features are of growing interest in tumor characterization in imaging. Nevertheless, based on results published to date on $^{18}$F-FDG PET, it is unclear which indices should be used, what they represent, and how they are related to conventional parameters such as SUVs, MTV, and TLG [89]. Therefore, in this thesis, we also investigate the relationships between tumor heterogeneity, assessed by texture analysis of $^{18}$F-FDG-PET images, metabolic parameters, and pathologic staging in patients with NSCLC.

## PET/CT Texture analysis in breast cancer

Currently, PET/CT is usually performed in breast cancer for recurrences detection or treatment response assessment, both in the context of neoadjuvant therapy or metastatic cancer treatment [94]. Changes in tumor metabolic activity, assessed by $^{18}$F-FDG PET/CT, are an early indicator of treatment effectiveness for breast cancer, both neoadjuvant and metastatic setting [95-98]. Likewise, a significant reduction in the metabolic activity of the tumor (i.e. treatment-sensible tumors) early during therapy, is associated with longer overall survival and progression-free survival in patients with this tumor [99-101]. Even when these metabolic changes have shown to be valuable to predict the treatment response in breast cancer and other malignant tumors; in the last years, there is an increasing interest in the clinical and prognostic utility of quantitative imaging analysis through radiomics [16, 25]. Although the radiomic methodology has been applied to several malignant entities. In breast cancer specifically, most of the radiomic studies have been carried out with MR images and in a neoadjuvant treatment context [102, 103]. However, some studies appearing more recently have explored the potential of radiomics with PET/CT, but none of them metastatic patients [104].

To build image biomarkers from radiomic features, a classical statistical approach could be used. However, an increasing number of radiomic studies are based on machine learning (ML) classifiers, which have shown great promise [105]. This approach relies on a pipeline, including extraction of numerous handcrafted imaging features; followed by feature selection and machine learning-based classification. Feature selection or reduction variable methodologies should be carried before any ML model construction, because these can identify redundant and irrelevant imaging features, allowing to remove them from further analysis and improving the ML classifier performance [106].

To the best of our knowledge, no radiomic studies with ML have been carried out in recurrent or metastatic breast cancer. However, we hypothesized that ML models using radiomic features extracted can help to predict the metabolic response to the systemic treatment in this kind of patient. Therefore, as part of our thesis, we proposed to use radiomic metrics extracted from $^{18}$F-FDG PET and CT to build ML models and identify optimal feature selection and ML methods for recognizing precociously patients that might have a treatment-sensible or resistant tumor phenotype, which could help to select or adjust a particular therapy.

In radiomics emerges two important questions, namely, these quantitative parameters consist of a very large image feature [20-25]; and above all, they don't have an intuitive interpretation or cannot be perceived directly by the medical imaging specialists [55], which can create confusion, avoiding its rapid transfer to the clinical setting. Fortunately, these problems can be overcome by incorporating modern decision systems, based on artificial intelligence, which can manage a large amount of information and do not consider the biological or clinical significance itself of the image features [44-46]. So then, we have moved from a simple qualitative inspection of medical images to guide the management of oncology patients towards the use of several image features to perform standard statistical modeling of relevant clinical outcomes, and more recently modern methods of image analysis and predictive modeling by using artificial intelligence. The ultimate goal is to improve the medical care of patients with cancer. Before any of these methodologies come to be routinely used in oncology, they should be explored in different clinical scenarios. Alike, any potential imaging biomarker must be previously validated. The validation involves a demonstration of the accuracy, precision, and feasibility.

## Motivation

This doctoral thesis has been carried out inside the ***Grupo de Física Nuclear*** from the ***Universidad Complutense de Madrid***. The $^{18}$F-FDG-PET/CT images, as well as the clinical information, were collected in the Nuclear Medicine Service of the General Hospital of Ciudad Real and Division of Nuclear Medicine, Department of Biomedical Imaging and Image-Guided Therapy, Medical University of Vienna.

Despite the great improvements in medical procedures for the diagnosis and treatment of cancer, the medical management of cancer patients is a great challenge. The arrival of new technologies, as well as methodologies for information processing, particularly the new field of "Radiomics" undoubtedly offers great possibilities to extend these improvements.

During the development of this thesis, we explored how $^{18}$F-FDG-PET/CT textural features of patients with NSCLC, which are surrogated of tumor heterogeneity, are related to metabolic and clinical parameters (e.g. SUVmax, MTV, TLG, and tumor staging). It is a fundamental step in the search for adequate image-based features of clinical utility. Also, the problem of determining whether an SPN is malignant or not was addressed through the construction of a multivariate logistic model, allowing us to confirm the importance of visual assessment, which contrasts with its enormous simplicity. However, the predictive model combining SUVmax and age improves diagnostic performance. The aforementioned work has paved the way for the implementation and use of radiomic and ML methodologies to predict the response to chemotherapy treatment in patients with recurrent or metastatic breast cancer. We compared different methods of feature selection and ML to look for the best combination with prediction performance. This last work has allowed us to obtain radiomic signatures with high predictive power, in a little-explored clinical context.

## Objectives

Taking into account the previously exposed considerations about the application of textural analysis and radiomics in oncology, the major objective of this thesis has been to investigate the relation of the metabolic and radiomic features derived from $^{18}$F-PET/CT images, with the clinical variables of oncologic patients. Likewise, a central aim was using this PET/CT image-based information for the construction of predictive models of malignity and treatment response.
The specific objectives can be summarized as follows:

- To investigate the relationships between tumor heterogeneity, assessed by texture analysis of $^{18}$F-FDG PET, metabolic parameters, and pathologic staging in patients with non–small cell lung carcinoma.
- To assess and compare different criteria, based on metabolic, morphologic, and clinical parameters, including a multivariate logistic model, for the correct classification as malignant or not of SPN.

- To use a radiomic approximation, by using image features extracted from $^{18}$F-FDG PET/CT images of patients with recurrent or metastatic breast cancer to construct predictive models of the metabolic response to the chemotherapy. Aiming to find the most suitable combination between feature selection methods and machine learning classifiers.

## Structure of the thesis

This thesis is divided into six chapters:

Chapters 1, 2, and 3 present some fundamental concepts related to the work developed in this thesis.

- In **Chapter 1**, the main concepts of tumor biology and medical images in oncology are introduced. The importance of tumor heterogeneity, and how it can be assessed through medical images, and used to predict treatment response is presented.

- In **Chapter 2**, the basic concepts of image-based biomarkers, textural analysis, and radiomics are presented.

- In **Chapter 3**, an overview of ML methods in the medical imaging context is shown.

Chapters 4 to 6 present each of the studies developed throughout this thesis.

- In **Chapter 4**, the relationships between tumor heterogeneity, assessed by texture analysis of $^{18}$F-FDG PET, metabolic parameters, and pathologic staging in patients with non–small cell lung carcinoma are investigated.

- In **Chapter 5**, the diagnostic performance of different metabolic, morphologic are assessed, and clinical criteria for the correct presurgical classification of SPN is explored.

- In **Chapter 6**, a radiomic procedure based on image features extracted from $^{18}$F-FDG PET/CT images of patients with recurrent or metastatic breast cancer is developed to predict the metabolic response to the chemotherapy. This aimed to find an optimal combination between feature selection methods and machine learning classifiers.

At the end of the manuscript, we present the general conclusions of the thesis and the publications and conference proceedings derived from this work.

# CHAPTER

# 1

# Medical background

## 1.1 Introduction

Cancer is a leading cause of morbidity and mortality around the world [1-2]. It can affect any part of the body and consists of malignant cells growing faster than normal cells. Cancerous cells can produce local destruction by the invasion of adjacent tissue, or in an advanced state; they can spread to other organs, which is the main cause of death. The prevention, diagnosis, and treatment of cancer is a very challenging task, a better understanding of the biological and clinical characteristics of cancer is necessary for the fight against this disease.

## 1.2 Clinical and biological background

### 1.2.1 Epidemiology of cancer

Cancer is the second leading cause of death globally after heart disease. According to the World Health Organization (WHO) in 2018, there were an estimated 18 million new cases of cancer and 10 million deaths from cancer worldwide. The predicted global burden will double to about 29–37 million new cancer cases by 2040 [2].

The most common cancers worldwide, together with their respective mortality, are shown in Figure 1.1. It can be seen that lung cancer together with breast cancer account for more than 20% of new cancers worldwide. Being lung cancer the one with the highest mortality rate (18.4%).

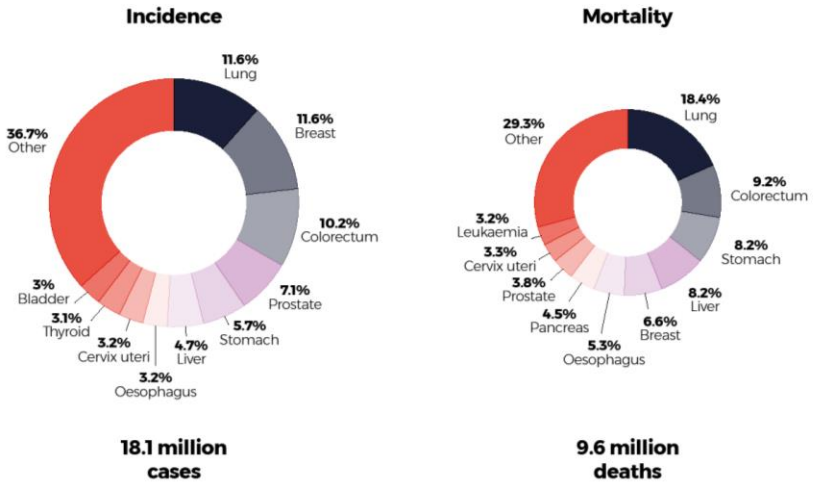



***Figure 1.1.*** Cancer incidence and mortality worldwide in 2018, according to WHO [1].

The rate of 5-year overall survival greatly varies between tumor sites, stages, and subtypes [1, 2]. It ranges from less than 10% in pancreatic cancer to more than 90% in prostate cancer in the USA [107]. Well established cancer risk factors include smoking, unhealthy diet, alcohol abuse, overweight and lack of physical activity, human papillomavirus (HPV) and hepatitis infection, exposure to ionizing and ultraviolet radiation as well as urban air pollution.

### 1.2.2 Hallmarks of cancer

Carcinogenesis is a multistep process, where a cell accumulates genetic mutations, which finally enables it to have an uncontrolled proliferation [52, 53]. In this multistep process, a normal cell develops into a precancerous lesion and subsequently evolves into a malignant tumor. Figure 1.2 shows a schematic representation of the natural progression of cancer.

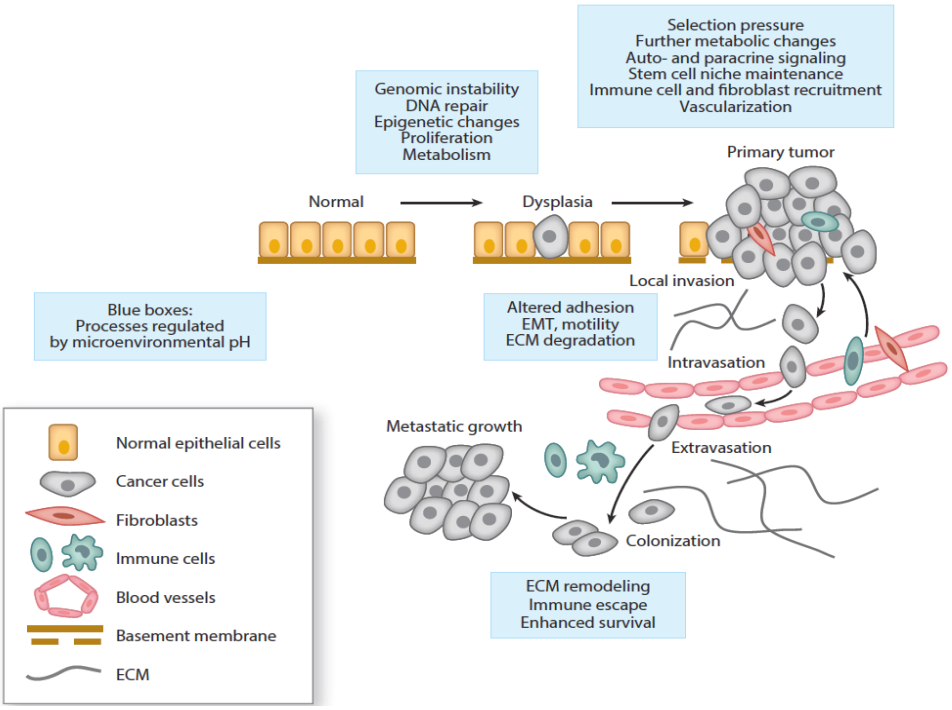



***Figure 1.2.*** The natural history of progression toward cancer, evolving from dysplasia to in-situ carcinoma and eventually to a malignant invasive tumor [from 54]. Cells from normal tissues after genetic damage change to dysplastic cells, which present increased proliferation and atypical alterations that, affect their size, shape, and organization. After a period of latency, these cells can develop all the characteristics of a malignant tumor. In this transformation process, there are many biological factors involved.

The hallmarks of cancer comprise several biological capabilities acquired by a normal cell to become cancerous. These consist of permanent proliferative signaling, evasion of growth suppressor agents, resistance to cell death or apoptosis; which enable the replicative immortality, induction of angiogenesis, and activation of local invasion of tissues and metastasis [52, 53]. Unlike normal cells, cancer cells are biologically resistant to the normal regulatory mechanism controlling cell populations. For their proliferation, cancer cells need energy. Depending on the tumor type and the microenvironmental conditions, energy production in cancer cells relies on a combination of glycolysis, oxidative phosphorylation, and other metabolic pathways. However, tumors have mainly an anaerobic glucose metabolism, even in the presence of normal oxygen concentration, which is called the "**Warburg effect**" [108].

In presence of normal concentrations of oxygen, most differentiated cells primarily metabolize glucose to pyruvate via glycolysis and posteriorly to carbon dioxide (CO2) and water through the tricarboxylic acid (TCA) cycle inside the mitochondria. This reaction produces NADH [nicotinamide adenine dinucleotide (NAD+), reduced], which then fuels oxidative phosphorylation to produce adenosine triphosphate (ATP). The NADH formed is an energy-rich molecule because each contains a pair of electrons having a high transfer potential. When these electrons are used to reduce molecular oxygen to water, a large amount of free energy is liberated, which is used to generate ATP. By using the cell this metabolic pathway as the main source of energy, lactate production is minimal. It is only under anaerobic conditions that differentiated cells produce large amounts of lactate (Figure 1.3) [109].

Because oxygen is required as the final electron acceptor to completely oxidize the glucose, it essential for this process, when it is limiting, cells can redirect the pyruvate generated by glycolysis away from mitochondrial oxidative phosphorylation by generating lactate (anaerobic glycolysis). This generation of lactate during anaerobic glycolysis results in minimal ATP production when compared with oxidative phosphorylation.

In the 1920s, Otto Warburg showed that cultured tumor tissues have high rates of glucose uptake and lactate secretion, even in the presence of oxygen (aerobic glycolysis). Those three metabolic properties—glucose uptake, lactate secretion, and oxygen availability—constitute the Warburg effect as he defined it [110].

Warburg originally hypothesized that cancer cells develop a defect in mitochondria that leads to impaired aerobic respiration and subsequent reliance on glycolytic metabolism [108]. However, subsequent work showed that mitochondrial function is not impaired in most cancer cells [109, 111]. They remain functional and some oxidative phosphorylation continues.

The anaerobic glycolysis is less efficient but a faster process to obtain energy. However, cancer cells can incorporate more glucose and compensate for this inefficiency [112, 113]. This independence of oxygen allows the tumoral cells to escape from the immune system and survive [111, 112].

Besides, elevated glucose metabolism decreases the pH in the microenvironment due to lactate secretion. An acid-mediated invasion hypothesis suggests that H+ ions secreted from cancer cells diffuse into the surrounding environment and alter the tumor-stroma interface allowing for enhanced invasiveness [111]

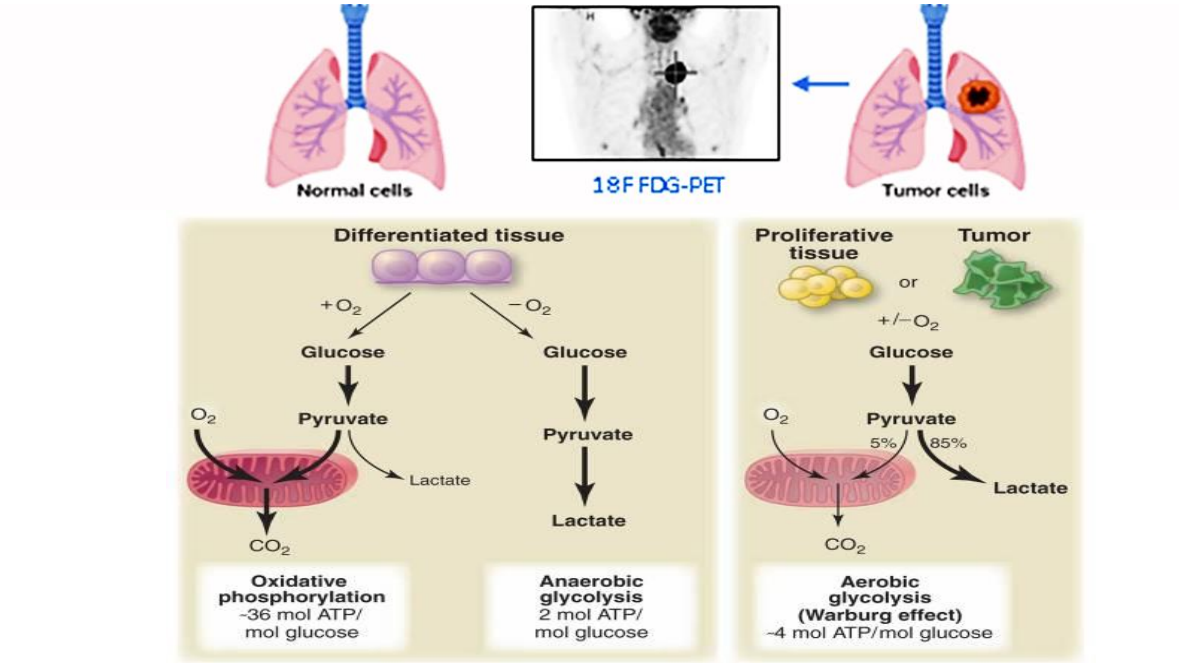



***Figure 1.3.*** Warburg effect in tumor cells and its exploitation to obtain PET image of a patient with lung cancer, which has increased the glucose uptake [adapted from 111]. In the presence of oxygen, normal cells first metabolize glucose to pyruvate via glycolysis. Then by oxidative phosphorylation produce ATP in the mitochondria, the lactate production is minimal. When oxygen is limiting, cells use anaerobic glycolysis, with minor production of pyruvate. Cancer cells tend to convert most glucose to lactate regardless of whether oxygen is present (aerobic glycolysis).

Positron emission tomography with $^{18}$F-FDG, a glucose analog, marked with the positron-emitting radionuclide fluorine-18, takes advantage of the "Warburg effect" to image the tumors, since it accumulates in several tumors, allowing it detection [113]. Figure 1.3 shows the "Warburg effect" and its exploitation to obtain $^{18}$F-FDG images of a patient with lung cancer.

Cancer cells demonstrate the upregulation of glucose metabolism, which means that the glucose or glucose-analogs uptake is increased [114]. The labeling of deoxy-glucose with the positron-emitting radionuclide $^{18}$F to form $^{18}$F-FDG makes detectable these cells by PET.

$^{18}$F-FDG is transported into the cells by the same carrier as glucose, the glucose transporter 1 (GLUT1), but at a much higher rate. Then it is phosphorylated to $^{18}$F-FDG-6-phosphate ($^{18}$F-FDG-6-P) by the action of hexokinase or glucokinase [114, 115]. This substance does not enter the standard metabolic pathways, because of the presence of fluorine instead of the hydroxyl group in glucose and can leave the cell only slowly by the action of glucose-6-phosphatase. Therefore, it is trapped and accumulated in the neoplastic cells [116]. This 'metabolic trapping' of $^{18}$F-FDG-6-P forms the basis of the analysis of PET data. Figure 1.4 shows a schematic representation of the metabolic trapping of $^{18}$F-FDG in tumor cells. Besides, there is an increase in glycolytic activity in cancer cells, under both aerobic and anaerobic conditions. The neoplastic transformation often determines an increase in the activity of glycolytic enzymes (e.g. hexokinase) and glucose transporters (e.g. GLUT1) [117].

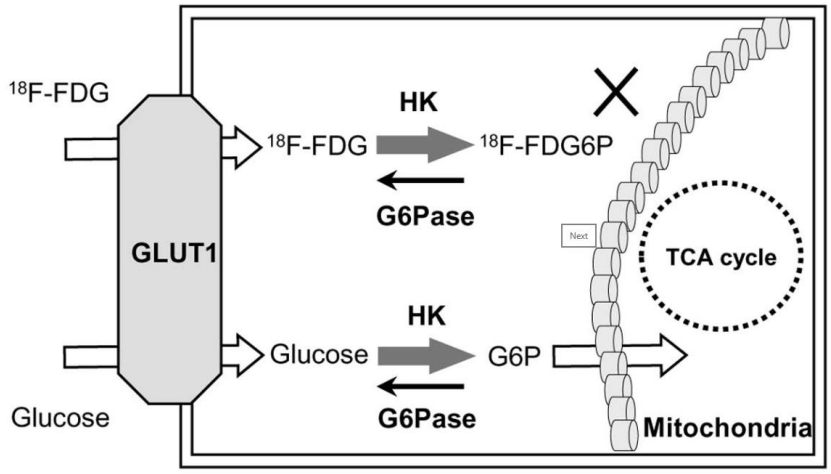



Figure 1.4. Schematic for the metabolic trapping of 18F-FDG. $^{18}$F-FDG6P: $^{18}$F-2-fluoro-2-deoxy-D-glucose-6-phosphate; GLUT1, glucose transporter 1; G6P, glucose-6-phosphate; G6Pase, glucose-6-phosphatase; HK, hexokinase; TCA, tricarboxylic acid (from [118])

A common characteristic of most tumors is a low oxygen level, which is called hypoxia. In highly proliferating tumor tissue, the distance between cells and the existing vasculature is constantly increased, which hamper the oxygen diffusion creating so a hypoxic microenvironment. At the same time, tumor-hypoxia leads to dysfunctional vascularization (blood vessels that are immature, tortuous, and hyperpermeable), cell mobility, and metastasis. Hypoxia alters cancer cell metabolism and contributes to therapy resistance. As tumors growing in size, distinctive regions inside them become apparent, each one with different metabolism, perfusion, and level of hypoxia [119, 120].

It has been demonstrated in cancer cell culture and animal models of cancer that hypoxic conditions played a critical role in inducing high $^{18}$F-FDG accumulation [121]. Likewise, spatial co-localization has been found between high $^{18}$F-FDG uptake and tumor hypoxia, and such regions had low blood perfusion. On the other hand, non-hypoxic regions displayed low $^{18}$F-FDG uptake [122-126]; therefore, $^{18}$F-FDG cannot map oxic cancer cells. This finding is critically important, especially for assessing the anti-cancer effect with $^{18}$F-FDG PET/CT; a negative $^{18}$F-FDG PET finding does not necessarily mean the absence of viable cancer cells, which simply indicates the absence of hypoxic cells. Tumor heterogeneity, $^{18}$F-FDG metabolic parameters, and high expression of hypoxia were found to be prognostic factors in NSCLC patients who were candidates for surgery [127]. However, a more recent study showed that SUVmax of $^{18}$F-FDG PET correlated weakly with the expression of hypoxia-inducible factor 1-alpha (HIF-1α) both in the overall sample and in tumor subgroups. Therefore, $^{18}$F-FDG PET cannot be used for the prediction of hypoxia in clinical practice [128].

In Figure 1.5, spatial variation in staining for angiogenesis (CD34), pimonidazole (hypoxia), and glucose transporter protein expression (Glut-1) in patients with NSCLC are shown [55]. As well as for Hematoxylin and eosin stain (H&E): one of the principal tissue stains used in histology. The hematoxylin stains cell nuclei blue, and eosin stains the extracellular matrix and cytoplasm pink, with other structures taking on different shades, hues, and combinations of these colors.

In general, hypoxia contributes to tumor aggressiveness and promotes the growth of many solid tumors, originating resistant to conventional therapies. To achieve successful therapeutic strategies targeting different cancer types, it is necessary to understand the molecular mechanisms and signaling pathways that are induced by hypoxia [120]. Since 18F-FDG is the most used radionuclide for PET/CT in daily clinical practice, an understanding between glucose metabolism and hypoxia is also necessary.

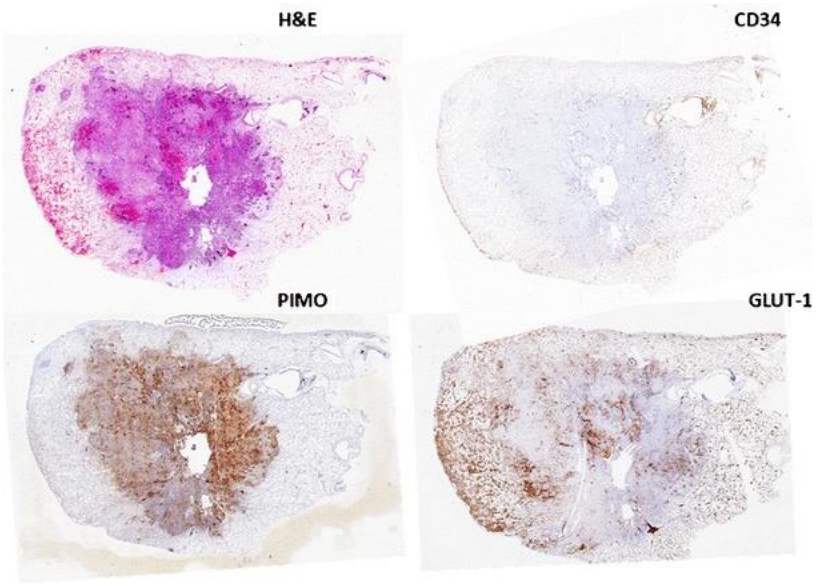



***Figure 1.5.*** Non-small-cell lung cancer showing spatial variation in staining for angiogenesis (CD34), pimonidazole (hypoxia), and glucose transporter protein expression (Glut-1), (from [55]). It can be seen that the spatial distribution of the markers of angiogenesis, hypoxia, and glucose metabolism is not homogeneous, which translates into heterogeneities in the distribution of the voxel intensities in their respective PET images.

### 1.2.3 Heterogeneity of cancer

Tumor heterogeneity refers to differences between tumor cells regarding cellular morphology, gene expression, metabolism, motility, proliferation, and metastatic potential [129]. This phenomenon occurs both between tumors (inter-tumor heterogeneity) and within tumors (intra-tumor heterogeneity). In the last case, the heterogeneity might result in a non-uniform distribution of genetically distinct tumor-cell subpopulations across and within disease sites (spatial heterogeneity) or temporal variations in the molecular features of cancer cells (temporal heterogeneity).

To explain the heterogeneity of tumor cells, cancer stem cell or clonal evolution model are used. The cancer stem cell model considers that within a tumor, there is only a small number of cells that are tumorigenic (tumor-forming). These cells are termed cancer stem cells (CSC) and are marked by the ability to both self-renew and differentiate into non-tumorigenic progeny. The CSC model postulate that the heterogeneity observed between tumor cells is the consequence of differences in the stem cells from which they originated. These cells are hypothesized to persist in tumors as a distinct

population, causing relapse and metastasis by giving rise to new tumors [130]. The clonal evolution model, consider that tumors arise from a single mutated cell, accumulating successively mutations as it progresses. These changes give rise to additional subpopulations, and each of these subpopulations can divide and mutate further. This heterogeneity may give rise to sub-clones that possess an evolutionary advantage over the others within the tumor environment, and these sub-clones may become dominant in the tumor over time.

While tumors are growing, their heterogeneity is also increased. This finally translates, as it was mentioned already, into the existence of tumor sub-regions inside the primary tumor, each one with different biological behavior (some more aggressive and treatment-resistant than others). At the macroscopic level, these sub-regions are characterized by a different pattern of vascularization, vessel permeability, metabolism, cell proliferation, cell death, and other features [8], which can be potentially captured by current medical imaging methods [8-10]. PET/CT images showing spatial heterogeneity of glucose metabolism and hypoxia in a patient with NSCLC are shown in Figure 1.6 [131]. These images were obtained by using the radiotracers $^{18}$F-FDG and $^{18}$F-HX4 (a 2-nitroimidazole nucleoside analog).

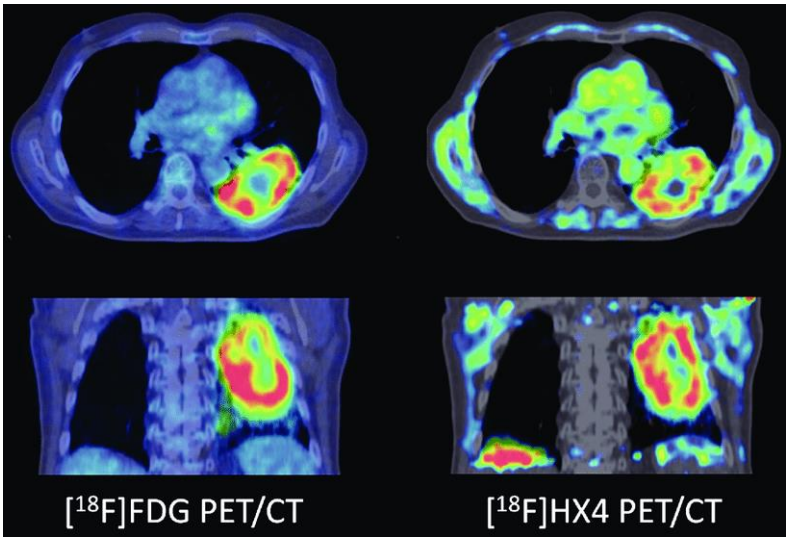



***Figure 1.6.*** Images of an NSCLC patient having both an $^{18}$F-FDG PET/CT scan (left) and a hypoxia $^{18}$F-HX4 PET/CT scan. The spatial tumor heterogeneity in both metabolic (FDG) and hypoxic (HX4) PET images are easily appreciated (from [131]). It can be easily appreciated that the spatial distribution of voxel intensities is heterogeneous in both images, with some degree of agreement between areas of increased glucose metabolism and hypoxic areas (red color).

The clinical importance of tumor heterogeneity is based on the fact that tumors with higher intratumoral heterogeneity have shown to have a poorer prognosis, which could be secondary to intrinsic aggressive biology or treatment resistance [132]. Tumor heterogeneity is one of its main resistance resources; therefore, an accurate assessment of tumor heterogeneity is essential for the development of effective therapies.

## 1.3 Cancer staging

Cancer staging plays a central role in cancer management. It allows us to estimate the patient prognosis, i.e. the likelihood of overcoming cancer once diagnosed and determine the best treatment approach. Besides, accurate staging is necessary to evaluate the results of treatments or clinical trials. The most clinically useful staging system is the tumor, node, and metastasis (TNM) staging system developed by the American Joint Committee on Cancer (AJCC) in collaboration with the Union for International Cancer Control (UICC) [30-32]. TNM classification has become the global standard for gathering, communicating, and exchanging cancer information worldwide and is widely used by clinicians and researchers.

The AJCC TNM system classifies cancers by the size and extent of the primary tumor (T), the involvement of regional lymph nodes (N), and the presence or absence of distant metastases (M), Tables 1.1, 1.2, and 1.3. Additionally, some evidence-based prognostic and predictive biomarkers have been added recently. Patients who generally have a similar prognosis, T, N, and M are grouped into prognostic stage groups.

TNM classification during the diagnostic workup time frame is named as the clinical stage and denoted as cT, cN, and cM. A clinical-stage record is important for all patients because, it is essential for selecting initial therapy, and comparison across patient cohorts when some have surgery as a component of initial treatment and others do not.

TNM classification after a surgical treatment is named pathological stage and denoted as pT, pN, and cM0, cM1, or pM1. There is also TNM classification after posttherapy or post neoadjuvant therapy, they are denoted respectively as ycTNM and ypTNM. The TNM criteria are defined separately for cancers in different anatomic locations and/or for different histologic types. Some general considerations are shown in Tables 2.1, 2.2, and 2.3.

***Table 1.1.*** Primary tumor (T) categories (from [32])

| Tumor category | Is assigned when there is… |
|---|---|
| **TX** | No information about the T category for the primary tumor, or it is unknown or cannot be assessed. |
| **T0** | No evidence of a primary tumor |
| **Tis** | Carcinoma *in situ* |
| **T1, T2, T3, or T4** | Primary invasive tumor, for which a higher category generally means<br>• an increasing size<br>• an increasing local extension, or<br>• both |

***Table 1.2.*** Regional lymph node (N) categories (from [32])

| Regional node category | Is assigned when there is… |
|---|---|
| **NX** | No information about the N category for the regional lymph nodes, or it is unknown or cannot be assessed |
| **N0** | No regional lymph node involvement with cancer |
| **N1, T2 or N3** | Evidence of regional node(s)<br>containing cancer, for which a higher category generally means<br>• involved more lymph node stations or<br>• or a greater number of affected nodes |

***Table 1.3.*** Distant metastasis (M) categories (from [32])

| Distant metastasis category | Is assigned when there is… |
|---|---|
| **M0** | No evidence of distant metastasis |
| **M1** | Distant metastasis |

## 1.4. Cancer treatment

There are many different approaches for treating cancer, depending on the type of cancer, how advanced it is, what types of treatment are available, and which will be the goals of treatment. Some treatments are local or loco-regional, such as surgery and radiation therapy, which are used to treat a tumor in a specific corporal localization. Drug treatments (such as chemotherapy, immunotherapy, or targeted therapy) are known as systemic treatments because they affect the entire body.

Depending on the goal, cancer treatments may be performed as:

- **Primary *treatment***: the goal is to completely remove cancer from the body, pretending to get the cure of the patient. The most common primary cancer treatment, in several types of cancer, is surgery. But, if the tumor is particularly sensitive to radiation therapy or chemotherapy, it could be used as a primary treatment

- ***Adjuvant treatment***: the goal is to kill any cancer cells that may remain after primary treatment to reduce the chance of recurrence. Common adjuvant therapies include chemotherapy, radiation therapy, and hormone therapy

- ***Neo-adjuvant therapy***: this treatment is performed before the primary treatment, looking to make the primary treatment easier or more effective

- ***Palliative treatment***: it helps to relieve the side effects of a previous treatment or signs and symptoms caused by cancer itself. It is applied when no cure is possible

The systemic treatment has evolved from the administration of non-specific cytotoxic drugs that damage both tumor and normal cells to more specific agents. Targeted therapies use agents that are directed again genes and proteins that are involved in the growth and survival of cancer cells, whereas immunotherapeutic treatments modulate the tumor immune response again cancer; both approaches aim to produce greater effectiveness with less toxicity. The development and use of such agents enable a more personalized cancer treatment, which the potential to reduce the side effects and increase the treatment effectiveness.

## 1.5. Some specific types of cancer

### 1.5.1 Non-Small Cell Lung Cancer (NSCLC)

Lung cancer is a leading cause of cancer-related death worldwide in both men and women [1-3]. Despite upgrades in imaging technologies and treatment over the past two decades, the median overall survival is 16·9, with a 5-year global survival rate as low as 23% in the younger age group [133]. The 5-year survival rate for stage IVA and IVB NSCLC (referred to as metastatic NSCLC, hereafter) is approximately 10 and <1%, respectively, whereas untreated lung cancer patients live 7.15 months

after diagnosis [134]. There are two main forms of lung cancer: NSCLC (85% of patients) and small cell lung cancer (SCLC) (15%) [135]. The WHO has classified NSCLC into three main types: adenocarcinoma, squamous cell carcinoma, and large cell. There are also several variants and combinations of clinical subtypes [29]. Adenocarcinoma is the most common type of NSCLC and accounts for approximately 40% of lung cancers. They arise from alveolar cells located in the smaller airway epithelium. Squamous cell carcinomas represent 25% to 30% of lung cancers; they tend to arise from cells located in the airway epithelium. Although NSCLCs are associated with cigarette smoke, adenocarcinomas may be found in patients who have never smoked. After the initial diagnosis, accurate staging of NSCLC using CT or $^{18}$F-FDG PET/CT is crucial for determining the appropriate therapy (see Figures 1.11 and 1.14). The TNM classification allows, not only to characterize a lesion according to the primary tumor (T), nodal status (N), and distant metastasis (M) but also to define prognostic stages (*I*, *II*, *III*, and *IV*), where higher stage mean worst prognosis [32]. Early-stage I tumors are small masses (less than 5 cm) contained within the lung without spread to the nearby lymph nodes. Cancers, which have spread to bronchial or hilar lymph nodes, are considered stage II. Stage III disease, or locally advanced disease, has spread to mediastinal lymph nodes. Tumors, which have metastasized outside of the lung, are considered stage IV disease. Figure 1.7 shows a patient with an NSCLC in the IIB stage, existing lymph node affectation in the lung, or near to bronchus.

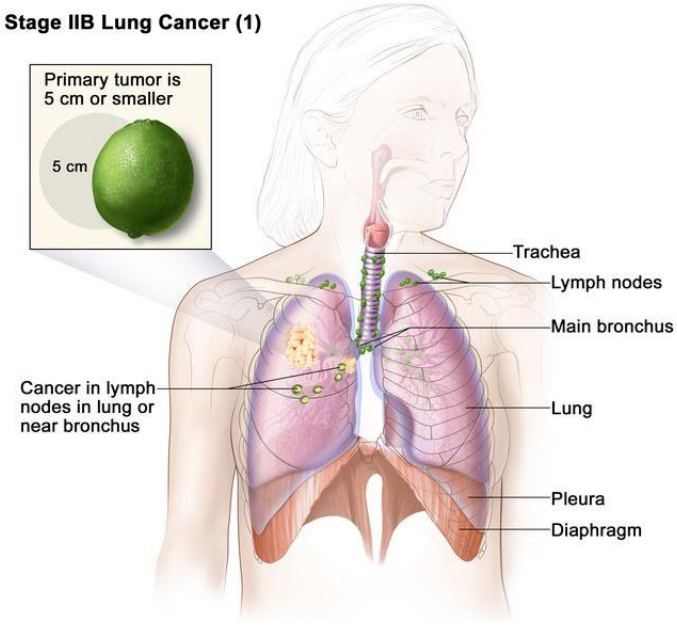



***Figure 1.7.*** Depicting a patient with an IIB stage NSCLC (from [136]). The primary tumor diameter is 5 cm or smaller, and there is an affectation of lymph nodes only in the lung or near the bronchus.

Treatment of NSCLC is stage-specific; patients with stage I or II should be treated with complete surgical resection when not contraindicated. Nonsurgical patients or advanced stage should be considered for conventional or stereotactic radiotherapy. Patients with advanced metastatic disease may achieve improved survival and palliation of symptoms with chemotherapy, targeted agents, and other supportive measures.

### 1.5.2 Breast cancer

Breast cancer is the most commonly diagnosed cancer and the second leading cause of death for cancer among women worldwide surpassed only by lung cancer [1-2]. Risk factors include older age, a personal or family history of breast or ovarian cancer, some genetic mutation for breast cancer, as well as some premalignant breast lesions such as atypical hyperplasia (AH) or lobular carcinoma in situ (LCIS). A history of radiation exposure in young women is associated with a higher prevalence of breast cancer. Other potentially modifiable risk factors for breast cancer include increased breast density, alcohol use, overweight, physical inactivity, and postmenopausal hormone therapy [135].

According to WHO classification, breast carcinomas are divided into invasive carcinomas (70–75%), lobular carcinomas (10–14%), and other carcinomas of special type [26]. After the initial diagnosis, it is important to define accurately the initial extent of the disease, because it will affect the treatment election. In breast cancer staging, the TNM classification is used. In the evaluation of tumor size, conventional techniques including mammography and ultrasound (US) give sufficient information, while magnetic resonance imaging (MRI) is the preferred method for the assessment of multifocal tumors [137]. For axillary nodal staging, in patients with clinically negative axilla, sentinel lymph node biopsy (SLNB) has become a standard approach, where histopathological analysis of the sentinel nodes is conducted intraoperatively [138].

In 2018, the AJCC [32] updated the breast cancer staging guidelines to add other cancer characteristics to the T, N, M system to determine the cancer stage. These include:

- **Histologic grade** (a measurement of how much the cancer cells look like normal cells).
    - Grade 1 or well-differentiated: the cells are slower growing, and look more like normal breast tissue.

  - Grade 2 or moderately differentiated: the cells are growing at a speed of and look like cells somewhere between grades 1 and 3.

  - Grade 3 or poorly differentiated

- **Hormone receptor status** (estrogen receptor [ER] and progesterone receptor [PR]), positive or not.

- **Human epidermal growth factor receptor-2 status** (HER2), positive or not.

- **A marker of cellular proliferation** (such as Ki-67 or a mitotic count)

- **For appropriate subgroups of tumors, a genomic prognostic panel** (such as Oncotype, Endopredict, and others).

Breast cancer is categorized into three major subtypes based on the presence or absence of molecular markers for ER, PR, and HER2: hormone receptor-positive/HER2 negative (80% of patients), HER2 positive (15%-20%), and triple-negative (15%). Triple-negative breast cancer is more likely to recur than the other two subtypes, with 85%. The two major pillars of breast cancer management are loco-regional treatment and systemic therapy; the histological and molecular characteristics of breast cancer largely influence treatment decisions. Local therapy for all patients with nonmetastatic breast cancer consists of surgical resection (Figure 1.8). Systemic therapy for nonmetastatic breast cancer is determined by subtype:

- Patients with hormone receptor-positive tumors receive endocrine therapy, and a minority receive additional chemotherapy.

- Patients with HER2-positive tumors receive HER2-targeted antibody or small-molecule inhibitor therapy combined with chemotherapy.

- Patients with triple-negative tumors receive chemotherapy alone.

- Metastatic breast cancer is treated according to subtype, with goals of prolonging life and palliating symptoms.

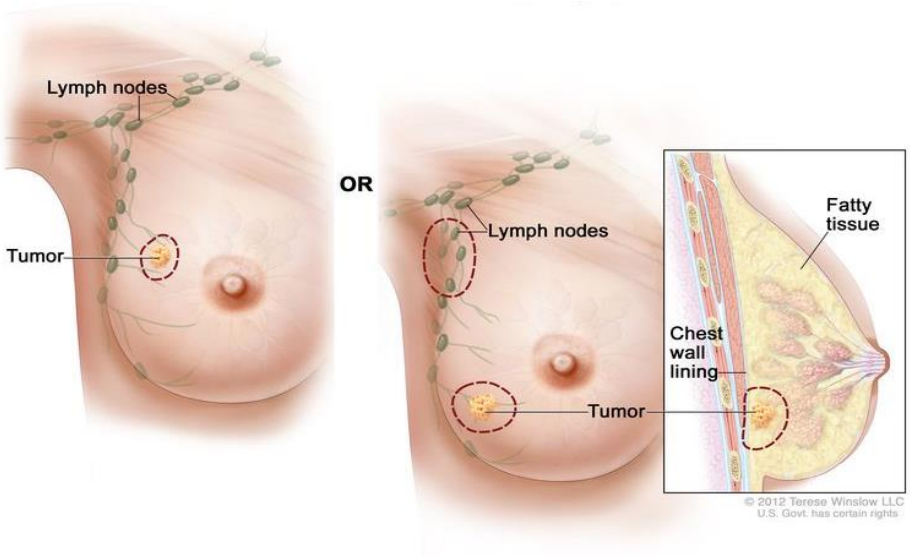



***Figure 1.8.*** Patient with no metastatic breast cancer. Breast-conserving surgery will be performed (only the tumor and some normal tissue around it are removed). Some lymph nodes in the axilla may be removed to an accurate pathological N staging. Part of the chest wall musculature may also be removed if the cancer is near it. Finally, the patient will receive local adjuvant radiotherapy (from [139])

## 1.6 Medical imaging methods in Oncology

Currently, medical imaging plays a central role in the management of cancer patients [140, 141]. It has a wide range of applications:

- ***Screening***: to detect disease in non-symptomatic patients.

- ***Detection and diagnosis****:* distinguish between malignant from benign disease, tumor localization, and guiding of the biopsy process for a pathological assessment.

- ***Cancer staging****:* determining tumor size and extent (local, loco-regional, or metastatic affectation).

- ***Guiding treatment and management decisions****:* determining suitability for treatment options (for instance operable or not) and tailoring the treatments to some specific characteristics of the patient.

- ***Guiding and verifying precise therapies**:* such as stereotactic irradiation, external beam radiotherapy, brachytherapy, particle therapy, intraluminal treatment delivery, and tissue ablation therapies.

- ***Enabling interventional placement**:* tumor fiducial markers, stents, catheters, and intravenous devices.

- ***Assessing treatment response**:* including complications.

Three main types of medical imaging methods and combinations of these are routinely used in clinical oncology: morphological, functional, and molecular methods. This section introduces the basics of medical imaging techniques used in this thesis: positron emission tomography and computed tomography, which are molecular and morphological imaging technic, respectively.

### 1.6.1 X-ray computed tomography (CT)

The most widely extended morphological imaging method is X-ray computed tomography or simply CT [142]. It uses X-rays to produce a cross-sectional slice of the body.
CT is based on the calculation of attenuation coefficients of an X-beam in the volume studied. It is performed by acquiring multiple X-ray views of the object and performing mathematical operations on digital data.

X-ray photons passing through the body either interact or pass unaffected. Interactions, either through scattering or absorption, attenuates the beam. The attenuation of monochromatic X-rays going through a homogeneous object is governed by [143]:

$$N = N_0 e^{-\mu x} \qquad (1.1)$$

Where $N$ is the X-ray intensity after passing the body, $N_0$ is the X-ray intensity before passing the object, $x$ is the length of the X-ray path through the object, and $\mu$ is the linear attenuation coefficient of the material for the X-ray energy employed. For inhomogeneous objects like the human body, the attenuation of x-rays consequently can be described by [143]:

$$N = N_0 e^{-\int \alpha(x)dx} \qquad (1.2)$$

In CT scanners, the X-ray attenuation is measured along with a variety of lines within a plane perpendicular to the long axis of the patient to reconstruct a map of the attenuation coefficients $\mu$ for this plane.

CT scanners use typically fan-beam projection geometries, with single or multiple detector arrays that are arranged in an arc relative to the x-ray tube (*Figure 1.9*) [144] The apex of the fan is the x-ray tube.

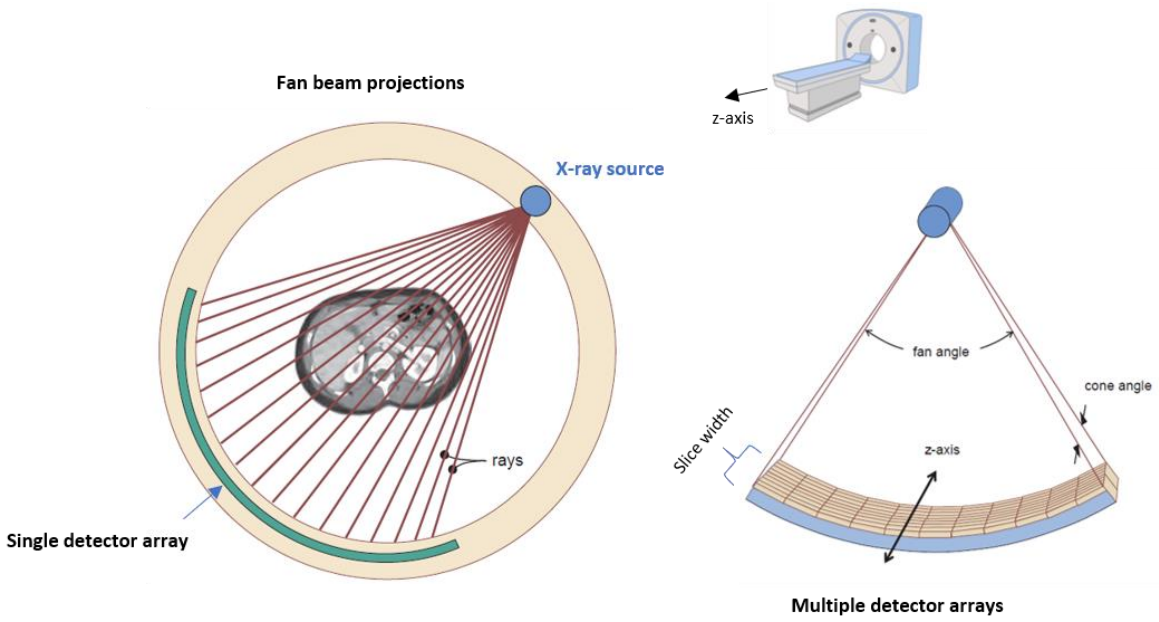



***Figure 1.9.*** Fan-beam projections and multiple detector array from a CT-scanner (from [144]). The X-ray source with a specific fan angle rotates around the patient. To the opposite side is a single or multiple detector array to collect the X-ray.

The source and detector rotate around the patient, as shown in Fig. 1.9 left. The individual rays correspond to each detector measurement. The collection of rays in this geometry is a fan beam projection. The use of multiple arrays of X-ray detectors (Fig. 1.9. Right) allows acquiring several slices simultaneously [144].

For image reconstruction, a slice can be considered as a matrix of rectangular boxes (voxels) of material or tissue (Figure 1.10). Conventionally, the X and Y directions are within the plane of the slice, whereas the Z direction is along with the patient (slice thickness direction) [145].

Thus, the objective of CT image reconstruction is to determine the attenuation of the narrow X-ray beam that occurs in each voxel of the reconstruction matrix. There are various algorithms used for CT image reconstruction, the more commonly utilized are filtered back projection and iterative algorithm.

The calculated attenuation values are represented as gray-scale values. The grayscale values in CT are called Hounsfield units (HUs). The HU is defined as [145]:

$$HU(x,y,z) = 1000\,\frac{(\mu(x,y,z) - u_w)}{u_w} \qquad (1.3)$$

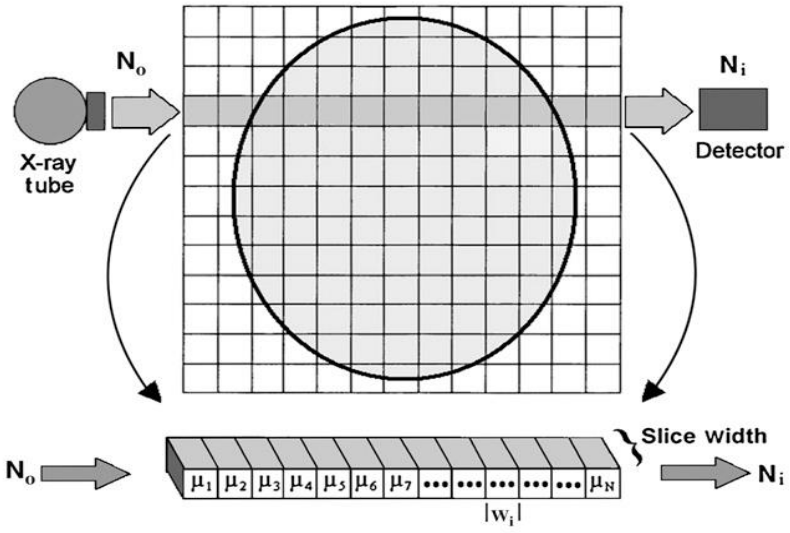



***Figure 1.10.*** Reconstruction matrix (from [145])

where $\mu(x,y,z)$ is the average linear attenuation coefficient for a voxel of tissue in the patient at the location $(x,y,z)$. *HU(x,y,z)* represents the grayscale CT images in the same $(x,y,z)$ spatial coordinates, and $\mu_w$ is the linear attenuation coefficient of water for the X-ray spectrum used. A cross-sectional slice of a patient with lung cancer is shown in Figure 1.11.

***Figure. 1.11.*** Thorax CT slice, where a lung tumor can be appreciated on the left side. In integrated PET/CT, a map of the attenuation coefficients of the patient can be obtained from the CT and used the perform attenuation correction in the PET image (personal collection).

### 1.6.2 Positron emission tomography

Positron emission tomography (PET) generates images depicting the distribution of a positron-emitting compound. Nowadays, PET systems are manufactured mainly with a coupled x-ray CT system, which is referred to as PET/CT.

PET imaging is based on the detection of pairs of photons resulting from the annihilation of a positron with an electron. It is an example of molecular imaging, where a radiopharmaceutical compound that emits $\beta+$ particles is administered to the patient. After a short distance traveling, this $\beta+$ particle will suffer an annihilation with a patient electron, following the emission of two photons in opposite directions and with an energy of 511 keV each one. Posteriorly, coincident detection of those photons, typically by a ring of detectors, allows defining the emission point, and thus to estimate the distribution of the radiopharmaceutical in the body.

There are several radionuclides available for PET, such as $^{18}$F, $^{11}$C, $^{13}$N, and $^{15}$O [146]. The PET radiotracer most widely used in oncology is $^{18}$F-FDG, which is an analog of glucose labeled with radioactive $^{18}$F. It helps to identify regions of pathological increase of glucose metabolism in the body. The basic principles of a PET system are shown in Figure 1.12 [147].

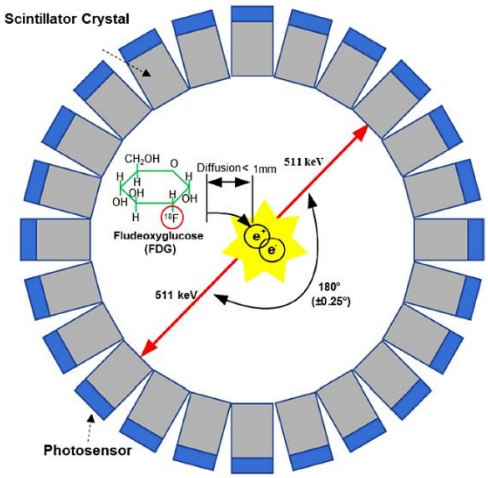



***Figure. 1.12.*** The basic principle of a PET system: a PET detector ring detects a pair of gamma photons with an energy of 511 keV (red arrows) which have resulted from the annihilation of an electron with a positron emitted by the radiotracer ($^{18}$F-FDG) (from [147]).

In general terms, a typical PET scanner consists of a ring of scintillation crystals attached to the photocathode of a photomultiplier tube (PMT). When an incoming photon excites electrons in the scintillation crystal by Compton scattering or photoelectric absorption, it causes the crystal to emit thousands of photons (around the visible part of the spectrum), which will be converted into an electrical signal by the PMT. Circuitry within the scanner identifies pairs of interactions occurring very near in the time, which is called annihilation coincidence detection (ACD). The circuitry of the scanner then determines the line in space connecting the locations of the two interactions, which are named line of response (LOR). Thus, ACD establishes the trajectories of detected photons.

Raw PET data collected in all detector pairs are used to reconstruct the image of the distribution of the activity concentration of PET radiotracers in the body. This can be achieved by employing different kinds of algorithms. However, to produce accurate and quantitative images, the acquired data have to be corrected for several physical processes involved in the emission, transmission, and detection of radiation, such as non-uniform response for a uniform source, attenuation scatters, and point spread function.

Both annihilation photons have to escape from the patient and reach the detectors to be registered as a coincidence in the PET scanner. The probability of both photons going through the patient without interaction is the product of the probabilities of each of them non-interacting [144]:

$$(e^{-\mu x}).e^{-\mu(d-x)} = e^{-ud} \qquad (1.4)$$

Where *d* is the total path length through the patient, *x* and (*d*-*x*) are the distances that photon must travel through the patient (Figure 1.13) [144].

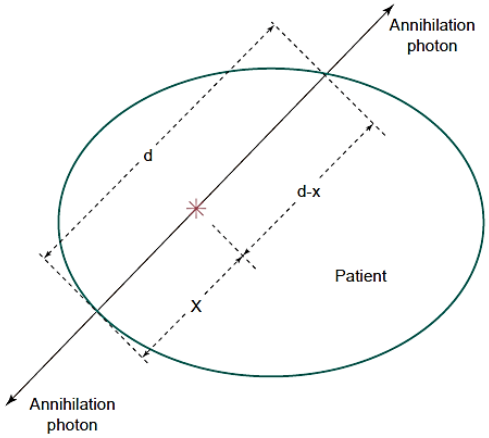



***Figure. 1.13.*** Attenuation in PET. The probability that both annihilation photons emitted along a particular LOR escape interaction in the patient is independent of location on the LOR (from [144]).

As a consequence of this, the attenuation correction in PET can be easily performed by measuring the attenuation through different LORs passing through the patient, by using an external source, either a positron emitter in a PET stand-alone scanner or using the CT component of the PET/CT. From this attenuation information, each LOR can be corrected by attenuation [148]. Other corrections such as random and scatter corrections, as well as corrections by positron range, improve the final image quality [149, 150].

The activity measured by the PET scanner depends on the patient's size and the injected activity. To have more comparable measurements, between patients, and for the same patient between different time points, the standardized uptake value (SUV) has been proposed for $^{18}$F-FDG acquisitions. It is a quantitative measure of radiopharmaceutical distribution, representing the ratio between the activity in a certain region or volume of interest and total injected activity, divided by normalizing factor, most commonly body weight, and is defined as [151, 152]:

$$SUV = \frac{Av}{A_i.w} \qquad (1.5)$$

Where: *Av* is the measured activity in voxel *i*, w is the bodyweight of patients and, $A_i$ is the injected activity corrected for decays during the uptake time.

SUV measurements can be influenced by a variety of biological and technological factors, such as body composition, blood glucose level, scanner variability, reconstruction parameter, use of contrast material, interobserver variability, and others [153]. For this reason, several other parameters have been proposed to quantify the tumoral lesion uptake in PET. The most commonly used metrics in the clinic are:

- **$SUV_{mean}$** – Average SUV in all pixels or voxels in a defined region or volume of interest.

- **$SUV_{max}$** – Represents the highest SUV score in the same region or volume of interest. For the PET-based monitoring of response to treatment, $SUV_{max}$ is known to be very sensitive to noise, because its definition relies on a single voxel measurement within the whole volume. On the other hand, it is less sensitive than $SUV_{mean}$, to the tumor contour delineation [154].

- **$SUV_{peak}$** – It has been promoted as a more robust metric of less susceptibility not only to tumor delineation but also to noise artifacts. It is calculated as the mean SUV in a spherical VOI of 1.2 cm diameter (volume of 1 ml) centered at the most active portion of the tumor [154].

Other parameters used for monitoring of treatment response are:

- **MTV** – (Metabolic tumor volume): the sum of all voxels in a volume defined by a segmentation method, which represents the metabolically active tumor tissue. The segmentation method used to define this group of voxels affects directly the value of this parameter.

- **TLG** – (Total lesion glycolysis): product of MTV and $SUV_{mean.}$

- **$SUL_{max}$** – $SUV_{max}$ corrected for lean body mass.
-
- **SULpeak** – $SUV_{peak}$ corrected for lean body mass.

## 1.7 Clinical applications of PET/CT

### 1.7.1 Cancer staging

To perform the best treatment possible and assess accurately the prognosis of a patient with cancer, a determination of the tumor stage as precise as possible must be performed. Cancer diagnosis requires the pathological confirmation of the tumor, whereas imaging is crucial to assess disease extension. As previously mentioned, the right choice of treatment depends on the correct TNM classification, which categorizes tumors based on the primary tumor characteristics as size and local invasion (T), regional lymph node involvement (N), and distant metastases (M). The addition of molecular imaging, mainly with $^{18}$F-FDG PET, to anatomical imaging such as CT or MR, has improved TNM staging and consecutively treatment outcome [155]. $^{18}$F-FDG PET/CT is a standard of care and an integral part of the clinical staging of patients with lung cancer [1]. NSCLC has shown to have a notable relationship between the FDG uptake, measured semi-quantitatively as the standardized uptake value, and the tumor size, histologic subtype, biologic aggressiveness, and prognosis [156]. An $^{18}$F-FDG PET/CT staging study is shown in Figure 1.14.

In breast cancer, whole-body $^{18}$F-FDG PET/CT imaging to assess the primary tumor characteristics and detect occult axillary lymph node metastases shows no benefit over the standard methods. However, the detection of extra-axillary lymph node involvement and distant metastases in the initial staging has been indicated as its major contribution [157]. $^{18}$F-FDG PET/CT imaging is used mainly for recidive detection and treatment response evaluation after neoadjuvant or systemic chemotherapy.

### 1.7.2 Response assessment

Assessment of tumor responses and treatment results is a fundamental part of an oncology treatment protocol. Although there are several methods to achieve this, medical imaging plays a critical role in objectively characterizing tumor response to therapy, and in a research context, defining trial endpoints for novel therapeutical agents.

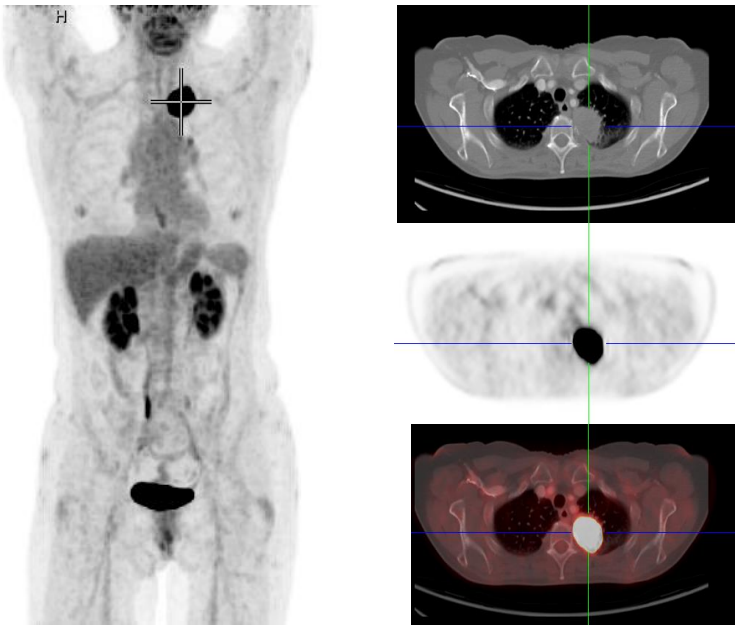


***Figure 1.14.*** $^{18}$F-FDG PET/CT a 62 years old female patient with NSCLC. The tumor has 4.3 cm of diameter and there are no lymph nodes or distant affectation (personal collection).

In radiology, the Response Evaluation Criteria in Solid Tumors (RECIST), version 1.1 [158], are routinely used. This uses unidimensional diameters of target lesions (measured in CT or MR), and the sum of measurements of all target lesions, as a quantitative measure of tumor burden. Changes of the quantitative tumor burden are assessed about the specific cutoff values to assign categorical response groups (complete, partial, stable, or progressive disease). These response evaluation criteria are limited by their dependence on only morphological changes to assess therapeutic response [159]. It is known, that functional and metabolic changes appear early in a course of treatment, and reflex better the tumor response [160]. In nuclear medicine exist the European Organization for Research and Treatment of Cancer PET response criteria (EORTC PET) and the PET response criteria in solid tumors (PERCIST) 1.0, which follow the model of RECIST, and define 4 response categories with similar names as RECIST – complete metabolic response (CMR), partial metabolic response (PMR), stable metabolic disease (SMD) and progressive metabolic disease [159]. PERCIST 1.0 recommends using SUV corrected for lean body mass (SUL) to avoid falsely high organ SUV in obese patients. It also recommends computing the SULpeak and total lesion glycolysis (TLG) values as response metrics. Whereas EORTC PET uses SUVmean, corrected by body surface area. A comparison between the three criteria mentioned above is shown in Table 1.4.

***Table 1.4.*** Response evaluation criteria for RECIST, EORTC, and PERCIST 1.0 (from [159]).

| Characteristic | RECIST 1.1 | EORTC | PERCIST 1.0 |
|---|---|---|---|
| Measurability of lesion at baseline | Lesions: longest diameter≥10 mm; lymph nodes: short axis≥15 mm | Lesions with high18F-FDG uptake | SUL peak of baseline lesions at least 1.5-fold greater than liver SUL mean + 2 × SD. If the liver is abnormal, the primary tumor should have uptake>2.0 × SUL mean of the blood pool |
| Objective response | **CR**: disappearance of all target lesions | **CMR**: complete resolution of $^{18}$F-FDGuptake within all lesions, making them indistinguishable from the surrounding tissue | **CMR**: complete resolution of18F-FDG uptake within all lesions to a level of less than or equal to that of the mean liver activity and indistinguishable from the background blood-pool levels |
| | **PR**: reduction of at least30 % in the sum of diameters of target lesions | **PMR**: reduction of at least 25 % in the sum of SUV | **PMR**: reduction of at least 30 % in SUL peak and an absolute drop of 0.8 SUL peak units |
| | **PD**: increase of at least 20 % in the sum of diameters of target lesions or appearance of new lesions | **PMD**: increase of at least 25 % in the sum of SUV or appearance of new $^{18}$F-FDG-avid lesions that are typical of cancer and not related to inflammation or infection | **PMD**: increase of at least 30 % in SUL peak and an absolute increase of 0.8 SUL peak units<br>OR: 75 % increase in TLG, with no decrease in SUL, or appearance of new $^{18}$F-FDG-avid lesions typical of cancer and not related to inflammation or infection |
| | **SD**: not CR, PR, or PD | **SMD**: not CMR, PMR, or PMD | **SMD**: not CMR, PMR, or PMD |

CR: complete response, PR: partial response, PD: progression disease, SD: stable disease, CMR: complete metabolic response, PMR: partial metabolic response, PMD: progression metabolic disease, SMD: stable metabolic disease.

EORTC criteria and PERCIST 1.0 have shown to be more sensitive and accurate than RECIST 1.1, especially for the detection of an early therapeutic response to chemotherapy. Changes in tumor metabolic activity are an early indicator of treatment effectiveness for breast cancer, mainly in the neoadjuvant setting.

The histopathologic response after chemotherapy has been used as the reference standard for assessment of the accuracy of $^{18}$F-FDG PET in predicting a response during systemic treatment. Figure 1.15 shows $^{18}$F-FDG PET/CT studies of a patient with locally advanced breast cancer, before and after completion of neoadjuvant chemotherapy [98]. A significant metabolic response can be appreciated. The treatment was performed with Carboplatin, Docetaxel, Trastuzumab, and Pertuzumab.

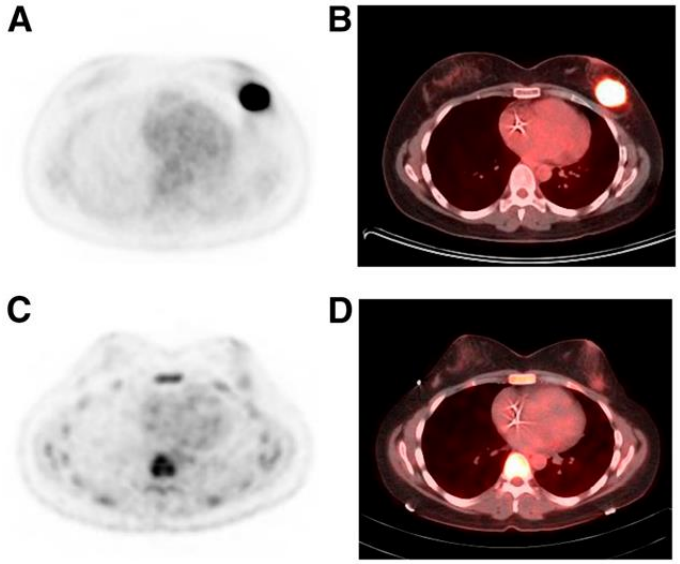



***Figure 1.15***. 37-years-old woman with HER2-positive ductal breast cancer (4.9 · 3.1 cm). (A and B) baseline $^{18}$F-FDG PET (A) and fused $^{18}$F-FDG PET/CT (B). (C and D) significant reduction in tumor $^{18}$F-FDG uptake, from SUV of 16.1 to SUV of 1.6, was seen on $^{18}$F-FDG PET (C) and fused $^{18}$F-FDG PET/CT (D). Histopathology after treatment showed minimal residual disease in the tumor bed (from [98])

Likewise, $^{18}$F-FDG-PET/CT has shown high accuracy in diagnosing metastatic breast cancer, and PERCIST criteria have shown higher predictive values than RECIST for prediction of progression-free survival [75, 94]. As recurrent and metastatic breast cancers are FDG-avid diseases [94], $^{18}$F-FDG-PET/CT is a valuable tool for response monitoring.

# CHAPTER

# 2

# PET/CT image biomarkers and Radiomics

## 2.1 Introduction

Cancer greatly affects our modern society. Not only because of the number of deaths but also because of its morbidity. Currently, sophisticated medical imaging methods, such as MR, CT, and PET/CT, allow the diagnosis, treatment, and follow-up of patients with cancer. Although some very basic quantitative or semi-quantitative parameters extracted from these images have been introduced, the evaluation of medical images has historically been a qualitative process, based mainly on the expertise of the medical specialist.

However, in the last years, a large number of scientific publications have shown the value of quantitative parameters extracted of the medical images (image features), for diagnosis, evaluation, and prognosis of the response to treatment, as well as prognosis of the long-term evolution of patients with cancer [19-27]. So then, we are moving from a simple qualitative inspection of medical images towards the use of modern image-based biomarkers. In this section, we concentrate on the quantitative imaging biomarkers (quantifiable features from medical images), and we briefly discuss one of the commonly-used quantitative imaging biomarkers in PET/CT.

## 2.2 Imaging biomarkers (IB)

An imaging biomarker (IB) can be defined as a characteristic derived from a medical image and can be used as an indicator of normal biological processes, a pathogenic process, or of responses to an exposure or intervention, including therapeutic interventions [161]. IBs have the advantage to provide a full view of the entire extent of the tumor, capturing regional variations in it (tumor heterogeneity).

They can be performed in a non-invasive way and repeatedly to monitor the development and progression of the disease as well as response to therapy. Besides, they have the potential to reveal tumor phenotypes (inside a patient population), associated with prognosis, diagnosis, therapy response, and risk stratification (Figure 2.1).

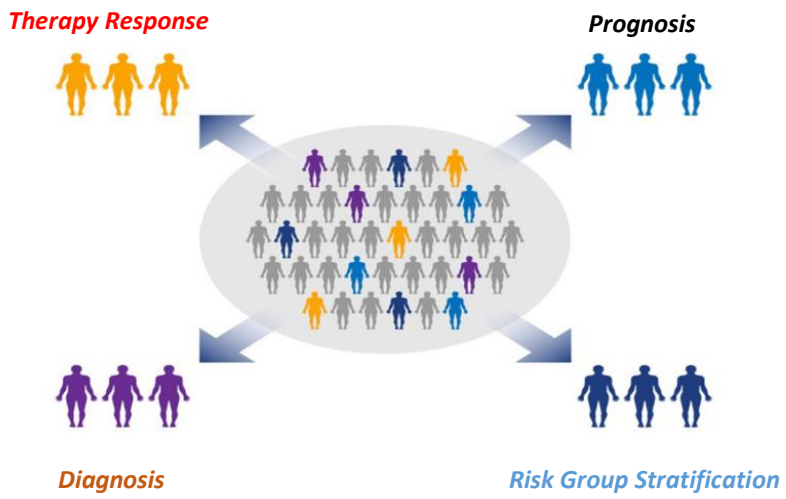



***Figure 2.1.*** Applications of cancer biomarkers (adapted from [162])

Although several IBs have been developed in the last few years, their clinical validation is a major hurdle. For this reason, simple biomarkers such as tumor size or SUV are still the most commonly used in everyday clinical practice. Some important definitions related to quantitative IB development are:

- Predictive biomarker: a biomarker intended to forecast disease course in the presence of a specific treatment

- Prognostic biomarker: a biomarker intended to forecast disease course in the absence of treatment

- Quantitative imaging: extraction and use of numerical/statistical features extracted from medical images

- Repeatability: the agreement between successive measurements made under the same conditions

- Reproducibility: the agreement between successive measurements made with varying conditions, such as location or operator

- Surrogate endpoint: a biomarker intended to substitute for a clinical endpoint

Before an imaging biomarker becomes routinely used in the clinic, it must be validated. The validation involves a demonstration of the accuracy, precision, and feasibility of biomarker measurement. On the other hand, if an imaging biomarker cannot be reliably measured, it will have little or no use as an indicator of a biological process or clinical outcome. IBs can be generally classified as structural, morphological, textural, functional, or physical [66, 161]. Some factors can intrinsically affect the measuring of quantitative imaging biomarkers and thus reducing their reliability, repeatability, and reproducibility; therefore, all these questions need to be properly investigated.

An IB routinely used in nuclear medicine, which previously was described in chapter 1, is the SUV of $^{18}$F-FDG PET images. Increased accumulation of FDG in tumors, relative to normal tissue, has shown to be a useful marker for the detection and staging of many cancers [98, 160]. And its temporal changes are useful for assessing response to therapy. Nevertheless, many factors can affect the SUV measure, for which other IB have been introduced. Positive tumor $^{18}$F-FDG uptake on visual analysis and intensity of uptake expressed as SUV, are associated with prognosis in patients with malignant diseases [163]. Usually, diagnosticians use a semiquantitative parameter, the maximum standardized uptake value (SUVmax), to characterize the tumoral lesion [152]. Also, although not routinely used, global or volume-based semiquantitative metabolic parameters of the primary tumor, such as SUVmean, MTV, and TLG, can be easily obtained from post-processed images. These parameters provide a more accurate assessment of the tumor burden, with potentially higher predictive and prognostic value than SUVmax in some cancers [17]. Unlike SUVmax, which measures $^{18}$F-FDG concentration in a single voxel of a metabolically active tumor, global semiquantitative parameters take into account all voxels inside the tumoral volume, thus probing a more general view of the tumor. Their drawback is the lack of a standardized method for volume definition, although several segmentation algorithms show good performance [154].

On the other hand, the biological heterogeneity of the tumors is an important factor implicated in poor treatment response, a higher chance of developing metastasis, and shorter progression-free and overall survival [6-8]. Generally, tumor FDG uptake shows uneven spatial distribution, at least partly due to intratumoral heterogeneity, which was already treated in chapter 2. Despite the clinical importance of tumor heterogeneity, established $^{18}$F-FDG PET/CT parameters such as SUV, MTV, SUVmean, and TLG do not reflect this property, raising the need for different analytic methods.

The conventional PET/CT analysis approach consists of the following: (1) PET/CT imaging acquisition and reconstruction; (2) tumor segmentation; (3) feature extraction such as SUVmax, SUVmean, SUVpeak, MTV, and TLG, and (4) statistical analysis. However, in recent years, there has been considerable effort in the medical imaging community to obtain correlations between image features and tumor heterogeneity, as a means to improve the clinical management of patients with cancer. It is thought that medical images have much more information than is directly appreciated by the human eye or through simple measurements [25, 98]. Through several mathematical methods, it is possible to extract spatio-temporal distribution patterns of the pixel/voxel intensities, as well as their relationship. This is the field of radiomics, which refers to the extraction and analysis of large amounts of advanced quantitative imaging features from medical images [25]. The hypothesis behind the radiomics approach is that these advanced imaging features change noticeably between different pathological processes or after therapeutic intervention. Therefore, radiomic features can be used to build descriptive and predictive models of a specific clinical outcome ("radiomic signature"). It is believed that radiomics has the potential to improve on traditional, manual interpretation by detecting features and patterns that otherwise would go unnoticed to the human eye [25].

Radiomics analysis can be performed on medical images from different modalities, allowing for an integrated cross-modality approach using the potential additive value of imaging information extracted, e.g., from magnetic resonance imaging (MRI), computed tomography (CT), and positron emission tomography (PET), instead of evaluating each modality by its own [67]. Radiomics has emerged from oncology, but can be applied to other medical problems where a disease is imaged.
The overall objective of radiomics is to build classification and/or regression models based on some quantitative features extracted from the imaging data. Currently, with the emergence of radiomics, there is a large amount of information available, powerful AI techniques can unlock clinically relevant information hidden in this massive amount of information [164], that cannot be directly evaluated by humans. This large amount of information and the availability of more powerful computers has stimulated the application of artificial intelligence (AI) to perform these classifications and/or regression tasks.

Radiomic analysis can be applied to any pathologic process, where medical images are used at any time along with its evolution. In oncology, the radiomic analyses are mainly performed in the tumor regions, however, it is possible also to perform it in normal tissues. On the other hand, this analysis can be carried out with PET, CT, or both images.

## 2.3 PET/CT radiomic methodology and workflow

The general process for building a PET/CT radiomic signature with prognostic/predictive value is rather independent of the underlying disease and consists of several steps [165, 166] (Figure 2.2). It is similar to the conventional PET/CT analysis but needs some additional steps such as intensity quantization and feature extraction.

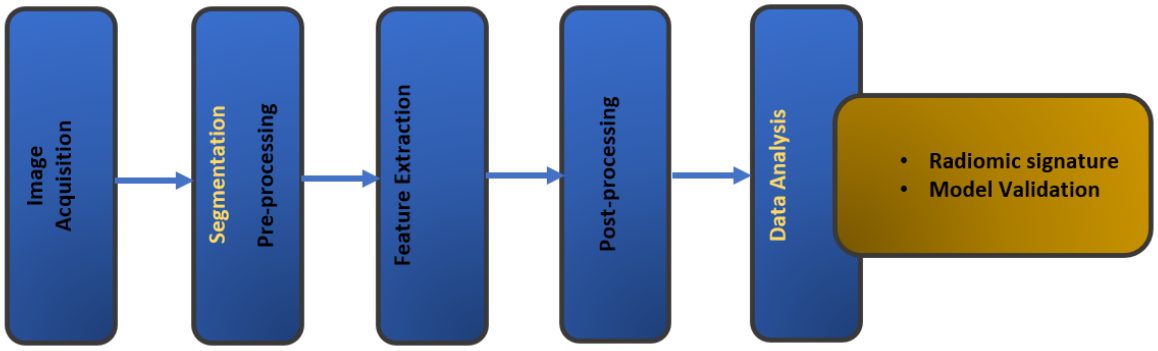



***Figure 2.2.*** The standard workflow in radiomics (adapted from [165]).

### 2.3.1 Acquisition

The radiomic workflow begins with the medical image acquisition, which can be performed in two, three, or four dimensions if time is taking into account. It includes both the examination itself and the patient preparation protocol. The output will be a medical image consist of pixels or voxels, whose intensity represents the physical, chemical, or biological properties of an equivalent area/volume in the patient.

A wide range of parameters intervenes in the acquisition process, among them tube current and voltage for the CT acquisition, spatial resolution (voxel size of CT and PET), reconstruction algorithm, and related settings both for CT and PET. All these parameters may have a significant impact on the radiomic features computed [167, 168], with certain features being affected more than others. Thus, robust radiomic features in reconstruction settings can be considered as good parameters in radiomic analysis. Additionally, the movement during the acquisition as well as respiratory movements influences significantly the values of the radiomic features [169].

Since the images can be obtained at different scales, and there are variations across different medical machines and acquisition protocols, recommendations to perform radiomic analysis have been published [170].

### 2.3.2 Tumor segmentation and preprocessing

The process of tumor segmentation can be categorized into two sequential phases; recognition and delineation [171]. In the recognition phase, a target lesion with high uptake is identified and distinguished from other target lesion-like entities in the PET image. In the delineation phase, the target lesion is precisely separated from the background and non-significant object. The nuclear medicine physician can perform the segmentation manually. However, it is highly subjective and time-consuming and has high intra- and inter-observer variability, whereby it is less reproducible [172]. Semiautomatic methods such as gradient-based and thresholding-based can be also used. The latter can use with an adaptive or fixed threshold, being a cutoff value of SUV of 2.5 commonly used for malignant tumor delineation, also, deviations of a reference uptake (e.g. in liver or mediastinum) can be used [173, 174]. Other methods are automatic, based on different algorithms such as C-Means (FCM) and fuzzy locally adaptive Bayesian (FLAB) [172, 175]. More recently, machine learning algorithms have been proposed [176]. Radiomics features are sensitive to different delineation methods [177]. Therefore, it is recommended to applicate multiple segmentations and assess the feature sensibility to them [24, 170].

Pre-processing is a crucial step in the radiomic workflow, affecting significantly the overall outcome of a radiomic signature [177]. It may involve spatial filtering, resampling, and intensity discretization. The objective of spatial filtering is either to reduce noise or emphasize features at different scales. Some filters commonly used are Butterworth smoothing, Gaussian filters, and Laplacian of Gaussian filters [165, 166].

Textural features values are sensitive to variations in voxel size [177, 178-182], which affects its reproducibility when images with different acquisition settings are used. There are two ways to reduce this effect, namely through reconstruction to the same voxel dimensions or by image interpolation. In most situations, re-reconstruction is not possible due to the lack of access to the original raw data, and image resizing is required.

Image resizing, which allows converting an image to another with a different voxel size, requires voxel interpolation. Commonly used interpolation algorithms include nearest neighbors, linear and cubic convolution, and cubic spline. In short, nearest-neighbor interpolation assigns the intensity of the most nearby voxel in the original grid to each voxel in the interpolation grid. Trilinear interpolation uses the intensities of the eight most nearby voxels in the original grid to calculate a new interpolated intensity using linear interpolation. tricubic convolution and tricubic spline interpolation draw upon a larger neighborhood to evaluate a smooth, continuous third-order polynomial at the voxel centers in the interpolation grid [183]. Since the interpolation result is an approximation, artifacts should be expected. These have been classified into four broad categories, namely ringing, aliasing, blocking, and blurring. [184].

While no consensus exists concerning the optimal choice of the interpolation algorithm, trilinear interpolation is usually seen as a conservative choice. It does not lead to the blockiness produced by nearest-neighbor interpolation that introduces a bias in local textures [185], nor does it lead to out-of-range intensities that may occur due to overshoot with tricubic and higher-order interpolations. The latter problem can occur in acute intensity transitions, where the local neighborhood itself is not sufficiently smooth to evaluate the polynomial within the allowed range. Tricubic methods, however, may retain tissue contrast differences better. Particularly when upsampling, trilinear interpolation may act as a low-pass-filter that suppresses higher spatial frequencies and cause artifacts in high-pass spatial-filters [186]. Interpolation algorithms and their advantages and disadvantages are treated in more detail in [184]. In a phantom study, Larue *et al.* [186] compared the nearest neighbor, trilinear and tricubic interpolation, and indicated that feature reproducibility is dependent on the selected interpolation algorithm, i.e. some features were more reproducible using one particular algorithm.

In routine clinical settings, most imaging modalities produce anisotropic voxels after scan reconstruction, where the thickness between axial slices is larger than the cross-sectional resolution (i.e. $\Delta z > (\Delta x, \Delta y)$). To establish conservation of scale in all three directions, and remove a directional bias in 3-dimensional (3D) features, it is recommended to resample images with 3D interpolation such that $\Delta z = \Delta x = \Delta y$ [178]. Texture features quantify spatial variation in voxel intensities, and interpolation either decreases (up-sampling) or increase (down-sampling) the spatial distance between voxels. Down-sampling to a larger voxel size leads to information loss, where-as up-sampling to a smaller voxel size creates artificial information at a higher resolution. Extreme down-sampling creates a poor-quality image, extreme up-sampling creates local homogeneity and image smoothing [183, 186].

The main reasons to use resampling are, to allow for comparison of datasets obtained from multiple centers (there are variations regarding protocols and reconstruction parameters that result in different voxel sizes), resampling of registered multimodal imaging (such as in PET/CT) to the same voxel resolution, and to acquire isotropic voxel dimensions for 3D feature extraction because they require isotropic voxel spacing to be rotationally invariant [183, 187].

Currently, there are no clear indications of whether upsampling or down-sampling schemes are preferable [183]. For example, Ligero *et al*. [182] studied the influence of voxel size for radiomic feature variability in CT-images. They found that voxel size resampling increased the mean percentage of robust CT radiomic features. The radiomic data were extracted from images resampled to isometric voxels of $1 \times 1 \times 1$ mm$^3$. Shafiq-ul-Hassan *et al*. [188] assessed the impact of slice thickness and pixel size on features acquired on CT phantom images with different acquisition and reconstruction parameters. Images were resampled to one voxel size $(1 \times 1 \times 2)$ mm$^3$ using linear interpolation to determine if this improved robustness. 42 out of 213 features studied improved significantly after resampling. Twenty-one features had large variations before and after resampling.

Whybra *et al*. [186] assessed the stability of radiomic features to interpolation processing and categorized features based on stable, systematic, or unstable responses in $^{18}$F-FDG-PET images of oesophageal cancer patients. Images were resampled by using a linear and spline method to 6 isotropic voxel sizes (1.5 mm, 1.8 mm, 2.0 mm, 2.2 mm, 2.5 mm, 2.7 mm) and 141 features were extracted from each VOI. They found that texture features are sensitive to interpolation and evaluated potential correction techniques for features showing a potential systematic voxel-size dependence. They also found large differences in many features extracted when interpolating using a linear method compared to spline. However, the robustness categorizations remained consistent for all features; stable features had stable responses for both interpolation methods. However, they did not investigate the impact of interpolation on predicting the performance of any clinical parameter. A feature with large variability between interpolation methods may still show strong predictive significance in a developed radiomic model. Thorough reporting of feature extraction settings including the interpolation method is a necessity for reproducibility and validation. They identified robust features showing stability to isotropic interpolation, but this does not necessarily correspond to any clinical application. However, clinically robust and thus relevant features are likely to be a subset of those that have a predictable interpolation response. Therefore, due to the abundance of features in radiomics and the need for

reduction techniques to limit overfitting, pruning features that have not shown required interpolation stability may be one of several selection steps to consider for all radiomic studies with multi-center datasets that requires resampling to common voxel size. Isotropic voxel size is thought to impact the predictive value of features [20]. Image intensities may require rounding after interpolation or the application of cut-off values. For example, in CT images intensities represent Hounsfield units, and these do not take non-integer values. Following voxel interpolation, interpolated CT intensities are thus rounded to the nearest integer. In this thesis, in line with some other clinical works [189, 190], we have preferred to resampling into an isotropic voxel size of (1x1x1) $cm^3$

VOI voxel intensity discretization or quantization is often required for the calculation of higher-order texture and heterogeneity features [166]. During the discretization process, the intensities or SUVs voxels are discretized into new values. It reduces the effect of noise in radiomic analysis, by changing the continuous voxel intensity scale to one discrete. In $^{18}$F-FDG PET/CT images, two different methods are commonly used for intensity or SUV quantization, namely the fixed bin-number and the fixed bin-size method.

Fixed bin-number discretization:

For a VOI with N voxels, the discretization into D number of bins (e.g., 16, 32, 64, 128, 256) can be performed by the following formula [166]:

$$I(x)_{new} = round\left(D\frac{I(x)-SUV_{max}}{SUV_{max}-SUV_{min}+1}\right) \quad (2.1)$$

Where, $I(x_i)$ is the intensity or SUV of the $i^{th}$ voxel, and $I(x)_{new}$ is the resampled voxel value.

An adequate number of bins is 32 or 64 because it allows having a resolution sufficiently fine of the SUV [20, 66]. Radiomic features are highly dependent on the number of bins. However, a fixed bin-

number discretization enables a direct comparison of feature values across multiple VOIs or multiple patients.

Fixed bin-size discretization:

For a VOI with N voxels, the discretization into a fixed-bin size B (e.g., 0.1, 0.25, 0.5, or 1 of SUV), can be performed by the following formula [166]:

$$I(x)_{new} = \left(\frac{I(x)}{B}\right) - min\left(round\left(\frac{I(x)}{B}\right)\right) + 1 \qquad (2.2)$$

The fixed bin-number and the fixed bin-size methods mostly result in discordant values of texture features to each other. However, the fixed bin-size method produces identical intensity resolution among different PET images, which enables direct comparison of values of texture features in different PET images, even in a clinical response setting that the SUV range of a tumor varies during treatment. It also seems to have better repeatability and lower sensitivity of texture features to the methods of delineation and reconstruction [166].

To be clinically useful, a radiomic feature must be robust to image processing steps, which has made robustness testing a necessity for many technical aspects of feature extraction. Therefore, it is necessary for the standardization of these methodologies for performing radiomic studies [191, 192].

### 2.3.3 Feature extraction

From these pre-processed VOIs several radiomic features are extracted, currently is possible to obtain thousands of different features by applying different mathematical concepts and transformations, or more recently via Deep Learning (section 3.5 in this thesis).

There are two big classes of features: “hand-crafted” and those based on Deep Learning. Hand-crafted features are obtained via some suitable mathematical functions. Most common among them are shape and texture features. By contrast, Deep Learning features are obtained implicitly by training on large datasets of images [165].

Without pretending to be exhaustive in definitions, below we mention some of the characteristics used in this thesis. For a complete definition of each feature and their formulation, please refer to the "*Image biomarker standardization initiative"* (IBSI) document [191, 192]. Some formulas and definitions are given in ***Table C1 of Appendix C.***

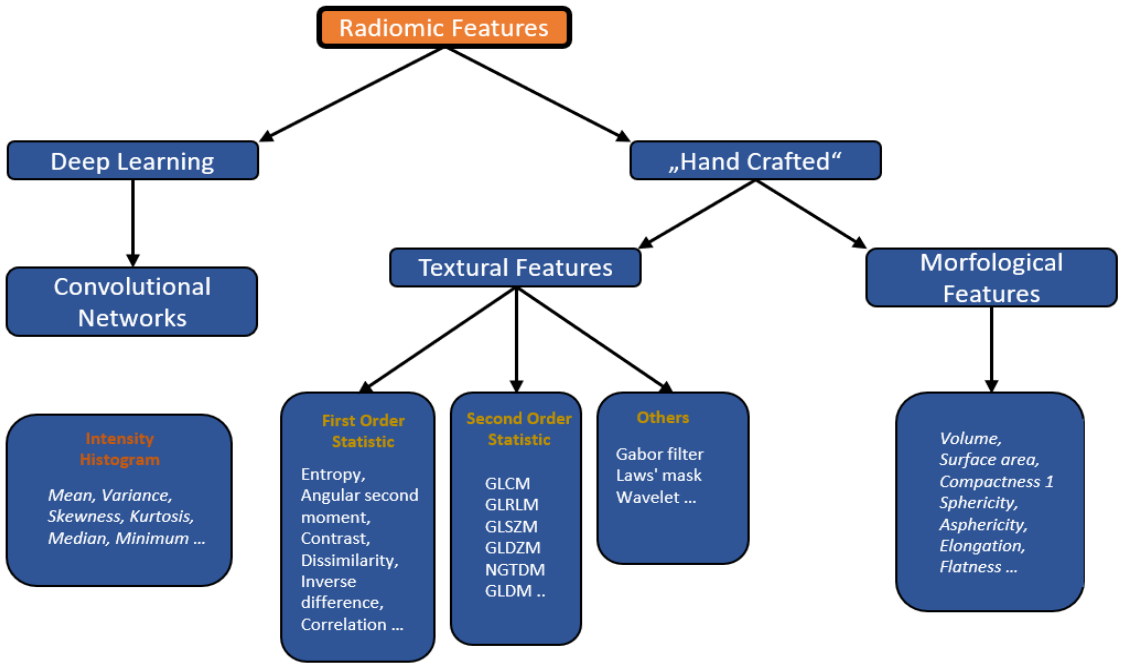



***Figure 2.3.*** Radiomic features classification (adapted from [165])

***Morphological or shape features*** describe geometric aspects of an ROI or VOI. These features are based on voxel representations of the volume, which can be set as coordinates of the voxel centers, or a surface mesh representation [191]. A list of some of them is presented in Table 2.1.

***Table 2.1.*** Morphological features

| |
|---|
| Volume, Approximate volume, Surface area, Surface to volume ratio, Compactness 1, Compactness 2, Spherical disproportion, Sphericity, Asphericity, Centre of mass shift, Maximum 3D diameter, Major axis length, Minor axis length, Least axis length, Elongation, Flatness. |

Texture features are categorized into first, second, and higher orders by the number of voxels involved in their texture matrix design. The first-order texture features are calculated from the intensity frequency histogram, which represents the frequency distribution of one-voxel intensity in the ROI or VOI of the tumor. Therefore, they describe only the global characters of a tumor on PET/CT images. The first-order texture features include:

***Intensity-based statistical features****, which* describe how voxel values within the ROI are distributed. These voxel-values do not need discretization. Below is a list of these features [191]. A list of some of them is presented in Table 2.2.

***Table 2.2.*** Intensity-based statistical features

| |
|---|
| Mean, Variance, Skewness, Kurtosis, Median, Minimum, Percentile 10, Percentile 90, Maximum, Interquartile range, Range, Mean absolute deviation, Robust mean absolute deviation, Median absolute deviation, Coefficient of variation, Quartile coefficient of dispersion, Energy, Root mean square. |

**Intensity histogram features**, which are calculated from a discretized histogram of the original voxel intensities. Below is a list of these features [191]. A list of some of them is presented in Table 2.3.

***Table 2.3.*** Intensity histogram-based features

| |
|---|
| Mean, Variance, Skewness, Kurtosis, Median, Minimum, Percentile 10, Percentile 90, Maximum, Mode, Interquartile range, Range, Mean absolute deviation, Robust mean absolute deviation, Median absolute deviation, Coefficient of variation, Quartile coefficient of dispersion, Entropy, Uniformity, Maximum gradient, Maximum gradient grey level, Minimum gradient, Minimum gradient grey level. |

***Intensity-volume histogram*** **features**, which use a cumulative intensity-volume histogram (IVH) of the voxel intensities in the ROI, which describes the relationship between discretized intensity *i* and the fraction of the volume containing at least intensity *i*. Below is a list of these features [191]. A list of some of them is presented in Table 2.4.

***Table 2.4.*** Intensity-volume histogram-based features

| |
|---|
| Volume fraction at 0.10 intensity, Volume fraction at 0.90 intensity, Intensity at 0.10 volume, Intensity at 0.90 volume, Difference volume fraction at 0.10 and 0.90 intensity, Difference intensity at 0.10 and 0.90 volume, The area under the IVH curve. |

The second-order texture features are calculated based on grey level co-occurrence matrices.

**Grey level co-occurrence matrix (GLCM)** determines how often (i.e., the probability) that a pixel of intensity *i* finds itself within a certain relationship to another pixel of intensity *j*. Namely, how two voxels "co-occur" with respect to one another; along a certain direction and with a certain distance [191]. GLCM with a neighboring distance of 1 has a 26-connected neighborhood in 3D and an 8-connected neighborhood in 2D, yielding 13 unique directions in 3D and four in 2D. Thus, for a 3D approach with a distance of one, an ROI has 13 unique GLCMs for every 13 direction. A list of some of them is presented in Table 2.5.

***Table 2.5.*** Grey level co-occurrence-based features

| |
|---|
| Joint maximum, Joint average, Joint variance, Joint entropy, Difference average, Difference variance, Difference entropy, Sum average, Sum variance, Sum entropy, Angular second moment, Contrast, Dissimilarity, Inverse difference, Inverse difference normalized, Inverse difference moment, Inverse difference moment normalized, Inverse variance, Correlation, Autocorrelation. |

The higher-order texture features are calculated from several texture matrices computed based on interrelationships of 3 or more voxels [191].

**Grey level run length-based Matrix (GLRLM)** quantifies gray level runs, which are defined as the length in the number of pixels, of consecutive pixels that have the same gray level value. In a gray level run length matrix $P(i,j|\Theta)$, the $(i,j)^{th}$ element describes the number of runs with gray level *i* and length *j* occur in the image along angle $\Theta$ [191]. A list of some of them is presented in Table 2.6.

***Table 2.6***. Grey level run length-based features

| |
|---|
| Short runs emphasis, Long runs emphasis, Low grey level run emphasis, High grey level run emphasis, Short run low grey level emphasis, Short run high grey level emphasis, Long run low grey level emphasis, Long run high grey level emphasis, Grey level non-uniformity, Grey level non-uniformity normalized, Run length non-uniformity, Run length non-uniformity normalized, Run percentage, Grey level variance, Run length variance, Run entropy. |

**Gray Level Size Zone-based matrix (GLSZM)** quantifies gray level zones in an image. A gray-level zone is defined as the number of connected voxels that share the same gray level intensity. In a gray level size zone matrix P*(i, j)*, the *(i, j)*$^{th}$ element equals the number of zones with gray level *i* and size *j* appear in the image [191]. A list of some of them is presented in Table 2.7.

***Table 2.7***. Gray Level Size Zone-based features

| |
|---|
| Short runs emphasis, Long runs emphasis, Low grey level run emphasis, High grey level run emphasis, Short run low grey level emphasis, Short run high grey level emphasis, Long run low grey level emphasis, Long run high grey level emphasis, Grey level non-uniformity, Grey level non-uniformity normalized, Run length non-uniformity, Run length non-uniformity normalized, Run percentage, Grey level variance, Run length variance, Run entropy. |

**Grey level distance zone matrix (GLDZM)** counts the number of groups (or zones) of linked voxels that share a specific discretized grey level value and possess the same distance to the ROI edge [191]. The GLDZM thus captures the relation between location and grey level. A list of some of them is presented in Table 2.8.

***Table 2.8.*** Grey level distance zone-based features

| |
|---|
| Small distance emphasis, Large distance emphasis, Low grey level zone emphasis, High grey level zone emphasis, Small distance low grey level emphasis, Small distance high grey level emphasis, Large distance high grey level emphasis, Grey level non-uniformity, Normalized grey level non-uniformity, Zone distance non-uniformity, Normalized zone distance non-uniformity, Zone percentage. |

**Neighborhood grey tone difference matrix (NGTDM)** quantifies the difference between a gray value and the average gray value of its neighbors within distance *d*. The sum of absolute differences for gray level *i* is stored in the matrix [191]. A list of some of them is presented in Table 2.9.

***Table 2.9.*** Neighborhood grey tone difference-based features

| |
|---|
| Contrast, Busyness, Complexity, Strength |

**Neighboring grey level dependence (GLDM)** quantifies gray level dependencies in an image. A gray level dependency is defined as the number of connected voxels within distance *d* that are dependent on the center voxel. A neighboring voxel with gray level *j* is considered dependent on the center voxel with gray level *i* if $|i-j| \leq \alpha$. In a gray level dependence matrix *P(i,j)* the $(i,j)^{th}$ element describes the number of times a voxel with gray level *i* with *j* dependent voxels in its neighborhood appears in the image [191]. A list of some of them is presented in Table 2.10.

***Table 2.10.*** Neighboring grey level dependence-based features

| |
|---|
| Low dependence emphasis, High dependence emphasis, Low grey level count emphasis, High grey level count emphasis, Low dependence low grey level emphasis, Low dependence high grey level emphasis, High dependence low grey level emphasis, High dependence high grey level emphasis, Grey level non-uniformity, Normalized grey level non-uniformity. |

### 2.3.4 Post-processing (features selection)

The obtained image features can undergo further processing to reduce redundancy and/or increase their discrimination capability. The most common approaches to this end are feature selection and feature generation. Feature selection consists of retaining a subset of the original features by selecting the most discriminative ones. This is crucial in radiomics, for some image features tend to be strongly correlated with one another [165]. Approaches to feature selection come in different varieties, such as correlation-based selection, reduction based on mutual information gain, recursive elimination, and Lasso regularization. Feature generation involves obtaining new features by combining the original ones through some suitable transformations, such as Linear Discriminant Analysis (LDA), Principal Component Analysis (PCA), and Multi-Dimensional Scaling (MDS) [165].

Without wishing to be exhaustive, we will now describe the methods used in this thesis, specifically in chapter 6:

### 2.3.4.1 Analysis of variance (ANOVA) with F-value

ANOVA is a parametric statistical hypothesis test for determining whether the means from two or more samples of data come from the same distribution or not [193]. The statistic parameter used in the ANOVA, known as the $F_{ratio}$, is the ratio between the variance of the group means and the average of the variance within the groups. This statistic follows a distribution known as "Fisher-Snedecor's F". ANOVA is used when one variable is numeric and one is categorical, such as numerical input radiomic features and a classification target variable (treatment response or not). The results of this test can be used for feature selection where those features that are independent of the target variable can be removed from the dataset.

### 2.3.4.2 Mutual information (MI)

MI between two random variables is a non-negative value, which measures the dependency between the variables [194]. The concept of MI is intimately linked to that of entropy (E) of a random variable, a fundamental notion in information theory that quantifies the expected "amount of information" held in a random variable. H is a measure of the uncertainty of a random variable. The uncertainty is related to the probability of occurrence of an event. Thus, high entropy means that each event has about the same probability of occurrence, while low entropy means that each event has a different probability of occurrence. The MI is a measure of the amount of information that one random variable has about another variable. This definition is useful within the context of feature selection because it gives a way to quantify the relevance of a feature subset with respect to the output vector. This value is equal to zero if and only if two random variables are independent, and higher values mean higher dependency.

### 2.3.4.3 Principal component analysis (PCA)

Principal Component Analysis (PCA), is a dimensionality-reduction method that allows reducing the dimensionality of large data sets, by transforming the variables into a smaller one, that still contains

most of the information of the large set. The goal of the PCA is to find a correlation, maximizing variance. A better description is performed in section 3.4.3.

#### 2.3.4.4 Independent component analysis (ICA)

ICA like PCA is a dimensionality reduction method that separates a multivariate signal into additive subcomponents that are maximally independent (unlike PCA, it maximizes independence among features). To do it, ICA finds a linear transformation of the feature space into a new feature space such that each of the individual new features are mutually independent, in a statistical sense.

#### 2.3.4.5 Least absolute shrinkage and selection operator (Lasso)

Lasso regression like linear models tries to find the relationship between predictors ($x1$,$x2$,...$xn$) and the response variable ($y$) as follows: $y=\beta0+\beta1x1+\beta2x2+\cdots+\beta nxn,$ where, the coefficients $\beta1,\cdots,\beta n$ correspond to the amount of expected change in the response variable for a unit increase/decrease in the predictor variables. $\beta0$ is the intercept and it corresponds to the variation that is not captured by the other coefficients in the model. Lasso regression applies a mathematical penalty on the predictor variables that are less important for explaining the variation in the response variable. This way, the strongest predictors for understanding how the response variable changes can be selected. This method uses a different penalization approach which allows some coefficients to be exactly zero (L1 regularization). Lasso has a regularization parameter, alpha, that controls how strongly coefficients are pushed toward zero. A lower alpha allowed us to fit a more complex model. The higher value of lambda indicates more regularization (i.e. reduction of the coefficient magnitude, or shrinkage). Log alpha = 0 corresponds to "no regularization" (i.e. regular linear model with a minimum residual sum of squares) [195].

#### 2.3.4.6 Clustering

Clustering is a Machine Learning technique that involves the grouping of data points. Given a set of data points, we can use a clustering algorithm to classify each data point into a specific group. Here we used hierarchical clustering algorithms to group the textural features. A better description is performed in section 3.4.2.

#### 2.3.4.7 Wilcoxon test

Wilcoxon signed-rank test is a non-parametric alternative to the student's t-test. This non-parametric test can be used to compare two related samples, matched samples, or repeated measurements on a single sample to assess whether their population means ranks differ. It can be used as an alternative to the paired Student's *t*-test [193]. The null hypothesis for this test is that the medians of the two samples are equal, which is rejected when $p < 0.05$. The results of this test can be used for feature selection where those features that independent of each other can be removed from the dataset.

### 2.3.5 Analysis of radiomic data

Data analysis comprises two separate steps: the first (model building), in which classification and/or regression model is generated; the second, where the model is used to make predictions about the case or cohort of patients under evaluation.

Model building involves (a) establishing the type of classifier or regressor to be used, and (b) feeding the model with a set of pre-classified cases—i.e., arrays of features/label pairs where the label indicates the clinical condition of the corresponding subject. This process of presenting the model with pre-classified cases is usually referred to as training. Crucial to this step, of course, is the availability of large enough datasets of pre-classified cases (ground truth).

As for the type classifier, there are several options available, which range from conceptually easy solutions, for instance, Linear Discriminant Analysis], K-nearest neighbors, and naive Bayes, to more involved ones, such as random forests and support vector machines, and logistic regression [165, 166]

## 2.4 Some important issues in radiomic analysis

A serious challenge with the calculation of radiomic features is the very large number of features and the different ways to calculate them. In addition, radiomic features exhibit variable sensitivity to differences in scanner model, acquisition protocols, quantitative corrections, reconstruction algorithm settings, and post-reconstruction processing. All of these affect the feature reproducibility and so, the integration of radiomic analysis to the daily clinic routine [177, 178, 196]. Several studies have explored the issues of the repeatability of texture features using test-retest data sets using the same

setting of image acquisition, reconstruction, segmentation, and quantization. A test-retest study shown that most GLCM texture features including entropy, homogeneity, and dissimilarity, and some GLSZM texture features including intensity variability and size zone variability had comparable reproducibility to SUVmax [197]. Another test-retest study proved that the majority of texture features had a high reproducibility, although GLSZM texture features showed the overall lowest reproducibility compared with global and GLCM texture features [172]. The image discretization employing a fixed bin size method seems to produce image features with better repeatability.

Most radiomic studies to date have been single-center based and retrospective in nature, in small cohorts of patients, and most radiomic models are not externally validated [198]. Chalkidou *et al*. [91] shown an inappropriate control of type I error in many radiomic studies that involved the calculation of a large number of imaging features in a small number of patients. The authors underlined the need for confirming observations and validating models by using independent patient cohorts in multicenter settings. Therefore, carrying out large multicentric studies would a very important step to bring the radiomic approach into the clinical setting. This requires large-scale radiomic data analysis, hence the need for integrating radiomic features extracted from images acquired in different centers. However, radiomic features values are notoriously sensitive to variability in scanner models, acquisition protocols, and reconstruction settings, such as the number of iterations or subsets, the scan duration per bed position, the post-reconstruction filter, and the voxel size [167, 168, 180, 199]. This variability of radiomic feature values implies that a radiomic model established using data from a given PET scanner might not be directly applicable to data from another PET scanner [200]. For this reason, harmonization strategies for obtaining compatible research results between different institutions have been proposed [92, 201].

Orlhac et al. [202] proposed a method to standardize features measured from PET images obtained with different imaging protocols, removing the center effect while preserving patient-specific effects. This work was on ComBat method, which consists of dealing with the variability of the distribution parameters so that they can be pooled together. It was initially described in genomics [203] to deal with the "batch effect", which refers to technical variation or non-biological differences between measurements of different groups of samples [204] (e.g. different laboratories, different technicians, different days), which are a source of variations in measurements, that potentially can mask individual variations. It is conceptually similar to variations induced in radiomic features by the scanner model, the acquisition protocol, and/or the reconstruction settings, sometimes called "center effect". ComBat identifies a batch-specific transformation to express all data in a common space devoid of center

effects [205]. By using the ComBat method, Da-ano *et al*. [205] successfully harmonized the data of different reconstructions settings by removing the estimated center effect without loss of the pathophysiological information.

The traditional approach to clinical image research is driven by standard medical statistics. Thus, in many radiomic studies, statistical tests are performed to find those radiomic features correlated with the target variable or construct predictive models. This approach gives rise to an important issue in radiomic analysis, it is the multiple comparison problems, which arise when multiple simultaneous statistical tests of the same dataset are performed. To correct the multiple comparison problems, the *p*-values of the statistical tests should be adjusted [206, 207]. Bonferroni correction is one of the methods most commonly used. Besides, multivariable analysis is recommended to evaluate the additional usefulness and independent significance of texture features [166].

## 2.5 Examples of PET/CT radiomic applications

Several studies have investigated the potential of PET/CT radiomics in lung and breast cancer for diagnose, prediction of treatment response, and prognosis [208, 209]. A PET/CT radiomic analysis can attempt to determine a specific characteristic at some point in time, for instance, benign vs. malignant lesions, or identify the histological subtype. But also, it can aim to predict the likely evolution of the disease over time, such as overall survival, disease-free survival, and/or response to treatment. Here we show some applications of PET/CT radiomics. Below, an example of the PET/CT radiomic application in breast cancer will be shown.

Ou *et al*. [210] assessed the ability of $^{18}$F-FDG PET/CT radiomic features to differentiate breast carcinoma from breast lymphoma using machine-learning. Breast lymphoma, as a rare type of extranodal lymphoma, which clinical and imaging presentations mimic those of breast carcinoma, leading to misdiagnosis. Commonly imaging techniques as mammography and ultrasonography have quite hard to differentiate breast lymphoma from breast cancer based on imaging features because both are shown as unilateral, solitary, and a palpable mass (Figure 2.4). In that study, sixty-five breast nodules from 44 patients diagnosed as breast carcinoma or breast lymphoma were included. SUV and radiomic features from CT and PET images were extracted.

Following the radiomic workflow (Figure 2.5), the constructed six discriminative models, including PETa (based on clinical, SUV, and radiomic features from PET images), PETb (SUV and radiomic

features from PET images), PETc (radiomic features only from PET images), CTa (clinical and radiomic features from CT images), CTb (radiomic features only from CT images), and SUV model. To this, they used the least absolute shrinkage and selection operator method and linear discriminant analysis. To evaluate the models, the areas under the receiver operating characteristic curve (AUCs), accuracy, sensitivity, and specificity were computed. They found that the PETa and CTa models had the best performance, in both training and validation groups (AUCs of 0.867 and 0.806 for the PETa model, AUCs of 0.891 and 0.759 for the CTa model, respectively). They concluded that models based on clinical, SUV, and radiomic features of $^{18}$F-FDG PET/CT images could accurately discriminate breast carcinoma from breast lymphoma.

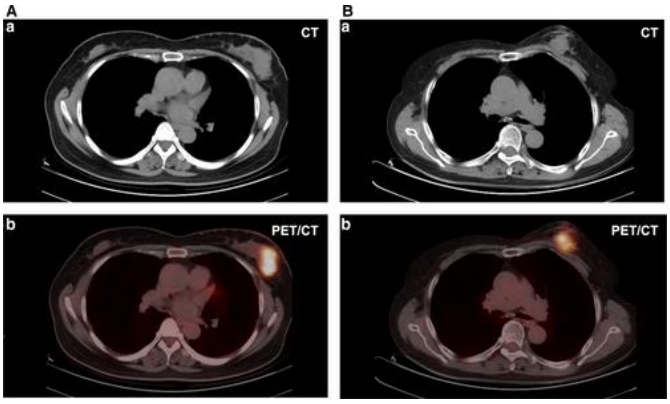



***Figure 2.4.*** Two cases of CT and PET/CT images from patients with breast lymphoma (A) and breast carcinoma (B) (from [210]).

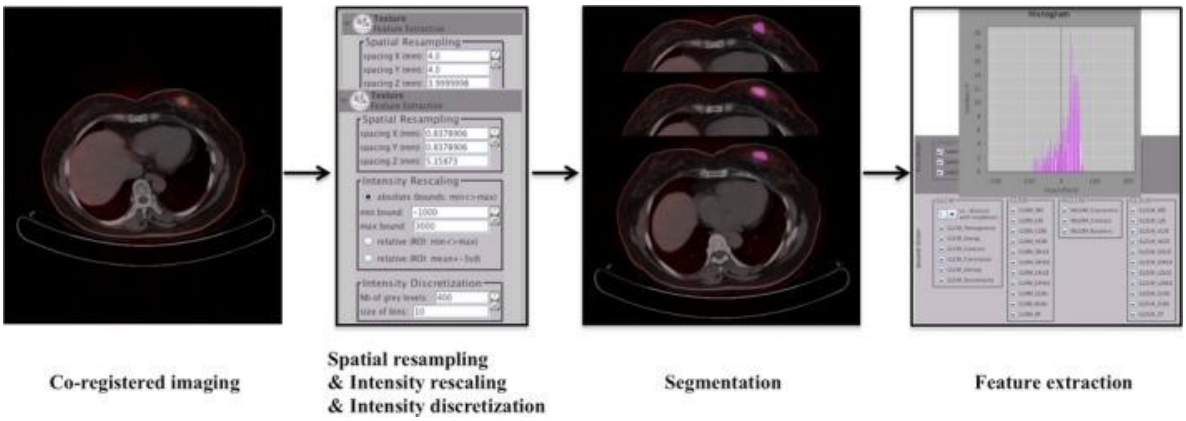



***Figure 2.5.*** The flowchart of radiomics. After images were co-registered, spatial resampling, intensity rescaling, and intensity discretization were set automatically. Tumor segmentation was manually contoured in 3D VOI. Radiomic features from this volume were extracted, including first- and second-order features (from [210]).

This study exemplifies one of the main applications of radiomics, its use to assist in the differential diagnosis of different pathologies. From the medical imaging expert's point of view, differentiating between a breast tumor and a breast lymph node is challenging. However, the characteristics of radiomics can account for differences not observed by the human eye.

Some patients with advanced breast cancer receive neoadjuvant chemotherapy (NAC) before an operation be performed. Subsequently, after the operation, the surgical specimen is examined to determine the response to the treatment received (pathologic response). If the tumor is not detected by the pathological study, the patient is considered to have achieved a pathologic complete response (pCR). This is the primary endpoint for neoadjuvant trials and a surrogate marker for disease-free survival (DFS) and overall survival (OS) [211]. Hence, it is important to identify those patients who will have a pCR after NAC so the toxicity of ineffective chemotherapy is avoided and other treatment options are considered. Li *et al*. [212] worked on predicting response to treatment, an area of great interest and scientific activity, especially with the emergence of personalized medicine. They used radiomic features from PET/CT in patients with breast cancer, as well as unsupervised and supervised machine learning models to identify prognostic radiomic predictors of therapeutic efficacy to NAC. They found that the PET/CT radiomic predictors achieved a prediction accuracy of 0.857 (AUC = 0.844) on the training split set and 0.767 (AUC = 0.722) on the independent validation set. When age was incorporated, the accuracy for the split set increased to 0.857 (AUC = 0.958) and 0.8 (AUC = 0.73) for the independent validation set and both outperformed the clinical prediction model. An additional step would be to apply radiomics in the context of patients receiving chemotherapy for recurrent or metastatic breast cancer. This is a topic treated in the last chapter of this thesis.

# CHAPTER

# 3

# Machine learning in Medical Images and model construction

## 3.1 Introduction

In daily clinical routine, PET/CT images are evaluated by a nuclear medicine/radiologist specialist, who by using qualitative descriptor and very simple semiquantitative or quantitative measures perform a diagnosis, performs a classification process, often binary, where the absence or presence of a characteristic is determined [213]. Specialist´s skill or "acquired intelligence" is obtained through years of performing the same task, receiving constant feedback from colleagues and final patient´s evolution, making mistakes, and performing its respective corrections. After a long experience, his/her diagnostic accuracy stabilizes in a specific percentage range, which will depend on the quality of the training received, as well as the limitations of the imaging technique used in each case.

On the other hand, radiomics generates a large amount of information, which along with the availability of more powerful computers has stimulated the application of artificial intelligence (AI) to perform such classification tasks [214]. AI refers to the capability of machines to emulate intelligent human behavior. The goals of artificial intelligence include learning, reasoning, and perception. AI is being applied in a great range of fields, ranging from games, automobile and aviation industry, economy, and health care, and biomedical research. In this thesis, we have only considered machine learning (ML) methods, which is the area of AI focused on computer systems and algorithms able to learn from the available data without explicit programming. ML allows automating the classification process in medical image analysis and potentially providing a significant improvement in the performance of this task, not only in terms of accuracy but also in reproducibility. In many applications, the performance of ML-based automatic detection and diagnosis systems already is comparable to that of a well-trained and experienced radiologist [213, 214]. In this chapter, we describe the main aspects of ML relevant to this thesis, including some of its applications in oncology. An exhaustive treatment of this subject is outside the scope of this thesis and can be found elsewhere [215]

## 3.2. Machine learning methods

Machine learning (ML) refers to computer systems and algorithms able to extract some knowledge, i.e. “learn”, from the available data (they must be trained). Usually, this is achieved by using some statistical analysis, so it is also known as analytics or statistical learning. The performance of most machine learning algorithms depends on the choice of various tuning parameters. Some of which,

called "hyperparameters", cannot be learned by the algorithm of ML directly during the training, and rather they must be set before the training starts.

ML methods are usually classified into three broad categories, depending on the type of task performed: supervised learning, unsupervised learning, and reinforcement learning. Each of these methods has its scope and applications, which are depicted in Figure 3.1.

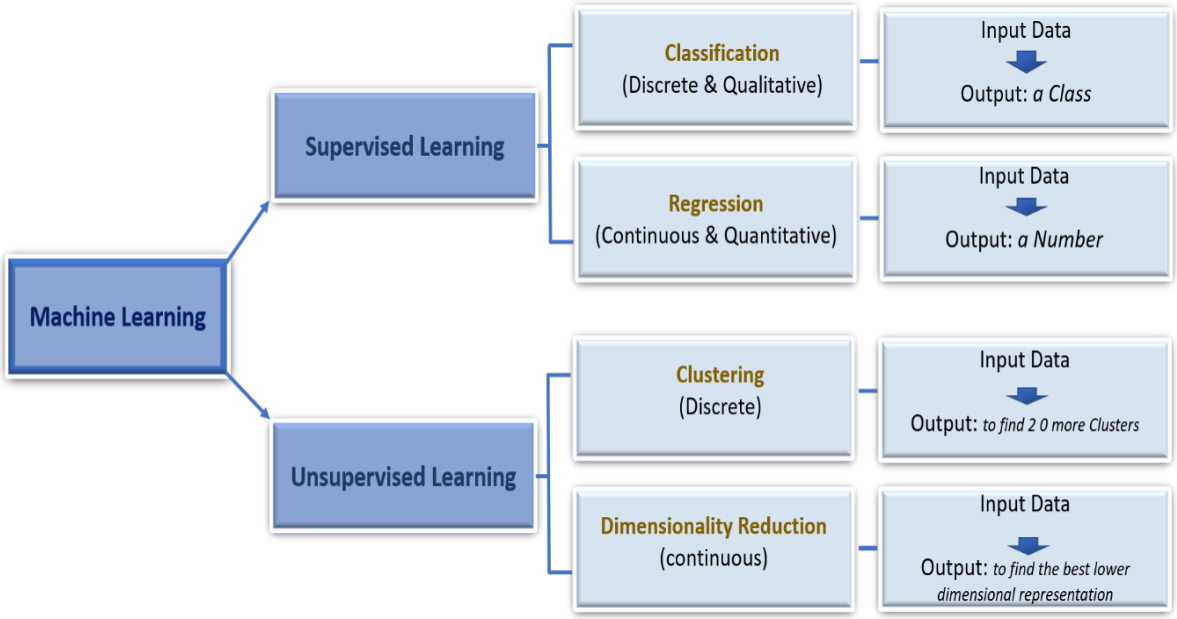



***Figure 3.1.*** Overview of supervised and unsupervised learning paradigms and subcategories. The two major categories of supervised learning are classification and regression [Adapted from 216].

### 3.2.1 Supervised Learning

Supervised ML methods are comprised of a series of algorithms, which build a mathematical model of the relationship between inputs (for instance, a PET image) and outputs (for instance, the diagnosis) based on some available data, known as "***training data***". The training data should contain both the inputs and the outputs. Once the algorithm has been trained and the model has been created, it can be used to estimate predictions of outputs from new input data sets. The training process of the ML model uses an iterative optimization procedure that minimizes the differences between the predictions and the actual outputs in the training data set. With proper regularization methods and using enough data, the model obtained can be made general enough to be applied to new data sets.

Supervised learning methods can perform both classification and regression tasks (Figure 3.1). In classification tasks, outputs are restricted to a limited set of values (class), while regression may generate numerical values within a specific range.

### 3.2.2 Unsupervised Learning

These methods refer to a set of algorithms where the only available information are the inputs. In this case, the algorithms obtain outputs based on some properties found in the data. Unsupervised learning can be applied to obtain groups or clusters of entities with shared similarities that may be previously unknown. These are referred to as clustering algorithms. Some of the most common methods employed in this approach include k-means clustering and anomaly detection. Another application of unsupervised learning is dimensionality reduction, with algorithms such as principal component analysis able to obtain the most distinct components in the data.

### 3.2.3 Reinforcement Learning Method

Reinforcement learning methods are related to algorithms that learn how to perform a series of actions in certain environments to maximize some kind of reward. Reinforcement learning is often used in game theory, operations research, control theory, information theory, multi-agent systems, simulation-based optimization, statistics, swarm intelligence, and genetic algorithms. For machine learning, the environment is typically represented by a Markov decision process. These algorithms do not necessarily assume knowledge, but instead are used when exact models are infeasible.

The most common use of reinforcement learning can be seen in complex and changing situations such as games where there are “computer players” or a player that is represented by the computer and plays against human opponents. In these “computer players” reinforcement learning enables them to respond in a way that is not exact and precise every time, but instead in a way that actually challenges the human. This way, games cannot be memorized and overcome, but instead feature some diversity and uncertainty to them.

The three learning categories above mentioned, as well as they use the data to learn, are depicted in Figure 3.2 [44].

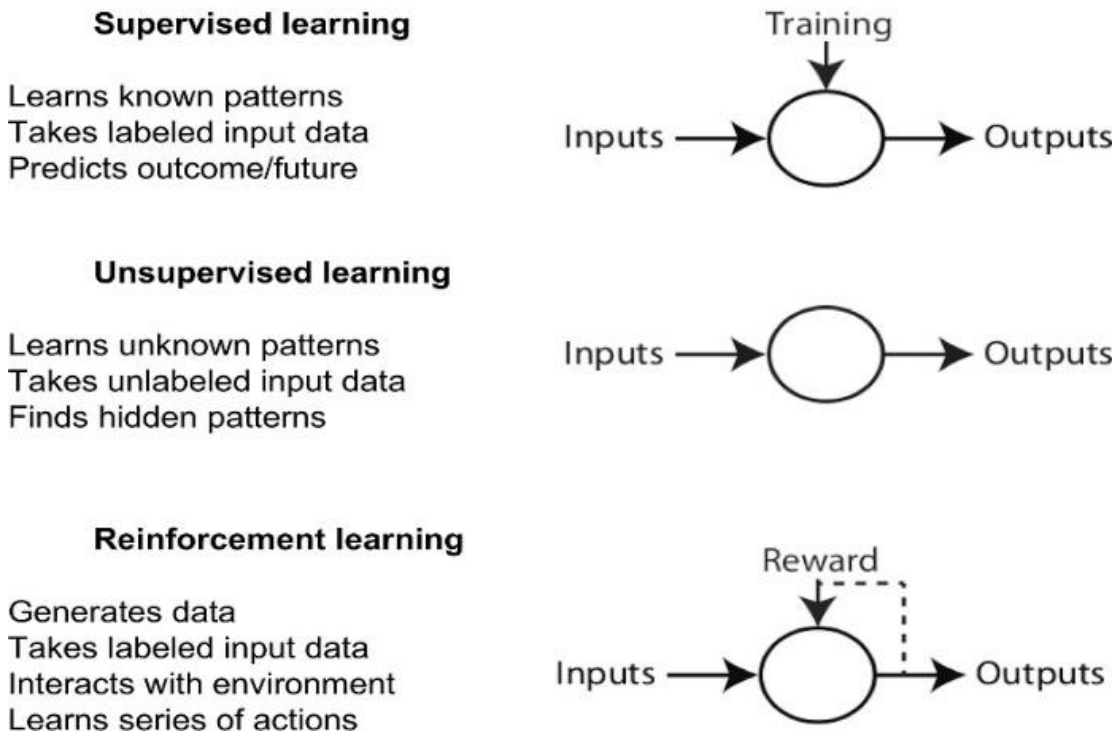



***Figure 3.2.*** Categories of machine learning, and how they learn from data (from [44]).

## 3.3 Supervised Machine Learning Algorithms

### 3.3.1. Logistic Regression

Logistic regression (LR) is a classification method that uses a logistic function (Figure 3.3) for predicting a binary dependent variable (target). A variation of this method is the multinomial logistic regression, which can be used to classify a target with more than two outcomes.

In this model, the probabilities describing the possible outcomes of a single trial are modeled using a logistic function. This function is useful in LR because it takes any input in the range of negative to positive infinity and maps it to output in the range of 0.0 to 1.0. This allows us to interpret the output value as a probability.

$$f(x) = \frac{\mathrm{L}}{1+ e^{-k(x-x_0)}} \qquad (3.1)$$

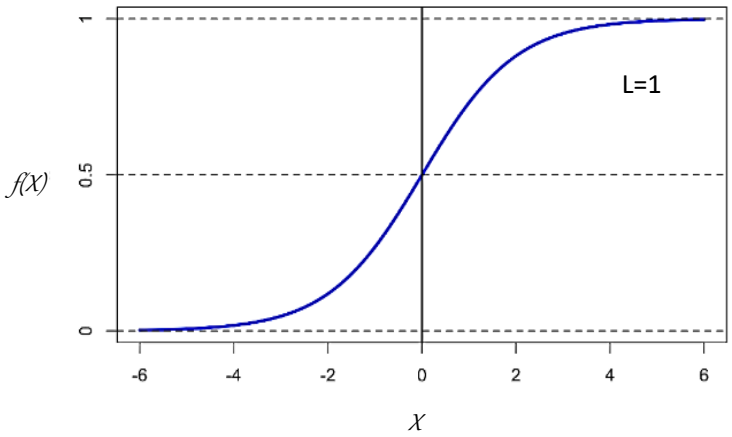



***Figure 3.3.*** Logistic function

Where:

***e***; the natural logarithm base

***Xo***; the value of the sigmoid's midpoint,

*L*; the curve's maximum value, and

***k***; the logistic growth rate of the curve

In the binary approach, the function yields a value of 0 or 1 which represents the negative (0) and the positive (1) case.

The model attempts to estimate the probability, P(y = 1| x), that is the probability of a positive outcome (y = 1) given data x. Using the Bayes rule, P (y = 1| x) can be expressed in the form of a logistic function [217]:

$$P(y = 1\,|x) = \frac{1}{1+e^{-\alpha}} \qquad (3.2)$$

where α is the log-odds ratio (the odds of a positive classification relative to the odds of a negative classification), which can be expressed as a linear function:

$$\alpha = \ln \frac{P(x|y=1\,|x)P(y=1)}{P(x|y=0\,|x)P(y=0)} + \beta_0 + \beta_1^T \boldsymbol{x} \qquad (3.3)$$

The weights ($\beta_0$, $\beta_1$) can be calculated using the maximum likelihood approach [218].

The log-likelihood expression serves as an error function and using gradient descent, the optimal weights can be iteratively solved to minimize error.

The advantage of logistic regression is that it is fast to train and can use discrete and continuous variables as inputs. The disadvantages include that it is a linear model. Therefore, complex data problems may pose difficulties.

### 3.3.2. Naive Bayes

Naive Bayes's methods are a set of supervised learning algorithms based on applying Bayes' theorem with the "naive" assumption of conditional independence between every pair of features given the value of the class variable [217]. Naive Bayes calculates the probability associated with each possible class conditional on a set of covariates, i.e., the product of the prior probability and the likelihood function. The classifier then selects the class with the highest probability as the "correct" class. The prior probability typically reflects one's belief about the outcome, either based on the study itself or from other published literature. The independence assumption in naive Bayes greatly simplifies the calculation by decomposing the likelihood function into a product of likelihood functions, one for each covariate. Even though the independence assumption is often wrong, the naive Bayesian classifier still performs very well in real applications, even with small training data sets. One advantage is that it returns not only the prediction but also the degree of certainty, which is often very useful. Also, it makes dealing with missing values easy. Due to its simplicity, the naive Bayes classifier is less prone to overfitting from artificial neural networks, for example.

### 3.3.3. k-Nearest Neighbours Classifier

The K-nearest neighbor (KNN) classifier is one of the simplest and most common classifiers, yet its performance competes with the most complex classifiers in the literature. The core of this classifier depends mainly on measuring the distance or similarity between the tested examples and the training examples [218]. Nearest neighbors are the foundation of many other learning methods. They can perform classification for data with discrete labels or regression for data continuous labels. The principle behind nearest neighbor methods is to find a predefined number of training samples closest in distance to the new point and predict the label from these. The number of samples can be a user-defined constant (k-nearest neighbor learning) or vary based on the local density of points (radius-based neighbor learning). The distance can, in general, be any metric measure: standard Euclidean distance is the most common choice.

### 3.3.4. Support Vector Machine

Support vector machines (SVMs) are a set of supervised learning methods used for classification, regression, and outliers detection [217, 219]. They only use a subset of training points in the decision function "support vectors", making they memory efficient, being often preferred over other machine learning algorithms, such as neural networks, because they are simpler and can achieve high accuracy with less computation. The algorithms work by finding a hyperplane in an n-dimensional space that distinctly separates the data points into two classes with the maximum marginal distance ($m$) to provide a robust decision boundary that can tolerate noisy test data. Thus, the SVM algorithm optimizes between maximum margin and training error to solve the ideal decision boundary.

By setting $m$ inversely proportional to decision boundary parameters, $m = \frac{1}{||\beta||}$, the soft margin SVM classifier can be formulated as a minimization problem, where $\boldsymbol{x}$ is the training data, $\boldsymbol{y}$ is the label, $\beta_0$ and $\boldsymbol{\beta}$ are decision boundary parameters, N is the number of training data, $\varepsilon$ is a slack variable to measure misclassification overlap and C is a penalization cost for misclassification:

$$\underset{\beta_0,\beta}{\text{minimize}}\frac{1}{2}\left|\left|\beta\right|\right|^2 + \ C\sum_{i=1}^{N}\epsilon_i subject\ to\ y_i\left(\beta_0 + \boldsymbol{\beta}.\boldsymbol{x_i^T}\right) \geq 1 - \ \epsilon_i.\epsilon_i \ \geq 0, i = 1, \ldots \ N \qquad (3.4)$$

The minimization problem is solved by computing the Lagrange Dual and performing quadratic optimization. Figure 3.4 shows a graphical example of SVM.

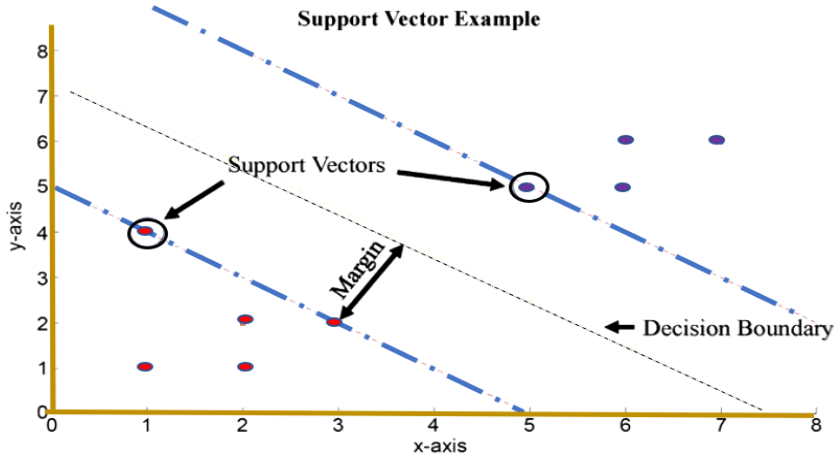



***Figure 3.4.*** Support vector machine example (adapted from [217]).

The two classes to be separated are represented with different colors, the decision boundary (hyperplane that lies in the middle) is represented by the yellow dotted line, the margin hyperplanes between the class and the decision boundary are represented by the red dotted lines. The Support Vectors (SVs), the data closest to the decision boundary, and lying on the margins, are circled data. This decision boundary is then used to evaluate new data based on the position of the data with respect to the decision boundary.

In the case of non-linearly separable data, SVM uses kernel functions to transform the data into a higher dimension, in which the data can be linearly separated [220], this is depicted in Figure. 3.5.

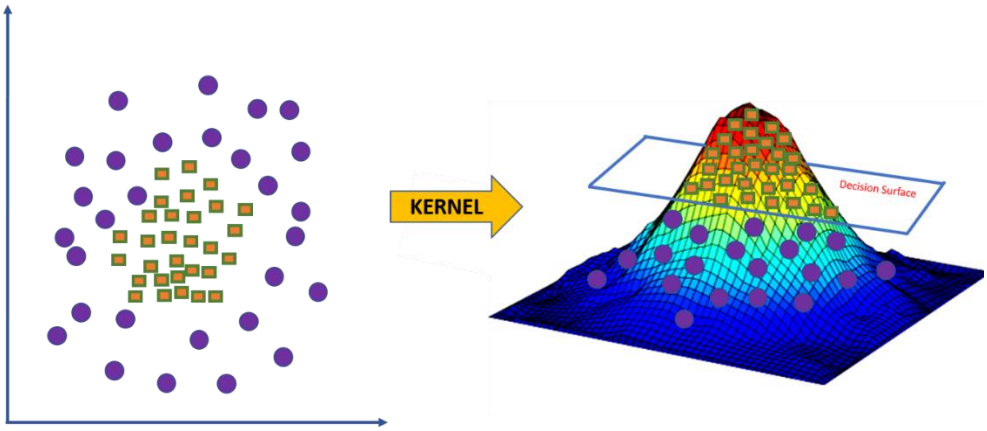



***Figure 3.5.*** A kernel function. It takes data in low dimensional input space, which are not laniary separable, and transforms it into a higher-dimensional space, where they are linearly separable.

The advantage of SVM is the simple mathematics behind the decision boundary and its application in higher dimensions. However, since SVM is essentially an optimization problem attempting to balance between errors in the training set with a larger margin decision boundary, it may be slow for large datasets, especially where the class separation is small. SVMs do not directly provide probability estimates, these must be calculated using cross-validation.

### 3.3.5. Decision Tree

Decision tree (DT) is also a classical ML algorithm. The DT divides the data based on features to determine the appropriate class. The features used to split the data are determined using the Information Gain provided by individual features [217]. To obtain the information gain, initially, the

entropy or Gini index of a dataset is computed. For a dataset (S) with two classes, the entropy would be calculated as:

$$Entropy\ (S) = \sum_{i=1}^{2} -P_i\ log_2\ Pi \qquad (3.5)$$

Individual feature-specific information gain is calculated by the difference between the entropy of the training set and the entropy of the feature. The information gain for feature A would be as follows:

$$Information\ Gain = Entropy(S) - \sum_{v\ \in\ Values(A)} \frac{|S_v|}{|S|} Entropy\ (S_v) \qquad (3.6)$$

where Values (A) is the set of all possible values for a feature A and S*v* corresponds to the subset of S where feature A has a value *v*. Features providing relevant and valuable information to separate the classes are then selected to be used in the DT. Features that provide the highest information gain split the data earlier in the tree, and features that provide less information gain are at lower stages in the DT. An example of a decision tree is depicted in Figure 3.6.

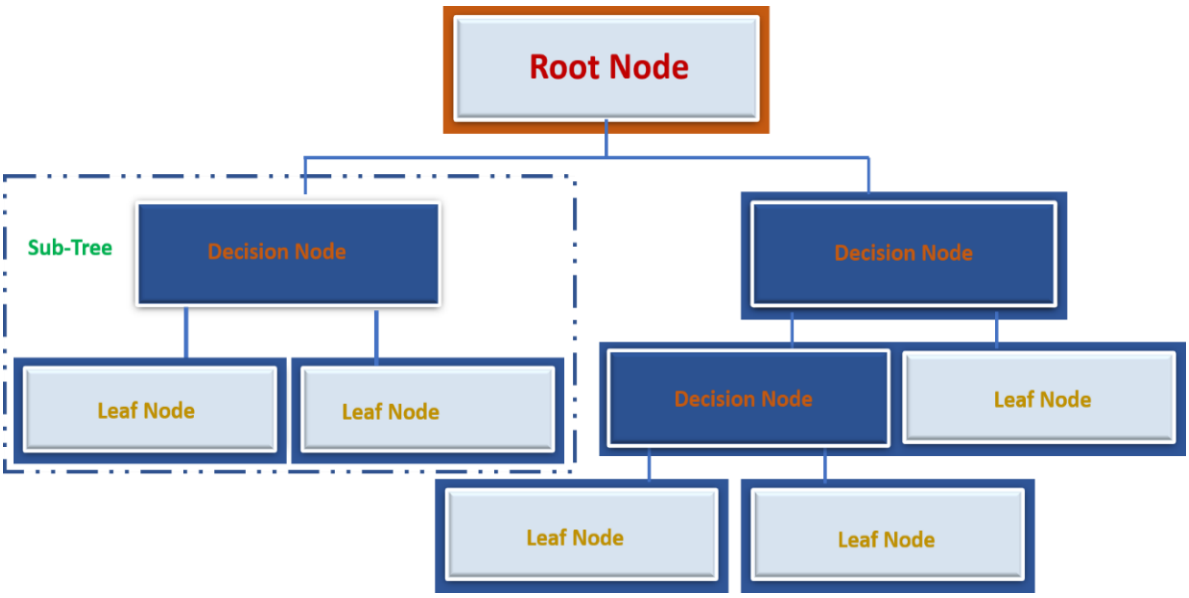



***Figure 3.6.*** *Decision tree nodes.* The root node is the first node and contains the whole dataset. It gets split into two or more *decision nodes* based on the feature value. Posteriorly, the *decision node* split into new decision nodes and *leaf nodes***,** which are the terminal nodes, as it cannot get split further. A *sub-Tree* or branch is a subdivision of a complete tree (adapted from [313]).

### 3.3.5.1. Random Forests

Random forests (RF) are a class of ensemble learners that combine several decision trees models. RF creates several decision trees and merges them to get a better prediction [221]. Implementation of an RF classifier is depicted in Figures 3.7. This is usually trained using bagging. In bagging, several models, in this case, decision trees, are trained on subsets of the training data where the data points in the subsets are selected randomly with replacement (bootstrapping).

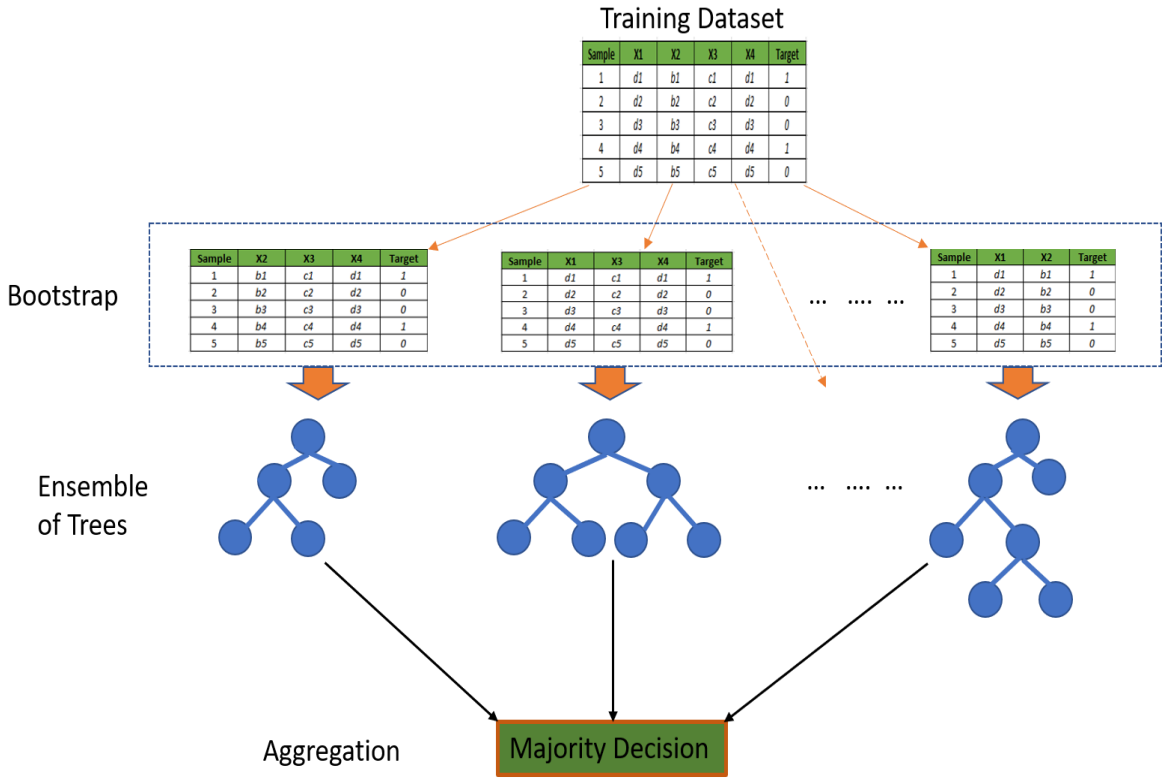



***Figure 3.7.*** Implementation of an RF classifier. The dataset has four features ($X_1$, $X_2$, $X_3$, and $X_4$) and two classes ($Y$ = 1 and 2). RF classifier is an ensemble method that trains several decision trees in parallel with bootstrapping followed by aggregation. Each tree is trained on different subsets of training samples and features.

Bagging improves prediction performance by reducing variance. Decision trees are sensitive to the training data. When the training data changes, the resulting decision tree can be very different and, hence, the result can be quite different too. Bagging improves the result by averaging the prediction from each model. The bagging algorithm works as follows:

Create random subsets of the training data set with replacement, meaning the same data can be reused. Train each decision tree with the random subsets. Using a new data set, calculate the average prediction of all decision trees.

The advantage of Decision Trees is that they are both easy to visualize and understand. The disadvantage is that feature selection plays a dominant role in the accuracy of the algorithm. One set of features can provide drastically different performance than a different set of features. Large Random Forests can be used to alleviate this problem.

## 3.4 Unsupervised Machine Learning Algorithms

### 3.4.1 k-means Clustering Algorithm

The k-means clustering algorithm is one of the simplest and most popular forms of an unsupervised machine learning algorithm that exists. The algorithm cluster data by trying to separate samples in **k** groups of equal variances, minimizing a criterion known as the *inertia* or within-cluster sum-of-squares. Clusters in a dataset refer to collections of data points that are connected due to specific similarities. Each member of the cluster has more in common with other members of the same cluster than with members of the other groups. The most representative point within the group is called the centroid. Usually, this is the mean of the values of the points of data in the cluster.

The k-means algorithm looks for centroids that minimize the inertia**,** or within-cluster sum-of-squares criterion [219]**:**

$$\sum_{i=0}^{n} \min_{u_j \in C} \left( ||x_i - u_j|| \right) \qquad (3.7)$$

k-means clustering is an iterative approach. The algorithm starts by randomly selecting k initial centroids from the data set, k has to be selected by the user and is not necessarily known at the beginning of a clustering project. Each instance of the data set is then assigned to a cluster, based on its distance to the centroid. Usually, the Euclidean distance is used. After each instance has been

assigned to a cluster, the centroids are recomputed by calculating the mean of each instance in a cluster. This process is repeated until the centroids do not change anymore.

### 3.4.2 Hierarchical clustering

Hierarchical clustering or hierarchical cluster analysis (HCA) is a family of clustering algorithms that build hierarchies of clusters. There are two basic approaches, agglomerative and divisive clustering. Hierarchical clustering does not only partition the data, it also depicts the relationships among the clusters, creating nested cluster over a variety of scales by merging or splitting them successively. Hierarchy clusters are represented as a tree (or dendrogram, see Figure 3.8). The root of the tree is the unique cluster that gathers all the samples, the leaves being the clusters with only one sample.

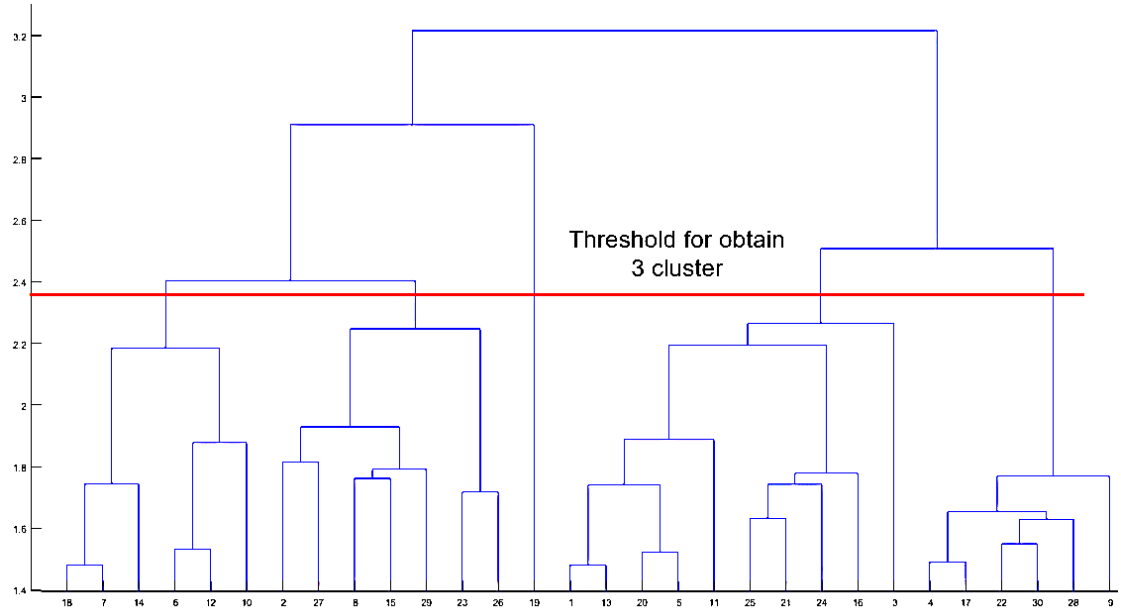



***Figure 3.8.*** Agglomerative hierarchical clustering for supervolxel of PET-CT images of patients with breast cancer. 3 big clusters can be separated with a threshold (cophenetic distance ) near 2.4.

Divisive clustering is a top-down approach and works oppositely. The clustering starts with one single cluster that is divided into subclusters. The subclusters are further subdivided into the next iterations until each data point is in its cluster. Divisive clustering does not need to go through all iterations and can halt once a stop criterion is met.

There are several ways to measure the distance between clusters to decide the rules for clustering, and they are often called Linkage Methods. The cophenetic distance between two observations that have been clustered is defined to be the intergroup dissimilarity at which the two observations are first combined into a single cluster.

### 3.4.3 Principal Component Analysis

Principal Component Analysis (PCA), is a dimensionality-reduction method that allows reducing the dimensionality of large data sets, by transforming variables into a smaller one, that still contains most of the information of the large set. Which makes it suitable for dimensionality reduction [136]. PCA reduces the dimensionality of a dataset while preserving as much variability (i.e. statistical information) as possible. It finds new variables that are linear functions of those in the original dataset, that successively maximize variance, and that are uncorrelated with each other. Finding such new variables, the principal components (PCs), reduces to solving an eigenvalue-eigenvector problem. Thus, PCA is defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by some scalar projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. The principal components are often computed by the eigendecomposition of the data covariance matrix of the data. PCA is sensitive to the scaling of the variables. Mean subtraction is necessary for performing classical PCA to ensure that the first principal component describes the direction of maximum variance. If mean subtraction is not performed, the first principal component might instead correspond more or less to the mean of the data. A mean of zero is needed for finding a basis that minimizes the mean square error of the approximation of the data [222].

## 3.5. Artificial Neural Networks and Deep Learning

Artificial neural networks (ANN) are the area of Machine Learning that has evolved more rapidly in the last few years. They encompass a large number of different models and learning methods. Their development was inspired by the human nervous system.

An ANN is composed of input, hidden, and output layers, with each layer composed of individual nodes [217]. An ANN is depicted in Figure 3.9. Nodes in different layers are connected by weights, represented by arrows in the figure. The values from each node in the previous layer are multiplied

by the corresponding weights and are summed at nodes in the next layer. Furthermore, a bias node and activation functions are included in the hidden layer to introduce non-linearity into the ANN. Outputs at a hidden layer node can be calculated as:

$$\text{Output } = \varphi(\sum_i w_i x_i + b) \quad (3.8)$$

where $w_i$ corresponds to weights connecting the inputs from the previous layer denoted as $x_i$, b is the bias and φ is the activation function.

The most popular activation for ANN is the sigmoid function because it outputs values between the range of 0 and 1 and has a simple derivative form.

$$\varphi(z) = \frac{1}{1+e^{-z}} \qquad (3.9)$$

$$\varphi^1(z) = \varphi(z)(1 - \varphi(z)) \qquad (3.10)$$

After the values from the input layer are traversed forward through the hidden layer(s), at the output layer a SoftMax function, also known as normalized exponential function is applied to calculate the ANN's confidence percentage in each class. During training, the error between the calculated class and the expected class is determined (typically using the sum-squared error function) and the error is backpropagated through the network to update the values of the weights. Backpropagation uses the derivative of the error to update the weights, which is why simple to derive activation functions, like the sigmoid function, are favorable. The algorithm iterates through all training data until the error of the network falls below a certain threshold to avoid overfitting.

An advantage of ANN is that although the mathematics behind the algorithm is simple, the non-linearities and weights allow the ANN to solve complex problems. Disadvantages of ANN include the training time required for numerous iterations over the training data, tendency to easily overfit on training data, and numerous tuning hyperparameters (parameter in the ANN configuration that is not directly learnable by training) are required for determining optimal performance. Some hyperparameters are the number of hidden layers, number of neurons in a layer, activation function,

number of epochs (i.e. number of times all training examples have been passed through the network during training), learning rate (i.e. step length for update the weights).

Different values of hyperparameters can have a major impact on the performance of the network. By optimizing the hyperparameters of ANN, its performance is improved and stable predictions are obtained. There are several ways, ranging from manual trial and error to sophisticated algorithmic methods to optimize an ANN [219, 221].

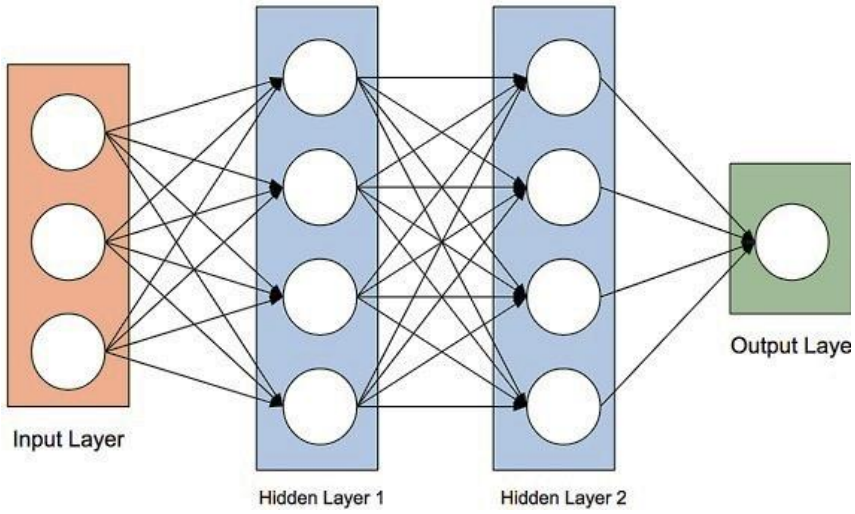



***Figure 3.9***. Artificial neural network (from [314]).

### 3.5.1. Deep Learning

Although deep learning was not used in this thesis, it is an important method that is worth considering in future works. Deep Learning (DL) methods are based on ANN but are composed of many additional layers to add complexity to the algorithm to learn features and representations automatically. Networks with three or more layers are generally considered deep. An example of DL is the Convolution Neural Network (CNN) model, which is very suitable for medical image processing. As illustrated in Figure 3.10, CNNs are composed of multiple convolutional and pooling layers with fully-connected layers for classification. In the convolutional layers, filters are convolved with the input to create a stack of filtered images. In the pooling layer, the stack of filtered images is simplified by reducing the size. CNN learns simple features such as edges and corners. These simple features are then used to learn more complex features at higher layers of the CNN. As in ANN, all weights in the CNN are randomly initialized and are updated throughout training and backpropagation until the error of the training set falls below a specified threshold.

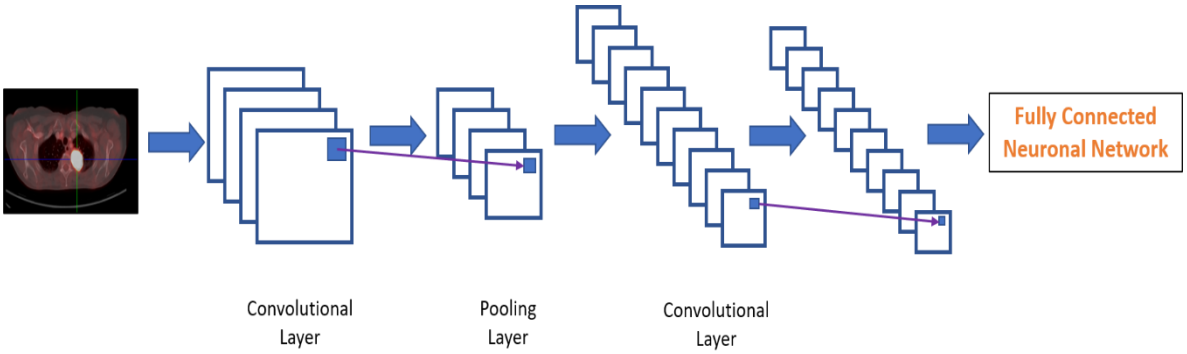



***Figure 3.10.*** Example of convolutional neural network (adapted from [217])

## 3.6 Machine learning workflows

In this thesis, we have focused on classification tasks. Classification problems aim to provide models that can predict the class of a new sample, based on the attributes or features describing that sample. Supervised classification learning procedures include three phases, namely training, validation, and testing, for each of which different subsets of the data are used (Figure 3.11).

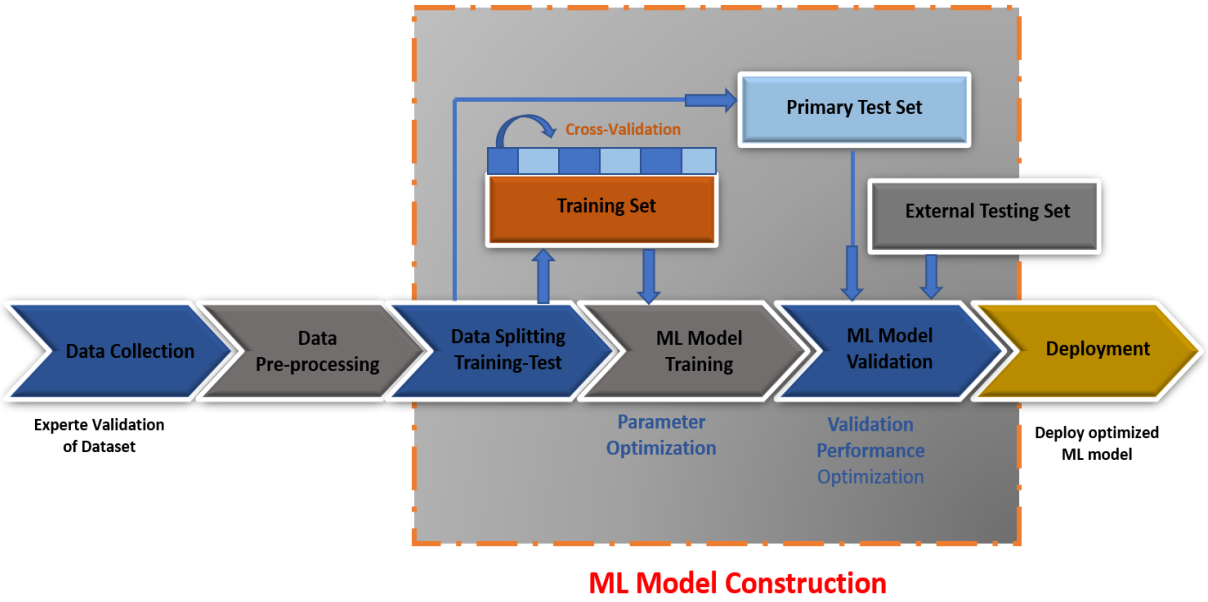



***Figure 3.11.*** Supervised machine learning model design overview. Steps for the deployment of a supervised machine learning model. From left to right, the figure shows the initial team of multidisciplinary experts defining a study design to address a need. Data are then collected, processed, trained tested, validated, and ultimately deployed (adapted from [216])

During the training phase, the parameters of the learning algorithm are tunned following different procedures depending, mainly cross-validation. There are algorithm-specific details to choose, for instance, the depth of a decision tree or the number of neighbors for the k-Nearest Neighbors. We already mentioned, these choices are commonly referred to as *hyper-parameters*, as opposed to the parameters used in the mathematical formula describing each model. These choices influence the prediction accuracy, reliability of the training procedure, interpretability of the model, computation time, and memory usage. It is usually not possible to know the best choice a priori. However, some recommended steps may be followed [216]. The goal of validation is to assess how well different models perform on a validation dataset. In this dataset, the true classification output (labels) of each sample is known, which allows us to compute the error made by the learning algorithm. Finally, the testing phase makes use of a test set to evaluate the performance of the model that had been chosen through validation. Classification models can learn with different parameter configurations, different input features, or from different samples, and several strategies can be implemented, depending mainly on data availability. However, if we expected robust and reliable results, it is necessary to awake about the quality of available data, data features, and type of learning method. So, during model construction, there are many questions to answer, for instance, what variables and which classification algorithm to use for the training process.

When dealing with ML algorithms, one should also be familiar with the concepts of underfitting and overfitting. In Figure 3.12 are shown examples of this problem.

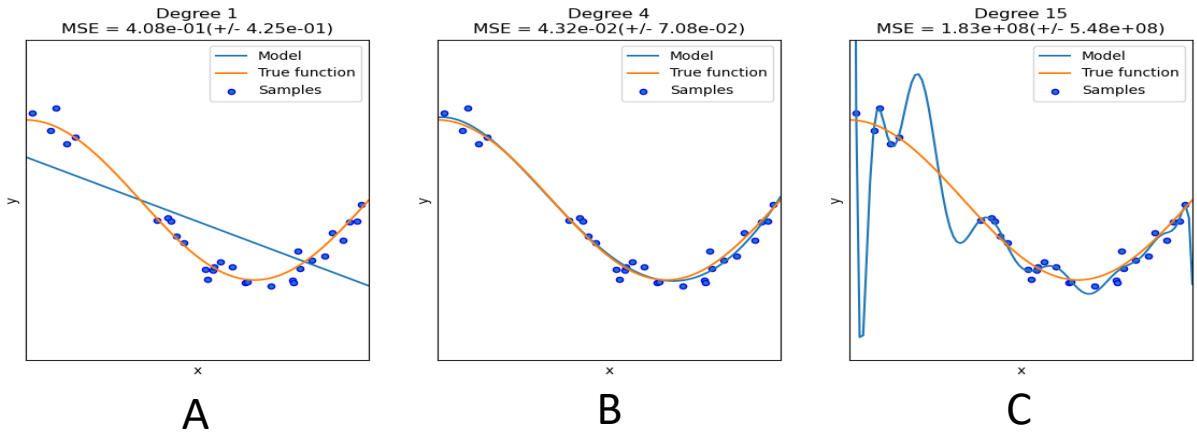



***Figure 3.12.*** A. we see that a polynomial with degree 1 is not sufficient to fit the training samples. This is called underfitting. B. A polynomial of degree 4 approximates the true function very well. C. a higher degree polynomial will overfit the training data [220].

Overfitting refers to a modeling error where the algorithm "learns" or reflects the training data too closely, using even noise or random fluctuations in the training data as concepts that may not apply to new datasets and consequently negatively impact algorithm performance in new datasets (or generalization of the model). On the contrary, underfitting refers to insufficient fitting to the training sample.

### 3.6.1 Data preprocessing

After data are collected, they must be cleaned and preprocessed. Data cleaning includes compensation for missing values, through imputation methods or removal of outliers and inconsistencies [216, 217]. These steps are important because the performance of a classifier could be biased by a non-representative dataset. It is known that missing or duplicate data, outliers, noise, and imbalanced class representation worsen the accuracy of a classifier.

### 3.6.2 Data splitting

The supervised ML model building phase usually includes splitting the data into an initial *training* and *testing* set that allows training of the model followed by testing for its initial validation phase (Figure 3.13). The training set will be used to train the model (allowing tunning of model hyper-parameters). It means, the model observes and learns from this data and optimizes its parameters. To minimize the overfitting of the models, certain model adjustments and incorporating cross-validation (CV) processes allow the empirical build of a large number of models whose performances can be subsequently assessed to find the most generalizable model. It can be performed by split the training set, into subgroups, and take all of them, except one, to train the model, then validation is carried out on this last one. For doing so, we avoid reducing the available number of samples available for model training, because a cross-validation group is not formally created. The test set consists of data used to provide an unbiased evaluation of a final model fit on the training dataset. It is only used once the model is completely trained by using the training and cross-validation sets. Therefore, the test set is used to replicate the type of situation that will be encountered once the model is deployed for real-time use. Typically, the splitting ratio is around 80:20 or 70:30, where 80% of the data will be used for training (and cross-validation) and the other 20-30% to test the model.

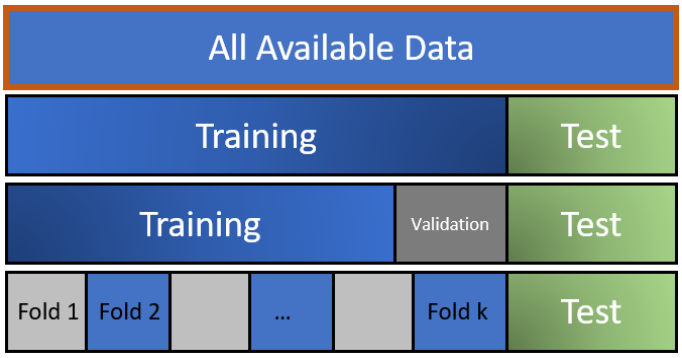



***Figure 3.13.*** Different methods of data splitting

### 3.6.3 Feature Selection and Dimensionality Reduction

ML models are often characterized by a high dimensional space of features. Moreover, features used to construct the model are not necessarily all relevant and of benefit for the learning task. Additionally, a high number of features may slow down the induction process while giving similar results as obtained with a much smaller feature subset. By only keeping the most relevant variables from the original dataset (this technique is called feature selection) is possible to improve the performance of a model. It does not involve any feature transformation, but rather concentrates on selecting the better features among the existing ones.

By finding a smaller set of new variables, each being a combination of the input variables, containing basically the same information as the input variables (this technique is called dimensionality reduction). The original feature space is mapped onto a new, reduced dimensionality space and the original examples are then represented in the new space. The mapping is usually performed either by selecting a subset of the original features or/and by constructing new features [48, 49, 223].

There are many reasons for using feature selection or dimensionality reduction in machine learning. For instance, it improves prediction performance, learning efficiency, provides a faster computation of the ML models, reduces the complexity of the learned results, and enables a better understanding of the underlying process.

This reduction of the number of features is invariably used in machine learning models that use radiomic features, because of the high dimensional space of features. Different methods and algorithms have been developed for feature selection or dimensionality reduction. Some of them are listed below.

- Univariate feature selection (by using a univariate statistical test)
- least absolute shrinkage and selection operator (Lasso)
- Cluster
- Principal Component Analysis (PCA)
- Independent Component Analysis (ICA)
- Feature importance (e.g. with random forest)

### 3.6.4. Model performance evaluation

The assessment of the model performance is a key element of ML, not just to determine the predictive performance of the final model, but also to select the most suitable pre-processing and classifier meta-parameters [39, 40, 224]. The misclassification error should be summarized by one or more representative metrics. Performance is mostly described by accuracy which is usually defined as the area under the curve (AUC) of a ROC curve. However, it can include other quantitative measures, such as sensitivity, specificity, error rate, and F-scores.

#### 3.6.4.1 Confusion Matrix

This matrix describes the complete performance of the model. From it, can compute the following metrics:

- True positives (TP): cases with prediction 1 and actual output is 1
- True negatives (TN): cases with prediction 0and actual output is 0
- False positives (FP): cases with prediction 1 and actual output is 0
- False negatives (FN): cases with prediction 0 and actual output is 1

Some important definitions are:

- The TPR (Sensitivity or recall) is defined as TP/ (FN+TP). TPR corresponds to the proportion of positive data points that are correctly considered as positive, with respect to all positive data points

- True Negative Rate (TNR) or specificity is defined as TN / (FP+TN). FPR corresponds to the proportion of negative data points that are correctly considered as negative, with respect to all negative data points

- FPR is defined as FP / (FP+TN). FPR corresponds to the proportion of negative data points that are mistakenly considered as positive, with respect to all negative data points

False Positive Rate and True Positive Rate both have values in the range [0, 1]. PR and TPR both are computed at varying threshold values.

### 3.6.4.2 Receiver operating characteristic curve (ROC curve)

ROC curves are graphical illustrating the diagnostic ability of a binary classifier system as its discrimination threshold is varied (see Figure 3.14).

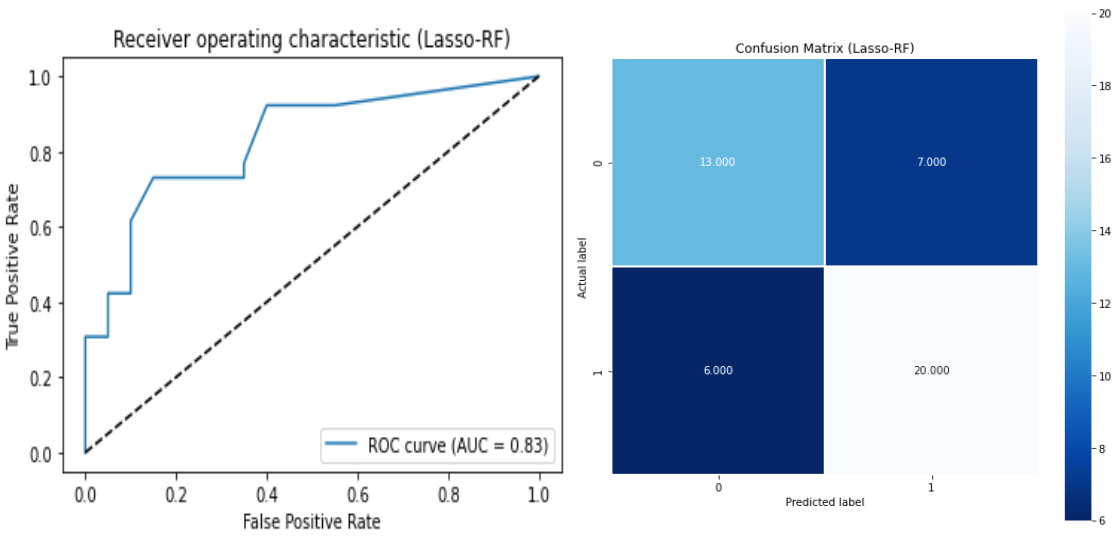



***Figure 3.14.*** Validation ROC curve and confusion matrix for RF classifier to predict treatment response in recurrent/metastatic breast cancer.

The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings [40]. The area under the ROC curve (AUC) is one of the most widely used metrics for evaluation. AUC of a classifier is equal to the probability that the classifier will rank a randomly chosen positive example higher than a randomly chosen negative example. AUC has a range of [0, 1]. The greater the value, the better is the performance of our model. Figure 3.14 shows an example of a ROC curve and a confusion matrix.

## 3.7. Application of ML for biomarkers development

The ultimate goal of the radiomic approach is to build predictive models for treatment outcome and risk assessment, based on quantitative phenotypic characteristics of the tumor computed from radiological images and other clinically available information [216, 217]. Due to its ever-growing high-dimensional nature, the radiomic process needs much more powerful analytic tools, and AI appears to be a potential candidate for this purpose, because of its great capabilities of handling a massive amount of data compared with the traditional statistical methods. AI algorithms are not only able to analyze the numeric data provided by the predefined or hand-crafted radiomic features but also able to directly analyze the images, without any need for human intervention, to automatically design its own radiomic features.

In essence, radiomics consists of converting images into a high-dimensional feature space that can be studied via statistical and machine learning methods. It should be noted that the extraction of texture or radiomic features by itself does not necessarily require AI. However, AI (ML) is used to construct prediction models that can learn from existing datasets and analyze and perform predictions on related but new datasets. Therefore, a radiomic pipeline may be constructed by combining a computerized image analysis software (for image analysis and feature extraction) and an ML approach (either classic ML or deep learning) for constructing prediction models. Alternatively, DL may be used to perform both tasks (i.e. image analysis and construction of prediction models). This is a clear advantage of DL and highlights the great interest in this technology for applications in medical imaging. However, the relative disadvantage is the larger datasets required for constructing reliable algorithms that may be a disadvantage for early studies and pilot investigations, especially on uncommon disease entities or disease entities requiring significant sub-stratification resulting in small patient numbers, as has been alluded to earlier.

Despite the great potential of the radiomics and ML approaches discussed so far, significant challenges remain, which must be overcome before this methodology get to be applied routinely in the clinical setting. One of the main challenges is the replicability of radiomic studies, a pre-requisite for widespread clinical implementation. The main sources of variations are those related to the image acquisition and reconstruction process as well as the image segmentation method. Although the feature extraction process, is the easiest to standardize, there are a large number of characteristics available, so that feature selection or reduction variable methodologies should be carried before any ML model construction, because these can identify redundant and irrelevant imaging features, allowing to remove them from further analysis and improving the ML classifier performance [106]. The radiomic and ML methodologies are influenced by the feature selection method as well as the ML algorithm used. Because, different combinations have different performances, and it depends possibly on the tumor and clinical setting, some authors have recommended performing this kind of study for each tumor and clinical context [225]. The identification of optimal ML methods for radiomic applications is a crucial step towards stable and clinically relevant radiomic biomarkers construction. In chapter 6 of this thesis, we have addressed this circumstance, becoming one of the most significant contributions of this work.

# CHAPTER

# 4

# Heterogeneity in $^{18}$F-FDG-PET/CT of Non–Small Cell Lung Carcinoma and Its Relationship to Metabolic Parameters and Pathologic Staging

A substantial portion of this chapter is written or based on the following publications:

**Heterogeneity in [$^{18}$F] Fluorodeoxyglucose Positron Emission Tomography / Computed Tomography of Non–Small Cell Lung Carcinoma and Its Relationship to Metabolic Parameters and Pathologic Staging**. **van Gómez López O**, García Vicente AM, Honguero Martínez AF, Soriano Castrejón AM, Jiménez Londoño GA, Udias JM, León Atance P. ***Mol Imaging.*** 2014;12. doi: 10.2210/8290.2014.00022.

## Summary

To investigate the relationships between tumor heterogeneity, assessed by texture analysis of $^{18}$F-Fluorodeoxyglucose positron emission tomography (FDG-PET) images, metabolic parameters, and pathologic staging in patients with non–small cell lung carcinoma (NSCLC). A retrospective analysis of 38 patients with histologically confirmed NSCLC who underwent staging FDG-PET/computed tomography ($^{18}$F-FDG-PET/CT) was performed. Tumor images were segmented using a standardized uptake value (SUV) cutoff of 2.5. Five textural features, related to the heterogeneity of gray-level distribution, were computed (energy, entropy, contrast, homogeneity, and correlation). Additionally, metabolic parameters such as SUVmax, SUVmean, metabolic tumor volume (MTV), and total lesion glycolysis (TLG), as well as pathologic staging, histologic subtype, and tumor diameter, were obtained. Finally, a correlation analysis was carried out. Of 38 tumors, 63.2% were epidermoid and 36.8% were adenocarcinomas. The mean ±standard deviation values of MTV and TLG were 30.47 ± 25.17 mL and 197.81 ± 251.11 g, respectively. There was a positive relationship of all metabolic parameters (SUVmax, SUVmean, MTV, and TLG) with entropy, correlation, and homogeneity and a negative relationship with energy and contrast. The T component of the pathologic TNM staging (*p*T) was similarly correlated with these textural parameters. Textural features associated with tumor heterogeneity were shown to be related to global metabolic parameters and pathologic staging.

## 4.1. Introduction

Lung cancer is a leading cause of cancer-related death worldwide in both men and women. Most often lung cancer occurs as non-small cell lung carcinoma (NSCLC) [135]. After the initial diagnosis, accurate staging of NSCLC using CT or $^{18}$F-FDG-PET/CT is crucial for determining the appropriate therapy. Despite upgrades in imaging technologies and treatment over the past two decades, the improvement in survival remains modest, with a five-year global survival rate as low as 16% [107], while untreated lung cancer patients live 7.15 months after diagnosis [226].

Because FDG is a glucose analog, tumoral metabolism can be assessed by PET/CT. Usually, diagnosticians use a semi-quantitative parameter, the maximum standard uptake value (SUVmax), to characterize the tumoral lesion [227]. In addition, although not routinely used, global or volume-based semi-quantitative metabolic parameters of the primary tumor such as mean standard uptake value (SUVmean), metabolic tumor volume (MTV), and total lesion glycolysis (TLG) (given as MTV x SUVmean) can be easily obtained from post-processed images. These parameters provide a more

accurate assessment of the tumor burden, with potentially higher predictive and prognostic value than SUVmax for lung tumors [228-230].

Unlike SUVmax, which measures FDG concentration in a single voxel of a metabolically active tumor, global semi-quantitative parameters consider all voxels inside the tumoral volume, thus probing a more general view of the tumor. Their drawback is the lack of a standardized method for volume definition, although several segmentation algorithms show good performance [231, 232].

On the other hand, the biological heterogeneity of the tumors is an important factor implicated in poor treatment response, a higher chance of developing metastasis, and shorter progression-free and overall survival [233-235]. An explanation is that tumor heterogeneity originates from several factors at molecular, cellular, and physiological levels. Some of these factors are cell proliferation and angiogenesis rates, hypoxia, receptor distribution, and metabolism, which have a non-homogeneous spatial distribution in tumor volume [236, 237]. In recent years, there has been a considerable effort in the medical imaging community to obtain correlations between image features and tumor heterogeneity [55, 56, 64, 65]. An approximation to this issue is the texture analysis because image texture gives us information about the spatial arrangement of voxel intensities (i.e. the spatial distribution of radiotracer) in an image or a selected region.

The metrics or features obtained by texture analysis can be of first, second, or higher orders. First-order features are statistics calculated from the image voxel intensities distribution, like variance and mean, and do not consider pixel neighbor relationships. Second- and higher-order features measure the relationships between groups of two or more pixels in the image and reflect the underlying spatial variation of voxel intensities, providing a measure of the image heterogeneity. They can be computed from grey-level co-occurrence matrices (GLCM). These matrices determine how often (i.e. the probability) a pixel of intensity *i* finds itself within a certain relationship to another pixel of intensity *j*. Formally, given the image $f(x,y)$ with a set of N discrete intensity levels, the co-occurrence matrix $P_{d\theta}(i,j)$ is defined such that its $(i,j)$th entry is equal to the number of times that $f(x1,y2) = i$ and $f(x2,y2) = j$, where $(x2,y2) = (x1,y1) + (d\cos\theta, d\sin\theta)$ . This yields a square matrix whose dimension equals the square of the number of intensity levels in the image, for each distance $d$ and orientation $\theta$. Second-order features based on co-occurrence matrices include entropy, energy, contrast, homogeneity, and correlation [238, 239]. The parameters of energy and entropy are defined by the equations that follow.

$$\text{Energy} = \sum_{i=1}^{N}\sum_{\text{j}=1}^{\text{N}} P_{d\theta}\,(i,j)^2 \qquad (4.1)$$

$$\text{Entropy} = -\sum_{i=1}^{N}\sum_{j=1}^{N} P_{d\theta}(i,j)\log\left(P_{d\theta}\,(i,j)\right) \qquad (4.2)$$

For instance, entropy and energy features assess the spatial heterogeneity in the radiotracer uptake. Entropy measures the randomness of the pixel or voxel distribution in the image; it will take a higher value for a more random distribution. Energy measures the homogeneity in an image, where higher values mean greater uniformity of the gray-level values in the co-occurrence matrix [56]. In short, the higher the entropy and the lower the energy, the higher the heterogeneity of the gray-level distribution of tumor image. On the other hand, contrast relates to the dynamic range of intensity levels in an image and the level of local intensity variation; homogeneity represents the uniformity of the co-occurrence matrix, and correlation is a measurement of gray-tone linear dependencies [55]. Similarly, this procedure can be extended to three-dimensional (3D) images.

In NSCLC, researchers have described relations between texture parameters assessed in CT or PET images, metabolic parameters like SUVmax and SUVmean with survival and treatment response [74, 240]. The relationship between textural features assessed in PET images with volume-based (MTV and TLG) analysis and tumor stage has not yet been established. However, in other tumors, such as esophageal carcinoma, this relationship has been shown [241]. Therefore, the objective of this study was to analyze the relationship between tumor heterogeneity assessed by 3D textural analysis of the metabolic parameters of PET images and pathologic staging in patients with NSCLC.

## 4.2. Methods

### 4.2.1. Patients

Thirty-eight patients (34 men and 4 women) undergoing combined whole-body $^{18}$FDG-PET/CT, between January 2007 and December 2011 were included in the study. Patients fasted for at least 4 hours before the PET/CT examination and had blood glucose levels less than 160 *mg/dL* prior to an intravenous administration of approximately 370 *MBq* of $^{18}$F-FDG.

All patients had a resectable tumor up to Stage IIIA. After surgery, tumor size (*cm*) according to the largest diameter of the specimen and pathological staging (*p*T, *p*N, and integrated TNM stage) were assigned according to the seventh edition of AJCC TNM classification [30]. The data analysis was carried out after approval by the institutional review board.

### 4.2.2. PET/CT Image Acquisition

The patients underwent $^{18}$F-FDG-PET/CT scans using an integrated PET/CT scanner (Discovery STE 16, GE Health Care). Prior to PET acquisition, helical CT was performed from the head to proximal thighs to provide attenuation correction, with acquisition parameters for the CT of 120 *kV* and modulated 80 *mA*. No oral or intravenous contrast agents were used. Emission scans from the head to the proximal thigh were acquired at 60-90 minutes after $^{18}$F-FDG administration. Images were acquired in three-dimensional (3D) modes, 3 minutes per table position. PET images were reconstructed using CT for attenuation correction with ordered-subset expectation maximization iterative reconstruction algorithm supplied by the scanner manufacturer. The PET and CT slice thickness was 3.8 *mm*.

### 4.2.3. Lesion Segmentation

The PET images in DICOM format were transferred to 3D SLICER software (Harvard University, Cambridge, MA) [33]. In the visualization tools of this software, a nuclear medicine specialist detects the primary tumoral lesion, and on the basis of this localization, the regions equal to or greater than SUV 2.5 were selected to automatically delineate the volume of interest (VOI). All parameters were subsequently extracted from this delineated volume (Figure 4.1).

### 4.2.4. Metabolic Parameters

By using the same software, semiquantitative metabolic parameters of the primary tumor, such as maximum SUVmax (*g/mL*), SUVmean (*g/mL*), and MTV (i.e., the tumoral volume in milliliters obtained by using a SUVmax cutoff of 2.5), were obtained. TLG (*g*), given as MTV 2 SUVmean, was also calculated. FDG uptake was quantified using the expression given below:

$$SUV = \frac{\text{Activity concentration in the tissue}(\frac{\text{Bq}}{\text{mL}})}{\text{Administered activity (Bq)/bodyweight (g)}} \qquad (4.3)$$

### 4.2.5. Texture Analysis

VOI was transferred from 3D SLICER to DICOM format, and then second-order 3D textural features were extracted from the segmented VOI using *MaZda* software (Lodz University of Technology, Lodz, Poland) [34]. The MaZda software was configured to evaluate the texture at distance of one voxel ($d = 1$) at 12 different angles $\Theta$; because there is a textural feature value per angle, the final textural feature value was an average over all directions. Using this method, we obtained five textural features: energy, entropy, contrast, correlation, and homogeneity.

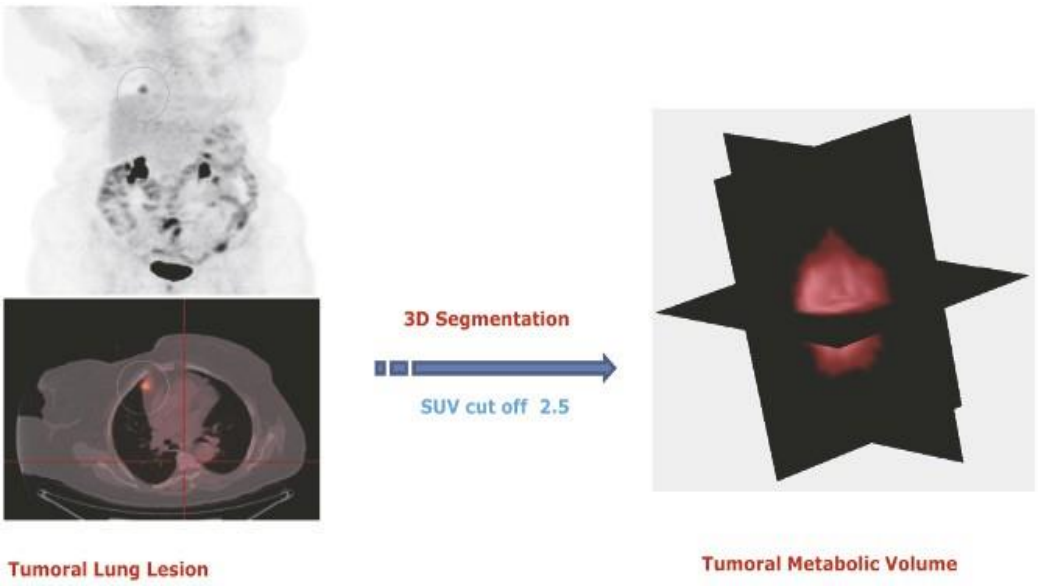



***Figure 4.1.*** PET/CT segmentation of the lesions. From the $^{18}$F-FDG-PET/CT study, the metabolic tumor volume was segmented by using a standardized uptake value (SUV) cutoff of 2.5. From these volumes, all second-order textural features and metabolic parameters were extracted (personal collection).

### 4.2.6. Statistical Analysis

We performed statistical analysis using *SPSS for Windows* version 19.0 (IBM, Armonk, NY) [35]. A Pearson correlation test assessed the relationship between the continuous variables (textural features of the PET images and metabolic parameters). We also performed linear regression. To analyze the correlation between textural parameters and categorical variables (TNM classification, histologic

type, and AJCC stage) [30], we used the nonparametric Spearman correlation test. All comparisons were two-sided, with statistical significance assessed at $p < 0.05$.

## 4.3. Results

Thirty-eight patients with NSCLC up to the IIIA stage were assessed (34 men and 4 women, with a mean age of 64.4 ± 8.3 years).

### 4.3.1. Pathologic Characteristics and Metabolic Parameters

Of the primary tumors, 24 (63.2%) were epidermoid and 14 (36.8%) were adenocarcinomas. The number of patients with tumor stage IA, IB, IIA, IIB, or IIIA were 4, 16, 8, 4, and 6, respectively. Their demographics and pathologic stage are summarized in Table 4.1. Mean ±standard deviation values of tumoral size, SUVmax, SUVmean, MTV, and TLG were 3.63 ± 1.51 *cm*, 13.85 ± 7.25 *g/mL*, 5.81 ± 1.98 *g/mL*, 30.47 ± 25.17 *mL*, and 197.81 ± 251.11 *g*, respectively. The distribution of pathologic and metabolic variables, together with the five textural features for each patient, is shown in Table 4.2. No statistically significant differences were found between the mean values of the tumor size, AJCC stage, and metabolic parameters between adenocarcinomas and epidermoid tumors. However, adenocarcinomas had a mean value of energy and homogeneity lower and higher ($p = 0.027$ and $p = 0.047$, respectively) than epidermoid tumors.

***Table 4.1.*** Patient Characteristics and Disease Stage

| | **Epidermoid** n = 24 (63.2%) | **Adenocarcinoma** n = 14 (36.8%) |
|---|---|---|
| Age, mean ± SD | 65.7 ± 7.6 | 61.8 ± 8.7 |
| Gender, n (%) | | |
| Male | 23 (95.8) | 11 (78.6) |
| Female | 1 (4.2) | 3 (22.4) |
| AJCC stage, n (%) | | |
| IA | 4 (20) | — |
| IB | 10 (40) | 6 (42.9) |
| IIA | 3 (12) | 5 (35.7) |
| IIB | 3 (12) | 1 (7.1) |
| IIIC | 4 (16) | 2 (14.3) |

AJCC = American Joint Committee on Cancer [30]

The SUVmax and SUVmean were correlated with the MTV and TLG (*r = 0.424, p = 0.008; r = 0.640, p < .0001 and r = 0.423, p = 0.008; r = 0.643, p < 0.0001*, respectively). The MTV and TLG were correlated (*r = 0.934, p < 0.0001*), as well as the SUVmax and SUVmean (*r = 0.931, p < 0.0001*).

### 4.3.2. Correlation between Texture and Metabolic Parameters

There was a positive relationship between all metabolic parameters (SUVmax, SUVmean, MTV, and TLG) with entropy, correlation, and homogeneity and a negative relationship with energy and contrast. Detailed values are given in Table 4.3. Linear regression fit of the usual metabolic parameters (SUVmax and SUVmean) to the global metabolic (TMV and TLG) parameters, entropy, and energy is shown in Figure 4.2 and Figure 4.3, respectively.

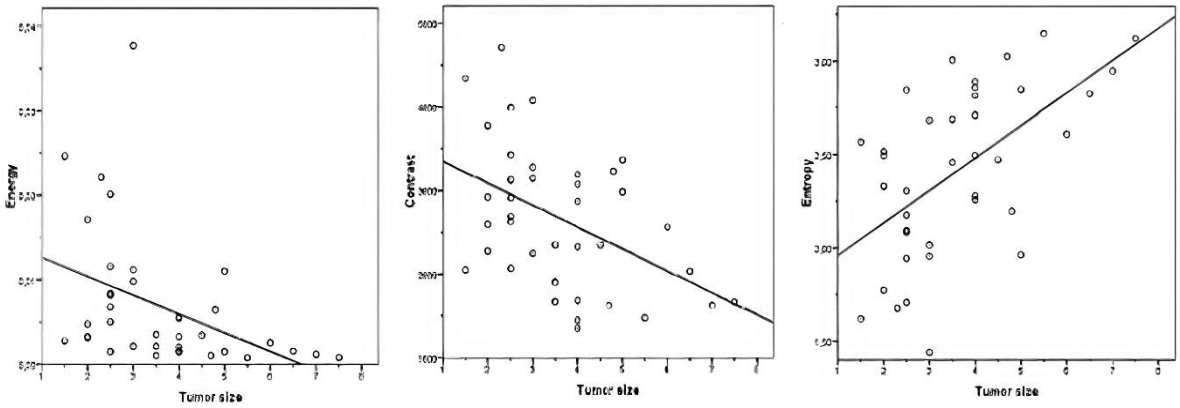



***Figure 4.2.*** Linear regression for energy, entropy, and contrast with tumoral size. Linear regression showed a negative correlation between metabolic tumor volume, tumor size, and energy and contrast (*r = -0.413, p = 0.009* and *r = -0.461, p = 0.004*, respectively). A positive correlation between tumor size and entropy (*r = 0.570, p < 0.0001*) was observed.

### 4.3.3. Correlation between Textural Parameters and Tumor Stage

Tumoral size was correlated with energy, contrast, correlation, entropy, MTV, and TLG (*r = -0.418, p = 0.009; r = -0.461, p = 0.004; r = 0.432, p = 0.007; r = 0.573, p < 0.0001; r = 0.596, p < 0.0001; r =0.500, p = 0.001*, respectively). The linear regressions for energy, entropy, and contrast are shown in Figure 4.4.

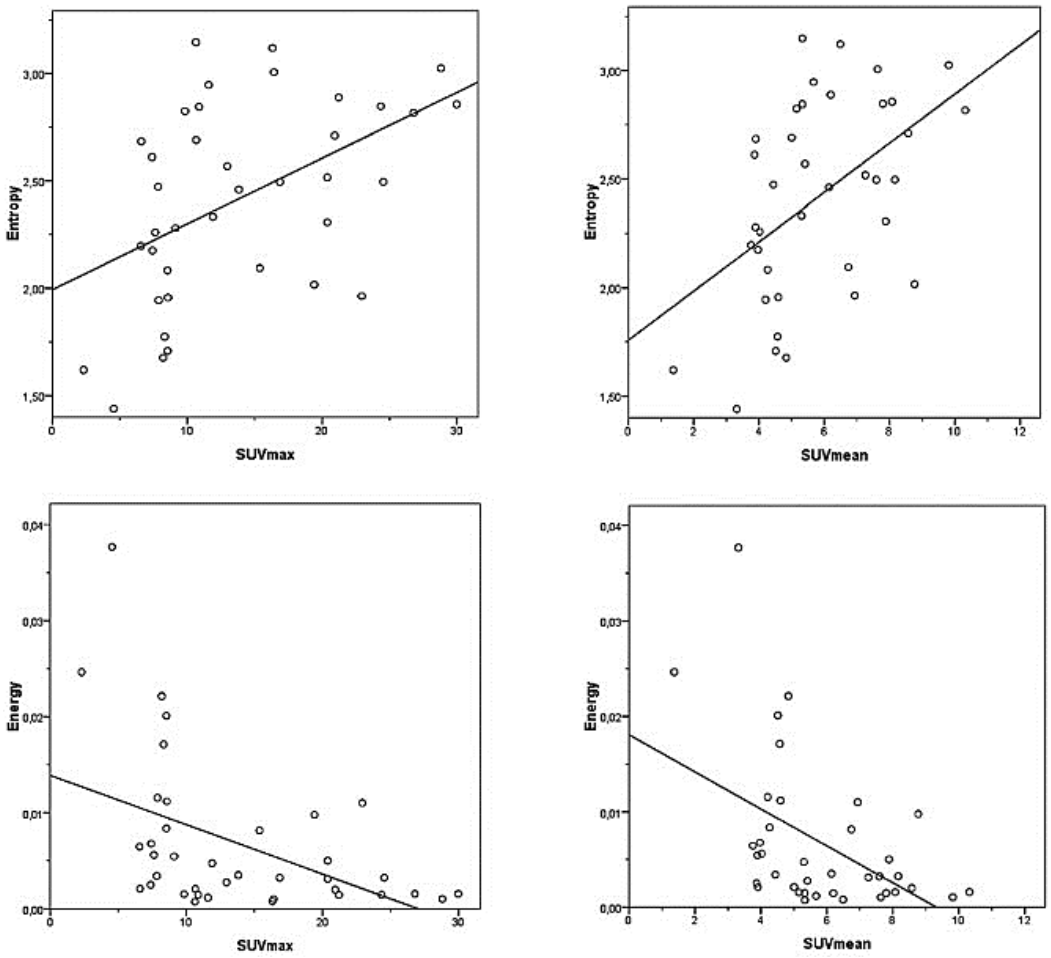



***Figure 4.3.*** Linear regression for energy and entropy with SUVmax and SUVmean. Linear regression showed a positive correlation between SUVmax and SUVmean with entropy (A and B); *r = 0.486, p = 0.002* and *r = 0.492, p = 0.002*, respectively. C and D show a negative correlation of the same two metabolic parameters with energy; *r = -0.465, p = 0.003* and *r = -0.479, p =0.002*, respectively.

A statistical correlation between the *p*T and energy, contrast, entropy, and MTV (*r = -0.376, p = 0.02; r = 0.319, p = 0.05; r = 0.360, p = 0.026; r = 0.376, p <0.02*) was found, but not between remaining AJCC components and textural or metabolic parameters.

However, by removing from the sample those patients with a low tumoral size but a high *p*T stage, according to the tumor location, we obtained a correlation between the AJCC stage and textural (energy, homogeneity, and entropy) and metabolic (MTV and TLG) parameters. On the other hand, *p*N becomes related to all five textural features, and MTV and TLG (Table 4.4 and Figure 4.5).

***Table 4.2.*** Summary of Patients' Demographics, Global Metabolic Parameters, and Textural Features of Primary Tumor

| Patient Number | Histology | AJCC Stage | SUVmax | SUVmean | MTV | TLG | Energy | Entropy | Contrast | Homogeneity | Correlation |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Ad | IB | 16.86 | 7.59 | 24.84 | 188.58 | 0.0032 | 2.50 | 2325.6 | 0.02 | 0.37 |
| 2 | Ad | IIA | 26.79 | 10.32 | 51.83 | 535.01 | 0.0016 | 2.82 | 3076.4 | 0.03 | 0.20 |
| 3 | SCC | IIB | 9.11 | 3.89 | 18.19 | 70.75 | 0.0054 | 2.28 | 2870.7 | 0.05 | 0.38 |
| 4 | Ad | IIIA | 28.82 | 9.81 | 76.38 | 749.54 | 0.0010 | 3.02 | 1625.9 | 0.04 | 0.57 |
| 5 | SCC | IB | 29.98 | 8.07 | 58.48 | 472.18 | 0.0016 | 2.86 | 1354.4 | 0.08 | 0.66 |
| 6 | SCC | IA | 24.54 | 8.16 | 25.72 | 209.96 | 0.0033 | 2.50 | 2272.4 | 0.03 | 0.44 |
| 7 | Ad | IB | 10.66 | 5.00 | 34.33 | 171.75 | 0.0021 | 2.69 | 1898.5 | 0.03 | 0.48 |
| 8 | Ad | IIA | 10.63 | 5.33 | 94.47 | 503.58 | 0.0007 | 3.15 | 1474.5 | 0.04 | 0.57 |
| 9 | Ad | IIA | 19.41 | 8.77 | 9.58 | 84.03 | 0.0098 | 2.02 | 3278.1 | 0.02 | 0.21 |
| 10 | SCC | IIIA | 16.42 | 7.63 | 68.56 | 523.28 | 0.0010 | 3.01 | 1672.3 | 0.04 | 0.53 |
| 11 | SCC | IIB | 10.87 | 5.32 | 64.45 | 343.05 | 0.0015 | 2.84 | 2068.8 | 0.04 | 0.39 |
| 12 | Ad | IIA | 4.54 | 3.32 | 3.23 | 10.70 | 0.0377 | 1.44 | 4080.3 | 0.01 | -0.03 |
| 13 | SCC | IB | 21.22 | 6.19 | 57.31 | 354.89 | 0.0014 | 2.89 | 1444.8 | 0.06 | 0.61 |
| 14 | SCC | IIA | 9.83 | 5.15 | 45.08 | 232.24 | 0.0015 | 2.82 | 2036.4 | 0.03 | 0.41 |
| 15 | SCC | IIA | 22.94 | 6.93 | 8.02 | 55.60 | 0.0110 | 1.96 | 3363.6 | 0.03 | 0.21 |
| 16 | SCC | IA | 8.56 | 4.59 | 7.82 | 35.91 | 0.0112 | 1.96 | 3145.6 | 0.02 | 0.20 |
| 17 | SCC | IIA | 11.59 | 5.67 | 59.46 | 337.25 | 0.0012 | 2.95 | 1626.7 | 0.04 | 0.53 |
| 18 | Ad | IIIA | 7.88 | 4.20 | 7.92 | 33.28 | 0.0116 | 1.94 | 3423.1 | 0.03 | 0.17 |
| 19 | SCC | IB | 7.82 | 4.44 | 21.61 | 95.93 | 0.0034 | 2.47 | 2348.1 | 0.03 | 0.34 |
| 20 | SCC | IB | 6.59 | 3.90 | 33.84 | 132.03 | 0.0021 | 2.68 | 2244.0 | 0.03 | 0.37 |
| 21 | SCC | IIIA | 16.32 | 6.49 | 88.70 | 575.51 | 0.0008 | 3.12 | 1668.4 | 0.04 | 0.54 |
| 22 | SCC | IIB | 12.96 | 5.41 | 26.60 | 143.83 | 0.0028 | 2.57 | 2051.4 | 0.03 | 0.46 |
| 23 | SCC | IIIA | 8.19 | 4.83 | 5.37 | 25.97 | 0.0221 | 1.68 | 4705.6 | 0.02 | -0.10 |
| 24 | Ad | IB | 6.56 | 3.76 | 12.91 | 48.51 | 0.0065 | 2.20 | 3230.4 | 0.02 | 0.20 |
| 25 | SCC | IIIA | 15.38 | 6.74 | 10.07 | 67.86 | 0.0082 | 2.09 | 3125.6 | 0.03 | 0.24 |
| 26 | SCC | IB | 11.89 | 5.30 | 16.53 | 87.61 | 0.0047 | 2.33 | 2922.8 | 0.03 | 0.27 |
| 27 | Ad | IIB | 8.53 | 4.26 | 9.68 | 41.25 | 0.0084 | 2.08 | 2683.9 | 0.03 | 0.32 |
| 28 | Ad | IIA | 7.38 | 3.86 | 31.39 | 121.29 | 0.0025 | 2.61 | 2563.2 | 0.03 | 0.33 |
| 29 | Ad | IB | 8.32 | 4.57 | 5.67 | 25.91 | 0.0171 | 1.77 | 3772.4 | 0.03 | 0.10 |
| 30 | SCC | IB | 13.82 | 6.14 | 20.83 | 127.90 | 0.0035 | 2.46 | 2347.5 | 0.03 | 0.39 |
| 31 | Ad | IB | 8.54 | 4.51 | 5.09 | 22.93 | 0.0201 | 1.71 | 3993.7 | 0.03 | 0.06 |
| 32 | SCC | IA | 20.38 | 7.26 | 24.64 | 178.84 | 0.0031 | 2.52 | 2593.8 | 0.04 | 0.36 |
| 33 | SCC | IB | 20.38 | 7.88 | 15.26 | 120.27 | 0.0050 | 2.31 | 2628.1 | 0.03 | 0.35 |
| 34 | SCC | IA | 7.41 | 3.97 | 12.22 | 48.52 | 0.0068 | 2.18 | 2912.6 | 0.03 | 0.27 |
| 35 | SCC | IB | 24.35 | 7.80 | 47.73 | 372.05 | 0.0015 | 2.85 | 2981.7 | 0.03 | 0.23 |
| 36 | SCC | IB | 7.62 | 4.02 | 14.47 | 58.18 | 0.0056 | 2.26 | 3189.4 | 0.02 | 0.25 |
| 37 | Ad | IB | 2.30 | 1.37 | 4.21 | 5.76 | 0.0246 | 1.62 | 4340.3 | 0.03 | 0.01 |
| 38 | SCC | IB | 20.94 | 8.57 | 35.60 | 305.25 | 0.0020 | 2.71 | 1688.2 | 0.04 | 0.53 |

Ad = adenocarcinoma; AJCC = American Joint Committee on Cancer; MTV = metabolic tumor volume ($cm^3$); SCC = squamous cell carcinoma; SUV = standardized uptake value; TLG = total lesion glycolysis (g).

***Table 4.3***. Correlations between Global Metabolic Parameters and Textural Features of the Primary Tumors

| | Energy | Entropy | Contrast | Correlation | Homogeneity |
|---|---|---|---|---|---|
| **SUVmax** | -0. 466 (0.003) | 0.487 (0.002) | -0.437 (0.006) | 0.475 (0.003) | 0.551 (<0.001) |
| **SUVmean** | -0.480 (0.002) | 0.493 (0.002) | -0.420 (0.009) | 0.424 (0.008) | 0.352 (0.030) |
| **MTV** | -0.627 (<0.001) | 0.905 (<0.001) | -0.781 (<0.001) | 0.749 (<0.001) | 0.574 (<0.001) |
| **TLG** | -0.578 (<0.001) | 0.842 (<0.001) | -0.701 (<0.001) | 0.684 (<0.001) | 0.588 (<0.001) |

MTV = metabolic tumor volume ($cm^3$); SUVmax = maximum standardized uptake value; SUVmean = mean standardized uptake value; TLG = total lesion glycolysis (g). Numbers indicate Pearson coefficient; p values are inside parentheses.

## 4.4. Discussion

FDG-PET/CT has been broadly used for the diagnosis and staging of malignancies and recently to determine early response to treatment. In daily routine, only semiquantitative measures of lesion activity, such as SUVs, are usually employed, even though volume-based metabolic parameters could be more representative of tumor characteristics and have more predictive value [228-230] because they consider all the voxels in the tumor image. However, spatial information about voxel intensity relationships, that is, about tumor heterogeneity of radiotracer uptake, is missing in volume-based parameters, which could limit their ability to describe the tumor characteristics and especially to predict therapy response or prognosis. Therefore, in this work, we explored the associations between metabolic parameters and textural features, which consider the spatial relationship of the image voxels.

Parameters obtained by texture analysis of images, reflecting the underlying spatial variation and heterogeneity of voxel intensities in the image of the tumor, provide a measure of the tumor heterogeneity [22]. For example, the spatial distribution of $^{18}$F-FDG uptake has been related to the intratumoral distribution of hypoxia [242, 243] and textural parameters derived from CT images of NSCLC have been correlated to tumor hypoxia and angiogenesis [244]. Although several metrics can be derived from texture analysis, only a limited number of them have shown robustness in a clinical setting because technical and physiologic factors such as PET limited resolution, partial volume effect, reconstruction algorithms, movement artifacts, and noise could affect the tumor heterogeneity quantification utilizing textural features [175, 197, 245]. We must be sure that the textural feature

measures underlying tumor properties, such as tumor heterogeneity, and not inaccuracies, such as blurring and artifacts, so a standard procedure of measurement must be adopted.

Most texture analyses have been performed on 2D CT images, although 3D analysis of the whole tumor seems to be more representative of tumor heterogeneity [246]. Instead of 2D analysis, we performed a 3D textural feature extraction of the PET component from PET/CT images, which allowed us to dispose of a higher number of voxels to perform a statistical textural computation. This methodology is particularly important in texture analysis of PET images, where fewer voxels are representing the lesion to perform statistical analysis because the voxel size is larger compared with CT. The better resolution and more favorable image characteristics make CT more suitable for texture analysis; for this reason, most of the studies of tumor texture analysis have used CT. However, PET images have a functional meaning, showing biological aspects of the tumors not expressed in CT images; thus, they should be more widely used for texture analysis.

Few studies have used PET images to evaluate the heterogeneity in radiotracer distribution, even when its distribution offers firsthand information about biological processes inside the tissues. The relationships between some textural features based on CT and tumor stage in NSCLC have been previously reported [247]. However, similar analyses have not been performed by deriving textural features from the PET image of lung tumors. To date, in NSCLC, texture analyses based on PET images have only been compared to SUV, not considering volume-based metabolic parameters. To our knowledge, no association of PET textural features with volume-based metabolic parameters and the AJCC stage has been previously reported. Table 4.5 provides a summary of the main studies and their results [248-251].

In our study, we computed the textural features energy, entropy, contrast, correlation, and homogeneity from the 3D tumoral volume extracted from $^{18}$-FDG-PET/CT images of the patients. These textural features showed correlations with all metabolic parameters, especially with the global metabolic parameters (MTV and TLG). Tumors with higher SUVmax, SUVmean, MTV, and TLG were more heterogeneous in the textural analysis.

The correlation between entropy and energy with MTV and TLG was not linear. The curve of the scatter plot might be better modeled by a nonlinear function, such as a quadratic; with such adjustment, an expected increase in the correlation coefficient is expected. The previous works have shown that intratumor heterogeneity increases as tumors grow [65, 175, 252]. Possibly because larger

tumors are composed of several different types of tissues and regions with variable uptake. On the other hand, smaller tumors may also have heterogeneity at the cellular and tissue levels, but it may be blurred in PET images due to the limited spatial resolution [253]. Since TLG and MTV are volume-based metabolic parameters, we can infer from the graph that when tumors are small (for example, less than 10 ml), a smaller change in tumor volume translates into a significant change in the value of the tumoral entropy and energy (i.e. tumor heterogeneity), while with larger tumors this change is less noticeable, especially with the energy. This is of great importance if we intend to establish the limits of application of the textural analysis of PET / CT images. So, an accurate determination of the tumoral volume should be performed.

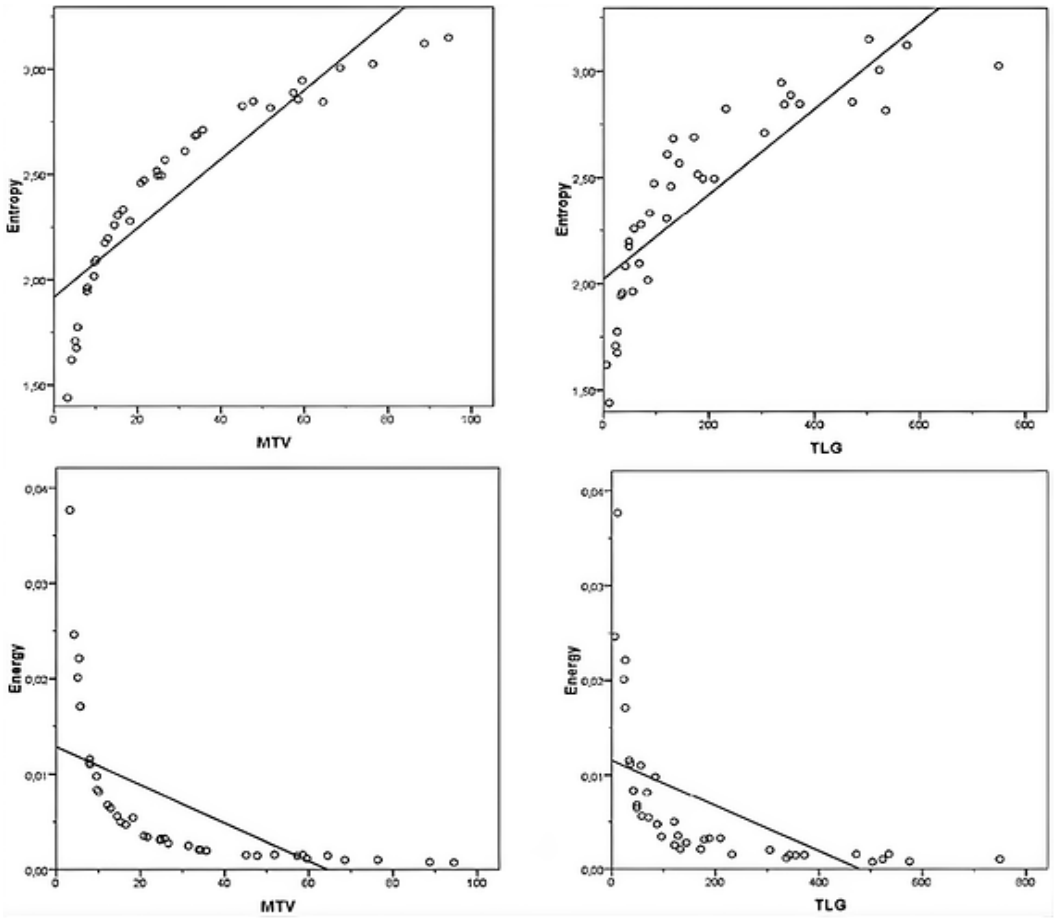



***Figure 4.4***. Linear regression for energy and entropy with metabolic tumor volume (MTV) and total lesion glycolysis (TLG). Linear regression showed a positive correlation between MTV and TLG with entropy (A and B); $r = 0.904$, $p < 0.0001$ and $r = 0.842$, $p < 0.0001$, respectively. C and D show a negative correlation between both global metabolic parameters with energy; $r = -0.627$, $p < 0.0001$ and $r = -0.578$, $p < 0.0001$, respectively.

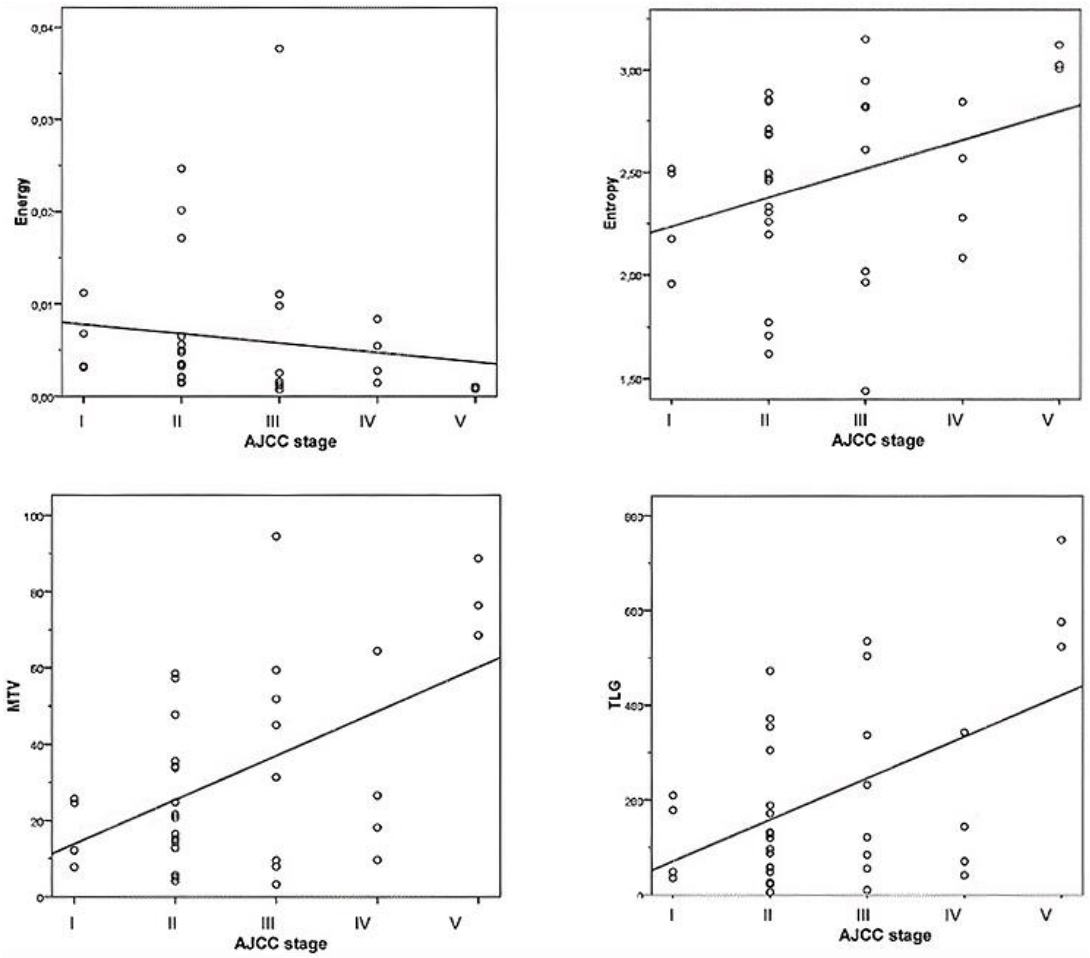



**Figure 4.5.** Spearman correlations between AJCC stage and energy, entropy, metabolic tumor volume (MTV), and total lesion glycolysis (TLG). Spearman correlations: (A) AJCC stage and energy (*r = -0.363, p = 0.038*); (B) AJCC stage and entropy (*r = 0.351, p = 0.038*); (C) AJCC stage and MTV (*r = 0.387, p = .022*); and (D) AJCC stage and TLG (*r = 0.329, p = 0.050*).

We found that adenocarcinomas had a lower mean value of energy and higher homogeneity than epidermoid tumors. This could indicate that adenocarcinomas are more metabolically heterogeneous than epidermoid tumors. Furthermore, although differences in SUVmax between adenocarcinomas and epidermoid tumors have been reported [254] we could not confirm such a finding in our data. There are several methods to assess the textural features of an image; however, there are not models that directly link these textural features of the image to the biological proprieties of the imaged tumor. It would be necessary, to formulate such models in order to be able to use texture analysis to identify tumor types and to differentiate benign from malignant tumors.

***Table 4.4.*** Correlation between Textural Parameters and Clinical Characteristics

| | MTV | TLG | Energy | Entropy | Contrast | Correlation | Homogeneity |
|---|---|---|---|---|---|---|---|
| **TD (cm)** | -0.59 (<0.001) | 0.50 (0.001) | -0.21 (0.188) | 0.42 (0.007) | -0.41 (0.009) | 0.41 (0.010) | 0.42 (0.007) |
| **AJCC stage** | 0.31 (0.05) | 0.31 (0.057) | -0.54 (0.476) | 0.33 (0.843) | -0.10 (0.537) | 0.14 (0.394) | 0.15 (0.341) |
| **pT** | -0.47 (<0.001) | -0.34 (0.033) | -0.27 (0.91) | 0.37 (0.019) | -0.33 (0.036) | 0.32 (0.045) | 0.37 (0.019) |

AJCC = American Joint Committee on Cancer; MTV = metabolic tumor volume ($cm^3$); TLG = total lesion glycolysis (g). Numbers indicate Spearman coefficient; p values are inside the parentheses.

***Table 4.5.*** Summary of Studies Investigating Textural Analysis of $^{18}$F-FDG-PET in NSCLC

| Study, Year | Analysis Type | Metabolic Parameters | PET Features | CT Features | Findings |
|---|---|---|---|---|---|
| **Hatt et al, 2011 [248]** | 3D | MTV | CV | Tumor volume | Direct relationship between CT volume and MTV and heterogeneity |
| **Vaidya et al, 2012 [249]** | 3D | TLG | Energy, contrast, entropy, homogeneity | Energy, Contrast, entropy, homogeneity | No correlation of metabolic entropy, and textural parameters with recurrence |
| **Bagci et al, 2013 [250]** | 3D | SUVmax | Energy, contrast, entropy, homogeneity, correlation, and others | Not assessed | Correlation of SUVmax with several textural features |
| **van Velden et al, 2011 [251]** | 3D | SUVmax | SUV-volume histogram | Not assessed | SUV-volume histogram better index of heterogeneity than SUVmax |

CT = computed tomography; CV = coefficient of variation; $^{18}$F-FDG = fluorodeoxyglucose; MTV = metabolic tumor volume; NSCLC = non-small cell carcinoma; PET = positron emission tomography; SUVmax = maximum standardized uptake value; TLG = total lesion glycolysis.

All of the tumoral metabolic parameters were correlated with each other. The higher correlation was between SUVmax and SUVmean and MTV and TLG (see Table 4.3). Thus, tumors with a higher SUVmax had a higher SUVmean, and as was expected, a larger MTV indicated a larger TLG. Tumoral size correlated with volume-based metabolic parameters (MTV and TLG). This should be expected because volume depends directly on the tumor size. On the other hand, the tumoral size did not correlate with the SUVs, even though this relationship has been previously published when

assessing, unlike in our study, patients with no operable stages [255]. In addition, we found that larger tumors were more heterogeneous (lower energy and higher entropy) and had lower contrast and higher correlation. In the case of contrast, even though larger tumors appear more homogeneous in a visual assessment (i.e., they exhibit lower contrast) they are more heterogeneous in the FDG spatial distribution. This could be because low contrast images have low spatial frequencies but not necessarily low heterogeneity. However, because the contrast is a textural feature with large variability depending on the acquisition modes and reconstruction parameters [167], these results must be reviewed in future work.

In our original series of 38 patients, we found a correlation only between the *p*T and tumor size, energy, contrast, entropy, and MTV. However, when we rejected three patients with small tumors and high AJCC staging (because the tumors affected the pleura or were located less than 2 cm from the carina), we also found a correlation with the AJCC staging. In the reduced series of 35 patients, the AJCC staging correlated with the tumor size, energy, homogeneity, entropy, MTV, and TLG; the most significant correlation was with the energy and MTV. We thought that this finding was because the lung cancer AJCC stage includes, in the same group, tumors with different sizes because it uses other prognostic features besides the tumor size. On the other hand, entropy and energy had no linear relationship with volume-based metabolic parameters, which finally depend on tumoral size.

Another important factor is that our population mostly consisted of males. It is because lung cancer incidence has historically been higher in men than women. However, the magnitude of this difference has decreased in the last years, because cigarette smoking in women has increased. We do not know if this unbalanced population could have affected our results. However, due to this bias, the generalizability of our study may be compromised so that, additional studies must be performed with larger numbers of patients (a more comparable number of male and female patients) to settle this issue.

We also found no correlation between the SUVmax and SUVmean with the AJCC stage and *p*T or *p*N stage or a statistically significant difference in the mean values, although these relationships have been reported in NSCLC [256] and other tumors [241]. One of the most important advantages of texture analysis is that its measurement is performed in the post-processing PET/CT image. Therefore, it could be easily included in the daily clinical routine. Parameters derived from texture analysis may have reproducibility similar to or better than that of simple SUV measurements [197] and be less susceptible to the noise or the reconstruction algorithms used [167]. Before performing a

3D texture analysis of PET/CT images, it is necessary to perform segmentation; however, there is no truly validated segmentation algorithm able to deal with tracer uptake heterogeneity. The threshold-based approximation has several drawbacks, especially in heterogeneous tumors; when the tumor uptake is more heterogeneous, the underestimation of the PET volumes by threshold-based techniques is larger [248]. We decided to use a cutoff value of 2.5 given the simplicity of the threshold method and the fact that previous authors have shown that the tumor length seen on an FDG-PET/CT image with a cutoff value of 2.5 was close to the gross tumor length [257] Therefore, more sophisticated approximations can be considered in future work. An important drawback in the measures of any image-based parameter in lung tumors is respiratory movement. Given that respiratory gating permits better characterization of the malignant lung lesion compared to the standard acquisition [258, 259], the assessment of textural changes linked to the respiratory cycle seems to be the more correct option. The assessment of FDG uptake heterogeneity by PET and its spatial and temporal variations could help us understand tumor biology. Therefore, SUV and other measurements of FDG tumor uptake can potentially be supplemented by additional imaging parameters derived from either the PET or the CT images [260]. One additional step could be the use of both CT and PET images to perform a joint assessment of the tumor heterogeneity. It is expected that textural metrics derived from staging PET/CT images can give us better information about patient survival than semiquantitative metabolic parameters [74, 175, 240], and also help us decide which initial therapy to adopt. By performing interim PET/CT (i.e., PET/CT in the middle of the treatment) and assessing the textural changes of the tumor after the treatment, the response can be assessed and the treatment can be adjusted. For instance, if we found a lung tumor with high entropy and low energy in a staging PET/CT, it could need more aggressive therapy from the beginning than others with low entropy and high energy, but this requires additional work to be validated.

## 4.5. Conclusion

Tumor heterogeneity in NSCLC assessed by texture analysis of the PET component of $^{18}$F-PET/CT images is correlated with global metabolic parameters, and both are associated with macroscopic tumor diameter and, under special considerations (no inclusion of a small tumor with high AJCC stage), with the AJCC stage. Some textural features have no linear relationship with volume-based metabolic parameters, making them more sensitive to tumor volume definition and defaulting to establish relationships with AJCC staging. These textural parameters have the potential to be used in clinical practice but require additional work to further validate their importance.

# CHAPTER

# 5

# $^{18}$F-FDG-PET/CT in the assessment of pulmonary solitary nodules: comparison of different analysis methods and risk variables in the prediction of malignancy

A substantial portion of this chapter is written or based on the following publications:

**18F-FDG-PET/CT in the assessment of pulmonary solitary nodules: comparison of different analysis methods and risk variables in the prediction of malignancy.**

***Ober van Gómez López***, Ana María García Vicente, Antonio Francisco Honguero Martínez, Germán Andrés Jiménez Londoño, Carlos Hugo Vega Caicedo, Pablo León Atance, Ángel María Soriano Castrejón. ***Transl Lung Cancer Res*** 2015;4(3):228-235.

## Summary

The goal of this chapter is to compare the diagnostic performance of different metabolical, morphological, and clinical criteria for the correct presurgical classification of the solitary pulmonary nodule (SPN). Fifty-five patients, with SPN, were retrospectively analyzed. All patients underwent preoperative $^{18}$F-fluorodeoxyglucose positron emission tomography/computed tomography ($^{18}$F-FDG PET/CT. Maximum diameter in CT, maximum standard uptake value (SUVmax), histopathologic result, age, smoking history, and gender were obtained. Different criteria were established to classify a SPN as malignant: (I) visually detectable metabolism, (II) SUVmax > 2.5 regardless of SPN diameter, (III) SUVmax threshold depending on SPN diameter, and (IV) ratio SUVmax/diameter greater than 1. For each criterion, statistical diagnostic parameters were obtained. Receiver operating characteristic (ROC) analysis was performed to select the best diagnostic SUVmax and SUVmax/diameter cutoff. Additionally, a predictive model of malignancy of the SPN was derived by multivariate logistic regression. The results show that fifteen SPN (27.3%) were benign and 40 (72.7%) malignant. The mean values ± standard deviation (SD) of SPN diameter and SUVmax were 1.93 ± 0.57 *cm* and 3.93 ± 2.67 respectively. Sensitivity (*Se*) and specificity (*Sp*) of the different diagnostic criteria were (I): 97.5% and 13.1%; (II) 67.5% and 53.3%; (III) 70% and 53.3%; and (IV) 85% and 33.3%, respectively. The SUVmax cut-off value with the best diagnostic performance was 1.95 (*Se*: 80%; *Sp*: 53.3%). The predictive model had a *Se* of 87.5% and *Sp* of 46.7%. The SUVmax was an independent variable to predict malignancy. In conclusion, the assessment by semiquantitative methods did not improve the Se of visual analysis. The limited *Sp* was independent of the method used. However, the predictive model combining SUVmax and age was the best diagnostic approach.


## 5.1. Introduction

A solitary pulmonary nodule (SPN) is defined radiologically as an intraparenchymal lung lesion of less than 3 cm in diameter, with no associated atelectasis or adenopathy [261, 262]. The management of SPN is clinically controversial and is mainly dependent on the perceived probability of malignancy [81]. The prevalence of lung cancer in patients with SPN varies widely, from 2-13% in screening studies to 46-82% in positron emission tomography (PET) studies [80, 263]. For a suspicious malignant SPN, percutaneous transthoracic biopsy, transbronchial needle aspiration biopsy or video-assisted thoracoscopic surgery provides histological information. However, these are invasive procedures, skill dependent, and with variable accuracy to the diagnosis of cancer [80, 264, 265].

PET with $^{18}$F-FDG has had an important impact on the diagnosis of benign and malignant nodules. Some reports have suggested that PET can reduce the number of patients with pulmonary nodules who undergo unnecessary surgical biopsies [36]. Therefore, PET using $^{18}$F-FDG is an accurate and non-invasive method for diagnosing SPNs, with an overall sensitivity (*Se*) of 95% and a specificity (*Sp*) of 82% [37]. However, surgical resection is still needed to differentiate lung cancer from benign lesions in a significant number of cases [37]. The combination of computed tomography (CT) and PET in the hybrid imaging, has shown an excellent performance in classifying SPN as benign or malignant, where the Se of CT and the Sp of PET, resulting in an overall significantly improved accuracy [38, 81].

To determine the management and treatment of the patient with an SPN is necessary to estimate the probability of malignity from clinical and imaging data. Some independent predictors of malignancy include age, current or past smoking history, previous extrathoracic malignancy, nodule diameter, spiculation, and upper lobe location [80, 266]. Although specific models exist for the calculation of the probability of malignancy of an SPN, they do not have enough accuracy to replace the clinician's judgment. On the other hand, adding metabolic parameters derived from PET studies has been shown to improve the prediction of malignancy in SPN [38, 267], however, it is necessary to increase the evidence that supports the use of such metabolic parameters. FDG uptake on PET has been qualitatively and semiquantitatively evaluated. Visual assessment is usually based upon a comparison of FDG lesion uptake with normal mediastinal blood pool [268] and is the simplest among all the analyses, but nodules with similar FDG uptake to the mediastinum are difficult to evaluate visually. In order to have a more objective assessment, a cut-off of the maximum standard uptake value (SUVmax) has been used for the establishment of malignancy. However, a great number of factors can affect the SUV, among them, body size, the blood glucose concentration, the time after injection, and the lesion diameter [269]. As a result, the SUVmax of an SPN could not reflect its true nature. In an attempt to improve the diagnostic accuracy of the presurgical evaluation of the SPN, the integration of risk variables into predictive models has been carried out, because, contrary to the clinical judgment, quantitative predictive models might have advantages in accuracy and reproducibility [80, 266-268, 270]. Even though, several CT derived parameters have been included in such predictive model, metabolic variables have been no included. The purposes of the present study were as follows: (I) to determine an optimum semiquantitative criterion that allows discriminating between malignant and benign nodules and comparing with the visual assessment and (II) to derivate a model to estimate the pretest probability of malignancy of a patient with SPN based on clinical and $^{18}$F-FDG PET/CT image variables.

## 5.2. Materials and methods

A retrospective evaluation of $^{18}$F-FDG PET/CT image data, final pathological classification, and risk clinical and demographic variables of patients with SPN was performed. The data analysis was carried out after approval by the Institutional Review Board.

### 5.2.1 Patients

Between January 2007 and December 2012, patients with a suspicious SPN underwent a combined whole-body $^{18}$F-FDG PET/CT imaging and surgical resection of the SPN was included. After surgery, a final histological diagnose was assigned. Other patient's characteristics as gender, age, and previous or current history of smoking were analyzed.

### 5.2.2. PET/CT image acquisition and interpretation

Patients fasted for at least 4 h and had blood glucose levels less than 160 mg/dL previous to an intravenous administration of 370 *MBq* of $^{18}$F-FDG. PET/CT scans were performed approximately 60 min after $^{18}$F-FDG administration using an integrated PET/CT scanner (Discovery STE 16, GE Healthcare). PET/CT was obtained from the head to the proximal thighs. Prior to PET acquisition, helical CT was performed to provide attenuation correction, with acquisition parameters for the CT of 120 *kV* and modulated 120 *mA*. No oral or intravenous contrast agents were used. Emission images were acquired in three-dimensional (3D) mode, 3 *min* per table position. PET images were reconstructed using CT for attenuation correction with ordered-subset expectation maximization iterative reconstruction algorithm. The PET and CT section thickness was 3.8 *mm*. Two experienced nuclear medicine physicians reviewed the $^{18}$F-FDG-PET/CT studies in consensus. In the visual analysis of the PET data, a lesion was defined as negative (no $^{18}$F-FDG uptake visually detected) or positive ($^{18}$F-FDG-avid SPN regardless of its intensity). For semiquantitative analysis, a circular region of interest was placed over the nodule location with the peak activity.

The maximum intensity of $^{18}$F-FDG uptake was defined by body-weight SUVmax measurement using the commercially available software provided by the manufacturer. On the other hand, the nodule diameter (mm) was assessed in axial projection on the CT image. Four metabolic criteria were used to consider an SPN as positive and therefore probably malignant:

(I) A visually detectable metabolism

(II) SUVmax > 2.5 regardless of nodule diameter

(III) SUVmax ≥ 1 if diameter ≤ 1 cm or SUVmax > 2.5 if diameter > 1 *cm*

(IV) Ratio SUVmax/SPN diameter >1

### 5.2.3. Final diagnosis

All patients underwent surgical resection of the SPN. A definitive pathologic diagnosis of the SPN, classifying the lesions as benign or malignant, was established.

### 5.2.4. Statistical analysis

Statistical analysis was performed using *SPSS for windows* version 19.0 (IBM, Armonk, New York, USA) [35]. All the comparisons were two-sided using a *p*-value less than *0.05* to indicate statistical significance. An independent *t*-test was used for comparing the age, diameter, and SUVmax of the benign and malignant nodules, while that chi-square was used for smoking history, and gender. The diagnostic accuracy was obtained for each of the four different diagnostic approaches. A positive SPN classified by any of the four criteria was considered malignant in the metabolic assessment. A receiver operating characteristic (ROC) curve analysis was performed to obtain the best cut-off of the SUVmax and SUVmax/diameter (diagnostic approaches III and IV), and the areas under curve (AUC) values were obtained with a confidence interval (CI) of 95%.

Finally, we developed a model to estimate the probability of malignancy of patients with SPN by using stepwise logistic regression, with the final diagnosis as the dependent variable and the following independent variables: age, gender, smoking history (never *vs*. ever), nodule size, and SUVmax. Using backward selection, we achieved a final reduced model by eliminating variables that were not statistically significant at a level of *0.05*. We used this final model to calculate the estimated probability of malignancy in each patient. We compared the predicted probability of malignancy with the final diagnosis and constructed a ROC curve. To describe the accuracy of the model for identifying malignancy in the patients, we reported the AUC with a CI of 95%.

## 5.3. Results

Fifty-five patients with SPN (45 men and 10 women, with a mean age of 62±11 years) were studied. The pathologic analysis classified 40 (72.7%) of SPN as malignant and 15 (27.3%) as benign. From malignant SPN, the most prevalent histologies were: 65% adenocarcinoma, 17.5% epidermoid, and 12.5% small cell carcinoma. Among the benign SPN, the most prevalent histologies were: 40% organizing pneumonia and 20% fibrosis. Patient demographics, smoking history, and SPN characteristics attending the final pathologic diagnosis of the SPN are shown in Tables 5.1 and 5.2.

Mean ± standard deviation (SD) values of SPN diameter and SUVmax were 1.93 ± 0.57 *cm* and 3.93 ± 2.67, respectively. There were statistically significant differences between the SUVmax values and patient age with the final histology of the SPN (malignant or benign). The mean ± SD of the SUVmax for benign nodules was 2.29 ± 1.31 and 4.54 ± 2.80 for malignant nodules ($p = 0.005$). The mean patient age was 58 ± 9 and 64 ± 11 for benign and malignant SPN respectively, ($p=0.045$). No statistically significant differences were found for the rest of the variables (Table 5.3).

*Se*, *Sp* and diagnostic accuracy for the different diagnostic criteria were (I): 97.5%, 13.1% and 74.5%; (II) 67.5%, 53.3% and 63.3%; (III) 70%, 53.3% and 64.5%; (IV) 85%, 33.3% and 70.9%, respectively. ROC analysis showed an AUC for SUVmax and SUVmax/diameter of 0.75 and 0.79 ($p <0.005$), respectively. The cutoff values with the best diagnostic performance were 1.95 (*Se*: 80%, *Sp*: 53.3%) and 1.04 (*Se*: 82.5%, *Sp*:53.3%) for SUVmax and SUVmax/diameter, respectively. Figure 5.1 shows the ROC curves.

***Table 5.1*** Patient's characteristics

| Characteristic | Benign, n (%) | Malignant, n (%) |
|---|---|---|
| **Number of patients** | 15 (27.3) | 40 (72.7) |
| **Age (mean ± SD)** | 58.0 ± 9.1 | 64.2 ± 11.1 |
| **Gender** | | |
| **Male** | 14 (25.5) | 31 (56.4) |
| **Female** | 1 (1.8) | 9 (16.4) |
| **Smoking history** | | |
| **Yes** | 15 (27.3) | 35 (63.6) |
| **No** | 0 | 5 (9.1) |

SPN, solitary pulmonary nodule; SD, standard deviation.

***Table 5.2*** Pathologic results

| Characteristic/Pathologic diagnosis | Number (%) |
|---|---|
| Malignant | 40 (72.7) |
| Adenocarcinoma | 26 (47.3) |
| Squamous cell carcinoma | 7 (12.7) |
| Large cell carcinoma | 1 (1.8) |
| Small cell carcinoma | 5 (9.1) |
| Carcinosarcoma | 1 (1.8) |
| Benign | 15 (27.3) |
| Organizing pneumonia | 5 (9.1) |
| Fibrosis | 3 (5.5) |
| Hamartoma | 2 (3.6) |
| Inflammatory pseudotumor | 2 (3.6) |
| Granuloma | 2 (3.6) |
| Lung abscess | 1 (1.8) |

***Table 5.3*** Univariate analysis of patients' data

| **Characteristic** | **Benign, n (%)** | **Malignant, n (%)** | ***p*** |
|---|---|---|---|
| **Patient age (years)** | 58 ± 9.1* | 64.2 ± 11.1* | 0.045 |
| **Gender** | | | 0.169 |
| **Male** | 14 (25.5) | 31 (56.4) | |
| **Female** | 1 (1.8) | 9 (16.4) | |
| **Smoking history** | | | |
| **No** | 0 | 5 (9.1) | 0.189 |
| **Yes** | 15 (27.3) | 35 (63.6) | |
| **Nodule diameter (cm)** | 1.93 ± 0.66* | 1.92 ± 0.53* | 0.960 |
| **SUV max** | 2.29 ± 1.31* | 4.54 ± 2.80* | 0.005 |

*, Values are mean ± standard deviation. SUV max, maximum standard uptake value.

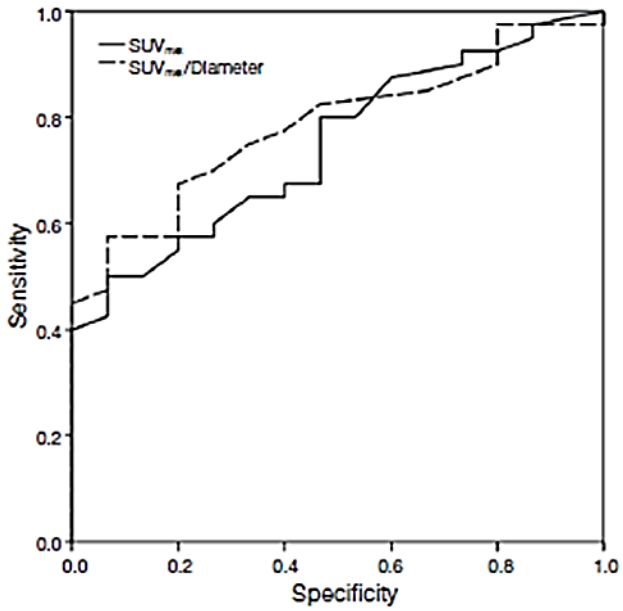



***Figure 5.1.*** ROC curve of SUVmax and SUVmax/diameter methodologies. ROC, receiver operating characteristic; SUVmax, maximum standard uptake value.

By using univariate analysis, we identified that age and SUVmax were associated with malignity (Table 5.3). However, only SUVmax was an independent predictor in the multivariate analysis, with an odds ratio of 1.6 and (95% CI, *1.01 - 2.417*), see Table 5.4. Although age was not an independent variable, it was included in the predictive model, because of its clinical importance, becoming to be an independent predictor in patients older than 60 years. All other variables were not predictors of malignity and therefore were not included in the final model. The prediction model is described by the following equations:

$$\text{Probability of malignancy of an SPN} = \frac{e^x}{1+e^x} \qquad (5.1)$$

Where $x = -3.767 + (4.89 \times SUVmax) + (0.052 \times Age)$, $e$ is the base of the natural logarithm, *Age* is the age in years and *SUV*max is the maximum uptake value on the PET. The accuracy of the model was good with an AUC of the ROC curve of 0.793 (95% CI, 0.676-0.911, $p < 0.001$), with *Se* and *Sp* of 87.5% and 46.7% respectively (Figure 5.2).

***Table 5.4*** Multivariate logistic regression analysis

| Factor | Regression coefficient | *P*-value | Odds ratio value | 95% CI | |
|---|---|---|---|---|---|
| | | | | Lower | Upper |
| **SUVmax** | 0.489 | 0.015 | 1.631 | 1.010 | 2.417 |
| **Age** | 0.052 | 0.100 | 1.053 | 0.988 | 1.123 |

CI, confidence interval; SUVmax, maximum standard uptake value.

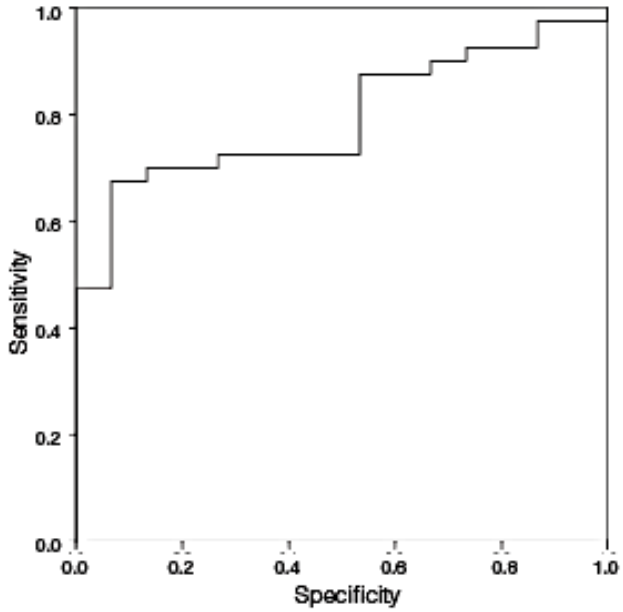



***Figure 5.2.*** ROC curve of clinical data of the patients using our predictive model. ROC, receiver operating characteristic.

## 5.4. Discussion

The diagnosis of SPN remains a major challenge in medical practice. Detecting and diagnosing SPN is critical, as early identification of malignant nodules improves the chance for successful treatment.

With regard to the FDG PET/CT imaging, some interpretation approaches have been assessed. Attending to visual assessment, a broad range of Se and Sp have been reported with values ranging from 69% to 100% and 63% to 85%, respectively [38, 81, 267, 271]. We found for the visual assessment (criterion I), a Se of 97.5%, which is in accordance with previously reported results, although the Sp (13.1%) was very limited, which is probably explained by the high prevalence of

malignancy in our sample of patients. It is well known that higher the prevalence is, lower the risk of false-positive results, and the prevalence will be higher as the inclusion of individuals in the screening program becomes more selective, focusing on higher clinical risk.

Abnormal $^{18}$F-FDG uptake is not specific for malignancy; some benign lesions such as bacterial pneumonia, active sarcoidosis, infectious granulomas, acute pyogenic abscesses, cryptogenic fibrosing alveolitis, and so forth have been known to produce false-positive readings on PET [272]. In our sample of patients, 27.3% of lesions were finally classified as benign, and from them, the most prevalent were organizing pneumonia (40%), fibrosis (20%), and granulomas (13.3%). The median SUVmax for the benign lesions were 2.29 ± 1.31, while that for malignant lesions were 4.54 ± 2.80 ($p < 0.001$).

In an attempt to improve the accuracy of the metabolic assessment some semiquantitative procedures have been developed. For instance, the uptake of the SPN (i.e., the glucose utilization) can be semiquantitatively assessed by the SUVmax and the uptake relative to the background activity in the uninvolved adjacent lung parenchyma and the mediastinum [273].

When we used a semiquantitative method, the *Sp* increased with a decrease in sensibility and accuracy. The criterion (II), using a SUV cut-off of 2.5 regardless of the nodule size, had a sensibility, accuracy, and *Sp* of 67.5%, 63.3%, and 53.3% respectively. However, these parameters have been reported to be higher. A meta-analysis reported pooled *Se* of 95% (95% CI, 0.93-0.98) and *Sp* of 82% (95% CI, 0.77-0.88) to malignant nodules [37]. Partial volume effect and motion during the scan acquisition affects the uptake values measurement, especially for lesions smaller than about three times the spatial resolutions of the equipment, so partial volume and motion corrections factors for standardized PET uptake values may significantly change the differential diagnosis of small pulmonary nodules [274]. In order to consider volume partial effect, we used two different approximations to consider a SPN as malign, (criterion III): a variable threshold of SUVmax depending on the SPN diameter, and (criterion IV): the value obtained by dividing the SUVmax between the diameter of the nodule. This approach is justified, because the SUVmax measure is affected by the nodule size, and although it is possible to use a recovery coefficient to have a more accurate measurement [275], we use the nodule size since it is proportional to the recovery coefficient. The respiratory movement reduces the Se to detect pulmonary lesions; however, the synchronized acquisition of PET with respiratory movement (4D PET) can reduce this inconvenient. When the 4D PET is used to evaluate faint pulmonary lesions there is an increase of SUVmax with respect to 3D

[258, 259]. Even when this modality of acquisition was not used in our patient group, we expect to apply it to develop future works.

We aimed to assess the diagnostic accuracy of $^{18}$F-FDG PET/CT, as well as to identify predictive factors of malignancy in SPN. With respect to the ROC analysis, the best cut-off for the SUVmax was 1.95 *vs.* 1.04 for the index SUVmax/diameter. Both values showed a *Se* of 80% and 82.5% respectively, with the same *Sp* (53.3%). There was an improvement in the diagnostic parameters, especially for the *Sp*. Our values of *Se* and *Sp* were similar to others published. For instance, Kim *et al*. [275] found that a SUVmax value of 2.5 had a *Se* and *Sp* of 89% and 51%, respectively, for all lesion sizes. Also, Grgic *et al*. [267] obtained a *Se* and *Sp* of 96% and 55%, respectively.

Age has been reported to be one important risk factor for SPN malignancy [80, 268]. In our study, we found a statistically significant association between age and malignancy, as has been described. However, it was not an independent predictor of malignity. Pulmonary nodules size is a very important predictor of malignancy, so it has to be measured as accurately as possible. To measure pulmonary nodules, the largest diameter, the mean diameter, or the volume can be used. The Fleischner Society states that the mean diameter is better for risk estimation [276]. Numerous studies have confirmed this finding, always associating lesion growth with its malignant potential. Nodules of more than 20 *mm* in diameter have a greater than 50% chance of being diagnosed as malignant [274, 277, 278]. However, we did not find a significant association between lesion diameter and malignancy. We believe that the nodule size of our patient sample strongly depended on the institutional dataset from which the investigated nodules have been collected. This sample consisted mainly of patients referred to our PET/CT service from thoracic surgeons and the nodules were sent to invasive diagnosis (surgery resection). Therefore, we had a very homogeneous group of pulmonary nodules with a diameter close to 20 mm, which represents a potential selection bias in this study, preventing reaching the statistical significance. Additionally, we performed only the measurement of the largest diameter of the node. We believe that the small size of the sample might also have influenced this result. Smoking has been found as an independent predictor of malignancy [80, 268]. In our population, the majority of patients (50 from 55) had a smoking history. Because of the low percentage of non-smokers, our population was biased. It might have influenced the results since we did not find a relation between smoking history with the SPN malignancy. An interesting fact is that even when the principal histological types related to smoking are squamous cell carcinoma and small cell carcinoma [279], we had a low prevalence of these histologic types.

The retrospective nature of the study and the selection criteria could affect our results especially the latter. The fact that all the included patients with $^{18}$F-FDG PET/CT were undergone surgery implied a high pre-test probability of malignancy that biases the PET/CT *Sp*. However, that warrantied the final histopathological confirmation of all lesions. Concerning our results, a significant statistical difference between the SUVmax and patient age with the final histology of SPN (benign and malignant) was found. This is in accordance with other studies [267, 274]. However, we found no statistically significant relation between malignancy and factors previously described as predictors of malignancy, such as smoking status, gender, and nodule diameter. Predictive models of SPN malignancy is of major interest to clinicians. We derived a model to predict the probability of malignancy by multivariate regression analysis and identified the SUVmax as the only independent predictor of malignancy of SPN. Our model had a *Se* and an *Sp* of 92.5% and 66.7%, respectively. Unlike other models [266, 268, 274], in which only clinical and morphological variables have been used, our model includes the SUVmax as a metabolic variable. The results obtained in this preliminary study allow us to conclude that the SUVmax is a good predictor of malignancy in an SPN and can be used in the diagnostic setting whenever available.

On the other hand, it will be necessary to develop new predictor models of malignancy based on clinical, morphological, and metabolic variables, and test their validity. The use of invasive diagnostic methods, such as fine-needle puncture, has risks to the patients, such as pneumothorax, bleeding, and dissemination of the tumor along the trajectory of the needle [280]. On the other hand, surgical lung biopsy has a mortality rate of around 0.6% [281]. An accurate, robust, and efficient predictive model for SPN malignity, could provide clinicians with reliable information to avoid the need for an invasive diagnostic method, allowing to limit the management of an SPN with safe clinical monitoring. Our predictive model of the SPN malignancy, unlike other models, used the metabolic variable SUVmax, showing that it is an independent variable to predict malignancy. The diagnostic performance of this model was higher than the visual and semiquantitative methodologies.

## 5.5. Conclusions

The assessment of SPN by semiquantitative methods did not improve the sensibility of visual analysis. The limited specificity was independent of the method used. However, the predictive model combining SUVmax and age was the best diagnostic approach, showing the SUVmax to be an independent variable to predict the malignancy of an SPN.

# CHAPTER

# 6

# Comparison of cross-combinations between feature selection and machine-learning classifier methods based on $^{18}$F-PET/CT radiomic features for prediction of the metabolic response in metastatic breast cancer

## Summary

The goal of this chapter is to identify an optimal combination between feature selection methods and machine learning classifiers based on $^{18}$F-PET/CT radiomic features, to predict metabolic response to the systemic treatment, in patients with recurrent or metastatic breast cancer. In this study, 48 patients with histologically confirmed recurrent or metastatic breast cancer, who received systemic treatment between 2010 and 2015 were enrolled. All patients had an $^{18}$F-FDG PET/CT before and after the systemic therapy administration. A total of 228 tumor lesions were identified in the pre-treatment PET/CT; from these 127 were classified as responders (complete or partial metabolic response) and 101 as non-responders (stable or progressive metabolic response), by using PERCIST criteria. For each lesion, 101 image features from PET and CT were extracted (a total of 202 features per lesion). These features along with clinical and pathological information were used to construct several prediction models of metabolic response by using several combinations of feature selection and classification methods. However, before building the models, the lesions were randomly divided into two groups with a ratio of 80:20. The bigger group was used to create the models and 6-fold cross-validation, and the other to validate. To this, seven feature selection methods: ANOVA with F-score, mutual information (MI), least absolute shrinkage and selection operator (LASSO), Wilcoxon test, hierarchical clustering (HC), principal component analysis (PCA), and independent component analysis (IPA); in cross-combination with other seven classification methods: support vector machines (SVM), random forest (RF), gaussian naive Bayes (GNB), logistic regression (LR), k-nearest Neighborhood (KNN), adaptative boosting (AdaBoost) and neural network (NN); were compared for their performance in predict the metabolic response to the treatment. the model performances were investigated via area under the receiver-operating characteristic curve (AUC) and accuracy (ACC) analysis. The validation cohort was used to validate our models in terms of AUC and ACC. The results show that the selection method LASSO + classifier SVM or RF, ICA + SVM had the highest AUC in the cross-validation, with 0.91±0.05, 0.90±0.02, 0.90±0.05 respectively. The selection method LASSO + classifier RF had the highest AUC and ACC in the validating set, with 0.83 and 0.80 respectively, followed by LASSO + KNN (AUC = 0.83, ACC = 0.71). MI + NB or AdaBoost, as well as Wilcoxon + NB or RF, had good performance with an AUC of 0.80. SVM had the best mean performance in the cross-validation and validation cohort (only accuracy). RF had the best mean of AUC in the validation cohort. In conclusion, this study showed that image features obtained from a pre-treatment $^{18}$F-FDG PET/CT could predict the metabolic response in recurrent or metastatic breast cancer, by their incorporation in a machine learning model (ML), which performance depends largely on the feature selection and ML classifier methods selected.

## 6.1. Introduction

Breast cancer is the most commonly diagnosed cancer and the second leading cause of death for cancer among women worldwide surpassed only by lung cancer [1, 282]. After the initial diagnosis, it is important to define accurately the initial extent of the disease, because it will affect the treatment election. For instance, patients with locally advanced stages could receive neoadjuvant chemotherapy, followed by a breast operation and adjuvant radiotherapy, and depending on the hormonal receptor status of the tumor, they could receive posteriorly an adjuvant targeted-therapy or hormonotherapy [283]. However, after the primary treatment, some percentage of patients will have a recidive, months, or years after the initial diagnosis and treatment. This recidive could be loco-regional or distant (metastatic disease). The metastatic disease is largely responsible for the majority of cancer patient deaths [284], and its treatment implies usually to use of systemic therapies such as chemotherapy, hormonal therapy if the cancer is hormone-receptor-positive, and targeted therapy [283, 285]. Usually, systemic therapies have several side effects, which should be avoided by considering changing or discontinue the therapeutic regimen. In this context, the prediction of response to treatment or depiction of treatment-resistant phenotypes is essential in clinical practice, especially in the new era of precision medicine [285, 286].

Currently, positron emission tomography/ computer tomography (PET/CT) is widely used in oncology, it is usually performed in breast cancer for recurrences detection or treatment response assessment, both in the context of neoadjuvant therapy or metastatic cancer treatment [94]. Changes in tumor metabolic activity, assessed by $^{18}$F-fluorodeoxyglucose positron emission tomography/computed tomography ($^{18}$F-FDG PET/CT), is an early indicator of treatment effectiveness for breast cancer, both neoadjuvant and metastatic setting [95-98]. Likewise, a significant reduction in the metabolic activity of the tumor (i.e. treatment-sensible tumors), early during therapy, is associated with longer overall survival and progression-free survival in this tumor [99-101]. Even when these metabolic changes have shown to be valuable to predict the treatment response in breast cancer and other malignant tumors; in the last years, there is an increasing interest in the clinical and prognostic utility of quantitative imaging analysis through radiomics [25, 286]. Radiomics refers to the extraction and analysis of quantitative imaging features from medical images, such as CT, PET, magnetic resonance (MR), and several others. It has been shown that these features reflect mechanisms occurring at genetic and molecular levels [58, 287]. From this point of view, radiomic features could find patterns in medical images, which could help to detected disease, understand the pathological process, or predict the medical evolution of patients. Specifically, in

oncology, radiomic features relating to tumor size, shape, voxel intensities, and texture, allow the tumor characterization, which has been called the radiomic signature of the tumor [58], and have shown their ability for diagnosing and predicting in several cancers [288, 289]. The radiomic process begins with the medical image acquisition, followed by the identification and segmentation of regions of interest (ROIs), which can include the whole tumor or only some parts of it. From these ROIs, several quantitative image features are extracted [290], which together with clinical variables, genomic profiles, serum markers, and/or histology data, can be used to build image-based biomarkers for diagnose, treatment response, recurrence prediction, or survival. Although the radiomic methodology has been applied to several malignant entities.

In breast cancer specifically, most of the radiomic studies have been carried out with MR images and in a neoadjuvant treatment context [102, 103, 291]. This could be justified by the fact that MRI of the breast is widely applied in screening of high-risk women, staging, evaluating treatment response, monitoring recurrence, and especially providing complementary information for uncertain findings on mammography and ultrasonography. Likewise, MRI has a higher resolution of soft tissues, different sequences of MRI could provide functional information of tumors, such as blood flow and breast density, and find the heterogeneity of tumor microenvironments [292-294]. However, some studies appearing more recently have explored the potential of radiomics with PET/CT, but none of them in patients with recurrent or metastatic breast cancer [104], which makes this study much more relevant, because the ability of PET/CT to early assess the treatment response.

To build image biomarkers from radiomic features, a classical statistical approach could be used. However, an increasing number of radiomic studies are based on machine learning (ML) classifiers, which have shown great promise [105]. This approach relies on a pipeline, including extraction of numerous handcrafted imaging features; followed by feature selection and machine learning-based classification. Feature selection or reduction variable methodologies should be carried before any ML model construction, because these can identify redundant and irrelevant imaging features, allowing to remove them from further analysis and improving the ML classifier performance [106]. To the best of our knowledge, no radiomic studies with ML have been carried out in recurrent or metastatic breast cancer. However, we hypothesized that ML models using radiomic features extracted can help to predict the metabolic response to the systemic treatment in this kind of patient. Therefore, this work proposes to use radiomic metrics extracted from $^{18}$F-FDG PET and CT to build

ML models and identify optimal feature selection and ML methods for recognizing precociously patients that might have a treatment-sensible or resistant tumor phenotype, which could help to select or adjust a particular therapy.

## 6.2. Methods

### 6.2.1 Patient Cohort

This was a retrospective per-lesion study. A chart review of 136 patients with a locoregional or distant recidive of breast cancer, who were diagnosed and treated at "Vienna General Hospital" (AKH-Vienna) between 2010 and 2015, was performed. Medical records, as well as pathologic and radiologic reports, were reviewed to identify a set of patients who met eligibility criteria: female gender, a biopsy-proved recurrent/metastatic breast cancer, to have received chemotherapeutic treatment and to have an $^{18}$F-PET/CT study, before and during/after completion of the full course of chemotherapy. Exclusion criteria included patients with incomplete information, very small tumor lesions, image artifacts, which would have precluded accurate texture analysis. Approval from the local institution was obtained. The collected histopathologic parameters for the primary and metastatic tumor were: histologic type and grade, status for estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor (HER2), as well as Ki-67 proliferation index. All pathology and histopathology analyses were performed using standard procedures in AKH. Besides, clinical variables at initial diagnostic, such as age, breast side affectation, and TNM classification [32] were recorded. Clinical and demographic information of patients can be found in Table 6.1. All patients received chemotherapy or target therapy. The treatment protocol and schedule followed the standard oncology treatment and procedures in AKH-Vienna.

### 6.2.2 PET/CT image acquisition

$^{18}$F-FDG PET-CT images were performed according to the standard PET/CT image protocol of the nuclear medicine division of AKH-Vienna. A whole-body $^{18}$F-PET/CT from mid cranium to the upper thigh was performed using a 64-row multi-detector PET/CT system (Biograph TruePoint 64; Siemens, Erlangen, Germany) with an axial field-of-view of 216 *mm*, a PET sensitivity of 7.*6 cps/kBq*, and a transaxial PET resolution of 4–*5 mm* (full-width at half-maximum, FWHM). Finally, most of the images had a voxel size of 4.07 × 4.07 ×3.00 $mm^3$ or 0.7 × 0.7 × 2.*0* $mm^3$ for PET and CT respectively.

***Table 6.1.*** Demographic and clinical information of patients at initial diagnose

| **Characteristic** | **n(%)** |
|---|---|
| **Total patients** | 48 |
| **Mean age (years)** | 48.1 |
| **Affected side** | |
| right | 26 (54.2%) |
| left | 22 (45.8.2%) |
| **Histologic type** | |
| ductal | 42 (87.5%) |
| lobular | 5 (10.4%) |
| other | 1 (2.1%) |
| **Tumor size[1] (T)** | |
| T1a-b | 12 (25%) |
| T1c | 15 (31.3%) |
| T2 | 11 (22.9%) |
| T3 | 5 (10.4%) |
| **Nodal affectation[1] (N)** | |
| N0 | 14 (29.2%) |
| N1 | 22 (45.8%) |
| N2a-b | 4 (8.3%) |
| N3a | 2 (4.2%) |
| N3b | 1 (2.1%) |
| **Mestatase[1] (M)** | |
| M0 | 20 (39.6%) |
| M1 | 1 (2.1%) |
| Mx | 22 (43.8%) |
| **Estrogen receptor positivity** | |
| negative | 17 (54.2%) |
| low | 4 (54.2%) |
| moderate | 11 (54.2%) |
| strong | 16 (54.2%) |
| **Progesterone receptor positivity** | |
| negative | 24 (50%) |
| low | 8 (16.7%) |
| moderate | 7 (14.62%) |
| strong | 9 (18.8%) |
| **Her2-new[2] positivity** | |
| 0 | 33 (68.8%) |
| 1 | 15 (31.3%) |
| **Histologic Grade[3]** | |
| well | 1 (2.1%) |
| moderate | 20 (41.7%) |
| poor | 26 (54.2%) |

[1]For 5 patients were not available information; [2]human epidermal growth factor receptor 2;

[3]For 1 patient was not available information

### 6.2.3 ROI Delineation

For each patient, one or several tumor volumes of interest (VOI) were included. The delineation process was performed using the Hermes Hybrid 3D software, version 2.0 (Hermes Medical Solutions, Stockholm, Sweden). First, a cuboid volume of interest (VOI, 5 × 5 × 5 voxels) was defined in the background area of each PET image. Then, the tumor lesions were delineated using a semiautomatic region-growing tool to generate corresponding PET-VOIs. By using a mask on the CT component, CT-VOIs were also obtained. To perform a posterior analysis, all VOIs were saved as a comma-separated values (CSV) file, each one with spatial coordinates and intensity, as well as several metabolic parameters and other basic image information.

### 6.2.4 Image Preprocessing

PET image intensity was converted to SUVs. To remove individual acquisition differences, and in-line with other radiomic studies [189, 190], the images were normalized and resampled into a voxel size of $1 \times 1 \times 1$ mm$^3$. Further image preprocessing was not performed. See section 2.3.2 of this thesis for a better description of the resampling process.

### 6.2.5 Metabolic parameters extraction

From the VOIs of each lesion, metabolic parameters of $^{18}$F-FDG uptake such as maximum, mean, minimum, and peak standardized uptake value (SUVmax), (SUVmean), (SUVmin) and (SUVpeak) respectively, as well as metabolic tumor volume (MTV), total lesion glycolysis (TLG) and SUL (lean body mass corrected SUV) peak were obtained.

### 6.2.6 PET/CT response assessment

PET/CT scans were reviewed for a trained nuclear medicine physician. FDG-avid target lesions were identified in each patient on initial PET and were followed on follow-up PET. All target lesions on initial PET were confirmed to be metastatic by morphological imaging. Semiquantitative analysis was performed by using the maximum standardized uptake value (SUVmax) calculated for each target lesion. The changes in SUVmax between initial and follow-up PET were recorded for the highest SUVmax targets. The percentage of change in SULpeak (defined as the average SUV within a 1 cm$^3$ spheric ROI, centering around the hottest point in the tumor, and corrected by lean body mass

of the patient) allowed the patient´s classification according to PET response criteria in solid tumors (PERCIST) criteria [42, 295] into the following groups:

- Complete metabolic response (CMR)
- Partial metabolic response (PMR)
- Stable metabolic disease (SMD)
- Progressive metabolic disease (PMD)

Then patients with CMR and PMR were considered as responders and patients with SMD and PMD as no-responders. For some patients, the SULpeak was not available, in these cases a subjective and SUVmax based response was obtained.

### 6.2.7 Radiomic features extraction

From the PET and CT delineated lesions radiomic features radiomic features were extracted. The features were extracted in the platform MATLAB by adapting an open-source radiomic analysis package with its several MATLAB functions [189]. This package follows the definitions of features from the Imaging Biomarker Standardization Initiative (IBSI) [192]. The radiomic features are summarized in ***Table B1 of Appendix B.***

#### 6.2.7.1 Texture Features

A total of 101 textural features were extracted for each contoured tumor on $^{18}$F-FDG-PET and CT images, respectively, i.e. 202 textural features in total. From these 110 features, 13 were intensity histogram features and 88 textural features ( of which 31 explore intratumoral heterogeneity). These features were then used to construct an ML model to classify the desired binary groups (responders versus non-responders) [27, 296-298]. A general vision of the radiomic procedure implemented in this work can be appreciated in Figure 6.1.

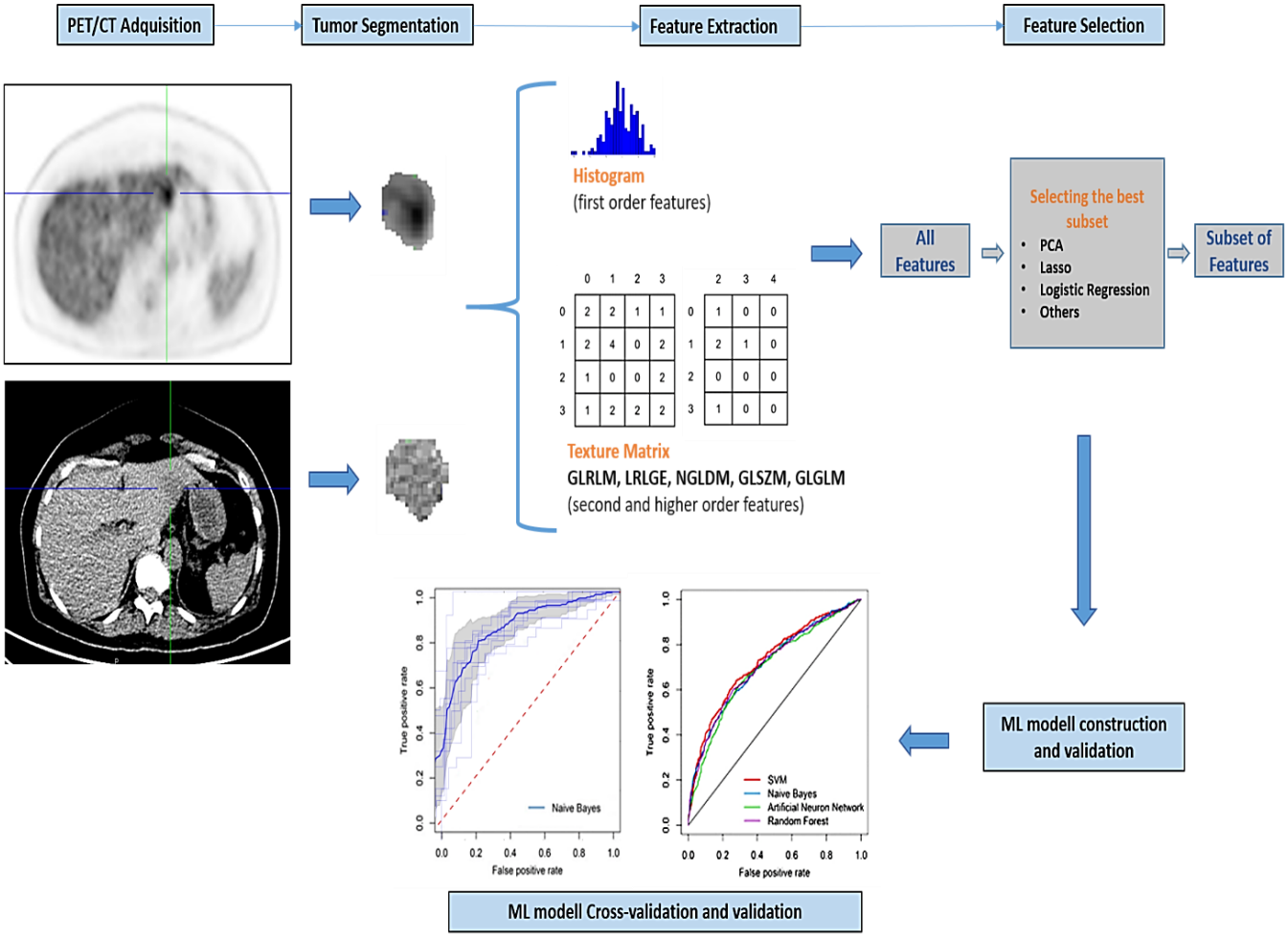



***Figure 6.1.*** Radiomic pipeline process

### 6.2.8 Univariate statistical analysis

Univariate analysis was performed to investigate associations of single features with the outcome (metabolic response). Firstly, all features were normalized via Z-score normalization to zero mean and unit variance. Then, for each clinical variable or image feature, a chi-squared or Mann-Whitney U statistic test was calculated. The significance level was set a $p < 0.005$. A Bonferroni test was applied to avoid overestimation due to multiple testing. Also, the Spearman correlation test was performed for each pair of image features or clinical variables. This statistical analysis was performed by using IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp [35].

### 6.2.9 Machine Learning Model

#### 6.2.9.1 Feature selection

Not all 202 features were included in the machine learning models implemented later in this article. Radiomics studies have hundreds of features, many of which are highly correlated with one another; this makes necessary feature selection methods to avoid collinearity, reduce dimensionality, minimize noise and so to minimize overfitting problems [106, 217, 299, 300]. Hence initially a data-preprocessing methodology was implemented to reduce the large set of features to a subset of the most significant features. A pairwise Spearman correlation matrix was used to identify pairs of highly correlated features ($|r| \geq 0.90$). Finally, from each pair, only those with the best association to the target variable (responders or not) were retained. This association was measured through a chi-squared or Mann-Whitney test between each variable and the response target variable. After this data-preprocessing step, seven popular feature selection methods were used to further reduce the number of features: ANOVA F-score, mutual information (MI), least absolute shrinkage and selection operator (LASSO), Wilcoxon test (WT), hierarchical clustering (HC), principal component analysis (PCA) and independent component analysis (IPA). These methods were chosen because of their popularity in several publications about radiomics and machine learning [301-304].

#### 6.2.9.2 Classification methods

In order to classify tumor lesions into responders and non-responder (0 or 1), we investigated seven popular machine-learning classifiers: support vector machines (SVM), Random Forest (RF), Gaussian Naives-Bayes (GNB), logistic regression (LR), k-nearest neighborhood (KNN), adaptative boosting (AdaBoost) and neural network (NN) [305, 306]. The acronym for each feature selection method and ML classification method are listed in Table 6.2. The feature selection and classification methods were implemented by using SciKit Learn package in python (scikit-learn version 0.21, python version 3.6.3), and using the open platform Google Colaboratory [220]. Each of the seven feature selection methods was combined with all the seven classification methods, and each classification method was combined with all the seven feature selection methods, yielding 48 cross-combinations of evaluated methods.

***Table 6.2.*** Feature selection and classification methods

| Feature selection method | ML Classifier |
|---|---|
| **AFV** (ANOVA F-value) | **SVM** (support vector machine) |
| **MI** (mutual information) | **GNB** (Gaussian naive Bayes) |
| **PCA** (principal component analysis) | **RF** (random forest) |
| **ICA** (independent component analysis) | **LR** (logistic regression) |
| **Lasso** (least absolute shrinkage and selection operator | **KNN** (k-nearest neighborhood) |
| **CL** (clustering) | **AdaBoost** (adaptive boosting) |
| **WT** (Wilcoxon test) | **NN** (neural network) |

### 6.2.9.3 Model construction

To create the radiomic-based model, we followed the next steps, which are recommended to perform a suitable model [225]:

1. Data imputation, by filling the empty data with the most frequent strategy.

2. Splitting of the data (80:20) into *X_train, X_test, y_train, y_test*, where *X* and *y* are predictive features (clinical and radiomic features) and target variable (responders or non-responders) respectively. Only the training set was used to construct the models, and the test set for validation purposes.

3. A synthetic minority over-sampling technique (SMOTE) [307] was performed for over-sampling the non-responder to have the same number of instances as the responder in the training procedure. Initially, the training dataset was unbalanced, with the non-responder and responder groups with 101 and 81 samples respectively, which might skew the model performance.

4. Data standardization, all variables are obligated to have mean zero and standard deviation of one.

5. Feature selection methods as the Wilcoxon test and hierarchical clustering were applied directly after the data-preprocessing methodology to obtain a smaller number of features.

However, ANOVA F-score, MI, LASSO, PCA, and IPA were accoupled to the seven different classifier methods, then an iterative process was implemented to find a group of features with the best performance in terms of ACC and AUC. Therefore, curves of the number of features selected versus model performance were obtained.

6. Hyperparameters tuning. The model performance in most machine learning algorithms are dependent upon the choice of various tuning parameters and selected feature number. Also, some tuning parameters take into account the number of predictors after feature selection. We performed parameter tuning for SVM, RF, KNN through cross-validation, and by using the class *GridSearchCV* of SciKit Learn. For GNB, LR, AdaBoost, and NN the default hyperparameter setting was used.

7. Finally, the 49 cross-combinations (each one with a specific subset of features, and an ML classifier with specific hyperparameters) were trained by using the training cohort.

#### 6.2.9.4 Model Performance Metrics and validation

The model performances were assessed with ACC and AUC metrics. A 6-fold cross-validation was done in the training group; it splits the data into six equal parts and used 5 parts for training and the rest for testing. The feature selection methods are included in the cross-validation algorithm so that their contribution to the final model is reflected in the performance metrics. Posteriorly, model validation was performed by applying the trained models to the test group. Receiver operating characteristic (ROC) curves were produced for each model.

## 6.3. Results

### 6.3.1 Clinical characteristics

Finally, forty-eight patients were identified to have a biopsy-proven recidive as well as available pretreatment and follow-up $^{18}$F-FDG PET-CT. A total of 228 tumor lesions were visualized on the pretreatment PET/CT and follow-up on the subsequent PET/CT. Patient and tumor characteristics are summarized in Table 6.1. The mean time elapsed between the initial and response PET/CT was of 149 days. A description of the treatment of each patient and places of affectation is giving in ***Table B2 of Appendix B***.

Seventy-two lesions showed a complete metabolic response, 55 partial response, 7 stable disease, and 94 progression. By considering as responder those lesions with CMR or PMR, and as no-responder lesions with SMD or PMD; 127 were considered as responders and 101 as no-responders, respectively.

### 6.3.2 Feature extraction and correlation

A total of 202 radiomic features (101 for each imaging modality), as well as 20 clinical and metabolic variables, were obtained and investigated in terms of their association with the metabolic response of the patients. The data-preprocessing filter removed 116 highly correlated features, leaving a set of only 106 predictors. Figure 6.2 shows the heatmap of the feature and clinical variable correlations, before and after the pre-processing filter application. From this, it is appreciated that clinical variables have low Spearman correlation between them. The results of the univariate analysis are presented in ***Table B3 of Appendix B***. However, clinical variables as tumor size, estrogen and progesterone receptor positivity, human epidermal growth factor receptor 2, and tumor grade, as well as some metabolic variables as SUVpeak, SUVmean, SUVmax had a statistically significant association with the target variable (metabolic response).

### 6.3.3 Feature reduction

After applying feature reduction with HC and WT to the original 222 features and clinical variables, only 58 and 106 parameters were maintained respectively. The relationship between selected feature number and performance of the models, for each combination between ANOVA F-score, MI, LASSO, PCA, and ML classifiers. More specifically, for each combination we iteratively incrementally the number of selected features or components (for PCA and ICA) that finally fed the ML classifier, which is subsequently trained and its performance assessed in each step through cross-validation. Figure 6.3 shows an example of how the classification performance for LASSO + SVM changes according to the number of selected features. Maximal AUC and ACC are found with 22 and 25 features respectively. Besides, hyperparameters for SVM, RF, and KNN were tuned during the iterative cross-validation. For LR, GNB, AdaBoost, and NN, the defect parameters of SciKit Learn were used.

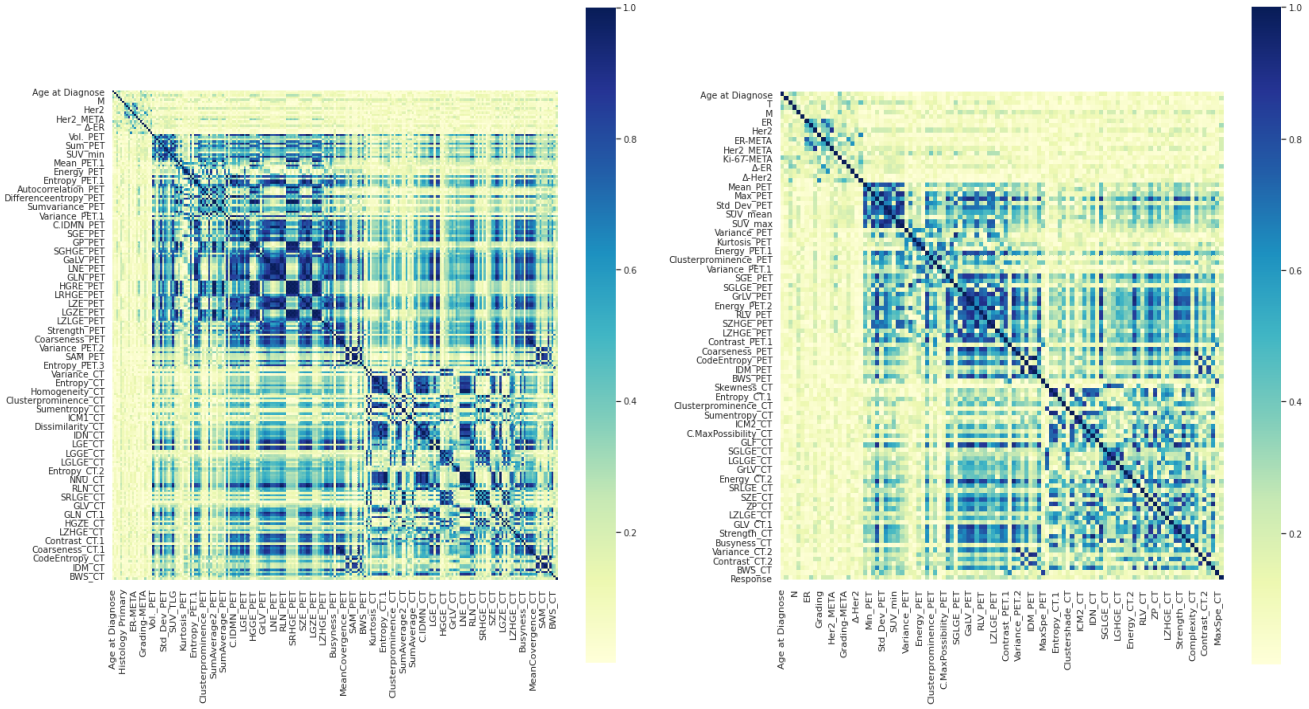


***Figure 6.2.*** Spearman correlation heatmap for the feature and clinical parameters; left: before preprocessing and right: after pre-processing

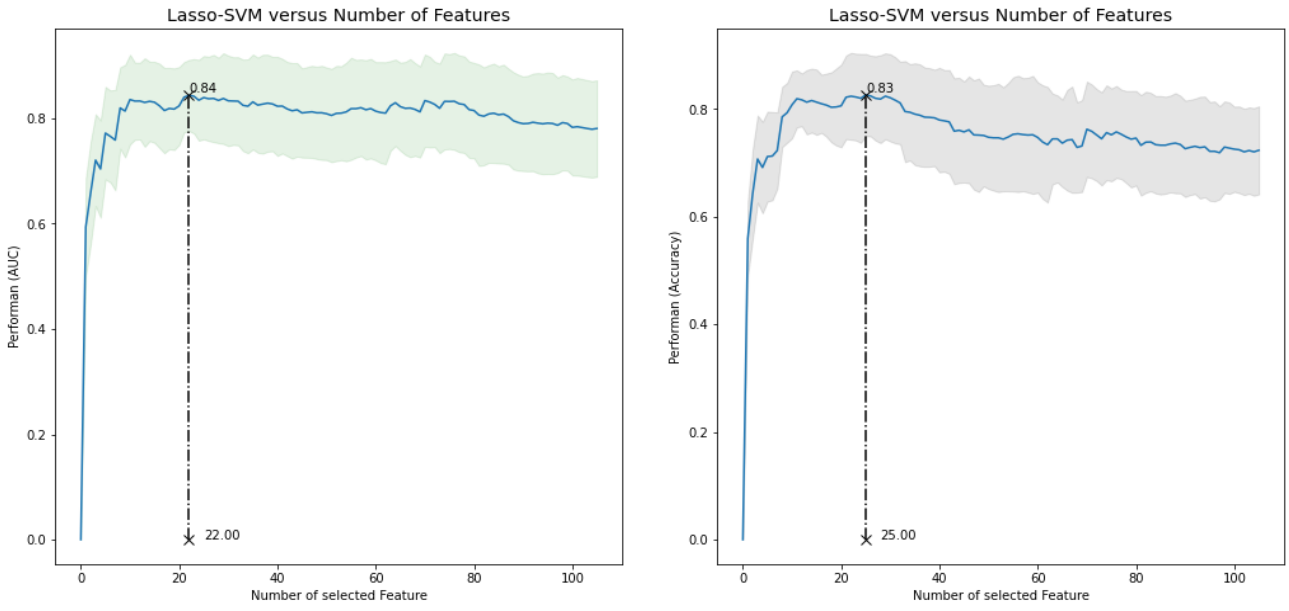



***Figure 6.3.*** Number of selected features and ML prediction performance (ACC and AUC)

### 6.3.4 Performance of feature selection methods and classifiers

The diagnostic performance of feature selection and classification methods were evaluated by repeated 6-fold cross-validation in the training group and validation in the test group. In this study, the diagnostic performance was quantified by the area under the curve (AUC) and accuracy (ACC). Additionally, the calibration of the best model found was assessed by comparing the probabilities as observed in the data and the model-based expected. It can be found in ***Figure B1 of Appendix B.***

### 6.3.5. Cross-validation

We examined 49 combinations of feature selection and classification methods. Table 6.2 reports the performance values, in terms of mean AUC and its standard deviation for the 6-fold cross-validation, for each pair feature selection (in rows) and ML classifier method (in columns). The combination of LASSO + SVM had the highest AUC of *0.91 ± 0.05*; follow by Lasso + RF and ICA + SVM, with *0.9 ± 0.02* and *0.9 ± 0.5* respectively. On average, both the GNB classifier and the F-Score feature selection method had each one individually, the lower performance.

***Table 6.3.*** Model performances in the cross-validation (AUC ± SD)

| | | **Classifier** | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **Model** | **SVM** | **Naive Bayes** | **RF** | **LR** | **KNN** | **AdaBoost** | **NN** |
| **Feature selection** | **F-Score** | 0.84 ±0.06 | 0.71±0.09 | 0.85±0.07 | 0.75±0.05 | 0.79±0.07 | 0.71±0.12 | 0.85±0.08 |
| | **Mutual information** | 0.85±0.08 | 0.68±0.08 | 0.86±0.06 | 0.74±0.06 | 0.82±0.08 | 0.80±0.07 | 0.76±0.07 |
| | **PCA** | 0.86±0.05 | 0.76±0.14 | 0.85±0.06 | 0.73±0.06 | 0.79±0.07 | 0.71±0.11 | 0.85±0.08 |
| | **ICA** | 0.90±0.05 | 0.70±0.07 | 0.85±0.07 | 0.77±0.06 | 0.86±0.04 | 0.77±0.05 | 0.74±0.07 |
| | **Lasso** | 0.91±0.05 | 0.80±0.06 | 0.90±0.02 | 0.74±0.04 | 0.89±0.04 | 0.68±0.05 | 0.73±0.10 |
| | **Clustering** | 0.80±0.07 | 0.65±0.07 | 0.80±0.05 | 0.76±0.06 | 0.68±0.07 | 0.74±0.07 | 0.85±0.06 |
| | **Wilcoxon** | 0.86±0.06 | 0.70±0.08 | 0.87±0.06 | 0.72±0.5 | 0.78±0.07 | 0.75±0.14 | 0.845±0.04 |
| **Mean** | | 0,86±0.06 | 0,71±0.08 | 0,85±0.06 | 0,74±0.12 | 0,80±0.06 | 0,74±0.09 | 0,80±0.07 |

The ROC curves of the 6-fold cross-validation for LASSO + SVM and Lasso + RF are plotted in Figure 6.4.

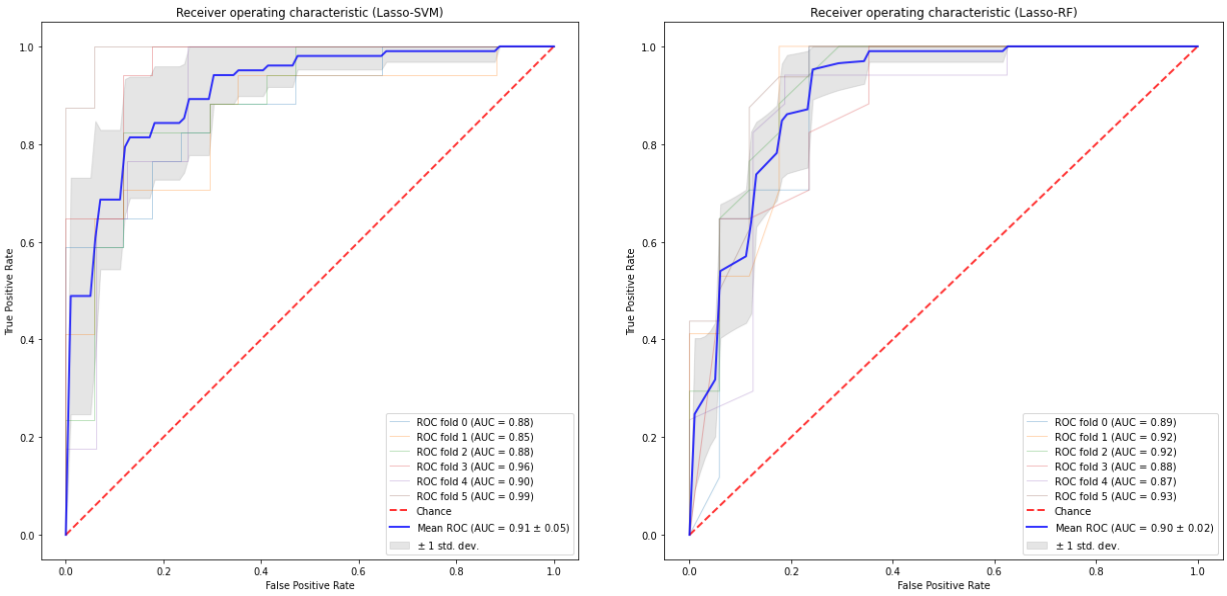



***Figure 6.4.*** ROC curves for cross-validation of Lasso + SVM and RF

### 6.3.6. Prediction performance (validation)

Tables 6.3 and 6.4 show respectively the AUC and ACC for the pair feature selection (in rows) and ML classifier method (in columns). Regarding AUC, the highest predictive performance was gotten by selection method LASSO + classifier RF or KNN with 0.83 each one. On average, any feature selection methods + KNN had the worst performance. When taking into account the ACC, the highest predictive performance was got by LASSO + KNN with 0.8, followed by Clustering + GNB with 0.74. The ROC curve and confusion matrix for the model LASSO + RF are showed in Figure 6.5.

## 6.4. Discussion

Currently, several therapeutic alternatives are available to treat metastatic breast cancer [283, 285]. But the existence of multiple possibilities also requires a judicious assessment of the response to the treatment administered, to avoid unnecessary side effects, especially when it is not working adequately, allowing so an early change to other potentially better therapeutic options. Even better

would be to have a method available to perform a pre-therapeutic prediction of the tumor response to a specific treatment. In this study, we demonstrated that such prediction is possible by using radiomic and machine learning. It is tempting to think of a method that allows predicting the probability of tumor response to each of the available treatments in patients with recurrent or metastatic breast cancer. This will allow starting treatment with that therapeutic regimen with a greater probability of success and fewer adverse effects.

***Table 6.4.*** Model performances in the validation (AUC)

| | | **Classifier** | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **Model** | **SVM** | **Naive Bayes** | **RF** | **LR** | **KNN** | **AdaBoost** | **NN** |
| **Feature selection** | **F-Score** | 0.67 | 0.77 | 0.72 | 0.70 | 0.73 | 0.73 | 0.76 |
| | **Mutual information** | 0.74 | 0.80 | 0.79 | 0.66 | 0.78 | 0.80 | 0.78 |
| | **PCA** | 0.74 | 0.59 | 0.72 | 0.71 | 0.66 | 0.67 | 0.77 |
| | **ICA** | 0.78 | 0.78 | 0.77 | 0.73 | 0.71 | 0.59 | 0.61 |
| | **Lasso** | 0.75 | 0.73 | 0.83 | 0.66 | 0.83 | 0.64 | 0.65 |
| | **Clustering** | 0.74 | 0.78 | 0.74 | 0.79 | 0.63 | 0.75 | 0.72 |
| | **Wilcoxon** | 0.72 | 0.80 | 0.80 | 0.76 | 0.68 | 0.70 | 0.70 |
| **Mean** | | 0,73 | 0,75 | 0,77 | 0,72 | 0,72 | 0,70 | 0,71 |

***Table 6.5.*** Model performances in the validation (ACC)

| | | **Classifier** | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **Model** | **SVM** | **Naive Bayes** | **RF** | **LR** | **KNN** | **AdaBoost** | **NN** |
| **Feature selection** | **F-Score** | 0.67 | 0.65 | 0.65 | 0.65 | 0.65 | 0.74 | 0.65 |
| | **Mutual information** | 0.72 | 0.59 | 0.74 | 0.67 | 0.72 | 0.70 | 0.76 |
| | **PCA** | 0.72 | 0.61 | 0.67 | 0.65 | 0.54 | 0.65 | 0.71 |
| | **ICA** | 0.76 | 0.70 | 0.71 | 0.76 | 0.65 | 0.61 | 0.58 |
| | **Lasso** | 0.67 | 0.70 | 0.71 | 0.65 | 0.80 | 0.63 | 0.67 |
| | **Clustering** | 0.72 | 0.74 | 0.63 | 0.70 | 0.61 | 0.72 | 0.71 |
| | **Wilcoxon** | 0.70 | 0.72 | 0.74 | 0.67 | 0.61 | 0.70 | 0.58 |
| **Mean** | | 0,71 | 0,67 | 0,69 | 0,68 | 0,65 | 0,68 | 0,67 |

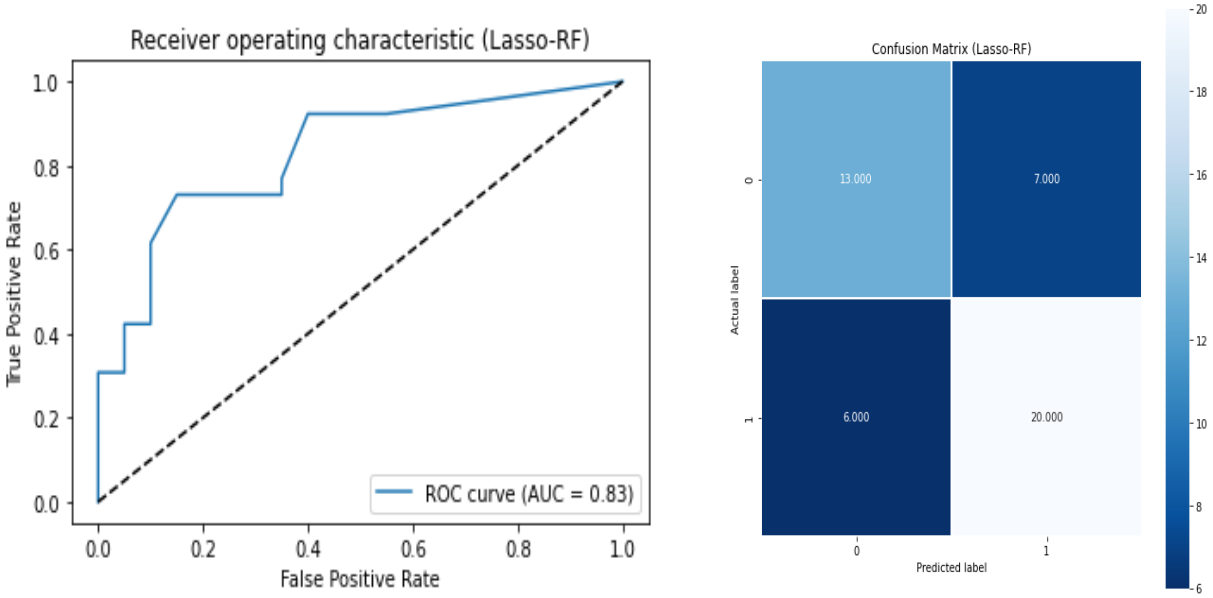



***Figure 6.5.*** Validation ROC curve and confusion matrix for model Lasso-RF

The emerging field of radiomics quantifies the phenotypic characteristics of tumor tissues on medical image features. Since $^{18}$F-FDG PET/CT is a valuable image method in oncology, and commonly used to assess the tumor response to the treatment in breast cancer, we investigated the ability of radiomic features of $^{18}$F-FDG PET/CT along ML algorithms to predict the metabolic response of tumor lesions to the systemic treatment in patients with recurrent or metastatic breast cancer.

To this aim, we developed and validated 49 predictive models, each one with different combinations of feature selection and ML methods. The most relevant set of features of each selection method, as well as the best hyperparameters of each ML classifier, were used for each particular model. Finally, the model performances were assessed by 6-fold cross-validation and validation in the testing group, by using AUC and ACC as metrics. This type of study tries to find an optimal configuration of feature selection and ML method, for a specific clinical setting.

The combination with the highest performance was the LASSO features selection method + RF as an ML classifier. For this combination, the AUC in the cross-validation was 0.91 ± 0.05, while in the validation; AUC and ACC were 0.83 and 0.71 respectively. However, other combinations also showed a good performance, so that they should not be simply discarded. These results show that a radiomic approach, by using ML models, might be able to predict the tumor metabolic response to

the systemic treatment in patients with recurrent or metastatic breast cancer. Most of the PET/CT studies for prediction of the treatment response in breast cancer not include radiomic analysis, and the have been performed in a neoadjuvant context [95, 308, 309]; only a few studies consider the treatment response in patients with metastatic cancer [99, 100, 310]. On the other hand, concerning the existing PET/CT radiomic studies in breast cancer, they have only considered a neoadjuvant context [23, 208, 311, 312]. To our best knowledge, there are not radiomic studies to predict the treatment response in metastatic breast cancer.

The radiomic and ML methodologies are influenced by the feature selection method as well as the ML algorithm used. Because, different combinations have different performances, and it depends possibly on the tumor and clinical setting, some authors have recommended performing this kind of study for each tumor and clinical context [225]. The identification of optimal ML methods for radiomic applications is a crucial step towards stable and clinically relevant radiomic biomarkers construction. We consider the importance of our study because of the lack of an ML-based radiomic approach to the assessment of metabolic response in patients with metastatic cancer.

In other pathologies and clinical contexts, several combinations of selection methods and ML classifiers are suitable for classification or prognostic purposes. For example, Dongyanh *et al.* [303] found the cross-combination fisher score (FSCR) plus KNN, SVM, or RF as suitable for differentiation between recurrence and inflammation (AUC 0.883, 0.867, and 0.892 respectively) by using PET/CT images of patients treated of nasopharyngeal carcinoma. In addition to FSCR, they used other feature selection methods as mutual information maximization, Fischer score, Relief-F, conditional mutual information maximization, Minimum redundancy maximum relevance, and Joint mutual information. Parmar *et al*. [306] investigated fourteen feature selection and twelve classification methods in terms of their performance for predicting overall survival in patients with non-small cell lung cancer (NSCLC). They used CT images and founded that the Wilcoxon test-based feature selection method and RF classification had the highest performance (AUC of $0.65 \pm 0.02$ and $0.66 \pm 0.03$ respectively). Yin *et al.* [302] aimed to identify optimal machine-learning methods for preoperative differentiation of sacral chordoma and sacral giant cell tumors based on 3D non-enhanced computed tomography (CT) and CT-enhanced (CTE) features. The selection method LASSO + classifier generalized linear models (GLM) had the highest AUC of 0.984 and ACC of 0.897 in the validating set. Three selection methods were used: Relief, LASSO, and RF. Moreover, SVM, GLM, and RF performed the classification.

There are some limitations to this study. Firstly, it is a retrospective study with a small cohort and heterogeneous group of patients, regarding clinical and pathological features as well as administered treatment, which was obtained from one institution. A prospective multicenter study with a larger cohort is necessary to confirm our results and improve the reliability and clinical application of this radiomic study. To improve the reproducibility and generalizability of this study, we used IBSI-based standardized radiomic features, which were normalized with the z-score method. We only compared seven commonly used feature selection methods and seven classification methods regarding their performance to predict metabolic response in patients with recurrent or metastatic breast cancer. Since there are many other methods, and therefore possible combinations, we cannot be sure that we have found the most suitable combination. Hyperparameters cannot be learned by the algorithm directly during the training, and rather they must be set before the training starts. In this study, four ML methods (LR, GNB, AdaBoost, and NN) were used with their default settings, whereas a hyperparameter tuning was performed for SVM, RF, and KNN, which might have resulted in enhanced performance of these last three methods.

The patients included in this work received different chemotherapy regimens, according to their clinical and pathological characteristics. This implies that there are many variables involved, depending on the patient, tumor, as well as of administered treatment. All of this affects the final metabolic response of the metastatic lesion. Despite everything, we have been able to predict the metabolic response to the treatment in these metastatic lesions, with an important AUC and ACC. We believe that, by recruiting a more homogeneous group of patients, with similar tumor biology and receiving more homogeneous therapeutic regimes, the performance of a radiomic ML-based model would be improved. On the other hand, by doing a long-term follow-up of the patients, such models could be expanded to predict outcomes as time-free of disease and survival.

## 6.5. Conclusion

In conclusion, we constructed models based on radiomic $^{18}$F-FDG PET/CT features and ML classifiers to predict the metabolic response to systemic therapy in patients with recurrent or metastatic breast cancer. The selection method LASSO + classifier RF had the highest performance in the validation cohort. Although, other combinations also showed high diagnostic performance. This comparative investigation may be an important reference in identifying reliable and effective machine-learning methods for radiomic-based prognostication in these kinds of patients.

# Conclusions of this thesis

The main contributions and conclusions of the thesis are summarized in this section.

## Lung cancer

In this thesis, we have used metabolic and textural features to investigate their clinical relevance in patients with lung and breast cancer.

We found that the tumor heterogeneity in NSCLC assessed by texture analysis of the PET component of $^{18}$F-FDG PET/CT images is correlated with metabolic parameters, and both are associated with macroscopic tumor diameter and, under special considerations AJCC-TNM stage.

Textural features, related to the heterogeneity such as (energy, entropy, contrast, homogeneity, and correlation, and metabolic parameters such as SUVmax, SUVmean, metabolic tumor volume (MTV), and total lesion glycolysis (TLG), as well as pathologic staging, histologic subtype, and tumor diameter, were correlated.

There was a positive relationship between all metabolic parameters (SUVmax, SUVmean, MTV, and TLG) with entropy, correlation, and homogeneity and a negative relationship with energy and contrast. The T component of the pathologic TNM staging (pT) was similarly correlated with these textural parameters. Textural features associated with tumor heterogeneity were shown to be related to global metabolic parameters and pathologic staging.

Therefore, SUV and other measurements of FDG tumor uptake can potentially be supplemented by additional imaging parameters derived from either the PET or the CT images. It is expected that such imaging parameters can be more suitable to guide and predict therapy response, as well as survival. For instance, if we found a lung tumor with high entropy and low energy in a staging PET/CT, it could need more aggressive therapy from the beginning than others with low entropy and high energy, but this requires additional work to be validated. However, some textural features have no linear relationship with volume-based metabolic parameters, making them more sensitive to tumor volume definition and defaulting to establish relationships with AJCC staging. Then, additional work is necessary to get to establish those image features with potential usability in the clinical practice.

## Solitary pulmonary nodules (SPN)

We investigated the diagnostic performance of different metabolic, morphologic, and clinical criteria for the correct presurgical classification of solitary pulmonary nodules (SPN). The assessment of SPN by semiquantitative methods did not improve the sensibility of visual analysis. However, a predictive logistic model combining SUVmax and age had the best diagnostic performance.

The different criteria of classification: (I) visually detectable metabolism, (II) SUVmax >2.5 regardless of SPN diameter, (III) SUVmax threshold depending of SPN diameter, (IV) ratio SUVmax/diameter greater than , had a sensibility/ specificity of (I): 97.5% and 13.1%; (II) 67.5% and 53.3%; (III) 70% and 53.3%; and (IV) 85% and 33.3%, respectively. However, the logistic predictive model had 87.5% and 46.7%. The results obtained in this study allow us to conclude that the SUVmax is a good predictor of malignancy in an SPN and can be used in the diagnostic setting whenever available. However, the specificity is improved by a logistic predictive. This gain in specificity is of great clinical importance, because the use of invasive diagnostic methods, such as fine-needle puncture, pretending an accurate diagnosis, has risks to the patients, such as pneumothorax, bleeding, and dissemination of the tumor along the trajectory of the needle. On the other hand, surgical lung biopsy has a mortality rate of around 0.6%. An accurate, robust, and efficient predictive model for SPN malignity, could provide clinicians with reliable information to avoid the need for an invasive diagnostic method, allowing to limit the management of an SPN with safe clinical monitoring.

## Breast cancer

The prognostic performance ML models, based on radiomic features, to predict the response the systemic therapy in patients with recurrent or metastatic breast The LASSO selection method + classifier RF had the highest performance. With an AUC of 0.90±0.02 in the cross-validation and 83% in the validation cohort. Although, other combinations also showed high diagnostic performance. This comparative investigation may be an important reference in identifying reliable and effective machine-learning methods for radiomic-based prognostication in these kinds of patients. 0.90±0.02 in the cross-validation with an AUC of 83% in the validation. The patients included in this work received different chemotherapy regimens, according to their clinical and pathological characteristics. This implies that there are many variables involved, depending on the patient, tumor, as well as of administered treatment. All of this affects the final metabolic response of the metastatic

lesion. Despite everything, we have been able to predict the metabolic response to the treatment in these metastatic lesions, with an important AUC and ACC. We believe that, by recruiting a more homogeneous group of patients, with similar tumor biology and receiving more homogeneous therapeutic regimes, the performance of a radiomic ML-based model would be improved.

An important contribution of this work is its novelty in predicting the response to systemic therapy in patients with breast cancer. Most of the PET/CT studies for prediction of the treatment response in breast cancer not include radiomic analysis, and they have been performed only in a neoadjuvant context. On the other hand, very few numbers of studies consider the treatment response in patients with metastatic cancer, by using only metabolic parameters. To our best knowledge, there are not radiomic studies to predict the treatment response in metastatic breast cancer. A systematic review of the use of radiomic in breast cancer has recently been published, but nothing has appeared regarding the use of it in patients with recurrent or metastatic breast cancer [208].

We are currently developing a study on the same group of patients to determine whether radiomic characteristics extracted from the sub-volume of the tumor are better predictors of response to therapy.

# Bibliography

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [published correction appears in CA Cancer J Clin. 2020 Jul;70(4):313]. CA Cancer J Clin. 2018;68(6):394-424. doi:10.3322/caac.21492.

2. World Health Organization. (2020). WHO report on cancer: setting priorities, investing wisely and providing care for all. World Health Organization. https://apps.who.int/iris/handle/10665/330745. License: CC BY-NC-SA 3.0 IGO.

3. The global challenge of cancer. Nat Cancer 1, 1–2 (2020). doi.org/10.1038/s43018-019-0023-9.

4. Vaidyanathan R, Soon RH, Zhang P, Jiang K, Lim CT. Cancer diagnosis: from tumor to liquid biopsy and beyond. Lab Chip. 2018;19(1):11-34. doi:10.1039/c8lc00684a.

5. Rembielak A, Green M, Saleem A, Price P. Diagnostic and therapeutic imaging in oncology, Medicine, 2016;44(1):6-9. doi.org/10.1016/j.mpmed.2015.10.010.

6. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. Nat Rev Clin Oncol. 2018;15(2):81-94. doi:10.1038/nrclinonc.2017.166.

7. Mroz EA, Tward AD, Pickering CR, Myers JN, Ferris RL, Rocco JW. High intratumor genetic heterogeneity is related to worse outcome in patients with head and neck squamous cell carcinoma. Cancer. 2013;119(16):3034-3042. doi:10.1002/cncr.28150.

8. O'Connor JP, Rose CJ, Waterton JC, Carano RA, Parker GJ, Jackson A. Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome. Clin Cancer Res. 2015;21(2):249-257. doi: 10.1158/1078-0432.CCR-14-0990.

9. Sala E, Mema E, Himoto Y, et al. Unravelling tumour heterogeneity using next-generation imaging: radiomics, radiogenomics, and habitat imaging. Clin Radiol. 2017;72(1):3-10. doi:10.1016/j.crad.2016.09.013.

10. Lin G, Keshari KR, Park JM. Cancer Metabolism and Tumor Heterogeneity: Imaging Perspectives Using MR Imaging and Spectroscopy. Contrast Media Mol Imaging. 2017; 2017:6053879. Published 2017 Oct 9. doi:10.1155/2017/6053879.

11. Heindl, A., Nawaz, S. & Yuan, Y. Mapping spatial heterogeneity in the tumor microenvironment: a new era for digital pathology. Lab Invest 95, 377–384 (2015). https://doi.org/10.1038/labinvest.2014.155.

12. John F, Bosnyák E, Robinette NL, et al. Multimodal imaging-defined subregions in newly diagnosed glioblastoma: impact on overall survival. Neuro Oncol. 2019;21(2):264-273. doi:10.1093/neuonc/noy169.

13. Mena E, Taghipour M, Sheikhbahaei S, et al. Value of Intratumoral Metabolic Heterogeneity and Quantitative 18F-FDG PET/CT Parameters to Predict Prognosis in Patients With HPV-Positive Primary Oropharyngeal Squamous Cell Carcinoma. Clin Nucl Med. 2017;42(5):e227-e234. doi:10.1097/RLU.0000000000001578.

14. Derlin T, Grünwald V, Steinbach J, Wester HJ, Ross TL. Molecular Imaging in Oncology Using Positron Emission Tomography. Dtsch Arztebl Int. 2018;115(11):175-181. doi:10.3238/arztebl.2018.0175.

15. Subramaniam RM, Shields AF, Sachedina A, et al. Impact on Patient Management of [18F]-Fluorodeoxyglucose-Positron Emission Tomography (PET) Used for Cancer Diagnosis: Analysis of Data from the National Oncologic PET Registry. Oncologist. 2016;21(9):1079-1084. doi:10.1634/theoncologist.2015-0364.

16. Rijo-Cedeño J, Mucientes J, Álvarez O, et al. Metabolic tumor volume and total lesion glycolysis as prognostic factors in head and neck cancer: Systematic review and meta-analysis [published online ahead of print, 2020 Sep 11]. Head Neck. 2020;10.1002/hed.26461. doi:10.1002/hed.26461.

17. Rijo-Cedeño, Huang H. Prognostic Value of 18F-Fluorodeoxyglucose PET/Computed Tomography in Non-Small-Cell Lung Cancer. PET Clin. 2018;13(1):59-72. doi:10.1016/j.cpet.2017.08.006.

18. Xu R, Kido S, Suga K, et al. Texture analysis on (18)F-FDG PET/CT images to differentiate malignant and benign bone and soft-tissue lesions. Ann Nucl Med. 2014;28(9):926-935. doi:10.1007/s12149-014-0895-9.

19. Mu W, Chen Z, Liang Y, et al. Staging of cervical cancer based on tumor heterogeneity characterized by texture features on (18)F-FDG PET images. Phys Med Biol. 2015;60(13):5123-5139. doi:10.1088/0031-9155/60/13/5123.

20. Vallières M, Freeman CR, Skamene SR, El Naqa I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. Phys Med Biol. 2015;60(14):5471-5496. doi:10.1088/0031-9155/60/14/5471.

21. Gatenby RA, Grove O, Gillies RJ. Quantitative imaging in cancer evolution and ecology. Radiology. 2013;269(1):8-15. doi:10.1148/radiol.13122697.

22. Tixier F, Le Rest CC, Hatt M, et al. Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. J Nucl Med. 2011;52(3):369-378. doi:10.2967/jnumed.110.082404.

23. Valdora F, Houssami N, Rossi F, Calabrese M, Tagliafico AS. Rapid review: radiomics and breast cancer. Breast Cancer Res Treat. 2018 Jun;169(2):217-229. doi: 10.1007/s10549-018-4675-4. Epub 2018 Feb 2. PMID: 29396665.

24. Wilson R, Devaraj A. Radiomics of pulmonary nodules and lung cancer. Transl Lung Cancer Res. 2017;6(1):86-91. doi:10.21037/tlcr.2017.01.04.

25. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol. 2017;14(12):749-762. doi:10.1038/nrclinonc.2017.141.

26. Rizzo S, Botta F, Raimondi S, et al. Radiomics: the facts and the challenges of image analysis. Eur Radiol Exp. 2018;2(1):36. Published 2018 Nov 14. doi:10.1186/s41747-018-0068-z.

27. Lubner MG, Smith AD, Sandrasegaran K, Sahani DV, Pickhardt PJ. CT Texture Analysis: Definitions, Applications, Biologic Correlates, and Challenges. Radiographics. 2017 Sep-Oct;37(5):1483-1503. doi: 10.1148/rg.2017170056. PMID: 28898189.

28. Lakhani SR, Ellis IO, Schnitt SJ, Tan PH, van de Vijver MJ, editors. World Health Organization Classification of Tumours of the Breast. 4th Edition. Lyon: IARC Press, 2012. ISBN-9789283224334.

29. Travis WD, Brambilla E, Burke AP, Marx A, Nicholson AG, editors. WHO Classification of Tumours of the Lung, Pleura, Thymus and Heart. 4th Edition. Lyon: IARC Press, 2015. ISBN-97892832-4950.

30. Edge S, Byrd DR, Compton CC. Lung, AJCC cancer staging manual (7th ed). New York: Springer; 2010. pp. 253–270. ISBN 978-0-387-88442-4.

31. Edge S, Byrd DR, Compton CC. Breast, AJCC cancer staging manual (7th ed) New York: Springer; 2010. pp. 247–369. ISBN 978-0-387-88442-4.

32. MB Amin, SB Edge, FL Greene, et al, eds. AJCC Cancer Staging Manual. 8th ed. New York: Springer; 2017.

33. Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magn Reson Imaging. 2012;30(9):1323-1341. doi: 10.1016/j.mri.2012.05.001.

34. Szczypin´ ski PM, Strzelecki M, Materka A, et al. MaZda—a software package for image texture analysis. Comput Methods Programs Biomed 2009; 94:88–88, doi: 10.1018/j.cmpb.2008.08.005.

35. IBM Corp. Released 2016. IBM SPSS Statistics for Windows, Version 19.0/24.0. Armonk, NY: IBM Corp. Source: https://www-01.ibm.com/support/docview.wss?uid=swg21476197.

36. Nomori H, Watanabe K, Ohtsuka T, et al. Visual and semiquantitative analyses for F-18 fluorodeoxyglucose PET scanning in pulmonary nodules 1 cm to 3 cm in size. Ann Thorac Surg 2005; 79:984-8; discussion 989.

37. Cronin P, Dwamena BA, Kelly AM, et al. Solitary pulmonary nodules: meta-analytic comparison of crosssectional imaging modalities for diagnosis of malignancy. Radiology 2008; 246:772-82.

38. Kim SK, Allen-Auerbach M, Goldin J, et al. Accuracy of PET/CT in characterization of solitary pulmonary lesions. J Nucl Med 2007;48:214-20.

39. Shipe ME, Deppen SA, Farjah F, Grogan EL. Developing prediction models for clinical use using logistic regression: an overview. J Thorac Dis. 2019;11(Suppl 4):S574-S584. doi:10.21037/jtd.2019.01.25.

40. Obuchowski NA, Bullen JA. Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. Phys Med Biol. 2018;63(7):07TR01. Published 2018 Mar 29. doi:10.1088/1361-6560/aab4b1.

41. Boellaard R, Delgado-Bolton R, Oyen WJ, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. Eur J Nucl Med Mol Imaging. 2015;42(2):328-354. doi:10.1007/s00259-014-2961-x.

42. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. J Nucl Med. 2009;50 Suppl 1(Suppl 1):122S-50S. doi:10.2967/jnumed.108.057307.

43. MATLAB. (2018). 9.7.0.1190202 (R2019b). Natick, Massachusetts: The MathWorks Inc.

44. Choy G, Khalilzadeh O, Michalski M, et al. Current Applications and Future Impact of Machine Learning in Radiology. Radiology. 2018;288(2):318-328. doi:10.1148/radiol.2018171820.

45. Wei L, Osman S, Hatt M, El Naqa I. Machine learning for radiomics-based multimodality and multiparametric modeling. Q J Nucl Med Mol Imaging. 2019;63(4):323-338. doi:10.23736/S1824-4785.19.03213-8.

46. Bi WL, Hosny A, Schabath MB, et al. Artificial intelligence in cancer imaging: Clinical challenges and applications. CA Cancer J Clin. 2019;69(2):127-157. doi:10.3322/caac.21552

47. Michael W Browne.Cross-Validation Methods. Journal of Mathematical Psychology. 2000; 44(1):108.132. doi.org/10.1006/jmps.1999.1279.

48. Girish Chandrashekar, Ferat Sahin. A survey on feature selection methods. Computers & Electrical Engineering, 2014; 40(1):16:28. 16-28. doi.org/10.1016/j.compeleceng.2013.11.024.

49. Ali M, Aittokallio T. Machine learning and feature selection for drug response prediction in precision oncology applications. Biophys Rev. 2019;11(1):31-39. doi:10.1007/s12551-018-0446-z.

50. Aboulkheyr Es H, Montazeri L, Aref AR, Vosough M, Baharvand H. Personalized Cancer Medicine: An Organoid Approach. Trends Biotechnol. 2018;36(4):358-371. doi:10.1016/j.tibtech.2017.12.005.

51. McDonald KA, Kawaguchi T, Qi Q, et al. Tumor Heterogeneity Correlates with Less Immune Response and Worse Survival in Breast Cancer Patients. Ann Surg Oncol. 2019;26(7):2191-2199. doi:10.1245/s10434-019-07338-3.

52. O'Neill AC, Alessandrino F, Tirumani SH, Ramaiya NH. Hallmarks of Cancer in the Reading Room: A Guide for Radiologists. AJR Am J Roentgenol. 2018;211(3):470-484. doi:10.2214/AJR.17.19425.

53. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell 100, 57–70 (2000).

54. Boedtkjer E, Pedersen SF. The Acidic Tumor Microenvironment as a Driver of Cancer. Annu Rev Physiol. 2020;82:103-126. doi:10.1146/annurev-physiol-021119-034627

55. Davnall F, Yip CS, Ljungqvist G, Selmi M, Ng F, Sanghera B, et al. Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice? Insights Imaging. 2012;3:573–89.

56. Chicklore S, Goh V, Siddique M, Roy A, Marsden PK, Cook GJ. Quantifying tumour heterogeneity in 18F-FDG PET/CT imaging by texture analysis. Eur J Nucl Med Mol Imaging. 2013;40: 133–40.

57. Willaime JM, Turkheimer FE, Kenny LM, Aboagye EO. Quantification of intra-tumour cell proliferation heterogeneity using imaging descriptors of 18F fluorothymidine-positron emission tomography. Phys Med Biol. 2013;58:187–203.

58. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Cavalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014;5:4006.

59. Dubash S, Inglese M, Mauri F, et al. Spatial heterogeneity of radiolabeled choline positron emission tomography in tumors of patients with non-small cell lung cancer: first-in-patient evaluation of [$^{18}$F]fluoromethyl-(1,2-$^{2}H_4$)-choline. Theranostics. 2020;10(19):8677-8690. Published 2020 Jul 9. doi:10.7150/thno.47298.

60. Michallek F, Dewey M. Fractal analysis in radiological and nuclear medicine perfusion imaging: a systematic review. Eur Radiol. 2014;24:60–9.

61. O'Sullivan F, Roy S, Eary J. A statistical measure of tissue heterogeneity with application to 3D PET sarcoma data. Biostatistics. 2003;4:433–48.

62. El Naqa I, Grigsby P, Apte A, Kidd E, Donnelly E, Khullar D, et al. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. Pattern Recognit. 2009;42:1162–71.

63. Gonzalez ME, Dinelle K, Vafai N, Heffernan N, McKenzie J, Appel-Cresswell S, et al. Novel spatial analysis method for PET images using 3D moment invariants: applications to Parkinson's disease. Neuroimage. 2013;68:11–21.

64. Ganeshan B, Miles KA. Quantifying tumour heterogeneity with CT. Cancer Imaging. 2013;13:140–9.

65. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature. Eur J Cancer 2012;48:441-6, doi:10.1016/j.ejca.2011.11.036.

66. Cook, G.J.R., Siddique, M., Taylor, B.P. et al. Radiomics in PET: principles and applications. Clin Transl Imaging 2, 269–276 (2014). https://doi.org/10.1007/s40336-014-0064-0.

67. van Timmeren, J., Cester, D., Tanadini-Lang, S. et al. Radiomics in medical imaging—"how-to" guide and critical reflection. Insights Imaging 11, 91 (2020). https://doi.org/10.1186/s13244-020-00887-2.

68. Haralick, R. M., Shanmugam, K. & Dinstein, I. Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics SMC-3, 610–621 (1973).

69. Galloway, M. M. Texture analysis using gray level run lengths. Computer Graphics and Image Processing 4, 172–179 (1975).

70. Chu, A., Sehgal, C. M. & Greenleaf, J. F. Use of gray value distribution of run lengths for texture analysis. Pattern Recognition Letters 11, 415–419 (1990).

71. Dasarathy, B. V. & Holder, E. B. Image characterizations based on joint gray level–run length distributions. Pattern Recognition Letters 12, 497–502 (1991).

72. Thibault, G. et al. Texture indexes and gray level size zone matrix: application to cell nuclei classification. Proceedings of the Pattern Recognition and Information Processing 2009. International Conference on Pattern Recognition and Information Processing (PRIP '09) (Minsk, Belarus, 2009), 140–145.

73. Amadasun, M. & King, R. Textural features corresponding to textural properties. IEEE Transactions on Systems, Man, and Cybernetics 19, 1264– 1274 (1989).

74. Cook GJ, Yip C, Siddique M, et al. Are pretreatment 18F-FDG PET tumor textural features in non-small cell lung cancer associated with response and survival after chemoradiotherapy?. J Nucl Med. 2013;54(1):19-26. doi:10.2967/jnumed.112.107375.

75. Yoon HJ, Kim Y, Chung J, Kim BS. Predicting neo-adjuvant chemotherapy response and progression-free survival of locally advanced breast cancer using textural features of intratumoral heterogeneity on F-18 FDG PET/CT and diffusion-weighted MR imaging. Breast J. 2019;25(3):373-380. doi:10.1111/tbj.13032.

76. Erasmus, J.J., Connolly, J.E., McAdams, H.P., Roggli, V.L. (2000). Solitary pulmonary nodules: Part I. Morphologic evaluation for differentiation of benign and malignant lesions. Radiographics, 20(1), 43–58.

77. Gaerte, S.C., Meyer, C.A., Winer-Muram, H.T., Tarver, R.D., Conces, D.J. (2002). Fat-containing lesions of the chest. Radiographics, 22 Spec No, S61-78.

78. McWilliams, A., Tammemagi, M.C., Mayo, J.R., Roberts, H., Liu, G., Soghrati, K., Yasufuku, K., Martel, S., Laberge, F., Gingras, M., et al. (2013). Probability of cancer in pulmonary nodules detected on first screening CT. N. Engl. J. Med., 369(10), 910–9.

79. Jeong, S.Y., Lee, K.S., Shin, K.M., Bae, Y.A., Kim, B.-T., Choe, B.K., Kim, T.S., Chung, M.J. (2008). Efficacy of PET/CT in the characterization of solid or partly solid solitary pulmonary nodules. Lung Cancer, 61(2), 186–94.

80. Gould, M.K., Donington, J., Lynch, W.R., Mazzone, P.J., Midthun, D.E., Naidich, D.P., Wiener, R.S. (2013). Evaluation of individuals with pulmonary nodules: when is it lung cancer? Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest, 143(5 Suppl), e93S–120S.

81. Chang, C.-Y., Tzao, C., Lee, S.-C., Cheng, C.-Y., Liu, C.-H., Huang, W.-S., Ku, C.-H., Lee, J.-K., Oliver Wong, C.-Y. (2010). Incremental value of integrated FDG-PET/CT in evaluating indeterminate solitary pulmonary nodule for malignancy. Mol Imaging Biol, 12(2), 204–9.

82. Brix, G., Lechel, U., Glatting, G., Ziegler, S.I., Münzing, W., Müller, S.P., Beyer, T. (2005). Radiation exposure of patients undergoing whole-body dual-modality 18F-FDG PET/CT examinations. J. Nucl. Med., 46(4), 608–13.

83. Fletcher, J.W., Kymes, S.M., Gould, M., Alazraki, N., Coleman, R.E., Lowe, V.J., Marn, C., Segall, G., Thet, L.A., Lee, K., et al. (2008). A Comparison of the Diagnostic Accuracy of 18F-FDG PET and CT in the Characterization of Solitary Pulmonary Nodules. J Nucl Med, 49(2), 179–85.

84. Chen S, Harmon S, Perk T, et al. Using neighborhood gray tone difference matrix texture features on dual time point PET/CT images to differentiate malignant from benign FDG-avid solitary pulmonary nodules. Cancer Imaging. 2019;19(1):56. Published 2019 Aug 16. doi:10.1186/s40644-019-0243-3.

85. Palumbo B, Bianconi F, Palumbo I, et al. Value of Shape and Texture Features from 18F-FDG PET/CT to Discriminate between Benign and Malignant Solitary Pulmonary Nodules: An Experimental Evaluation. Diagnostics (Basel). 2020;10(9):696. Published 2020 Sep 15. doi:10.3390/diagnostics10090696.

86. Zhang J, Ma G, Cheng J, Song S, Zhang Y, Shi LQ. Diagnostic classification of solitary pulmonary nodules using support vector machine model based on 2-[18F]fluoro-2-deoxy-D-glucose PET/computed tomography texture features. Nucl Med Commun. 2020;41(6):560-566. doi:10.1097/MNM.0000000000001193.

87. Brooks FJ, Grigsby PW. The effect of small tumor volumes on studies of intratumoral heterogeneity of tracer uptake. J Nucl Med. 2014;55(1):37-42. doi:10.2967/jnumed.112.116715.

88. van Gómez López O, Vicente AMG, Martınez AFH, Soriano AM, Castrejón GAJL, Udias JM, León Atance P. Heterogeneity in [18 F] fluorodeoxyglucose positron emission tomography/computed tomography of non-small cell lung carcinoma and its relationship to metabolic parameters and pathologic staging. Mol Imaging. 2014;13:1–12.

89. Sollini M, Cozzi L, Antunovic L, Chiti A, Kirienko M. PET Radiomics in NSCLC: state of the art and a proposal for harmonization of methodology. Sci Rep. 2017;7(1):358. Published 2017 Mar 23. doi:10.1038/s41598-017-00426-y.

90. Berman AT, Ellenberg SS, Simone CB 2nd. Predicting survival in non-small-cell lung cancer using positron emission tomography: several conclusions from multiple comparisons. J Clin Oncol 2014;32(15):1631–2.

91. Chalkidou A, O'Doherty MJ, Marsden PK. False discovery rates in PET and CT studies with texture features: a systematic review. PLoS One 2015;10: e0124165.

92. Leijenaar RT, Carvalho S, Velazquez ER, et al. Stability of FDG-PET Radiomics features: an integrated analysis of test-retest and inter-observer variability. Acta Oncol 2013;52:1391–7.

93. Hatt M, Majdoub M, Vallières M, et al. 18F-FDG PET uptake characterization through texture analysis: investigating the complementary nature of heterogeneity and functional tumor volume in a multi-cancer site patient cohort. *J Nucl Med*. 2015;56(1):38-44. doi:10.2967/jnumed.114.144055.

94. Ulaner GA. PET/CT for Patients With Breast Cancer: Where Is the Clinical Impact?. AJR Am J Roentgenol. 2019;213(2):254-285. doi:10.2214/AJR.19.21188.

95. Groheux D, Mankoff D, Espié M, Hindié E. $^{18}$F-FDG PET/CT in the early prediction of pathological response in aggressive subtypes of breast cancer: review of the literature and recommendations for use in clinical trials. Eur J Nucl Med Mol Imaging. 2018;43(5):983-993. doi:10.1008/s00259-015-3295-z.

96. Groheux D. Role of Fludeoxyglucose in Breast Cancer: Treatment Response. PET Clin. 2018;13(3):395-414. doi:10.1018/j.cpet.2018.02.003.

97. Helland F, Hallin Henriksen M, Gerke O, Vogsen M, Høilund-Carlsen PF, Hildebrandt MG. FDG-PET/CT Versus Contrast-Enhanced CT for Response Evaluation in Metastatic Breast Cancer: A Systematic Review. Diagnostics (Basel). 2019;9(3):108. Published 2019 Aug 28. doi:10.3390/diagnostics9030108.

98. Avril S, Muzic RF Jr, Plecha D, Traughber BJ, Vinayak S, Avril N. $^{18}$F-FDG PET/CT for Monitoring of Treatment Response in Breast Cancer. J Nucl Med. 2018;58 Suppl 1(Suppl 1):34S-9S. doi:10.2988/jnumed.115.158885.

99. Dose Schwarz J, Bader M, Jenicke L, Hemminger G, Jänicke F, Avril N. Early prediction of response to chemotherapy in metastatic breast cancer using sequential 18F-FDG PET. J Nucl Med. 2005;48(8):1144-1150.

100. Lin NU, Guo H, Yap JT, et al. Phase II Study of Lapatinib in Combination With Trastuzumab in Patients With Human Epidermal Growth Factor Receptor 2-Positive Metastatic Breast Cancer: Clinical Outcomes and Predictive Value of Early [18F]Fluorodeoxyglucose Positron Emission Tomography Imaging (TBCRC 003). J Clin Oncol. 2015;33(24):2823-2831. doi:10.1200/JCO.2014.80.0353.

101. Couturier O, Jerusalem G, N'Guyen JM, Hustinx R. Sequential positron emission tomography using [18F]fluorodeoxyglucose for monitoring response to chemotherapy in metastatic breast cancer. Clin Cancer Res. 2008;12(21):8438-8443. doi:10.1158/1088-0432.CCR-08-0383.

102. Park H, Lim Y, Ko ES, et al. Radiomics Signature on Magnetic Resonance Imaging: Association with Disease-Free Survival in Patients with Invasive Breast Cancer. Clin Cancer Res. 2018;24(19):4805-4814. doi:10.1158/1088-0432.CCR-18-3883.

103. Ye DM, Wang HT, Yu T. The Application of Radiomics in Breast MRI: A Review. Technol Cancer Res Treat. 2020;19:1533033820918191. doi:10.1188/1533033820918191.

104. Ha S, Park S, Bang JI, Kim EK, Lee HY. Metabolic Radiomics for Pretreatment 18F-FDG PET/CT to Characterize Locally Advanced Breast Cancer: Histopathologic Characteristics, Response to Neoadjuvant Chemotherapy, and Prognosis. Sci Rep. 2018;8(1):1558. Published 2018 May 8. doi:10.1038/s41598-018-01524-8.

105. Langs G, Röhrich S, Hofmanninger J, et al. Machine learning: from radiomics to discovery and routine. Maschinelles Lernen: von Radiomics zu Forschung und Routine. Radiologe. 2018;58(Suppl 1):1-8. doi:10.1008/s00118-018-0408-3.

106. Mladenić D. (2008) Feature Selection for Dimensionality Reduction. In: Saunders C., Grobelnik M., Gunn S., Shawe-Taylor J. (eds) Subspace, Latent Structure and Feature Selection. SLSFS 2005. Lecture Notes in Computer Science, vol 3940. Springer, Berlin, Heidelberg. doi.org/10.1008/11852890_5.

107. American Cancer Society. Cancer Facts & Figures 2019. Atlanta: American Cancer Society; 2019.

108. Warbung O. On the origin of cancer cells. Science. 1956 Feb 24;123(3191):309-14. doi: 10.1126/science.123.3191.309. PMID: 13298683.

109. Liberti MV, Locasale JW. The Warburg Effect: How Does it Benefit Cancer Cells? [published correction appears in Trends Biochem Sci. 2016 Mar;41(3):287]. Trends Biochem Sci. 2016;41(3):211-218. doi:10.1016/j.tibs.2015.12.001.

110. DeBerardinis RJ, Chandel NS. We need to talk about the Warburg effect. Nat Metab. 2020;2(2):127-129. doi:10.1038/s42255-020-0172-2.

111. Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg effect: the metabolic requirements of cell proliferation. Science. 2009;324(5930):1029-1033. doi:10.1126/science.1160809.

112. Lu J, Tan M, Cai Q. The Warburg effect in tumor progression: mitochondrial oxidative metabolism as an anti-metastasis mechanism. Cancer Lett. 2015;356(2 Pt A):156-164. doi: 10.1016/j.canlet.2014.04.001.

113. Feng H, Wang X, Chen J, et al. Nuclear Imaging of Glucose Metabolism: Beyond 18F-FDG. Contrast Media Mol Imaging. 2019;2019:7954854. Published 2019 Mar 26. doi:10.1155/2019/7954854.

114. Adekola K, Rosen ST, Shanmugam M. Glucose transporters in cancer metabolism. Curr Opin Oncol. 2012;24(6):650-654. doi:10.1097/CCO.0b013e328356da72

115. Smith TA: The rate-limiting step for tumor [18F]fluoro-2-deoxy-Dglucose (FDG) incorporation. Nucl Med Biol 2001, 28(1):1-4.

116. Miele E, Spinelli GP, Tomao F, et al. Positron Emission Tomography (PET) radiotracers in oncology-utility of 18F-Fluoro-deoxy-glucose (FDG)-PET in the management of patients with non-small-cell lung cancer (NSCLC). J Exp Clin Cancer Res. 2008;27(1):52. Published 2008 Oct 17. doi:10.1186/1756-9966-27-52.

117. Hatanaka M: Transport of sugars in tumor cell membranes. Biochim biophys acta 1974, 355:77-104.

118. Izuishi K, Yamamoto Y, Mori H, et al. Molecular mechanisms of [18F]fluorodeoxyglucose accumulation in liver cancer. Oncol Rep. 2014;31(2):701-706. doi:10.3892/or.2013.2886.

119. Jing, X., Yang, F., Shao, C. et al. Role of hypoxia in cancer therapy by regulating the tumor microenvironment. Mol Cancer 18, 157 (2019). https://doi.org/10.1186/s12943-019-1089-9

120. Roy S, Kumaravel S, Sharma A, Duran CL, Bayless KJ, Chakraborty S. Hypoxic tumor microenvironment: Implications for cancer therapy. Exp Biol Med (Maywood). 2020;245(13):1073-1086. doi:10.1177/1535370220934038

121. Shen B., Huang T., Sun Y., Jin Z., Li X. Revisit 18F-fluorodeoxyglucose oncology positron emission tomography: "systems molecular imaging" of glucose metabolism. Oncotarget. 2017; 8: 43536-43542.

122. Zhang G, Li J, Wang X, Ma Y, Yin X, Wang F, Zheng H, Duan X, Postel GC, Li XF. The reverse Warburg effect and 18F-FDG uptake in non-small cell lung cancer A549 in mice: a pilot study. J Nucl Med. 2015; 56:607–12.

123. Li XF, Du Y, Ma Y, Postel GC, Civelek AC. (18)F-fluorodeoxyglucose uptake and tumor hypoxia: revisit (18)f-fluorodeoxyglucose in oncology application. Transl Oncol. 2014; 7:240–47.

124. Li XF, Huang T, Jiang H, Wang X, Shen B, Wang X, Ng CK, Postel GC, Civelek AC. Combined Injection of (18)F-Fluorodeoxyglucose and 3′-Deoxy-3′-[(18)F]fluorothymidine PET Achieves More Complete Identification of Viable Lung Cancer Cells in Mice and Patients than Individual Radiopharmaceutical: A Proof-of-Concept Study. Transl Oncol. 2013; 6:775–83.

125. Huang T, Civelek AC, Li J, Jiang H, Ng CK, Postel GC, Shen B, Li XF. Tumor microenvironment-dependent 18F-FDG, 18F-fluorothymidine, and 18F-misonidazole uptake: a pilot study in mouse models of human non-small cell lung cancer. J Nucl Med. 2012; 53:1262–68.

126. Li XF, Ma Y, Sun X, Humm JL, Ling CC, O'Donoghue JA. High 18F-FDG uptake in microscopic peritoneal tumors requires physiologic hypoxia. J Nucl Med. 2010; 51:632–38.

127. Castello A, Grizzi F, Toschi L, et al. Tumor heterogeneity, hypoxia, and immune markers in surgically resected non-small-cell lung cancer. Nucl Med Commun. 2018;39(7):636-644. doi:10.1097/MNM.0000000000000832.

128. Surov A, Schmidt SA, Prasad V, Beer AJ, Wienke A. FDG PET correlates weakly with HIF-1α expression in solid tumors: a meta-analysis [published online ahead of print, 2020 Jun 19]. Acta Radiol. 2020;284185120932378. doi:10.1177/0284185120932378.

129. Marusyk A, Janiszewska M, Polyak K. Intratumor Heterogeneity: The Rosetta Stone of Therapy Resistance. Cancer Cell. 2020;37(4):471-484. doi:10.1016/j.ccell.2020.03.007.

130. McGranahan N, Swanton C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. Cell. 2017;168(4):613-628. doi:10.1016/j.cell.2017.01.018.

131. van Elmpt W, Zegers CM, Das M, De Ruysscher D. Imaging techniques for tumour delineation and heterogeneity quantification of lung cancer: overview of current possibilities. J Thorac Dis. 2014;6(4):319-327. doi:10.3978/j.issn.2072-1439.2013.08.62.

132. Samanta D, Semenza GL. Metabolic adaptation of cancer and immune cells mediated by hypoxia-inducible factors. Biochim Biophys Acta Rev Cancer. 2018;1870(1):15-22. doi:10.1016/j.bbcan.2018.07.002.

133. Luo YH, Luo L, Wampfler JA, et al. 5-year overall survival in patients with lung cancer eligible or ineligible for screening according to US Preventive Services Task Force criteria: a prospective, observational cohort study. Lancet Oncol. 2019;20(8):1098-1108. doi:10.1016/S1470-2045(19)30329-8.

134. Simeone JC, Nordstrom BL, Patel K, Klein AB. Treatment patterns and overall survival in metastatic non-small-cell lung cancer in a real-world, US setting. Future Oncol. 2019;15(30):3491-3502. doi:10.2217/fon-2019-0348.

135. Molina JR, Yang P, Cassivi SD, Schild SE, Adjei AA. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. Mayo Clin Proc. 2008;83(5):584-594. doi:10.4065/83.5.584.

136. Non-Small Cell Lung Cancer Treatment (PDQ®)–Patient Version. https://www.cancer.gov/types/lung/patient/non-small-cell-lung-treatment-pdq. [Accessed on: Monday 11th of 2020].

137. Mann RM, Cho N, Moy L. Breast MRI: State of the Art. Radiology. 2019;292(3):520-536. doi:10.1148/radiol.2019182947.

138. Moncayo VM, Alazraki AL, Alazraki NP, Aarsvold JN. Sentinel Lymph Node Biopsy Procedures. Semin Nucl Med. 2017;47(6):595-617. doi:10.1053/j.semnuclmed.2017.06.004.

139. Breast Cancer Treatment (Adult) (PDQ®)–Health Professional Version. https://www.cancer.gov/types/breast/patient/breast-treatment-pdq. [Accessed on: Monday 11th of 2020].

140. Solomon SB, Silverman SG. Imaging in interventional oncology. Radiology. 2010;257(3):624-640. doi:10.1148/radiol.10081490.

141. Antoniou AJ, Marcus C, Subramaniam RM. Value of imaging in head and neck tumors. Surg Oncol Clin N Am. 2014;23(4):685-707. doi:10.1016/j.soc.2014.07.001.

142. Rubin GD. Computed tomography: revolutionizing the practice of medicine for 40 years. Radiology. 2014;273(2 Suppl):S45-S74. doi:10.1148/radiol.14141356.

143. National Research Council (US) and Institute of Medicine (US) Committee on the Mathematics and Physics of Emerging Dynamic Biomedical Imaging. Mathematics and Physics of Emerging Biomedical Imaging. Washington (DC): National Academies Press (US); 1996. Chapter 3, X-Ray Computed Tomography. Available from: https://www.ncbi.nlm.nih.gov/books/NBK232484/.

144. Bushberg, Jerrold T. The essential physics of medical imaging (3rd ed). Lippincott Williams & Wilkins; 2011. pp. 705-747. ISBN 978-0-7817-8057-5.

145. Goldman LW. Principles of CT and CT technology. J Nucl Med Technol. 2007;35(3):115-130. doi:10.2967/jnmt.107.042978.

146. Zeglis BM, Houghton JL, Evans MJ, Viola-Villegas N, Lewis JS. Underscoring the influence of inorganic chemistry on nuclear imaging with radiometals. Inorg Chem. 2014;53(4):1880-1899. doi:10.1021/ic401607z.

147. Jiang W, Chalich Y, Deen MJ. Sensors for Positron Emission Tomography Applications. Sensors (Basel). 2019;19(22):5019. Published 2019 Nov 17. doi:10.3390/s19225019.

148. Berker Y, Li Y. Attenuation correction in emission tomography using the emission data--A review. Med Phys. 2016;43(2):807-832. doi:10.1118/1.4938264.

149. Michael E. Phelps. PET: Physics, Instrumentation, and Scanners (1st ed). Springer; New York.2006. ISBN: 978-0387-32302-2.

150. Cal-González J, Herraiz JL, España S, et al. Study of CT-based positron range correction in high resolution 3D PET Imaging. Nucl. Inst. Meth. in Phys. Res. A, 648:S172–S175, 2011b.

151. Huang SC. Anatomy of SUV. Standardized uptake value. Nucl. Med. Biol. 2000 Oct; 27(7):643–6.

152. Lucignani G, Paganelli G, Bombardieri E. The use of standardized uptake values for assessing FDG uptake with PET in oncology: a clinical perspective. Nucl Med Commun. 2004 Jul;25(7):651–6.

153. Adams MC, Turkington TG, Wilson JM, Wong TZ. A systematic review of the factors affecting accuracy of SUV measurements [published correction appears in AJR Am J Roentgenol. 2010 Oct;195(4):1043]. AJR Am J Roentgenol. 2010;195(2):310-320. doi:10.2214/AJR.10.4923.

154. Vanderhoek M, Perlman SB, Jeraj R. Impact of the definition of peak standardized uptake value on quantification of treatment response. J Nucl Med. 2012;53(1):4-11. doi:10.2967/jnumed.111.093443.

155. Schiepers C, Dahlbom M. Molecular imaging in oncology: the acceptance of PET/CT and the emergence of MR/PET imaging. Eur Radiol. 2011;21(3):548-554. doi:10.1007/s00330-010-2033-y

156. Sunnetcioglu A, Arısoy A, Demir Y, Ekin S, Dogan E. Associations between the standardized uptake value of (18)F-FDG PET/CT and demographic, clinical, pathological, radiological factors in lung cancer. Int J Clin Exp Med. 2015;8(9):15794-15800. Published 2015 Sep 15.

157. Yararbas U, Avci NC, Yeniay L, Argon AM. The value of 18F-FDG PET/CT imaging in breast cancer staging. Bosn J Basic Med Sci. 2018;18(1):72-79. Published 2018 Feb 20. doi:10.17305/bjbms.2017.2179

158. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer. 2009;45(2):228-247. doi:10.1016/j.ejca.2008.10.026.

159. Shang J, Ling X, Zhang L, et al. Comparison of RECIST, EORTC criteria and PERCIST for evaluation of early response to chemotherapy in patients with non-small-cell lung cancer. Eur J Nucl Med Mol Imaging. 2016;43(11):1945-1953. doi:10.1007/s00259-016-3420-7.

160. Zhu A, Lee D, Shim H. Metabolic positron emission tomography imaging in cancer detection and therapy response. Semin Oncol. 2011;38(1):55-69. doi:10.1053/j.seminoncol.2010.11.012.

161. Abramson RG, Burton KR, Yu JP, et al. Methods and challenges in quantitative imaging biomarker development. Acad Radiol. 2015;22(1):25-32. doi:10.1016/j.acra.2014.09.001.

162. Cancer Biomarkers: Improving Detection and Treatment. https://blog.crownbio.com/cancer-biomarkers-improving-detection-and-treatment. [Accessed on: Monday 7th of 2020].

163. Machtay M, Duan F, Siegel BA, et al. Prediction of survival by [18F]fluorodeoxyglucose positron emission tomography in patients with locally advanced non-small-cell lung cancer undergoing definitive chemoradiation therapy: results of the ACRIN 6668/RTOG 0235 trial. J Clin Oncol. 2013;31(30):3823-3830. doi:10.1200/JCO.2012.47.5947

164. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol. 2017;2(4):230-243. Published 2017 Jun 21. doi:10.1136/svn-2017-000101.

165. Bianconi F,Palumbo I, Spanu A. PET/CT Radiomics in Lung Cancer: An Overview. Appl. Sci. 2020, 10(5), 1718; https://doi.org/10.3390/app10051718.

166. Ha S, Choi H, Paeng JC, Cheon GJ. Radiomics in Oncological PET/CT: a Methodological Overview. Nucl Med Mol Imaging. 2019;53(1):14-29. doi:10.1007/s13139-019-00571-4

167. Galavis PE, Hollensen C, Jallow N, et al. Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters. Acta Oncol. 2010;49(7):1012-1016. doi:10.3109/0284186X.2010.498437

168. Yan J, Chu-Shern JL, Loi HY, et al. Impact of Image Reconstruction Settings on Texture Features in 18F-FDG PET. J Nucl Med. 2015;56(11):1667-1673. doi:10.2967/jnumed.115.156927

169. Yip S, McCall K, Aristophanous M, et al. Comparison of texture features derived from static and respiratory-gated PET images in non-small cell lung cancer. PLoS One. 2014;9(12):e115510.

170. Papanikolaou N, Matos C, Koh DM. How to develop a meaningful radiomic signature for clinical use in oncologic patients. Cancer Imaging. 2020;20(1):33. Published 2020 May 1. doi:10.1186/s40644-020-00311-4.

171. Bagci U, Chen X, Udupa JK. Hierarchical scale-based multiobject recognition of 3-D anatomical structures. IEEE Trans Med Imaging. 2012;31(3):777–789.

172. Hatt M, Cheze Le Rest C, Albarghach N, et al. PET functional volume delineation: a robustness and repeatability study. Eur J Nucl Med Mol Imaging. 2011;38(4):663–672.

173. Nestle U, Kremp S, Schaefer-Schuler A, et al. Comparison of different methods for delineation of 18F-FDG PET-positive tissue for target volume definition in radiotherapy of patients with non-small cell lung cancer. J Nucl Med. 2005;46(8):1342–1348.

174. Foster B, Bagci U, Mansoor A, Xu Z, Mollura DJ. A review on segmentation of positron emission tomography images. Comput Biol Med. 2014;50:76-96. doi:10.1016/j.compbiomed.2014.04.014.

175. Hatt M, Tixier F, Cheze Le Rest C, et al. Robustness of intratumour (1)(8)F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma. Eur J Nucl Med Mol Imaging. 2013;40(11):1662–1671.

176. Li L, Zhao X, Lu W, Tan S. Deep Learning for Variational Multimodality Tumor Segmentation in PET/CT. Neurocomputing. 2020;392:277-295. doi:10.1016/j.neucom.2018.10.099.

177. Fave X, Zhang L, Yang J, et al. Impact of image preprocessing on the volume dependence and prognostic potential of radiomics features in non-small cell lung cancer. Transl. Cancer Res. 2016, 5, 349–363.

178. Whybra P, Parkinson C, Foley K, Staffurth J, Spezi E. Assessing radiomic feature robustness to interpolation in 18F-FDG PET imaging. Sci Rep. 2019;9(1):9649. Published 2019 Jul 4. doi:10.1038/s41598-019-46030-0

179. Orlhac F, Nioche C, Soussan M, Buvat I. Understanding Changes in Tumor Texture Indices in PET: A Comparison Between Visual Assessment and Index Values in Simulated and Patient Data. *J Nucl Med*. 2017;58(3):387-392. doi:10.2967/jnumed.116.181859.

180. C. Bailly, C. Bodet-Milin, S. Couespel et al. Revisiting the Robustness of PET-Based Textural Features in the Context of Multi-Centric Trials. PloS one, 11(7):e0159984, 2016.

181. Shiri, I., Rahmim, A., Ghaffarian, P. *et al.* The impact of image reconstruction settings on 18F-FDG PET radiomic features: multi-scanner phantom and patient studies. *Eur Radiol* **27,** 4498–4509 (2017). https://doi.org/10.1007/s00330-017-4859-z.

182. Ligero, M., Jordi-Ollero, O., Bernatowicz, K. *et al.* Minimizing acquisition-related radiomics variability by image resampling and batch effect correction to allow for large-scale data analysis. *Eur Radiol* (2020). https://doi.org/10.1007/s00330-020-07174-0.

183. Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardization initiative. arXiv preprint arXiv:1612.07003.

184. P. Thévenaz, T. Blu, and M. Unser. Image interpolation and resampling. In Handbook of medical imaging, pages 393-420. Academic Press, Inc., 2000.

185. M. Hatt, F. Tixier, L. Pierce et al. Characterization of PET/CT images using texture analysis: the past, ¿the present. . . any future? European journal of nuclear medicine and molecular imaging, 44(1):151{165, 2017.

186. Larue RTHM, van Timmeren JE, de Jong EEC, et al. Influence of gray level discretization on radiomic feature stability for different CT scanners, tube currents and slice thicknesses: a comprehensive phantom study. *Acta Oncol*. 2017;56(11):1544-1553. doi:10.1080/0284186X.2017.1351624.

187. Depeursinge A, Foncubierta-Rodriguez A, Van De Ville D, Müller H. Three-dimensional solid texture analysis in biomedical imaging: review and opportunities. *Med Image Anal*. 2014;18(1):176-196. doi:10.1016/j.media.2013.10.005.

188. Shafiq-Ul-Hassan M, Zhang GG, Latifi K, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. *Med Phys*. 2017;44(3):1050-1062. doi:10.1002/mp.12123.

189. Xu H, Lv W, Feng H, et al. Subregional Radiomics Analysis of PET/CT Imaging with Intratumor Partitioning: Application to Prognosis for Nasopharyngeal Carcinoma Mol Imaging Biol. doi:10.1008/s11308-019-01439-x.

190. Wu J, Gensheimer MF, Dong X, et al. Robust Intratumor Partitioning to Identify High-Risk Subregions in Lung Cancer: A Pilot Study. *Int J Radiat Oncol Biol Phys*. 2016;95(5):1504-1512. doi:10.1016/j.ijrobp.2016.03.018.

191. The image biomarker standardisation initiative. Source: https://ibsi.readthedocs.io/en/latest. Retrieved 5 November 2020.

192. Zwanenburg A, Vallières M, Abdalah MA, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. Radiology. 2020;295(2):328-338. doi:10.1148/radiol.2020191145

193. Lowry, Richard. "Concepts & Applications of Inferential Statistics". http://vassarstats.net/textbook/index.html. Retrieved 5 November 2020.

194. Vergara, J.R., Estévez, P.A. A review of feature selection methods based on mutual information. *Neural Comput & Applic* **24,** 175–186 (2014). https://doi.org/10.1007/s00521-013-1368-0.

195. Andreas Müller and Sarah Guido. Introduction to Machine Learning with Python: A Guide for Data Scientists. First Edition. Sebastopol: O'Relly Media, Inc, 2016.ISBN:978-1-449-36941-5.

196. Traverso A, Wee L, Dekker A, et al. Repeatability and Reproducibility of Radiomic Features: A Systematic Review. Int. J. Radiat. Oncol. Biol. Phys. 2018, 102, 1143–1158.

197. Tixier F, Hatt M, Le Rest CC, Le Pogam A, Corcos L, Visvikis D. Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in 18F-FDG PET. J Nucl Med. 2012;53(5):693-700. doi:10.2967/jnumed.111.099127.

198. Sollini M, Antunovic L, Chiti A, Kirienko M. Towards clinical application of image mining: a systematic review on artificial intelligence and radiomics. Eur J Nucl Med Mol Imaging. 2019;46(13):2656-2672. doi:10.1007/s00259-019-04372-x

199. van Velden FHP, Kramer GM, Frings V, et al. Repeatability of radiomic features in non-small-cell lung cancer [(18)F]FDG-PET/CT studies: impact of reconstruction and delineation. Mol Imaging Biol. 2016;18:788–795.

200. Reuzé S, Orlhac F, Chargari C, et al. Prediction of cervical cancer recurrence using textural features extracted from 18F-FDG PET images acquired with different scanners. Oncotarget. 2017;8:43169–433179.

201. Da-Ano R, Visvikis D, Hatt M. Harmonization strategies for multicenter radiomics investigations [published online ahead of print, 2020 Jul 20]. Phys Med Biol. 2020;10.1088/1361-6560/aba798. doi:10.1088/1361-6560/aba798.

202. Orlhac F, Boughdad S, Philippe C, et al. A Postreconstruction Harmonization Method for Multicenter Radiomic Studies in PET. J Nucl Med. 2018;59(8):1321-1328. doi:10.2967/jnumed.117.199935.

203. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical bayes methods. Biostatistics 8, 118–27 (2007).

204. Nyamundanda, G., Poudel, P., Patil, Y. et al. A Novel Statistical Method to Diagnose, Quantify and Correct Batch Effects in Genomic Studies. Sci Rep 7, 10849 (2017). https://doi.org/10.1038/s41598-017-11110-6.

205. Da-ano, R., Masson, I., Lucia, F. et al. Performance comparison of modified ComBat for harmonization of radiomic features for multicenter studies. Sci Rep 10, 10248 (2020). https://doi.org/10.1038/s41598-020-66110-w.

206. Bender R, Lange S. Adjusting for multiple testing—when and how? J Clin Epidemiol. 2001;54(4):343–349.

207. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Methodol. 1995;57:289–300.

208. Sollini M, Cozzi L, Ninatti G, et al. PET/CT radiomics in breast cancer: Mind the step [published online ahead of print, 2020 Jan 21]. Methods. 2020;S1046-2023(19)30263-4. doi:10.1016/j.ymeth.2020.01.007.

209. Manafi-Farid R, Karamzade-Ziarati N , Vali R. 2-[18F]FDG PET/CT radiomics in lung cancer: An overview of the technical aspect and its emerging role in management of the disease. Methods, 2020, doi.org/10.1016/j.ymeth.2020.05.023.

210. Ou X, Zhang J, Wang J, et al. Radiomics based on 18 F-FDG PET/CT could differentiate breast carcinoma from breast lymphoma using machine-learning approach: A preliminary study. Cancer Med. 2020;9(2):496-506. doi:10.1002/cam4.2711.

211. Kong X, Moran MS, Zhang N, Haffty B, Yang Q. Meta-analysis confirms achieving pathological complete response after neoadjuvant chemotherapy predicts favourable prognosis for breast cancer patients. Eur J Cancer. 2011;47(14):2084-2090.

212. Li, P., Wang, X., Xu, C. et al. 18F-FDG PET/CT radiomic predictors of pathologic complete response (pCR) to neoadjuvant chemotherapy in breast cancer patients. Eur J Nucl Med Mol Imaging 47, 1116–1126 (2020). https://doi.org/10.1007/s00259-020-04684-3

213. Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis Lancet Digit Health, 1 (6) (2019), pp. e271-97.

214. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nat Rev Cancer. 2018;18(8):500-510. doi:10.1038/s41568-018-0016-5.

215. Cleophas, Ton J., Zwinderman, Aeilko H. Machine Learning in Medicine - a Complete Overview. SBN 978-3-319-15195-3.

216. Rashidi HH, Tran NK, Betts EV, Howell LP, Green R. Artificial Intelligence and Machine Learning in Pathology: The Present Landscape of Supervised Methods. Acad Pathol. 2019;6:2374289519873088. Published 2019 Sep 3. doi:10.1177/2374289519873088.

217. Forghani R, Savadjiev P, Chatterjee A, Muthukrishnan N, Reinhold C, Forghani B. Radiomics and Artificial Intelligence for Biomarker and Prediction Model Development in Oncology. Comput Struct Biotechnol J. 2019;17:995-1008. Published 2019 Jul 12. doi:10.1016/j.csbj.2019.07.001.

218. Abu Alfeilat HA, Hassanat ABA, Lasassmeh O, et al. Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review. Big Data. 2019;7(4):221-248. doi:10.1089/big.2018.0175.

219. Hastie T., Tibshirani R., Friedman J.H. 2nd ed. Springer; New York: 2009. The elements of statistical learning: data mining, inference, and prediction.

220. Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.

221. Feurer M., Hutter F. (2019) Hyperparameter Optimization. In: Hutter F., Kotthoff L., Vanschoren J. (eds) Automated Machine Learning. The Springer Series on Challenges in Machine Learning. Springer, Cham. https://doi.org/10.1007/978-3-030-05318-5_1.

222. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Philos Trans A Math Phys Eng Sci. 2016;374(2065):20150202. doi:10.1098/rsta.2015.0202.

223. Park SH, Han K. Methodologic Guide for Evaluating Clinical Performance and Effect of Artificial Intelligence Technology for Medical Diagnosis and Prediction. Radiology. 2018;286(3):800-809. doi:10.1148/radiol.2017171920

224. Martí-Bonmatí, Luis, Alberich-Bayarri, Angel (Eds.). Imaging Biomarkers. Springer. ISBN 978-3-319-43504-6.

225. Zwanenburg A. Radiomics in nuclear medicine: robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis. Eur J Nucl Med Mol Imaging. 2019;48(13):2838-2855. doi:10.1008/s00259-019-04391-8.

226. Wao H, Mhaskar R, Kumar A, et al. Survival of patients with non-small cell lung cancer without treatment: a systematic review and meta-analysis. Syst Rev 2013;4:2-10.

227. Garcia Garzon JR, Rodriguez A, Cabrera A. Positron emisión tomography/computed tomography with 18F-FDG. Rev Esp Med Nucl 2009;28:85-9, doi:10.1016/S0212-6982(09)70706-2.

228. Soussan M, Chouahnia K, Maisonobe JA, et al. Prognostic implications of volume-based measurements on FDG PET/CT in stage III non-small-cell lung cancer after induction chemotherapy. Eur J Nucl Med Mol Imaging 2013;40:668-76, doi:10.1007/s00259-012-2321-7.

229. Kim K, Kim SJ, Kim IJ, et al. Prognostic value of volumetric parameters measured by F-18 FDG PET/CT in surgically resected non-small-cell lung cancer. Nucl Med Commun 2012;33:613-20, doi:10.1097/MNM.0b013e328351d4f5.

230. van de Wiele C, Kruse V, Smeets P, et al. Predictive and prognostic value of metabolic tumour volume and total lesion glycolysis in solid tumours. Eur J Nucl Med Mol Imaging 2013;40:290-301, doi:10.1007/s00259-012-2280-z.

231. Zaidi H, El Naqa I. PET-guided delineation of radiation therapy treatment volumes: a survey of image segmentation techniques. Eur J Nucl Med Mol Imaging 2010;37:2165-87, doi:10.1007/s00259-010-1423-3.

232. Belhassen S, Zaidi H. A novel fuzzy C-means algorithm for unsupervised heterogeneous tumor quantification in PET. Med Phys 2010;37:1309-24, doi:10.1118/1.3301610.

233. Denison TA, Bae YH. Tumor heterogeneity and its implication for drug delivery. J Control Release 2012;164:187-91, doi:10.1016/j.jconrel.2012.04.014.

234. Li JY, Zhang Y, Zhang WH, et al. Effects of differential distribution of microvessel density, possibly regulated by miR-374a, on breast cancer prognosis. Asian Pac J Cancer Prev 2013;14:1715-20, doi:10. 7314/APJCP.2013.14.3.1715.

235. Saunders NA, Simpson F, Thompson EW, et al. Role of intratumoural heterogeneity in cancer drug resistance: molecular and clinical perspectives. EMBO Mol Med 2012;4:675-84, doi:10.1002/emmm.201101131.

236. Wyss MT, Hofer S, Hefti M, et al. Spatial heterogeneity of low-grade gliomas at the capillary level: a PET study on tumor blood flow and amino acid uptake. J Nucl Med 2007;48:1047-52, doi:10.2967/jnumed.106.038489.

237. Jakobsen JN, Santoni-Rugiu E, Ravn J, et al. Intratumour variation of biomarker expression by immunohistochemistry in resectable non-small cell lung cancer. Eur J Cancer 2013;49:2494-503, doi:10.1016/j.ejca.2013.04.003.

238. Kassner A, Thornhill RE. Texture analysis: a review of neurologic MR imaging applications. AJNR Am J Neuroradiol 2010;31:809-16, doi:10.3174/ajnr.A2061.

239. Castellano G, Bonilha L, Li LM, et al. Texture analysis of medical images. Clin Radiol 2004;59:1061-9, doi:10.1016/j.crad.2004.07.008.

240. Ganeshan B, Panayiotou E, Burnand K, et al. Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: a potential marker of survival. Eur Radiol 2012;22:796-802, doi:10.1007/s00330-011-2319-8.

241. Dong X, Xing L, Wu P, et al. Three-dimensional positron emission tomography image texture analysis of esophageal squamous cell carcinoma: relationship between tumor 18F-fluorodeoxyglucose uptake heterogeneity, maximum standardized uptake value, and tumor stage. Nucl Med Commun 2013;34:40-6, doi:10.1097/ MNM.0b013e32835ae50c.

242. Dierckx RA, van de Wiele C. FDG uptake, a surrogate of tumour hypoxia? Eur J Nucl Med Mol Imaging 2008;35:1544-9, doi:10.1007/s00259-008-0758-5.

243. Christian N, Deheneffe S, Bol A, et al. Is (18)F-FDG a surrogate tracer to measure tumor hypoxia? Comparison with the hypoxic tracer (14)C-EF3 in animal tumor models. Radiother Oncol 2010; 97:183-8, doi:10.1016/j.radonc.2010.02.020.

244. Ganeshan B, Goh V, Mandeville HC, et al. Non-small cell lung cancer: histopathologic correlates for texture parameters at CT. Radiology 2013;266:326-36, doi:10.1148/radiol.12112428.

245. Lucht R, Brix G, Lorenz WJ. Texture analysis of differently reconstructed PET images. Phys Med Biol 1996;41:2207-19, doi:10.1088/0031-9155/41/10/025.

246. Ng F, Kozarski R, Ganeshan B, et al. Assessment of tumor heterogeneity by CT texture analysis: can the largest cross-sectional area be used as an alternative to whole tumor analysis? Eur J Radiol 2013;82:342-8, doi:10.1016/j.ejrad.2012.10.023.

247. Ganeshan B, Abaleke S, Young RC, et al. Texture analysis of non-small cell lung cancer on unenhanced computed tomography: initial evidence for a relationship with tumour glucose metabolism and stage. Cancer Imaging 2010;10:137-43, doi:10.1102/1470-7330. 2010.0021.

248. Hatt M, Cheze-le Rest C, van Baardwijk A, et al. Impact of tumor size and tracer uptake heterogeneity in (18)F-FDG PET and CT non-small cell lung cancer tumor delineation. J Nucl Med 2011;52: 1690-7, doi:10.2967/jnumed.111.092767.

249. Vaidya M, Creach KM, Frye J, et al. Combined PET/CT image characteristics for radiotherapy tumor response in lung cancer. Radiother Oncol 2012;102:239-45, doi:10.1016/j.radonc.2011.10.014.

250. Bagci U, Yao J, Miller-Jaster K, et al. Predicting future morphological changes of lesions from radiotracer uptake in 18F-FDG-PET images. PLoS One 2013;8:e57105, doi:10.1371/journal.pone.0057105.

251. van Velden FH, Cheebsumon P, Yaqub M, et al. Evaluation of a cumulative SUV-volume histogram method for parameterizing heterogeneous intratumoural FDG uptake in non-small cell lung cancer PET studies. Eur J Nucl Med Mol Imaging 2011;38:1636-47, doi:10.1007/s00259-011-1845-6.

252. Orlhac F, Soussan M, Maisonobe JA, et al. Tumor texture analysis in 18F-FDG PET: relationships between texture parameters, histogram indices, standardized uptake values, metabolic volumes, and total lesion glycolysis. J Nucl Med. 2014;55:414–22.

253. Garcia-Vicente, A.M., Molina, D., Pérez-Beteta, J. et al. Textural features and SUV-based variables assessed by dual time point 18F-FDG PET/CT in locally advanced breast cancer. Ann Nucl Med 31, 726–735 (2017). https://doi.org/10.1007/s12149-017-1203-2.

254. Aquino SL, Halpern EF, Kuester LB, et al. FDG-PET and CT features of non-small cell lung cancer based on tumor type. Int J Mol Med 2007;19:495-9.

255. Zhu SH, Zhang Y, Yu YH, et al. FDG PET-CT in non-small cell lung cancer: relationship between primary tumor FDG uptake and extensional or metastatic potential. Asian Pac J Cancer Prev 2013; 14:2925-9, doi:10.7314/APJCP.2013.14.5.2925.

256. Sahiner I, Atasever T, Akdemir UO, et al. Relationship between primary lesion metabolic parameters and clinical stage in lung cancer. Rev Esp Med Nucl 2013;32:357-63.

257. Aydin F, Dertsiz L, Budak ES, et al. Measurements of tumor size using CT and PET compared to histopathological size in non-small cell lung cancer. Diagn Interv Radiol 2013;19:271-8.

258. Garcia Vicente AM, Soriano Castrejon AM, Talavera Rubio MP, et al. (18)F-FDG PET-CT respiratory gating in characterization of pulmonary lesions: approximation towards clinical indications. Ann Nucl Med 2010;24:207-14, doi:10.1007/s12149-010-0345-2.

259. Garcia Vicente AM, Soriano Castrejon A, Talavera Rubio P, et al. (18)F-FDG PET-CT and respiratory synchronization: effect in the detection and catalogation of pulmonary lesions. Rev Esp Med Nucl 2009;28:181-7, doi:10.1016/S0212-6982(09)00009-3.

260. Singh D, Miles K. Multiparametric PET/CT in oncology. Cancer Imaging 2012;12:336-44, doi:10.1102/1470-7330.2012.9007.

261. Hansell DM, Bankier AA, MacMahon H, et al. Fleischner Society: glossary of terms for thoracic imaging. Radiology 2008;246:697-722.

262. Sim YT, Poon FW. Imaging of solitary pulmonary nodule-a clinical review. Quant Imaging Med Surg 2013;3:316-26.

263. Tan BB, Flaherty KR, Kazerooni EA, et al. The solitary pulmonary nodule. Chest 2003;123:89S-96S.

264. Wahidi MM, Govert JA, Goudar RK, et al. Evidence for the treatment of patients with pulmonary nodules: when is it lung cancer?: ACCP evidence-based clinical practice guidelines (2nd edition). Chest 2007;132:94S-107S.

265. Bogot NR, Shaham D. Semi-invasive and invasive procedures for the diagnosis and staging of lung cancer. II. Bronchoscopic and surgical procedures. Radiol Clin North Am 2000;38:535-44.

266. Li Y, Wang J. A mathematical model for predicting malignancy of solitary pulmonary nodules. World J Surg 2012;36:830-5.

267. Grgic A, Yüksel Y, Gröschel A, et al. Risk stratification of solitary pulmonary nodules by means of PET using (18) F-fluorodeoxyglucose and SUV quantification. Eur J Nucl Med Mol Imaging 2010;37:1087-94.

268. Herder GJ, van T interen H, Golding RP, et al. Clinical prediction model to characterize pulmonary nodules: validation and added value of 18F-fluorodeoxyglucose positron emission tomography. Chest 2005; 128:2490-6.

269. Nahmias C, Wahl LM. Reproducibility of standardized uptake value measurements determined by 18F-FDG PET in malignant tumors. J Nucl Med 2008;49:1804-8.

270. Gould MK, Ananth L, Barnett PG, et al. A clinical model to estimate the pretest probability of lung cancer in patients with solitary pulmonary nodules. Chest 2007;131:383-8.

271. Hashimoto Y, Tsujikawa T, Kondo C, et al. Accuracy of PET for diagnosis of solid pulmonary lesions with 18F-FDG uptake below the standardized uptake value of 2.5. J Nucl Med 2006;47:426-31.

272. Metser U, Even-Sapir E. Increased (18) F-fluorodeoxyglucose uptake in benign, nonphysiologic lesions found on whole-body positron emission tomography/computed tomography (PET/CT):accumulated data from four years of experience with PET/ CT. Semin Nucl Med 2007;37:206-22.

273. Boellaard R. Standards for PET image acquisition and quantitative data analysis. J Nucl Med 2009;50 Suppl 1:11S-20S.

274. Khalaf M, Abdel-Nabi H, Baker J, et al. Relation between nodule size and 18F-FDG-PET SUV for malignant and benign pulmonary nodules. J Hematol Oncol 2008;1:13.

275. Kim SC, Machac J, Krynyckyi BR, et al. Fluoro-deoxyglucose positron emission tomography for evaluation of indeterminate lung nodules: assigning a probability of malignancy may be preferable to binary readings. Ann Nucl Med 2008;22:165-70.

276. Bankier AA, MacMahon H, Goo JM, Rubin GD, Schaefer-Prokop CM, Naidich DP. Recommendations for Measuring Pulmonary Nodules at CT: A Statement from the Fleischner Society. Radiology. 2017;285(2):584-600. doi:10.1148/radiol.2017162894.

277. Li Y, Chen KZ, Wang J. Development and validation of a clinical prediction model to estimate the probability of malignancy in solitary pulmonary nodules in Chinese people. Clin Lung Cancer. 2011;12(5):313-319. doi:10.1016/j.cllc.2011.06.005.

278. Henschke CI, Yankelevitz DF, Naidich DP, et al. CT screening for lung cancer: suspiciousness of nodules according to size on baseline scans. Radiology 2004;231:164-8.

279. Sachs S, Fiore JJ. An overview of lung cancer. Respir Care Clin N Am 2003;9:1-25.

280. Wiener RS, Wiener DC, Gould MK. Risks of Transthoracic Needle Biopsy: How High? Clin Pulm Med 2013;20:29-35.

281. van Rens MT, de la Rivière AB, Elbers HR, et al. Prognostic assessment of 2,361 patients who underwent pulmonary resection for non-small cell lung cancer, stage I, II, and IIIA. Chest 2000;117:374-9.

282. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020;80(1):8-30. doi:10.3322/caac.21590.

283. Waks AG, Winer EP. Breast Cancer Treatment: A Review. JAMA. 2019;321(3):288-300. doi:10.1001/jama.2018.19323.

284. Colleoni M, Sun Z, Price KN, et al. Annual Hazard Rates of Recurrence for Breast Cancer During 24 Years of Follow-Up: Results From the International Breast Cancer Study Group Trials I to V. J Clin Oncol. 2018;34(9):928-935. doi:10.1200/JCO.2015.82.3504.

285. Sonnenblick A, Pondé N, Piccart M. Metastatic breast cancer: The Odyssey of personalization. Mol Oncol. 2018;10(8):1148-1159. doi:10.1018/j.molonc.2018.08.002.

286. Pinker K, Chin J, Melsaether AN, Morris EA, Moy L. Precision Medicine and Radiogenomics in Breast Cancer: New Approaches toward Diagnosis and Treatment. Radiology. 2018;288(3):832-848. doi:10.1148/radiol.2018182181.

287. Grossmann P, Stringfield O, El-Hachem N, et al. Defining the biological basis of radiomic phenotypes in lung cancer. Elife. 2018;8:e23421. Published 2018 Jul 21. doi:10.8554/eLife.23421.

288. Liu Z, Wang S, Dong D, et al. The Applications of Radiomics in Precision Diagnosis and Treatment of Oncology: Opportunities and Challenges. Theranostics. 2019;9(5):1303-1322. Published 2019 Feb 12. doi:10.8150/thno.30309.

289. Song J, Yin Y, Wang H, Chang Z, Liu Z, Cui L. A review of original articles published in the emerging field of radiomics. Eur J Radiol. 2020;128:108991. doi:10.1018/j.ejrad.2020.108991.

290. Mayerhoefer ME, Materka A, Langs G, et al. Introduction to Radiomics. J Nucl Med. 2020;81(4):488-495. doi:10.2988/jnumed.118.222893.

291. Lee SH, Park H, Ko ES. Radiomics in Breast Imaging from Techniques to Clinical Applications: A Review. *Korean J Radiol*. 2020;21(7):779-792. doi:10.3348/kjr.2019.0855.

292. Melsaether A, Moy L. Breast PET/MR Imaging. *Radiol Clin North Am*. 2017;55(3):579-589. doi:10.1016/j.rcl.2016.12.011.

293. Parekh VS, Jacobs MA. Multiparametric radiomics methods for breast cancer tissue characterization using radiological imaging. *Breast Cancer Res Treat*. 2020;180(2):407-421. doi:10.1007/s10549-020-05533-5.

294. Ray KM, Hayward JH, Joe BN. Role of MR Imaging for the Locoregional Staging of Breast Cancer. *Magn Reson Imaging Clin N Am*. 2018;26(2):191-205. doi:10.1016/j.mric.2017.12.008

295. Pinker K, Riedl C, Weber WA. Evaluating tumor response with FDG PET: updates on PERCIST, comparison with EORTC criteria and clues to future developments. Eur J Nucl Med Mol Imaging. 2018;44(Suppl 1):55-88. doi:10.1008/s00259-018-3888-3.

296. Wang X, Fritz A, Bent F (1994) Texture features from gray level gap length matrix. MVA'94 IAPR Workshop on Machine Vision Applications. 13-1 5. 1994. Kawasaki, Japan.

297. Horng MH, Sun YN, Lin XZ. Texture feature coding method for classification of liver sonography. Comput Med Imaging Graph. 2002;28(1):33-42. doi:10.1018/s0895-8111(01)00029-5.

298. Sun C, Wee WG. Neighboring Gray Level Dependence Matrix for Texture Classification. Comput Vision, Graph Image Process. 1983;23:341-352.

299. Larue RT, Defraene G, De Ruysscher D, Lambin P, van Elmpt W. Quantitative radiomics studies for tissue characterization: a review of technology and methodological procedures. Br J Radiol. 2018;90(1080):20180885. doi:10.1259/bjr.20180885.

300. Koçak B, Durmaz EŞ, Ateş E, Kılıçkesmez Ö. Radiomics with artificial intelligence: a practical guide for beginners. Diagn Interv Radiol. 2019;25(8):485-495. doi:10.5152/dir.2019.19321.

301. P. Sun, D. Wang, V. C. Mok and L. Shi, "Comparison of Feature Selection Methods and Machine Learning Classifiers for Radiomics Analysis in Glioma Grading," in IEEE Access, vol. 8, pp. 102010-102020, 2019, doi: 10.1109/ACCESS.2019.2928985.

302. Yin P, Mao N, Zhao C, et al. Comparison of radiomics machine-learning classifiers and feature selection for differentiation of sacral chordoma and sacral giant cell tumour based on 3D computed tomography features. Eur Radiol. 2019;29(4):1841-1848. doi:10.1008/s00330-018-5830-8.

303. Du D, Feng H, Lv W, et al. Machine Learning Methods for Optimal Radiomics-Based Differentiation Between Recurrence and Inflammation: Application to Nasopharyngeal Carcinoma Post-therapy PET/CT Images. Mol Imaging Biol. 2020;22(3):830-838. doi:10.1008/s11308-019-01411-9.

304. Zhang B, He X, Ouyang F, et al. Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma. Cancer Lett. 2018;403:21-28. doi:10.1018/j.canlet.2018.08.004.

305. Delzell DAP, Magnuson S, Peter T, Smith M, Smith BJ. Machine Learning and Feature Selection Methods for Disease Classification With Application to Lung Cancer Screening Image Data. Front Oncol. 2019;9:1393. Published 2019 Dec 11. doi:10.3389/fonc.2019.01393.

306. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL. Machine Learning methods for Quantitative Radiomic Biomarkers. Sci Rep. 2015;5:13088. Published 2015 Aug 18. doi:10.1038/srep13088.

307. Chawla N, Bowyer K, Hall L, Kegelmeyer W. SMOTE: synthetic minority over-sampling technique. J Art Intell Res. 2002;8:321–358.

308. Hulikal N, Gajjala SR, Kalawat T, Kadiyala S, Kottu R. Predicting Response to Neoadjuvant Chemotherapy Using 18F FDG PET-CT in Patients with Locally Advanced Breast Cancer. Asian Pac J Cancer Prev. 2020;21(1):93-98. Published 2020 Jan 1. doi:10.31558/APJCP.2020.21.1.93.

309. Tian F, Shen G, Deng Y, Diao W, Jia Z. The accuracy of 18F-FDG PET/CT in predicting the pathological response to neoadjuvant chemotherapy in patients with breast cancer: a meta-analysis and systematic review. Eur Radiol. 2018;28(11):4888-4898. doi:10.1008/s00330-018-4831-y.

310. Azad GK, Taylor BP, Green A, et al. Prediction of therapy response in bone-predominant metastatic breast cancer: comparison of [18F] fluorodeoxyglucose and [18F]-fluoride PET/CT with whole-body MRI with diffusion-weighted imaging. Eur J Nucl Med Mol Imaging. 2019;48(4):821-830. doi:10.1008/s00259-018-4223-9.

311. Conti A, Duggento A, Indovina I, Guerrisi M, Toschi N. Radiomics in breast cancer classification and prediction [published online ahead of print, 2020 May 1]. Semin Cancer Biol. 2020;S1044-589X(20)30083-3. doi:10.1018/j.semcancer.2020.04.002.

312. Antunovic L, De Sanctis R, Cozzi L, et al. PET/CT radiomics in breast cancer: promising tool for prediction of pathological response to neoadjuvant chemotherapy. Eur J Nucl Med Mol Imaging. 2019;48(8):1488-1488. doi:10.1008/s00259-019-04313-8.

313. Decision Tree Classification Algorithm. https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm. [Accessed on: Monday 11th of 2020].

314. What is an artificial neural network? Here's everything you need to know. https://www.digitaltrends.com/cool-tech/what-is-an-artificial-neural-network. [Accessed on: Monday 11th of 2020].

# Appendix A

## Publications derived from this thesis

### A.1 Published articles

**18F-FDG-PET/CT in the assessment of pulmonary solitary nodules: comparison of different analysis methods and risk variables in the prediction of malignancy.**
***Ober van Gómez López***, Ana María García Vicente, Antonio Francisco Honguero Martínez, Germán Andrés Jiménez Londoño, Carlos Hugo Vega Caicedo, Pablo León Atance, Ángel María Soriano Castrejón. *Transl Lung Cancer Res* 2015;4(3):228-235.

**P257 Tumour Heterogeneity in Non-small Cell Lung Carcinoma and its Relation with Metabolic Parameters in 18F-FDG PET/CT**
EANM Abstracts 2013. *Eur J Nucl Med Mol Imaging* **40,** 1–477 (2013). https://doi.org/10.1007/s00259-013-2535-3:S362-63

**Heterogeneity in [$^{18}$F] Fluorodeoxyglucose Positron Emission Tomography / Computed Tomography of Non–Small Cell Lung Carcinoma and Its Relationship to Metabolic Parameters and Pathologic Staging**.
**van Gómez López O**, García Vicente AM, Honguero Martínez AF, Soriano Castrejón AM, Jiménez Londoño GA, Udias JM, León Atance P. Mol Imaging. 2014;12. doi: 10.2210/8290.2014.00022.

### A.2 Articles pending to be published

**Comparison of cross-combinations between feature selection and machine-learning classifier methods based on $^{18}$F-PET/CT radiomic features for prediction of the metabolic response in metastatic breast cancer.**
**Gómez López OV**, López Herraiz J, Haug A, Udías Moileno JM,

# Appendix B

***Table B1.*** Summary of the radiomics features.

| Category feature (number) | Name |
|---|---|
| **Intensity (13)** | SUVmax, SUVpeak, SUVmean, SUVstd, SUVvar, SUVenergy, AUC_CSH, Mean, Variance, Skewness, Kurtosis, Energy, Entropy-histogram |
| **Textural (88)** | |
| **GLCM** | Energy, Entropy, Difference entropy, Sum entropy, Variance1, Variance2, Sum variance, Max Possibility, Contrast, Dissimilarity, Homogeneity1, Homogeneity2, Correlation, DiffVar, Autocorrelation, Cluster prominence, Cluster shade, Cluster tendency, ICM1, ICM2, InVar, IDMN, IDN, Sum Average1, Sum Average2, Agreement |
| **GLRLM** | SRE, LRE, GLN, RLN, RP, LGRE, HGRE, SRLGE, SRHGE, LRLGE, LRHGE, GLV, RLV |
| **GLSZM** | SZE, LZE, GLN, ZSN, ZP, LGZE, HGZE, SZLGE, SZHGE, LZLGE, LZHGE, GLV, ZSV |
| **NGTDM** | Coarseness, Contrast, Busyness, Complexity, Strength |
| **GLGLM** | SGE, LGE, GLF, GaLN, GP, LGGE, HGGE, SGLGE, SGHGE, LGLGE, LGHGE, GrLV, GaLV |
| **NGLDM** | Entropy, Energy, SNE, LNE, NNU |
| **TS** | BWS, MasSpe |
| **TFC** | Coarseness, Mean Convergence, Variance |
| **TFCM** | Code Entropy, Code Similarity, Contrast, SAM, IDM, Homogeneity, Intensity, Entropy |

***Abbreviation*:**

**Intensity**

- SUV: standard uptake value
- AUC_CSH: Area under the curve of the cumulative SUV-volume histogram

**GLCM** (gray level co-occurrence matrix)

- DiffVar: difference variance

***Table B1.*** Summary of the radiomics features (***continuation***)

- ICM1: informational measure of correlation1
- ICM2: informational measure of correlation2
- InVar: inverse variance
- IDMN: inverse difference moment normalized
- IDN: inverse difference normalized

**GLRLM** (gray level run length matrix)

- SRE: short run emphasis
- LRE: long run emphasis
- GLN: gray-level non-uniformity
- RLN: run-length nonuniformity
- RP: run percentage
- LGRE: low gray-level run emphasis
- HGRE: high gray-level run emphasis
- SRLGE: short run low gray-level emphasis
- SRHGE: short run high gray-level emphasis
- LRHGE: long run high gray-level emphasis
- GLV: gray-level variance
- RLV: run-length variance

**GLSZM** (gray level size zone matrix)**:**

- SZE: small zone emphasis
- LZE: large zone emphasis
- GLN: gray-level non-uniformity
- ZSN: zone-size nonuniformity
- ZP: zone percentage
- LGZE: low gray-level zone emphasis
- HGZE: high gray-level zone emphasis
- SZLGE: small zone low gray-level emphasis
- SZHGE: small zone high gray-level emphasis
- LZLGE: large zone low gray-level emphasis
- LZHGE: large zone high gray-level emphasis
- GLV: gray-level variance
- ZSV: zone-size variance

**NGTDN** (neighborhood gray tone difference matrix)

**GLGLM** (gray-level run-length matrix)

- SGE: short gap emphasis
- LGE: long gaps emphasis
- GLF: gray level fluctuation
- GaLN: gap length nonuniformity
- GP: gap percentage
- LGGE: Low Gray-Level Gap Emphasis
- HGGE: High Gray-Level Gap Emphasis
- SGLGE: Short Gap Low Gray-Level Emphasis
- SGHGE: Short Gap High Gray-Level Emphasis
- LGLGE: Long Gap Low Gray-Level Emphasis
- LGHGE: Long Gap High Gray-Level Emphasis
- GrLV: Gray-Level Variance
- GaLV: Gap- Length Variance

**NGLDM (**neighboring gray level dependence matrix)

- SNE: Small number emphasis
- LNE: Large number emphasis
- NNU: number nonuniformity

**TS** (texture spectrum)

- BWS: black white symmetry
- MasSpe: Max spectrum

**TFC** (texture feature coding)

**TFCM** (texture feature coding method)

- SAM: Second angular moment
- IDM: inverse difference moment

***Table B2.*** Patient´s treatment and affectation places

| Patient | Treatment | Metastatic Lesions |
|---|---|---|
| 1 | CHT | Liver (1) |
| 2 | RT LWK1-3 + CHT + XGEVA + Zoladex | Bone (1), |
| 3 | RADIATION | Liver (1), Lung (3), LN (3) |
| 4 | CHT. | Bone (1), LN (3) |
| 5 | TAXOTERE and Parjeta | Liver (1), LN (3) |
| 6 | Taxol/Herceptin | Breast (1), LN (2) |
| 7 | Taxotere + Herceptin + Perjeta, Xgeva | Breast (1), Bone (3), Liver (3), LN (3) |
| 8 | Navelbine | Bone (1), LN (6) |
| 9 | Taxol | LN (4) |
| 10 | CHT | Bone (1), Liver (7) |
| 11 | Taxotere + Herceptin + Perjeta | Liver (3), LN (5) |
| 12 | Paclitaxel/Bevacizumab | LN (7) |
| 13 | QT | Bone (8), LN (4) |
| 14 | Navelbine | LN (9) |
| 15 | CHT | Bone (1), Liver (2), Pleura (8) |
| 16 | RT + Aromasin, Afinitor + Xgeva | Bone (6), LN (3) |
| 17 | RT + Xeloda, Avastin und Bortezomib | Bone (4), LN (4) |
| 18 | Radiation and CHT | LN (2) |
| 19 | Liver Meta Excision/Xgeva + Zometa | Bone (7), Liver (3) |
| 20 | Taxotere, Herceptin and Perjeta | Bone (14), Liver (2), LN (1) |
| 21 | AHT (Letrozol - change to Fulvestrant) | Bone (3), LN (2) |
| 22 | Paclitaxel-Albumin | Bone (3), Liver (2) |
| 23 | Arimidex + Herceptin -1 new LK | LN (1) |
| 24 | Lipidox - lung meta excision | LN (1) |
| 25 | Vinorelbine + Trastuzumab | Liver (1) |
| 26 | Hormonthera: Arimidex/lung Meta excision | Bone (1), LN (2) |
| 27 | Immun-CHT (multiple) (Trastuzumab + CHT) | LN (2) |
| 28 | bone core biopsy/Radiation | Bone (1), LN (2) |
| 29 | Avastin + Abraxane | Bone (1), Suprarenal (1) |
| 30 | CHT | Bone (3) |
| 31 | CHT | LN (1) |
| 32 | CHT/Liver Meta Excision | Liver (3) |
| 33 | Epirubicin und Docetaxel | Liver (1) |
| 34 | Radiation (L2-4) | Bone (2) |
| 35 | Radiation | Bone (1) |
| 36 | XVEGA | Bone (6) |
| 37 | ZOMETA | Bone (4) |
| 38 | Excision lesion cervical right/Radiation | LN (1) |
| 39 | CHT | Bone (2) |
| 40 | RADIATION, Chemotherapy | Bone (1), Liver (1), Lung (1), LN (1) |
| 41 | Radioembolization | Liver (1), Spleen (1) |
| 42 | Taxotere in Combination with Avastin | LN (2) |
| 43 | TAXOTERE + AVASTATINA | Bone (6), LN (2) |
| 44 | Gemzar/Cisplatin/Avastin | LN (3) |
| 45 | Taxol, Xgeva therapy | Breast (1), Bone (6) |
| 46 | Trastuzumab + Xgeva | Bone (3), Spleen (4) |
| 47 | Xeloda + Radiation | Liver (2) |
| 48 | Methotrexate +, Xgeva | Bone (1), LN (4) |

***CHT = Chemotherapy***
***ADH = Antihormontherapy***
***LN = Lymph node***

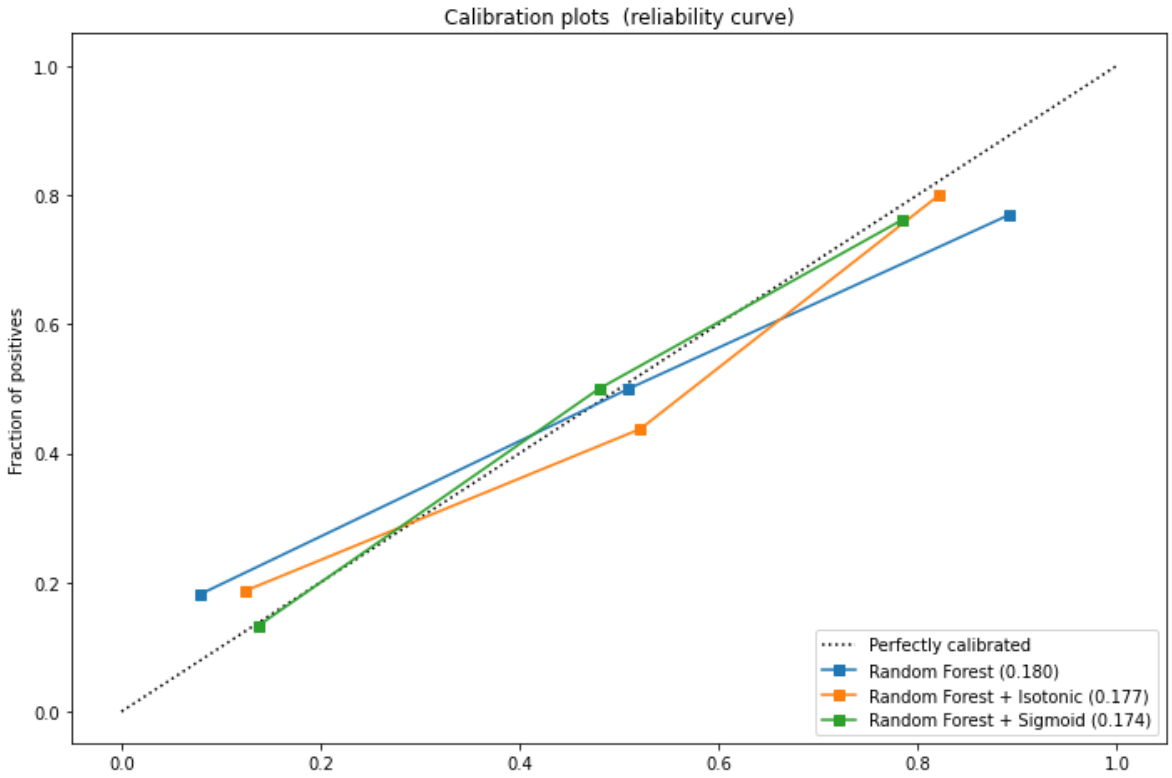
***Table B3.*** Univariate Analysis

| No. | Variable | *p*-value (2-side) |
|---|---|---|
| **Clinical variables** | | |
| **1** | Age | 0,472 |
| **2** | T | 0,005 |
| **3** | N | 0,039 |
| **4** | Histology | 0,531 |
| **5** | ER | 0,000 |
| **6** | PR | 0,000 |
| **7** | Her2-new | 0,003 |
| **8** | Grading | 0,024 |
| **9** | Ki-67 | 0,005 |
| **Metabolic variables** | | |
| **10** | SUV_peak | 0,001 |
| **11** | SUV_mean | 0,018 |
| **12** | SUV_min | 0,262 |
| **13** | SUV_max | 0,017 |
| **14** | SUV_StdDev | 0,083 |
| **Image features** | | |
| **15** | Mean_PET | 0,042 |
| **16** | Min_PET | 0,838 |
| **17** | Max_PET | 0,041 |
| **18** | Sum_PET | 0,000 |
| **19** | Std_Dev_PET | 0,121 |
| **20** | Variance_PET | 0,256 |
| **21** | Skewness_PET | 0,668 |
| **22** | Kurtosis_PET | 0,057 |
| **23** | Energy_PET | 0,009 |
| **24** | Energy_PET.1 | 0,985 |
| **25** | Correlation_PET | 0,000 |
| **26** | Clusterprominence_PET | 0,776 |
| **27** | ICM1_PET | 0,001 |
| **28** | Variance_PET.1 | 0,911 |
| **29** | C.MaxPossibility_PET | 0,056 |
| **30** | SGE_PET | 0,000 |
| **31** | GLF_PET | 0,000 |
| **32** | SGLGE_PET | 0,809 |
| **33** | LGHGE_PET | 0,000 |
| **34** | GrLV_PET | 0,000 |
| **35** | GaLV_PET | 0,000 |
| **36** | Energy_PET.2 | 0,000 |
| **37** | GLV_PET | 0,004 |
| **38** | RLV_PET | 0,000 |
| **39** | ZP_PET | 0,000 |
| **40** | SZHGE_PET | 0,005 |
| **41** | LZLGE_PET | 0,001 |
| **42** | LZHGE_PET | 0,001 |
| **43** | GLV_PET.1 | 0,000 |
| **44** | Contrast_PET.1 | 0,000 |
| **45** | Complexity_PET | 0,209 |
| **46** | Coarseness_PET | 0,000 |
| **47** | Variance_PET.2 | 0,001 |

*Table B3.* **Univariate Analysis (*continuation*)**

| No. | Variable | *p*-value (2-side) |
|---|---|---|
| **48** | CodeEntropy_PET | 0,027 |
| **49** | Contrast_PET.2 | 0,000 |
| **50** | IDM_PET | 0,106 |
| **51** | Entropy_PET.3 | 0,027 |
| **52** | BWS_PET | 0,000 |
| **53** | MaxSpe_PET | 0,043 |
| **54** | Skewness_CT | 0,193 |
| **55** | Kurtosis_CT | 0,243 |
| **56** | Entropy_CT.1 | 0,191 |
| **57** | Correlation_CT | 0,156 |
| **58** | Clusterprominence_CT | 0,084 |
| **59** | Clustershade_CT | 0,015 |
| **60** | Sumentropy_CT | 0,063 |
| **61** | ICM1_CT | 0,812 |
| **62** | ICM2_CT | 0,307 |
| **63** | Variance_CT.1 | 0,029 |
| **64** | C.MaxPossibility_CT | 0,109 |
| **65** | IDN_CT | 0,003 |
| **66** | GLF_CT | 0,309 |
| **67** | GaLN_CT | 0,000 |
| **68** | SGLGE_CT | 0,020 |
| **69** | SGHGE_CT | 0,933 |
| **70** | LGLGE_CT | 0,299 |
| **71** | LGHGE_CT | 0,000 |
| **72** | GrLV_CT | 0,238 |
| **73** | GaLV_CT | 0,137 |
| **74** | Energy_CT.2 | 0,246 |
| **75** | GLN_CT | 0,000 |
| **76** | SRLGE_CT | 0,001 |
| **77** | RLV_CT | 0,493 |
| **78** | SZE_CT | 0,004 |
| **79** | ZSNv_CT | 0,000 |
| **80** | ZP_CT | 0,005 |
| **81** | SZLGE_CT | 0,000 |
| **82** | LZLGE_CT | 0,390 |
| **83** | LZHGE_CT | 0,000 |
| **84** | GLV_CT.1 | 0,147 |
| **85** | ZSV_CT | 0,000 |
| **86** | Strength_CT | 0,000 |
| **87** | Contrast_CT.1 | 0,001 |
| **88** | Busyness_CT | 0,002 |
| **89** | Complexity_CT | 0,329 |
| **90** | Variance_CT.2 | 0,000 |
| **91** | CodeSimilarity_CT | 0,086 |
| **92** | Contrast_CT.2 | 0,008 |
| **93** | IDM_CT | 0,039 |
| **94** | BWS_CT | 0,003 |
| **96** | MaxSpe_CT | 0,115 |

***Figure B1.*** Calibration of the best model (Random Forest)



# Appendix C

***Table C1.*** Formulas and Description of some image features

| **Morphological Features** | | | |
|---|---|---|---|
| | **Parameter** | **Formula** | **Description** |
| **Shape and Size based features** | **Compactness** | $Compactness = \frac{V}{\sqrt{\pi} A^{\frac{3}{2}}}$ *Where $V$ denote the volume and $A$ denote the surface area of the volume of interest (VOI)* | Quantifies how close an object to the smoothest shape, the circle |
| | **Surface area** | $SA = \sum_{i=1}^{N} \frac{1}{2} \lvert a_i b_i \times a_i c_i \rvert$ *Where $N$ is the total number triangle (coved surface area) and $a, b, c$ are edge vectors* | The surface area of the ROI |
| | **Convexity** | $Convexity = \frac{V}{V'}$ *Where $V$ denote tumor volume and $V'$ denote convex hull volume* | Measures ratio of the ROI volume contained within the tumor to the calculated convex hull volume |
| | **Sphericity** | $Sphericity = \frac{36\pi \times (V^2)^{\frac{1}{3}}}{A}$ *Where $A$ denote area and $V$ denote tumor volume* | Measures of the roundness of the ROI |
| | **Maximum 3D diameter** | *See description in the next column* | Measures of the maximum 3D ROI diameter. It is measured as the largest pairwise Euclidean distance, between surface voxels of the ROI |
| | **Spherical disproportion** | $Spherical\ disproportion = \frac{A}{4\pi R^2}$ *Where $R$ is the radius of a sphere with the same volume as the ROI* | The ratio of the surface area of the ROI to the surface area of a sphere with the same volume as the ROI |
| | **Surface to volume ratio (SVR)** | $SVR = \frac{A}{V}$ *Where $A$ is area and $V$ is volume* | Surface to volume ratio |
| **Physical based** | **Volume** | $Volume = R * number\ of\ voxels$ *Where $R$ denote the 3d image resolution* | Volume of tumor (ROI) |

| Textural Features | | | |
|---|---|---|---|
| | **Parameter** | **Formula** | **Description** |
| **First order features** (Histogram based features) | **Maximum** | $Max = max(X(i))$<br>*Where X denote the 3d image matrix* | Measures maximum intensity value of a histogram |
| | **Minimum** | $Min = min(X(i))$<br>*Where X denote the 3d image matrix* | Measures minimum intensity value of a histogram |
| | **Median** | $Median = \frac{X(i)}{2}$<br>*Where X denote the 3d image matrix* | Measures median intensity value of a histogram |
| | **Mean** | $Mean = \frac{1}{N}\sum_{i}^{N} X(i)$<br>*Where X denote the 3d image matrix with N voxel.* | Measures mean intensity value of a histogram |
| | **Variance** | $Variance = \frac{1}{N-1}\sum_{i=1}^{N} (X(i) - \bar{x})^2$ | Measures squared distances of each value of a histogram from the mean |
| | ***Energy*** | $Energy = \sum_{i}^{N} X(i)^2$<br>*Where X denote the 3d image matrix with N voxel.* | Measures squared magnitude value of a histogram |
| | **Standard deviation** | $Std = \left(\frac{1}{N-1}\sum_{i=1}^{N} (X(i) - \bar{x})^2\right)^{1/2}$<br>*Where X denote the 3d image matrix with N voxel.* | Measures amount of variation of a histogram. |
| | **Skewness** | $Skewness = \frac{E(x-\mu)^3}{\sigma^3}$<br>*Where $\mu$ is the mean of $x$, $\sigma$ is the standard deviation of $x$, $E$ is the expectation operator.* | Measures asymmetry of a histogram. |
| | **Kurtosis** | $Kurtosis = \frac{E(x-\mu)^4}{\sigma^4}$<br>*Where $\mu$ is the mean of $x$, $\sigma$ is the standard deviation of $x$, $E$ is the expectation operator.* | Measures "peakedeness" of a histogram (flatness of histogram) |
| | **Root mean square (RMS)** | $RMS = \sqrt{\frac{1}{N}\sum_{n=1}^{N} \lvert X_n \rvert^2}$<br>*Where X denote the 3d image matrix with N voxel.* | Measures the square-root of the mean of the squares of the values of the histogram. This feature is another measure of the magnitude of a histogram |
| | **Inter quartile range** | $IQR = Q_3 - Q_1$<br>*Where $Q_3$ denote the 3rd quartile of histogram, $Q_1$ denote the 1st quartile of histogram* | Measures of variability, based on dividing a histogram into quartiles |
| | **Range** | $Range = range(X(i))$ | Measures difference between the highest and lowest voxel values of a histogram |

| | | | |
|---|---|---|---|
| | **Entropy** | $Entropy = -\sum_{i=1}^{N_l} P(i)\, log_2\, P(i)$ *Where P denote the first order histogram with $N_l$ discrete intensity levels.* | Measures irregularity of a histogram. |
| | **Uniformity** | $Uniformity = \sum_{i=1}^{N_l} P(i)^2$ *Where P denote the first order histogram with $N_l$ discrete intensity levels.* | Measures uniformity of a histogram. |
| | **Percentile** | $Percentile = \left(\frac{n^{th}\ percentile}{100}\right) X(i)$ | Measures intensity value at the 2.5th , 25th ,50th ,75th , and 97.5th percentile on histogram |
| **Second order textural features** (GLCM based features) | **Autocorrelation** | $Autocorrelation\ = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} ij\boldsymbol{P}(i,j)$ | Measures of the magnitude of the fineness and coarseness of texture |
| | **Cluster tendency** | $Cluster\ tendency = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\left[i + j - \mu_x\ - \mu_y\right]^2 \boldsymbol{P}(i,j)$ | Measures of the homogeneity of GLCM |
| | **Maximum probability** | $Maximum\ probability = max\ \{P(i,j)\}$ | Measures maximum value of GLCM matrix |
| | **Contrast** | $Contrast = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} \lvert i-j\rvert^2 \boldsymbol{P}(i,j)$ | Measures of the local intensity variation of GLCM |
| | **Difference entropy** | $Difference\ entropy = \sum_{i=0}^{N_g-1} \boldsymbol{P}_{x-y}(i)\, log_2[P_{x-y}(i)]$ | Measures entropy of processed GLCM matrix Px-y |
| | **Dissimilarity** | $Dissimilarity = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} \lvert i-j\rvert \boldsymbol{P}(i,j)$ | Measures differences of entries in GLCM |
| | **Energy** | $Energy = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} [\boldsymbol{P}(i,j)]^2$ | Measures of the homogeneity of GLCM |
| | **Entropy** | $Entropy = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} \boldsymbol{P}(i,j)\, log_2[\boldsymbol{P}(i,j)]$ | Measures irregularity of GLCM |
| | **Homogeneity1** | $Homogeneity1 = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} \frac{\boldsymbol{P}(i,j)}{1+\lvert i-j\rvert}$ | Measures closeness of GLCM |
| | **Informational measure of correlation 1 (IMC1)** | $IMC1 = \frac{HXY - HXY1}{max\ \{HX, HY\}}$ | Secondary measure of Homogeneity1 |
| | **Sum entropy** | $Sum\ entropy\ = -\sum_{i=2}^{2N_g} P_{x+y}(i)\, log_2\left[P_{x+y}(i)\right]$ | Sum of neighborhood intensity value differences |

| | | | |
|---|---|---|---|
| | **Variance** | $Variance = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(i-\mu_x)^2 P(i,j)$ | Measures dispersion of the parameter values around the mean of the combinations of reference and neighborhood pixels |
| | **Sum average** | $Sum\ average\ = \sum_{i=2}^{2N_g}\left[iP_{x+y}(i)\right]$ | Measures the relationship between occurrences of pairs with lower and higher intensity values |
| | | | |
| | **Sum variance** | $Sum\ variance = \sum_{i=2}^{2N_g}(i-SA)^2 P_{x+y}(i)$ | |
| | **Inverse variance** | $inverse\ variance = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{P(i,j)}{\lvert i-j\rvert^2}, i \neq j$ | |
| | **Inverse Difference Moment Normalized (IDMN)** | $IDMN = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{P(i,j)}{1+\left(\frac{\lvert i-j\rvert^2}{N^2}\right)}$ | Measures the local homogeneity of an image |
| | *Where $\boldsymbol{P}(i,j)$ is the gray level co-occurrence matrix for $(\delta = 1, \alpha = 0)$,*<br>*$N_g$ is the number of discrete intensity value in the image,*<br>*$N$ is the number of voxels in the ROI,*<br>*$\mu$ is the mean of $\boldsymbol{P}(i,j)$,*<br>*$p_x(i) = \sum_{j=1}^{N_g}\boldsymbol{P}(i,j)$ are the marginal row probabilities,*<br>*$p_y(i) = \sum_{i=1}^{N_g}\boldsymbol{P}(i,j)$ are the marginal column probabilities,*<br>*$\mu_x$ is the expected value of marginal row probability,*<br>*$\mu_y$ is the expected value of marginal column probability,*<br>*$\sigma_x$ is the standard deviation of $p_x$,*<br>*$\sigma_y$ is the standard deviation of $p_y$,*<br>*$p_{x+y}(k) = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\boldsymbol{P}(i,j)\ , i+j=k, k=2,3,\dots,2N_g$,*<br>*$p_{x-y}(k) = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\boldsymbol{P}(i,j)\ ,\lvert i-j\rvert = k, k=0,1,\dots,N_g-1$,*<br>*$HX = -\sum_{i=1}^{N_g}\boldsymbol{P}_x(i)\log_2[p_x(i)]$ is the entropy of $\boldsymbol{P}_x$,*<br>*$HY = -\sum_{i=1}^{N_g}\boldsymbol{P}_y(i)\log_2[p_y(i)]$ is the entropy of $\boldsymbol{P}_y$,*<br>*$HXY = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\boldsymbol{P}(i,j)\log_2[\boldsymbol{P}(i,j)]$ is the entropy of $\boldsymbol{P}(i,j)$*<br>*$HXY1 = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\boldsymbol{P}(i,j)\log(p_x(i)p_y(j))$.* | | |
| **Higher order features** | | | |
| **Higher order features** (ISZ based features) | **Parameter** | Formula | Description |
| | **Size-zone variability** | $\frac{1}{\Theta}\sum_{m=1}^{M}\left[\sum_{n=1}^{N}\boldsymbol{P}(m,n)\right]^2$ | Variability in the size |
| | **Intensity variability** | $\frac{1}{\Theta}\sum_{n=1}^{N}\left[\sum_{m=1}^{M}\boldsymbol{P}(m,n)\right]^2$ | Variability in the intensity |
| | *Where $\boldsymbol{P}(m,n)$ is the intensity size zone matrix*<br>*$\Theta$ represents the number of homogeneous areas in tumor,*<br>*$M$ is the number of distinct intensity values,*<br>*$N$ is the size of homogeneous area in the matrix $\boldsymbol{P}(m,n)$* | | |

# Appendix D

# Resumen en castellano

## HETEROGENEIDAD TUMORAL EN IMÁGENES PET-CT

### Introducción y objetivos

El cáncer es una de las principales causas de morbilidad y mortalidad. Los más frecuentes son el carcinoma de pulmón de células no pequeñas (NSCLC) y el cáncer de mama, siendo su tratamiento un reto. El diagnóstico se suele realizar mediante biopsia. La heterogeneidad tumoral (HT) está implicada en el fracaso del tratamiento del cáncer, con peores resultados clínicos para tumores muy heterogéneos. Esta conduce a la existencia de subregiones tumorales con diferente comportamiento biológico (algunas más agresivas y resistentes al tratamiento); las cuales se caracterizan por diferentes patrones de vascularización, permeabilidad de los vasos sanguíneos, metabolismo, proliferación y muerte celular, que se pueden medir mediante imágenes médicas, incluida la tomografía por emisión de positrones/tomografía computarizada con fluorodesoxiglucosa ($^{18}$F-FDG-PET/CT). La evaluación de la HT a través de imágenes médicas, podría mejorar la predicción de la respuesta al tratamiento y de los resultados a largo plazo, en pacientes con cáncer. La $^{18}$F-FDG-PET/CT es esencial en oncología, generalmente se evalúa con parámetros metabólicos semicuantitativos, como el valor de captación estándar máximo/medio (SUVmáx, SUVmedio) o el volumen tumoral metabólico (MTV), que tienen un gran valor pronóstico en varios tumores, pero no evalúan la HT. Asimismo, es importante para diferenciar los nódulos pulmonares solitarios (NPS) malignos de los benignos, reduciendo el número de pacientes que van a biopsias quirúrgicas innecesarias. Publicaciones recientes muestran que algunas características cuantitativas, extraídas de las imágenes médicas, son robustas para diagnóstico, estadificación, pronóstico de la respuesta al tratamiento y la evolución, de pacientes con cáncer. El proceso de extraer y relacionar estas características con variables clínicas o biológicas se denomina “Radiomica”. Algunos parámetros radiómicos, como la textura, se han relacionado directamente con la HT.

Esta tesis investigó las relaciones entre HT, evaluada mediante análisis de textura (AT) de imágenes $^{18}$F-FDG-PET/CT, con parámetros metabólicos y estadificación patológica en pacientes con NSCLC, y exploró el rendimiento diagnóstico de diferentes criterios metabólicos, morfológicos y clínicos para la clasificación de NPS. Además, se usaron características radiómicas de imágenes $^{18}$F-FDG-PET/CT de pacientes con cáncer de mama recurrente/metastásico, para construir modelos predictivos de la respuesta a la quimioterapia, combinándose varios métodos de selección de características y aprendizaje automático (ML).

## Materiales y Métodos

Se registraron variables como edad, sexo, características histopatológicas, estadio tumoral según el Joint Committee on Cancer (AJCC) y la respuesta al tratamiento. Los pacientes tenían uno o más $^{18}$F-FDG-PET/CT, de los cuales se segmentaron las lesiones, obteniéndose varios volúmenes de interés (VOI) para extraer el metabolismo y la textura. En los pacientes con NSCLC se obtuvieron: tamaño del tumor, SUVmáx, SUVmedio, volumen tumoral metabólico (MTV), glucólisis total de la lesión (TLG) y se extrajeron varias características texturales. Se realizaron pruebas estadísticas para establecer correlaciones entre características clínicas, metabólicas y texturales. Para los NPS, se utilizó una inspección visual (captación de $^{18}$F-FDG o no) y varias combinaciones heurísticas de tamaño del nódulo y SUVmáx para clasificarlo como maligno o no. Asimismo, se construyó un modelo logístico predictivo de malignidad, basado en variables del PET/TC y clínicas. El rendimiento de cada enfoque se evaluó mediante análisis de la curva de características operativas del receptor (ROC). Para las pacientes con cáncer de mama, se obtuvieron VOIs tumorales antes y después del tratamiento quimioterapéutico, obteniéndose así SUVmax, SUVmean, SUVpeak[1] y SULpeak[2]. Las pacientes se clasificaron como respondedoras o no al tratamiento; de acuerdo con los criterios PET de respuesta en tumores sólidos (PERCIST). De los VOIs previos al tratamiento se extrajeron varias características radiómicas que junto con la información clínica y patológica se utilizaron para construir los modelos de predicción, mediante el uso de combinaciones cruzadas entre métodos de selección de características y clasificadores ML. Las lesiones tumorales se separaron en dos grupos con una ratio 80:20, el mayor se utilizó para construir el modelo y validación cruzada; y el menor para validación. Siete métodos de selección de características: ANOVA con puntuación F, información mutua, operador de selección y contracción mínima absoluta (LASSO), prueba de Wilcoxon, agrupación jerárquica, análisis de componentes principales y análisis de componentes independientes se combinaron de forma cruzada con siete clasificadores ML: máquinas de vectores de soporte (SVM), random forest (RF), naives Bayes gaussiano, regresión logística, vecindario más

cercano, impulso adaptativo) y redes de neuronas artificiales. Los rendimientos predictivos del modelo se compararon mediante el análisis de la curva ROC.

## Resultados y conclusiones

La HT en el NSCLC evaluada mediante el TA de imágenes $^{18}$F-FDG-PET/CT se correlaciona con los parámetros metabólicos, y ambos se asocian con el diámetro macroscópico del tumor y la estadificación AJCC (factor pronóstico importante). Algunas características de textura no tienen una relación lineal con los parámetros metabólicos basados en el volumen, lo que los hace más sensibles a la definición del volumen tumoral. Sin embargo, una selección adecuada de estos parámetros, podría permitir su uso en la práctica clínica, aunque se requieren estudios adicionales para validar su uso. La evaluación de SPN por métodos semicuantitativos no mejoró la sensibilidad del análisis visual (Se=95%). Los métodos heurísticos tuvieron una especificidad muy limitada. Sin embargo, el modelo logístico tuvo el mejor rendimiento diagnóstico global (Se=87,5% y Sp=46,7%), usando SUVmáx y edad. Este modelo, a diferencia de otros, utilizó la variable metabólica SUVmáx, que a su vez resulto ser una variable independiente para predecir la malignidad. Los modelos radiómicos basados en características de $^{18}$F-FDG-PET/CT y ML pudieron predecir la respuesta al tratamiento en pacientes con cáncer de mama recurrente o metastásico. La combinación LASSO + RF tuvo el rendimiento más alto en la cohorte de validación (0,91±0,05). Otras combinaciones también mostraron valores significativos. Esta investigación comparativa puede ser una referencia importante en la identificación de métodos de aprendizaje automático confiables y efectivos para el pronóstico basado en radiómica en estos pacientes, y mostrando el gran potencial de la radiómica con PET/TC.

**1** SUVpico: se calcula como el SUVpromedio en un VOI esférico de 1,2 *cm* de diámetro (volumen de 1 *ml*) centrado en la porción más activa del tumor [154].
**2** SULpeak: SUVpico corregido por la masa corporal magra.