



ELSEVIER

Contents lists available at ScienceDirect

## Critical Reviews in Oncology / Hematology

journal homepage: [www.elsevier.com/locate/critrevonc](http://www.elsevier.com/locate/critrevonc)

## Radiomics: Principles and radiotherapy applications

I. Gardin<sup>a,b,\*</sup>, V. Grégoire<sup>c</sup>, D. Gibon<sup>d</sup>, H. Kirisli<sup>d</sup>, D. Pasquier<sup>e</sup>, J. Thariat<sup>f</sup>, P. Vera<sup>a,b</sup><sup>a</sup> Department of Nuclear Medicine, Centre Henri-Becquerel, France<sup>b</sup> LITIS EA4108, Normandie University, Rouen, France<sup>c</sup> Department of Radiation Oncology, Centre Léon Bérard, France<sup>d</sup> Research and Innovation Department, AQUILAB, Loos Les Lille, France<sup>e</sup> Department of Radiation Oncology, Centre Oscar Lambret, CRISAL UMR CNRS 9189, Lille University, Lille, France<sup>f</sup> Radiotherapy Department, Centre François Baclesse, Caen, France

## ARTICLE INFO

## Keywords:

Radiomics

Radiotherapy

Chemoradiotherapy

Medical imaging

## ABSTRACT

Radiomics is defined as the extraction of a large quantity of quantitative image features. The different radiomic indexes that have been proposed in the literature are described as well as the various factors that have an impact on the robustness of these indexes. We will see that several hundred quantitative features can be extracted per lesion and imaging modality. The ever-growing number of features studied raises the question of the statistical method of analysis used.

This review addresses the research supporting the clinical use of radiomics in oncology in the staging of disease, discrimination between healthy and pathological tissues, the identification of genetic features, the prediction of patient survival, the response to treatment, the recurrence after radiotherapy and chemoradiotherapy and the side effects.

Based on the existing literature, it remains difficult to identify features that should be used for current clinical practice.

## 1. Concept of radiomics

Medical imaging, such as computed tomography (CT), positron emission tomography (PET) using FDG radioactive glucose analogue, and magnetic resonance imaging (MRI) are used routinely in the clinical management of cancer patients. They tend to play a leading role in personalised treatment based on tumour imaging. Personalised medicine has been largely developed using invasive techniques based on genomics and proteomics. However, there is a spatial and temporal heterogeneity of tumour features (Marusyk et al., 2012) that makes it difficult to repeat biopsies, both in space and time, where imaging can readily give a spatial representation of the lesion and has the advantage of being easy to repeat.

Radiomics is defined as the extraction of a large quantity of quantitative features from the image (Kumar et al., 2012). It allows highlighting the heterogeneity, texture, and shape of the lesion corresponding to tissue characteristics such as necrosis, metabolic activity, tissue density, and vascularity (Alic et al., 2014). Medical interpretation of images is generally based on a simple visual interpretation of the contrast. Although this type of interpretation has proven to be very effective in the management of cancer patients, it remains qualitative

and subjective, while quantitative features are needed for reasons of objectivity and reproducibility.

There are two ideas underlying the concept of radiomics. The first is that tumour features at the tissue, cellular, and/or genomic level would also be highlighted in medical imaging (Yip and Aerts, 2016). It is considered that image features are strongly correlated with clinical and/or biological features. The second rationale is that the information obtained by the image would be complementary to the other data sources, thus increasing the number of features (Lambin et al., 2012) and the knowledge of the tumour.

The aim is to highlight recent results of the contribution of radiomics to radiotherapy in recent years and to assess the current state of the literature in this field. In the first part of this paper, we will describe the different radiomic features that have been proposed in the literature, the various factors that have an impact on the robustness of these indexes, and the most salient clinical results. We will see that several hundred quantitative features can be extracted by lesion and imaging modality, requiring the use of statistical tools to determine which of these features are relevant. We will conclude with the current limitations of radiomics and the perspectives that this new concept reveals.

\* Corresponding author at: Centre Henri-Becquerel 1, rue d'Amiens, 76038, Rouen Cedex, France.

E-mail address: [Isabelle.gardin@chb.unicancer.fr](mailto:Isabelle.gardin@chb.unicancer.fr) (I. Gardin).

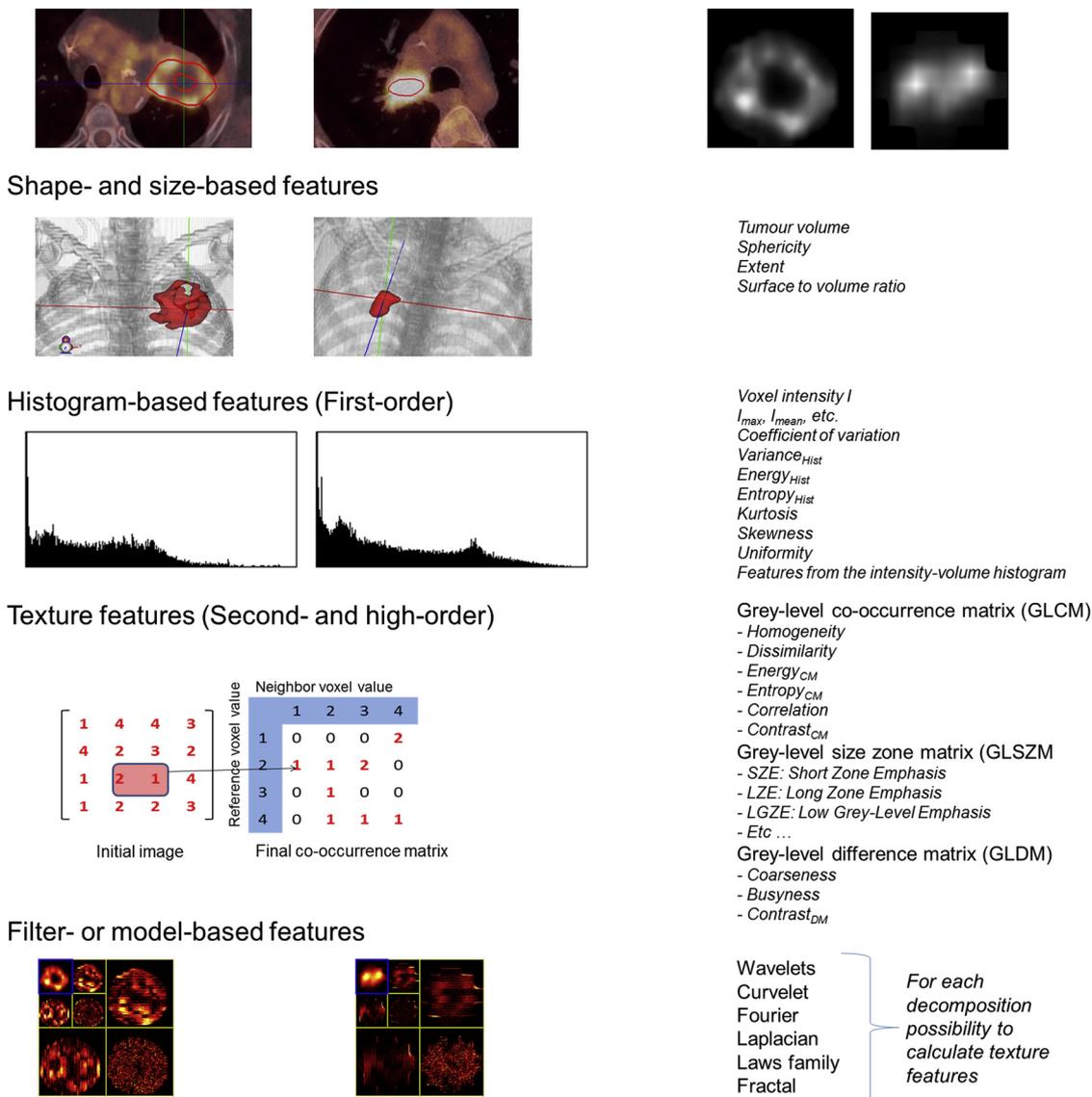


Fig. 1. Different categories of radiomic features and non-exhaustive list of associated features. Example with two lung lesions in PET/CT imaging. The filter-based example corresponds to wavelet decomposition.

## 2. Image features

Numerous image features have been proposed in the literature (Sollini et al., 2017). Features based on the shape and size of the lesion, histogram features based on first-order statistics, texture features, and filter- and model-based features are generally distinguished (see Fig. 1).

### 2.1. Shape- and size-based features

In oncology, it is generally considered that a small tumour volume (TV) leads to a better prognosis than a large lesion. The reduction in lesion volume in CT imaging, assessed by measuring the longest length of the lesion, is a predictive factor for treatment response (RECIST 1.1 criteria (Eisenhauer et al., 2009)). A review of the literature from Van de Wiele et al. (van De Wiele et al., 2013) lists numerous publications showing the predictive and prognostic value of the TV determined using FDG PET images for solid tumours. Tumour volume is thus an imaging feature of the first choice.

Similarly, in oncology, spherical lesions with sharp edges are frequently considered less aggressive than diffuse and infiltrating tumours. Thus, imaging features corresponding to a metric of this phenomenon

have been investigated, such as the sphericity and the extent. Hofheinz et al. showed the prognostic value of the sphericity on pre-therapeutic FDG-positive lesions in PET imaging in a cohort of 37 patients with head and neck tumours (H&N) (Hofheinz et al., 2014).

El Naqa et al. showed that the extent, in combination with other FDG PET imaging features, was involved in predicting the response to chemoradiotherapy (CRT) of H&N tumours (El Naqa et al., 2009).

### 2.2. Histogram-based features

In imaging, the basic feature is the intensity (I) of the voxel. In CT, this corresponds to the Hounsfield units (HU) linked to the tissue density travelled by the X-rays, which is widely used for the delineation of the gross tumour volume (GTV) and the organs at risk (OAR). In PET imaging, I corresponds to a normalised FDG uptake, called SUV for standard uptake value, representing the consumption of glucose by cells based on the fact that tumour cells need more energy than healthy cells. The maximum uptake value,  $SUV_{max}$ , is the most common feature used in PET imaging. FDG-positive tissues are used to delineate the biological target volume proposed for dose painting in RT in the setting of clinical trial (Differding et al., 2015). Furthermore, high FDG uptake

areas seen on pre-treatment PET/CT scans can identify tumour sub-volumes at greater risk of relapse in patients with non-small cell lung cancer (NSCLC) treated using concomitant CRT (Calais et al., 2015). In MRI, the interpretation of I is more complex due to the diversity of possible sequences and the multiple signal origins (proton density, relaxation times, diffusion coefficient, vascularity, etc.). In particular, in diffusion-weighted imaging (DWI), the intensity corresponds to the apparent diffusion coefficient (ADC) based on the fact that tumour tissues are composed of anarchic cell proliferation, where water diffusion is usually reduced. On the other hand, dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) allows the visualisation of tumour vascularisation (Kumar et al., 2012; Asselin et al., 2012).

Descriptive statistics, also called first-order statistics, describe voxel intensities in the volume of interest, here the TV, corresponding to the measure of central tendency and variability. The main imaging feature used to measure the central tendency is the mean ( $I_{\text{mean}}$ ), while the measure of variability includes the standard deviation (ST) or the variance, the coefficient of variation (COV) defined as ST divided by  $I_{\text{mean}}$ . Bundschuh et al. showed that the tumour heterogeneity of FDG uptake measured by the coefficient of variation was predictive of the response to CRT in rectal cancer (Bundschuh et al., 2014).

The frequency histogram is a simple and concise summary of first-order statistical information. Its shape reveals the image homogeneity in the TV. The histogram of a homogeneous lesion is distributed on a narrow peak, whereas a heterogeneous image corresponds to a broad peak of intensities. The general features of intensity distribution are asymmetry (skewness), flattening (kurtosis), energy, and entropy. By analogy with dose-volume histograms used in RT, El Naqa et al. have proposed an intensity-volume histogram to characterise tumour heterogeneity, enabling several indexes to be extracted, such as  $I_x$  (the minimum intensity to the x% highest intensity volume) and  $V_x$  (the percentage volume with at least x% intensity value) (El Naqa et al., 2009). In addition to the feature of the form corresponding to extent, El Naqa et al. showed in their small cohort of patients (9) with H&N cancer treated by CRT that  $V_{90}$  measured on the pre-treatment FDG PET image had the highest univariate predictive power of survival. King et al. (2013) found in H&N cancer that tumours that responded poorly to CRT showed a significant increase in the mean ADC values on DWI during treatment.

### 2.3. Texture features

The benefit of first-order indexes is that they are simple to calculate when the TV is delineated. However, they do not provide any information about the voxel intensity relative to each other. For this reason, it was proposed to use higher-order statistical methods to carry out a more in-depth analysis of the relationships between voxels, such as texture features. Four main texture matrices have been proposed in the literature. The grey-level co-occurrence matrices (GLCM) characterise the intensity relationships between pairs of neighbouring voxels in all spatial directions (Haralick et al., 1973). The grey-level difference matrix (GLDM) (Amadasun and King, 1989) characterises the intensity differences between neighbours. The grey-level run length (GLRLM) (Galloway, 1975) and the grey-level size zone (GLSZM) matrices (Thibault et al., 2009) characterise the range of intensities in one direction or in all directions, respectively. The GLRLM features are highly correlated with the GLSZM features and therefore do not provide complementary information (Tixier et al., 2011). Thus, in 3D imaging, the GLRL matrix is not of particular interest and is generally not used.

Tixier et al. (2011) studied the predictive value of the response to CRT of 38 texture features in addition to first-order features in a cohort of 41 patients with oesophageal cancer. Their results showed that features of textures enable to differentiate patients better than those of the first order.

### 2.4. Filter- or model-based features

Other approaches based on prior image transformation have been proposed in the literature, mainly on CT and MRI data, defined in spatial domains such as filters (Gabor filters, Law's filters, and Laplacian transform of Gaussian filters (LoG filters) or in the frequency domains (Fourier and wavelet transform) (Kumar et al., 2012). Other features derived from complex mathematical models have also been proposed in the literature, such as fractal dimension measuring image complexity (Lopes and Betrouni, 2009).

In order to study the predictive value of distant metastasis (DM) and the prognostic values of radiomic features, Coroller et al. (2015) extracted 635 indexes from planning CT in 98 patients treated using CRT for lung cancer for analysis and 84 patients for validation. Of the 635 features, 180 were extracted from LoG filters and 384 from wavelet transform maps. The authors found 12 prognostic features and 35 predictive features of DM occurrence, including several LoG filters and wavelet transform features.

## 3. Medical applications

Various reviews of the literature (Marusyk et al., 2012; Kumar et al., 2012; Yip and Aerts, 2016; Sollini et al., 2017; Scalco and Rizzo, 2017) show that many studies support the clinical interest of radiomics in oncology, both in MR, CT, and FDG PET imaging, particularly for patients treated using radiotherapy for many localisations (NSCLC, H&N, oesophagus, breast, cervix, etc.). Before treatment, this concerns the staging of the disease and the identification of genetic features. For treatment planning, this concerns the tissue discrimination between healthy and pathological tissues, the delineation of the GTV, or a boost. For therapeutic monitoring, this concerns the prediction of the patient's survival, the response to treatment, and the occurrence of distant metastases and side effects.

Image features can be extracted from diagnostic or patient monitoring images, as well as from CT (Coroller et al., 2015) or MR treatment planning images (Shiradkar et al., 2016). (Fave et al. (2015a) also showed in 10 NSCLC patients treated using RT that a radiomic analysis could be performed from cone-beam CT (CBCT) images using medical Linac on-board imaging systems. Of 68 features studied, 37 were robust to noise and poor image quality.

### 3.1. Genetic features

Radiomic analysis was used to describe different tumour phenotypes corresponding to distinct genotypes. For instance, Buch et al. (2015) showed, on contrast-enhanced CT, statistically significant differences in some texture features between human papillomavirus-positive and human papillomavirus-negative oropharyngeal tumours.

### 3.2. Segmentation

Nailon et al. (2008) have proposed an automatic segmentation of the GTV and OAR in bladder cancer based on 27 features of CT imaging. The authors showed that this approach offers significant accuracy using only 3 texture features. Yu et al. combined a radiomic analysis of PET and CT images in H&N tumours, highlighting that some image features had a better discriminating power between healthy and pathological tissues than experts, the UH values (CT), or the SUV (PET imaging) (Yu et al., 2009a, b).

Radiomic features can also be used for boost planning. Shiradkar et al. (2016) developed a radiomic-based software platform for the treatment of prostate cancer using brachytherapy and external beam radiation therapy (Rad-TRaP, radiation treatment planning for prostate). This platform allows the extraction of 154 radiomic features from T2-weighted and ADC MR images. From a database of 11 patients, the authors identified 11 features (5 from the T2-weighted image and 6

from the ADC image), making it possible to discriminate, within the gland, the voxels belonging to the lesion and to calculate the probability of the presence of the tumour for each voxel to define a boost from this probability map.

### 3.3. Prediction of treatment response and recurrence of disease

Aerts et al. (2014) studied the prognostic value of 440 radiomic features (first-order, form, and texture features (GLCM, GLRLM, and wavelets)) extracted from CT images on 3 cohorts of patients corresponding to a total of 1019 patients with lung cancer or H&N tumours. The authors showed that the results obtained using a radiomic signature generated from 4 features of the image were better than those obtained using TNM staging or the TV alone. The 4 features were energy (first-order statistics), compactness (shape feature), and 2 features of grey-level non-uniformity (texture and wavelets, respectively). In addition, this signature appeared to be associated with gene expression profiles.

In a database of 92 patients, Li et al. (2017) showed that features extracted from pre-treatment planning CT have a prognostic value of recurrence after stereotactic body radiotherapy (SBRT) for NSCLC. Because of the difficulty of differentiating between recurrence of lung cancer after stereotactic radiotherapy treatment and benign radiation-induced lung injury (RILI), Mattonen et al. (2016) showed the contribution of radiomics in the rapid prediction of the recurrence of cancerous disease.

Gnep et al. (2017) showed that pre-therapeutic features extracted from the co-occurrence matrix of T2-weighted images in patients treated with RT for prostate cancer were strongly correlated with a biochemical recurrence of the disease.

### 3.4. Prediction of side effects

Radiomic analysis has also been proposed to predict the occurrence of side effects due to ionising radiation on OAR in NSCLC and oesophageal cancers, as well as salivary glands and parotids in the treatment of H&N tumours using RT (Scalco and Rizzo, 2017).

Moran et al. (2017) showed that CT-based radiomic features were significantly correlated with radiation oncologist-scored post-SBRT lung injury, suggesting the potential of radiomics to measure post-SBRT lung injury. Cunliffe et al. (2015) showed, in a database of 106 patients treated with RT for their oesophageal cancer, a significant change in 20 radiomic features between the pre-RT CT scan and the post-RT CT scan as a function of the radiation dose. Twelve features changed significantly in patients who developed radiation pneumonitis (RP), allowing the authors to develop a classifier based on these CT features. The same team (Anthony et al., 2017) showed the additive value of the SUV standard deviation on pre-RT PET images indicating the potential of a multimodality radiomic analysis. In a prospective study, Jahangiri et al. (2019) showed the feasibility of FDG PET/CT imaging to quantify radiation-induced lung inflammation in locally advanced NSCLC receiving proton or photon radiotherapy. This study suggests less induction of inflammatory response in both ipsilateral and contralateral lungs of patients treated with proton compared to photon or combined proton-photon RT.

From a radiomic study conducted on a retrospective treatment planning CT database of 249 patients treated using RT for H&N tumours, van Dijk et al. (2017) showed the benefits of radiomic analysis in predicting xerostomia and sticky saliva 12 months after RT. With the same goal, Pota et al. (2017) developed a model for classifying patients at risk of developing parotid gland shrinkage and 12-month xerostomia. Features came from several information sources (clinical data, dosimetry, and CT).

## 4. Factors influencing radiomic features

The very encouraging results of radiomics, however, raise the issue

of the robustness of features with respect to the experimental conditions and image processing performed before their extraction as well as the manner of extraction.

### 4.1. Feature definition

In the literature, there are differences in the definition of radiomic features due to a lack of consensus (Orlhac et al., 2014). Sometimes, 2 texture features with different definitions are referred to using the same name. Features are also found under different names, but with the same mathematical expression. Calculation methods can also be different. Some texture features can be extracted from a single slice of the tumour (2D texture) or from the entire tumour volume (3D texture) (Ng et al., 2013; Fave et al., 2015b). Similarly, some authors considered the different directions of the GLC matrices individually, while others preferred to extract features from the average matrix in 13 directions (Hatt et al., 2015). These findings show that an effort to harmonise and define features is still essential (Sollini et al., 2017; Buvat et al., 2015; Lambin et al., 2017).

### 4.2. Image acquisition and processing

Several authors have studied the robustness of radiomic features according to test-retest studies (Balagurunathan et al., 2014a, b; Tixier et al., 2012; Desseroit et al., 2017), modifying acquisition (Fave et al., 2015b), image reconstruction parameters (Yan et al., 2015; Galavis et al., 2010), and lesion segmentation (Hatt et al., 2011). These studies showed that some features are less robust than others, leading some authors to eliminate the least robust features from the radiomic analysis.

These results show that it is necessary to harmonise image acquisition and perform the most automatic image processing and segmentation as possible to increase the robustness and relevance of the radiomic analysis (RTHM et al., 2017).

### 4.3. Lesion size

Brooks et al. (Brooks and Grigsby, 2014) studied the variation in entropy (GLCM) as a function of the number of voxels in TV based on a PET database of 70 patients with cervical cancer. The results showed that entropy is strongly correlated with TV if the latter is less than 45 cm<sup>3</sup> (700 voxels). Below this threshold, the index was 5 times more sensitive to volume differences. The authors have suggested that it is not appropriate to study this feature for a tumour with a volume less than 45 cm<sup>3</sup>, which in oncology corresponds to a large volume and eliminates a significant number of patients.

Hatt et al. (2015) studied the influence of tumour volume on 4 texture features using PET images based on a database of 555 patients with different cancers (lung, breast, and others). The authors found that the correlation between the radiomic features and the TV tends to decrease when TV increases.

### 4.4. Resampling image intensity values

An important methodological factor is the image discretisation corresponding to resampling image intensity values. When the number of intensity levels in the TV is large, resampling is necessary. Its aim is to group the very large number of intensity levels into a limited number. This step reduces the computation time of the texture matrices, as well as the noise in the image, but to the detriment of a loss of information (Tixier et al., 2012).

Two approaches have been proposed in the PET imaging literature. The first is based on relative resampling using a fixed number (e.g., 16, 32, or 64) of discrete resampled values or bins to divide the SUV range into equally spaced intervals before calculating textural features (Tixier et al., 2011; Orhac et al., 2014). Orhac et al. showed that texture

features are not very sensitive to the number of intensities used when the intensity level is greater than or equal to 32 (Orlhac et al., 2014). Hatt et al. concluded that this type of resampling by more than 64 intensity levels did not provide any additional information and advised limiting the number of resampling to this level (Hatt et al., 2015).

More recently, an absolute resampling process was proposed to resample SUV with a fixed bin size in units of SUV (for example, 0.1 and 0.5), maintaining a constant intensity resolution across all tumour images (Leijenaar et al., 2015; Orhac et al., 2015). It has been shown in PET imaging that absolute resampling has a better predictive power than relative resampling (Brooks and Grigsby, 2014; Leijenaar et al., 2015; Orhac et al., 2015; Desbordes et al., 2017).

## 5. Methods of radiomic analysis

The ever-growing number of features studied raises the question of the statistical method of analysis used and its quality, especially as patient databases can be limited.

### 5.1. Conventional statistical analysis

The majority of the first articles from the literature studying the predictive value of radiomic features were based on the Mann-Whitney *U* test when studying the ability of the feature to differentiate patients into 2 classes (for example, responding and not responding to treatment) or the Kruskal-Wallis test for more than 2 classes. When the feature corresponds to a continuous numerical value, the threshold value of the feature used to optimise the classification of the population into groups is determined using receiver operating characteristic (ROC) curve analysis to calculate the area under the curve, the sensitivity, and the specificity. For prognostic studies of patient survival, Kaplan-Meier's univariate analysis combined with a log-rank test followed by Cox's multivariate analysis is generally used.

The number of features extracted from the image does not necessarily mean that the amount of information increases. Many authors have shown that some features are highly correlated using ranking tests and should be excluded from radiomic analysis based on conventional statistical analysis, leading to a reduction in the number of characteristics studied (Orlhac et al., 2014; Desbordes et al., 2017). Orhac et al. (2015) showed on PET images that there was a significant correlation between texture features and TV when a relative resampling is used. This correlation does not exist with absolute resampling, but there is a correlation between texture features and SUV<sub>max</sub>. To further reduce the number of features, only those that are the most robust with regard to the experimental conditions and image processing are used (Balagurunathan et al., 2014a). Despite these strategies for limiting the number of features, radiomic studies based on conventional statistics present many methodological pitfalls and limitations when handling a large number of indexes (Chalkidou et al., 2015).

### 5.2. Machine learning methods

Given the number of features, the use of machine learning methods is of great interest. Based on a learning database, the goal of machine learning is to develop algorithms that generate a decision rule, also called a model, to predict new patient outcomes. Numerous machine learning algorithms have been proposed in the literature (see Table 1). Examples include logistic regression, multivariate Cox proportional hazards regression models (Aerts et al., 2014), methods based on Bayesian statistics, support vector machine searching the optimal hyperplane separating patient outcomes (Cortes and Vapnik, 1995), as well as the decision tree method and its extension to random forests (Breiman, 2001). Another group of methods corresponds to Artificial Neural Network and deep learning algorithms simulating neuronal functioning (Hosny et al., 2018).

The advantages of machine learning methods are that they were

created, from the start, to handle the processing of a large number of features. Although they require a learning database, which is quite important because the number of features studied is so large, some are better suited than others to the low ratio between the number of patients and the number of features. Depending on the method, they make few or no assumptions about the relationship between feature and prediction, nor about linear or rank relationships between features.

Many feature selection methods have been proposed in the literature to remove features that are highly correlated (Orlhac et al., 2014) or that have limited contribution (Desbordes et al., 2017; Parmar et al., 2015) and avoid overfitting. Here, unlike conventional statistical methods, highly correlated features, in the sense of a rank test, can provide additional information.

These methods are recent and widely used in all areas of big data processing. Therefore, they have a strong potential for methodological development in the future. Their diversity allows clinicians to make a choice depending on the problem, such as the balance between the different classes, the presence of features of different formats (binary, discrete, or continuous), and others.

Ypsilantis et al. (2015) studied the use of several machine learning methods (ANN, SVM, and RF, among others) to predict the response to neoadjuvant chemotherapy in 107 patients with oesophageal cancer. For this purpose, 103 radiomic features were extracted from FDG PET images. ANN gave the best performances for predicting treatment response with a good classification accuracy of 73.4% ± 5%, followed by RF (65.7% ± 6%) and SVM (60.5% ± 8%). The results were superior to those using SUV<sub>max</sub> alone, with an accuracy of 41.0% ± 5%.

From a database of 65 patients with oesophageal cancer treated using CRT and 61 clinical and FDG PET features, Desbordes et al. (2017) showed the superiority of RF over conventional statistical analysis. The RF classifier improved the predictive and prognostic values compared to the Mann-Whitney *U* test and the univariate Kaplan-Meier survival analysis when applied to several tens of features in a limited patient database.

## 6. Conclusions and perspectives

Numerous articles support the clinical use of radiomic analysis in oncology, as well as on MRI, CT than on FDG PET images, in staging of the disease, significantly differentiating the early and advanced stages of the disease, tissue discrimination between healthy and pathological tissues, demonstrating genetic characteristics, predicting the response to treatment, patient survival, and side effects. These results from the literature mainly concern solid tumours for many localisations (pulmonary, H&N, oesophagus, breast, and cervix, among others). They potentially concern any type of treatment (chemotherapy, radiotherapy, and concomitant treatments).

This study of the literature reveals very important perspectives within the context of personalised medicine and highlights many scientific issues not yet resolved, which leaves a large margin of progression for the concept of radiomics.

Numerous factors have been reported to influence the robustness of features, such as acquisition conditions, data reconstruction, tumour volume segmentation, and image resampling. Radiomics requires a harmonisation of practices, which raises the issue of the quality of the database, especially multicentric databases, in a context where the results must be robust to be generalised. Some features are more robust than others, but there is currently no real consensus to establish a list of preferred features or tumour signatures to separate lesions of good and poor prognosis.

Given the number of features studied, the size of databases and their constitution remain issues. In this context, conventional statistical analyses seem too limited and many methodological pitfalls and limitations were described (Leijenaar et al., 2015) that may explain the lack of the emergence of radiomic signatures. This is likely due to the multiplication of features. It can be noted that very few articles

**Table 1**  
Main machine learning algorithms used in radiomics with their advantages and limitations.

Algorithm	Advantages	Limitations
Logistic regression (Ypsilantis et al., 2015)	<ul style="list-style-type: none"> <li>- easy to interpret</li> <li>- Does not require input features to be scaled</li> <li>- Does not require any tuning</li> <li>- Easy to regularize</li> <li>- Easy to implement and very efficient to train</li> <li>- A good baseline to compare the performances with those of other more complex algorithms</li> </ul>	<ul style="list-style-type: none"> <li>- Cannot solve non-linear problems</li> <li>- Features must be independent</li> <li>- Risk of overfitting</li> </ul>
Multivariate Cox proportional hazards regression model (Aerts et al., 2014)	<ul style="list-style-type: none"> <li>- Does not require the division of the studied features into discrete classes (keep the information as precise as possible)</li> </ul>	<ul style="list-style-type: none"> <li>- Limited by its very restrictive application conditions (hazards should be proportional)</li> </ul>
Bayesian statistics (Parmar et al., 2015)	<ul style="list-style-type: none"> <li>- Provides a natural and principled way of combining prior information with data</li> <li>- Provides interpretable answers</li> </ul>	<ul style="list-style-type: none"> <li>- Requires the choice of prior</li> <li>- Can produce posterior distributions that are strongly influenced by the priors</li> <li>- Computationally expensive</li> </ul>
Support vector machine (Parmar et al., 2015; Ypsilantis et al., 2015)	<ul style="list-style-type: none"> <li>- Possible regularization parameter to avoid overfitting</li> <li>- Possible integration of expert knowledge about the problem through kernel engineering</li> <li>- No local minima</li> </ul>	<ul style="list-style-type: none"> <li>- kernel models can be sensitive to overfitting</li> <li>- Difficult to interpret results</li> <li>- Computationally expensive</li> </ul>
Random forests (Desbordes et al., 2017; Hosny et al., 2018; Parmar et al., 2015; Ypsilantis et al., 2015)	<ul style="list-style-type: none"> <li>- Powerful and accurate</li> <li>- Training on even small samples</li> </ul>	<ul style="list-style-type: none"> <li>- Sensitive to overfitting</li> </ul>
Artificial neural network (Hosny et al., 2018; Parmar et al., 2015; Ypsilantis et al., 2015)	<ul style="list-style-type: none"> <li>- Ability to work with incomplete knowledge</li> <li>- Bias tolerance in data</li> <li>- Distributed memory using several neurons in parallel</li> </ul>	<ul style="list-style-type: none"> <li>- Computationally expensive and needs a lot of computer resources</li> <li>- Output model difficult to understand (black-box)</li> <li>- Determination of proper network structure: fine tuning of the model hyperparameters</li> </ul>

currently address the contribution of multimodality imaging, genomics, and proteomics, corresponding to very important fields of investigation. These issues can be addressed using machine learning methods. Although many methods of selecting features have been proposed in the literature to reduce the number of features, machine learning methods make it possible to manage a large number of features. They can also incorporate correlated features according to a rank test as being relevant information.

Despite the enthusiasm for radiomics, it requires the definition of robust signatures to be used in clinical routine. The ultimate goal of such an approach would be to use these signatures individually for each patient. These radiomic signatures must be coupled with other relevant information sources (clinical, biological, genomics and biopathological) in order to improve the prognostic and predictive value of the treatment towards personalized medicine.

#### Financial disclosure statement

This work was supported by grant from the Cancéropôle Nord-Ouest, France.

#### References

Aerts, H.J.W.L., Velazquez, E.R., Leijenaar, R.T.H., et al., 2014. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* 5, 4006.

Alic, L., Niessen, W.J., Veenland, J.F., 2014. Quantification of heterogeneity as a biomarker in tumor imaging: a systematic review. *PLoS One* 9, e0110300.

Amadasun, M., King, R., 1989. Textural features corresponding to textural properties. *IEEE Trans. Syst. Man Cybern.* 19, 1264–1273.

Anthony, G.J., Cunliffe, A., Castillo, R., et al., 2017. Incorporation of pre-therapy 18F-FDG uptake data with CT texture features into a radiomics model for radiation pneumonitis diagnosis. *Med. Phys.* 44, 3686–3694.

Asselin, M.C., O'Connor, J.P., Boellaard, R., Thacker, N.A., Jackson, A., 2012. Quantifying heterogeneity in human tumours using MRI and PET. *Eur. J. Cancer* 48, 447–455.

Balagurunathan, Y., Gu, Y., Wang, H., et al., 2014a. Reproducibility and prognosis of quantitative features extracted from CT images. *Transl. Oncol.* 7, 72–87.

Balagurunathan, Y., Kumar, V., Gu, Y., et al., 2014b. Test-retest reproducibility analysis of lung CT image features. *J. Digit. Imaging* 27, 805–823.

Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.

Brooks, F.J., Grigsby, P.W., 2014. The Effect of Small Tumor Volumes on Studies of Intratumoral Heterogeneity of Tracer Uptake. *J. Nucl. Med.* 55, 37–42.

Buch, K., Fujita, A., Li, B., Kawashima, Y., Qureshi, M.M., Sakai, O., 2015. Using texture analysis to determine human papillomavirus status of oropharyngeal squamous cell carcinomas on CT. *Am J Neuroradiol* 36, 1343–1348.

Bundschuh, R.A., Dinges, J., Neumann, L., et al., 2014. Textural parameters of tumor heterogeneity in 18F-FDG PET/CT for therapy response assessment and prognosis in patients with locally advanced rectal cancer. *J. Nucl. Med.* 55, 891–897.

Buvat, I., Orlhac, F., Soussan, M., 2015. Tumor texture analysis in PET: where do we stand? *J. Nucl. Med.* 56, 1642–1644.

Calais, J., Thureau, S., Dubray, B., et al., 2015. Areas of high 18F-FDG uptake on pre-radiotherapy PET/CT identify preferential sites of local relapse after chemoradiotherapy for non-small cell lung cancer. *J. Nucl. Med.* 56, 196–203.

Chalkidou, A., O'Doherty, M.J., Marsden, P.K., 2015. False discovery rates in PET and CT studies with texture features: a systematic review. *PLoS One* 10, e0124165.

Coroller, T.P., Grossmann, P., Hou, Y., et al., 2015. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. *Radiother. Oncol.* 114, 345–350.

Cortes, C., Vapnik, V., 1995. Support-vector networks. *Mach. Learn.* 20, 273–297.

Cunliffe, A., Armato, S.G., Castillo, R., Pham, N., Guerrero, T., Al-Hallaq, H.A., 2015. Lung texture in serial thoracic computed tomography scans: correlation of radiomics-based features with radiation therapy dose and radiation pneumonitis development. *Int. J. Radiat. Oncol. Biol. Phys.* 91, 1048–1056.

Desbordes, P., Ruan, S., Modzelewski, R., et al., 2017. Predictive value of initial FDG-PET features for treatment response and survival in esophageal cancer patients treated with chemo-radiation therapy using a random forest classifier. *PLoS One* 12, e0173208.

Desseroit, M.C., Tixier, F., Weber, W.A., et al., 2017. Reliability of PET/CT shape and heterogeneity features in functional and morphologic components of non-small cell lung Cancer tumors: a repeatability analysis in a prospective multicenter cohort. *J. Nucl. Med.* 58, 406–411.

Differding, S., Hanin, F.X., Grégoire, V., 2015. PET imaging biomarkers in head and neck cancer. *Eur. J. Nucl. Med. Mol. Imaging* 42, 613–622.

Eisenhauer, E.A., Therasse, P., Bogaerts, J., et al., 2009. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur. J. Cancer* 45 (2), 228–247.

El Naqa, I., Grigsby, P.W., Apte, A., et al., 2009. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern Recognit.* 42, 1162–1171.

Fave, X., Mackin, D., Yang, J., et al., 2015a. Can radiomics features be reproducibly measured from CBCT images for patients with non-small cell lung cancer? *Med. Phys.* 42, 6784–6797.

Fave, X., Cook, M., Frederick, A., et al., 2015b. Preliminary investigation into sources of uncertainty in quantitative imaging features. *Comput. Med. Imaging Graph.* 44, 54–61.

Galavis, P.E., Hollensen, C., Jallow, N., Paliwal, B., Jeraj, R., 2010. Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters. *Acta Oncol.* 49, 1012–1016.

Galloway, M.M., 1975. Texture analysis using gray level run lengths. *Comput Graph Image Process* 4, 172–179.

Gnep, K., Fargeas, A., Gutiérrez-Carvajal, R.E., et al., 2017. Haralick textural features on T2-weighted MRI are associated with biochemical recurrence following radiotherapy for peripheral zone prostate cancer. *J. Magn. Reson. Imaging* 45, 103–117.

Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* 3, 610–621.

Hatt, M., Visvikis, D., Albarghach, N.M., Tixier, F., Pradier, O., Cheze-Le Rest, C., 2011. Prognostic value of 18F-FDG PET image-based parameters in oesophageal cancer and impact of tumour delineation methodology. *Eur. J. Nucl. Med. Mol. Imaging* 38,

- 1191–1202.
- Hatt, M., Majdoub, M., Vallières, M., et al., 2015. 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.* 56, 38–44.
- Hofheinz, F., Lougovski, A., Zöphel, K., et al., 2014. Increased evidence for the prognostic value of primary tumor asphericity in pretherapeutic FDG PET for risk stratification in patients with head and neck cancer. *Eur. J. Nucl. Med. Mol. Imaging* 42, 429–437.
- Hosny, A., Parmar, C., Coroller, T.P., et al., 2018. Deep learning for lung cancer prognostication: a retrospective multi-cohort radiomics study. *PLoS Med.* 15 (11 November), e1002711.
- Jahangiri, P., Pournazari, K., Torigian, D.A., et al., 2019. A prospective study of the feasibility of FDG-PET/CT imaging to quantify radiation-induced lung inflammation in locally advanced non-small cell lung cancer patients receiving proton or photon radiotherapy. *Eur. J. Nucl. Med. Mol. Imaging* 46, 206–216.
- King, A.D., Chow, K.-K., Yu, K.-H., et al., 2013. Head and neck squamous cell carcinoma: diagnostic performance of diffusion-weighted MR imaging for the prediction of treatment response. *Radiology* 266, 531–538.
- Kumar, V., Gu, Y., Basu, S., et al., 2012. Radiomics: the process and the challenges. *Magn. Reson. Imaging* 30, 1234–1248.
- Lambin, P., Rios-Velazquez, E., Leijenaar, R.T.H., et al., 2012. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur. J. Cancer* 48, 441–446.
- Lambin, P., Leijenaar, R.T.H., Deist, T.M., Peerlings, J., De Jong, E.E.C., Van Timmeren, J., et al., 2017. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol [Internet]* 14 (12), 749–762.
- Leijenaar, R.T.H., Nalbantov, G., Carvalho, S., et al., 2015. The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis. *Sci. Rep.* 5, 11075.
- Li, Q., Kim, J., Balagurunathan, Y., et al., 2017. Imaging features from pretreatment CT scans are associated with clinical outcomes in nonsmall-cell lung cancer patients treated with stereotactic body radiotherapy. *Med. Phys.* 44, 4341–4349.
- Lopes, R., Betrouni, N., 2009. Fractal and multifractal analysis: a review. *Med. Image Anal.* 13, 634–649.
- Marusyk, A., Almendro, V., Polyak, K., 2012. Intra-tumour heterogeneity: a looking glass for cancer? *Nat. Rev. Cancer* 12, 323–334.
- Mattonen, S.A., Palma, D.A., Johnson, C., et al., 2016. Detection of local Cancer recurrence after stereotactic ablative radiation therapy for lung Cancer: physician performance versus radiomic assessment. *Int. J. Radiat. Oncol. Biol. Phys.* 94, 1121–1128.
- Moran, A., Daly, M.E., Yip, S.S.F., Yamamoto, T., 2017. Radiomics-based Assessment of Radiation-induced Lung Injury After Stereotactic Body Radiotherapy. *Clin. Lung Cancer* 18, e425–e431.
- Nailon, W.H., Redpath, A.T., McLaren, D.B., 2008. Characterisation of radiotherapy planning volumes using textural analysis. *Acta Oncol.* 47, 1303–1308.
- Ng, F., Kozarski, R., Ganeshan, B., Goh, V., 2013. 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.* 82, 342–348.
- Orlhac, F., Soussan, M., Maisonobe, J.-A., Garcia, C.A., Vanderlinden, B., Buvat, I., 2014. 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.* 55, 414–422.
- Orlhac, F., Soussan, M., Chouahnia, K., Martinod, E., Buvat, I., 2015. 18F-FDG PET-derived textural indices reflect tissue-specific uptake pattern in non-small cell lung cancer. *PLoS One* 10, e0145063.
- Parmar, C., Grossmann, P., Bussink, J., Lambin, P., HJWL, Aerts, 2015. Machine Learning methods for Quantitative Radiomic Biomarkers. *Sci. Rep.* 5, 13087.
- Pota, M., Scalco, E., Sanguineti, G., Farneti, A., Cattaneo, G.M., Rizzo, G., Esposito, M., 2017. Early prediction of radiotherapy-induced parotid shrinkage and toxicity based on CT radiomics and fuzzy classification. *Artif. Intell. Med.* 81, 41–53.
- RTHM, Larue, Defraene, G., De Ruyscher, D., Lambin, P., Van Elmpt, W., 2017. Quantitative radiomics studies for tissue characterization: a review of technology and methodological procedures. *Br. J. Radiol.* 90 (1070), 20160665.
- Scalco, E., Rizzo, G., 2017. Texture analysis of medical images for radiotherapy applications. *Br Inst Radiol* 90 (1070), 20160642.
- Shiradkar, R., Podder, T.K., Algohary, A., Viswanath, S., Ellis, R.J., Madabhushi, A., 2016. Radiomics based targeted radiotherapy planning (Rad-TRaP): a computational framework for prostate cancer treatment planning with MRI. *Radiat. Oncol.* 11, 148.
- Sollini, M., Cozzi, L., Antunovic, L., Chiti, A., Kirienco, M., 2017. PET Radiomics in NSCLC: state of the art and a proposal for harmonization of methodology. *Sci. Rep.* 7, 358.
- Thibault, G., Fertil, B., Navarro, C., Pereira, S., Cau, P., Levy, N., et al., 2009. Texture indexes and gray level size zone matrix application to cell nuclei classification. *Pattern Recognit Inf Process* 5, 140–145.
- Tixier, F., Cheze-Le Rest, C., Hatt, M., et al., 2011. Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal. *Cancer J Nucl Med* 52, 369–378.
- Tixier, F., Hatt, M., Cheze-Le Rest, C., Le Pogam, A., Corcos, L., Visvikis, D., 2012. Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in 18F-FDG PET. *J. Nucl. Med.* 53, 693–700.
- van De Wiele, C., Kruse, V., Smeets, P., Sathekge, M., Maes, A., 2013. Predictive and prognostic value of metabolic tumour volume and total lesion glycolysis in solid tumours. *Eur. J. Nucl. Med. Mol. Imaging* 40, 290–301.
- van Dijk, L.V., Brouwer, C.L., van der Schaaf, A., et al., 2017. CT image biomarkers to improve patient-specific prediction of radiation-induced xerostomia and sticky saliva. *Radiother. Oncol.* 122, 185–191.
- Yan, J., Chu-Sherm, J.L., Loi, H.Y., et al., 2015. Impact of image reconstruction settings on texture features in 18F-FDG PET. *J. Nucl. Med.* 56, 1667–1673.
- Yip, S.S.F., Aerts, H.J.W.L., 2016. Applications and limitations of radiomics. *Phys. Med. Biol.* 61, R150–R166.
- Ypsilantis, P.-P., Siddique, M., Sohn, H.-M., Davies, A., Cook, G.J.R., Goh, V., Montana, G., Predicting, 2015. Response to neoadjuvant chemotherapy with PET imaging using convolutional neural networks. *PLoS One* 10, e0137036.
- Yu, H., Caldwell, C., Mah, K., Mozeg, D., 2009a. Coregistered FDG PET/CT-based textural characterization of head and neck cancer for radiation treatment planning. *IEEE Trans. Med. Imaging* 28, 374–383.
- Yu, H., Caldwell, C., Mah, K., Poon, I., Balogh, J., MacKenzie, R., et al., 2009b. Automated radiation targeting in head-and-Neck Cancer Using region-based texture analysis of PET and CT images. *Int. J. Radiat. Oncol. Biol. Phys.* 75, 618–625.