HHS Public Access

Author manuscript

*Phys Med Biol*. Author manuscript; available in PMC 2017 July 07.

Published in final edited form as:

*Phys Med Biol*. 2016 July 7; 61(13): R150–R166. doi:10.1088/0031-9155/61/13/R150.

# Applications and limitations of radiomics

**Stephen SF Yip**[1] and **Hugo J.W.L. Aerts**[1,2]

Hugo J.W.L. Aerts: Hugo_Aerts@dfci.harvard.edu

[1]Department of Radiation Oncology, Brigham and Women's Hospital, Dana-Farber Cancer Institute and Harvard Medical School, Boston MA

[2]Department of Radiology, Dana-Farber Cancer Institute and Harvard Medical School, Boston MA

## Abstract

Radiomics is an emerging field in quantitative imaging that uses advanced imaging features to objectively and quantitatively describe tumour phenotypes. Radiomic features have recently drawn considerable interest due to its potential predictive power for treatment outcomes and cancer genetics, which may have important applications in personalized medicine. In this technical review, we describe applications and challenges of the radiomic field. We will review radiomic application areas and technical issues, as well as proper practices for the designs of radiomic studies.

## 1. Introduction

Non-invasive medical imaging, such as magnetic resonance (MR) imaging, computed tomography (CT), and positron emission tomography (PET), is routinely used for assessing tumour and anatomical tissue characteristics for cancer management (Buckler *et al* 2011, Kurland *et al* 2012). Furthermore, imaging can potentially provide valuable information for personalized medicine that aims to tailor treatment strategy based on the characteristics of individual patients and their tumours.

Molecular characterization using genomics, proteomics, and metabolomics information has been the main focus of personalized therapy. However, spatial and temporal intratumoural heterogeneity that arises from regional variations in metabolism, vasculature, oxygenation, and gene expression is a common feature of malignant tumours (Maley *et al* 2006, Marusyk *et al* 2012, Chicklore *et al* 2013, Fisher *et al* 2013). Random samples of tumour tissues acquired through invasive biopsy for molecular characterization may thus fail to accurately represent the landscape of the biological variation within tumours (Gerlinger *et al* 2012). On the other hand, the entire tumour can be sampled non-invasively and repeatedly with medical imaging.

In particular, studies have hypothesized that tumour characteristics at the cellular and genetic levels are reflected in the phenotypic patterns that can captured with medical images (Henriksson *et al* 2007, Diehn *et al* 2008, Basu *et al* 2011, Yang and Knopp 2011). Several

Correspondence to: Hugo J.W.L. Aerts, Hugo_Aerts@dfci.harvard.edu.

studies have shown these associations across imaging modalities. For example, in MR, growing brain tumours that cause a shift in midline structures due to normal tissue compression, known as the mass effect, are found to be strongly correlated with proliferation gene-expression (Diehn *et al* 2008). Yamamoto *et al* (2014) found that lung tumours with anaplastic lymphoma kinase (ALK) mutations appeared to have larger pleural effusion and no pleural tails on CT images (Yamamoto *et al* 2014). Contrast-enhanced CT images revealed that the mutation status of von Hippel-Lindau (VHL) in renal cell carcinoma is significantly correlated with the "gross appearance of intratumoural vascularity", "well-defined tumour boundaries", and "enhancement of nodular tumour" (Karlo *et al* 2014). In PET imaging, [$^{18}$F]FDG uptake is related to the number of viable cancer cells, tumour histopathology, and a number of biological processes that support the continuous growth of the tumour (Higashi *et al* 1993, Haberkorn *et al* 1994, Rajendran *et al* 2004, Fanchon *et al* 2015). Studies have therefore proposed that tumour heterogeneity may be associated with the non-uniform distribution of [$^{18}$F]FDG (Henriksson *et al* 2007, Tixier *et al* 2011, Chicklore *et al* 2013).

Despite the promise of medical imaging to assess tumour heterogeneity (or genetic), imaging features are often assessed visually and described qualitatively by radiologists or nuclear medicine physicians. Subjective descriptions of tumour imaging phenotypes (e.g. "large necrotic core", "highly speculated", and "moderate heterogeneity"). However, these visual assessments can suffer from a large intra and inter-observer variability (de Jong *et al* 1995, Mussurakis *et al* 1996, Wetzel *et al* 2002, Davnall *et al* 2012, Tixier *et al* 2014c). Therefore, it is important to objectively and reproducibly quantify various imaging features that may reveal the underlying biology of tumours.

Radiomic uses the high-throughput extraction of advanced quantitative features to objectively and quantitatively describe tumour phenotypes (Figure 1). These features, termed radiomic features, are extracted from medical images using advanced mathematical algorithms to uncover tumour characteristics that may fail to be appreciated by the naked eye (Lambin *et al* 2012, Chicklore *et al* 2013, Aerts *et al* 2014, Cook *et al* 2014, Buvat *et al* 2015, Rahmim *et al* 2016). Radiomic may thus provide great potential to capture important phenotypic information, such as intra-tumour heterogeneity, subsequently providing valuable information for personalized therapy. In this review, we will review radiomic applications and technical limitations, as well as proper practices for the designs of radiomic studies.

## 2. Potential applications of radiomic

Numerous radiomic features, such as size and shape based–features, descriptors of the image intensity histogram, descriptors of the relationships between image voxels (e.g. gray-level co-occurrence matrix (GLCM), run length matrix (RLM), size zone matrix (SZM), and neighborhood gray tone difference matrix (NGTDM) derived textures), textures extracted from wavelet and Laplacian of Gaussian filtered images, and fractal features, can be extracted from the medical images (Haralick *et al* 1973, Galloway 1975, Pentland 1984, Amadasun and King 1989, Davnall *et al* 2012, Thibault *et al* 2013, Aerts *et al* 2014, Rahmim *et al* 2016).

Radiomic features not only provide an objective and quantitative way to assess tumour phenotype, they have also found a wide-range of potential applications in oncology. For example, radiomic features have shown promise in the prediction of treatment response, differentiating benign and malignant tumours, and assessing cancer genetics in many cancer types. We will review the potential application of the radiomic features.

### 2.1. Prediction of treatment response and outcomes

MR studies have shown that intensity histogram-based radiomic features are potentially useful for predicting cancer response to treatment (Johansen *et al* 2009, Baek *et al* 2012, Shukla-Dave *et al* 2012, King *et al* 2013, Peng *et al* 2013). In pre-clinical model, (Foroutan *et al* 2013) observed that mice with sarcomas treated with combinations of MK1775, a cell cycle checkpoint inhibitor, and gemcitabine showed a substantial change in the (apparent diffusion coefficient) ADC histogram skewness, kurtosis, entropy, and average ADC shortly after treatment compared to the untreated control group. In human patients with head-and-neck cancer, tumours that responded poorly to chemoradiotherapy demonstrated a significantly greater increase in average ADC and higher values in kurtosis and skewness on mid-treatment diffusion weighted MR (DW-MR) than tumours with a better therapeutic response (King *et al* 2013). K-trans is a measure derived from dynamic contrast-enhanced MR images and measures the diffusion of an intravascular contrast agent into the extracellular space. The skewness of K-trans was found to be a promising predictor of progression free survival and overall survival of patients with stage IV head-and-neck cancer (Shukla-Dave *et al* 2012). The findings of these aforementioned studies may support the notion that therapy induced changes in tumour microenvironment and composition can be potentially described by changes in the intensity-histogram shape.

In PET imaging, standardized uptake value (SUV) measures, such as the maximum SUV ($SUV_{max}$) and mean SUV obtained within a tumour ($SUV_{mean}$), are commonly used for tumour characteristic quantification. High baseline SUV uptake is often thought to be associated with aggressive tumour behavior and poor prognosis (Rizk *et al* 2006, Zhang *et al* 2011, Higgins *et al* 2012). However, as previously mentioned, $SUV_{max}$ and $SUV_{mean}$ are inadequate for describing the heterogeneous distribution of [$^{18}$F]FDG uptake (van Velden *et al* 2011, Marusyk *et al* 2012, Cheng *et al* 2013b).

Recently, radiomic textural features have drawn considerable interest due to their potential to describe distinctive tumour phenotype (“appearance”) that may be driven by underlying genetic and biological variability. In particular, they were demonstrated to outperform simple SUV measures, such as $SUV_{max}$ and $SUV_{mean}$, in treatment outcome prediction (Eary *et al* 2008, El Naqa *et al* 2009, Tixier *et al* 2011, Yang *et al* 2013). For example, Cook *et al* (2013) compared the predictive power of maximum and mean SUV and four NGTDM derived textures in fifty-three non-small cell lung cancer (NSCLC) patients (Cook *et al* 2013). They found that NGTDM derived coarseness, busyness, and contrast could better differentiate between responders and nonresponders to chemoradiotherapy than the aforementioned SUV measures. Furthermore, coarseness was found to be an independent predictor of patient overall survival. (Zhang *et al* 2014) built several multivariate models to predict pathologic response to preoperative chemoradiotherapy in twenty esophageal cancer

patients. They found that models constructed with combined radiomic features significantly improved the pathologic response prediction compared to models built with maximum SUV, metabolically active tumour volume and longest diameter.

For CT imaging, Aerts *et al* (2014) assessed the prognostic values of 440 shape- and intensity-based and textural features. They identified features that were predictive of patients' survival on a discovery dataset consisting of >420 lung cancer patients. The prognostic value of features were then validated on three independent datasets, including one lung cancer (225 patients) and two head-and-neck cancer (231 patients) cohorts. Their results not only confirmed the potential use of radiomic features in outcome prediction and describing intratumoural heterogeneity, but also showed that prognostic ability may be transferred from one disease type to another (i.e. from lung to head-and-neck cancer). On the other hand, Parmar *et al* (2015) observed that not all radiomic features that significantly predicted lung cancer patients' survival also predicted survival in head-and-neck cancer patients and vice versa (Parmar *et al* 2015b). Their results thus suggested that some radiomic features may be cancer-specific.

Studies have found a strong association between contrast-enhanced CT (CE-CT) and heterogeneity of the tumour vasculature (Tateishi *et al* 2002, Kim *et al* 2005). (Tixier *et al* 2014a) observed that tumour blood flow measured on CE-CT was significantly correlated with the metabolically active tumour volume.CE-CT radiomic therefore provides great potential for quantifying complex tumour phenotype arising from angiogenesis in cancer. For example, Hayano *et al* (2014) hypothesized that if the fractal dimension extracted from CE-CT is useful in describing tumour heterogeneity, then the measure may also be useful for predicting patient survival in hepatocellular carcinoma (Hayano *et al* 2014). They found that the patients with longer survival often had lower fractal dimensional on the arterial phase CE-CT image.

Furthermore, radiomic features can be potentially applied to assess the metastatic potential of tumours. Coroller *et al* (2015) identified thirty-five CT radiomic features to be significant predictors of distant metastasis and six features to be predictors of survival in 182 lung cancer patients (Coroller *et al* 2015). They concluded that the radiomic features they identified may be useful for early indication of cancer patients that will have a high risk of developing distant metastasis, thus allowing physicians to better adapt treatment plans for individual patients. Recently, Vallieres *et al* (2015) showed that the combination of MR and [$^{18}$F]FDG-PET textural features can better predict the risk of lung metastases in soft-tissue sarcomas than the features acquired from a single modality (Vallieres *et al* 2015).

### 2.2. Tumour staging

Many radiomic features were shown to be able to significantly differentiate between early and advanced stage diseases. For example, in a PET radiomic study by (Dong *et al* 2013), forty esophageal cancer patients were staged according to the American Joint Committee on Cancer (7$^{th}$ edition). $SUV_{max}$, GLCM-entropy, and GLCM-energy were found to be significantly correlated with T and N stage. In particular, a GLCM-entropy value > 4.70 could accurately identify tumours with stages above stage IIb (Dong *et al* 2013). In a recent study, Mu *et al* (2015) classified forty-two cervical cancer patients into early stage (stage I

and II) and advanced stage (stage III and IV) using PET-based radiomic features (Mu *et al* 2015). RLM-run percentage texture was found to be most associated with cervical tumour stage. Moreover, CT-based fine textures derived from Laplacian of Gaussian (LoG) filtered CT images were found to predict lung tumour stages above stage II (Ganeshan *et al* 2010). Early identification of tumour stage using radiomic features may assist physician to better stratify patients, subsequently selecting the best treatment for individual patients.

### 2.3. Tissue identification

Radiomic features have also been shown to be useful in discriminating between malignant and other tissues in many disease types. In the 1990s, GLCM textures derived from a 2D slice of T1- and T2-weighted MR images were first reported to be potentially useful for tissue differentiation, with the ability to differentiate brain tumour tissue, edema, cerebrospinal fluid (CSF), white matter, and gray matter, in patients with brain cancer (Lerski *et al* 1993, Kjær *et al* 1995). Mahmoud-Ghoneim *et al* (2003) demonstrated that GLCM textures computed within a 3D volume of the MR images outperformed 2D textures in separating necrosis and edema from solid tumours (Mahmoud-Ghoneim *et al* 2003). Besides brain tumours, Nie *et al* (2008) showed that combining shape-based, volume-based, and GLCM textural features of MR using an artificial neural network (ANN) may be used to differentiate malignant from benign tumours in breast cancer (AUC≥0.80) (Nie *et al* 2008). Furthermore, they also observed that benign tumours had smoother boundaries, rounder shape, and lower image intensity than malignant tumours.

CT-based radiomic features have been used to classify a pulmonary nodule as benign or malignant (McNitt-Gray *et al* 1999, Kido *et al* 2002, Petkovska *et al* 2006, Way *et al* 2006). Pulmonary nodules could be due to other diseases (e.g. tuberculosis and fungal infection) than cancer. Kido *et al* (2002) showed that the fractal dimensions for bronchogenic carcinomas were significantly smaller than pneumonias and tuberculomas ($p<0.0001$). Petkovska *et al* (2006) showed that GLCM textures extracted from contrast-enhanced CT can accurately identify malignant from benign nodules, while visual inspection by three experienced radiologists performed worse in malignant-benign nodule differentiation. Combining shape-, size, and histogram-based features has been show to improve the differentiation between malignant and benign nodules (Way *et al* 2006).

Furthermore, a study by (Xu *et al* 2014) developed a computer aided diagnosis method with combined CT- and PET-based textures for differentiating malignant and benign lesion in various tumour sites. [$^{18}$F]FDG uptake of malignant lesions were observed to be more heterogeneous than the benign tissues. Compared with the histological diagnosis of the lesions, the classification results of their texture-based diagnosis method achieved sensitivity, specification, and accuracy >75% (Xu *et al* 2014). (Yu *et al* 2009) assessed the ability of 14 PET and 13 CT-based textures in delineating primary and nodal tumours from normal tissues. The sensitivity, specificity, and accuracy of the delineation results based on the radiomic textural features were >95% comparing to the tumour contours manually segmented by three radiation oncologists (Yu *et al* 2009).

### 2.4. Assessment of cancer genetics

Many studies have shown that there is a strong relationship between imaging features and the underlying tumour genetics, which may provide a biological basis for the clinical applications of radiomic. MR-based volumetric features are often observed to be associated with somatic mutations and genetic expression of brain tumours (Diehn *et al* 2008, Ellingson *et al* 2013, Naeini *et al* 2013, Gutman *et al* 2015). For instance, Ellingson *et al* (2013) observed that MGMT unmethylated glioblastoma (GBM) usually had smaller volumes of TI-contrast enhanced and T2-FLAIR hyperintensity than methylated GBM. In a recent study by Gutman *et al* (2015), volumetric measures, such as contrast enhancing volume, necrosis volume, and total tumour volume, were found to significantly predict GBM mutations, including TP53, NF1, EGFR, RB1, and PDGFRA.

In CT imaging, Aerts *et al* (2014) found that radiomic features related to shape and wavelet features describing the heterogeneous phenotype of lung tumours were found to be significantly associated with cell cycle pathway, suggesting that highly proliferative tumours demonstrate complex imaging patterns. Moreover, various biological mechanisms may be described by different radiomic features as the features were found to be related with different biological gene sets, including DNA recombination and regulation of DNA metabolic processes (Aerts *et al* 2014).

Nair *et al* (2012, 2014) investigated the association between PET-SUV histogram radiomic features and various NSCLC genes and gene expressions in cohorts consisting of >300 patients (Nair *et al* 2012, Nair *et al* 2014). Features such as skewness, $SUV_{max}$, $SUV_{mean}$, median of the SUV histogram were strongly correlated with several gene signatures and expressions (e.g. NF-κB) that are related to patient survival (Nair *et al* 2012, Nair *et al* 2014).

As numerous radiomic features can be extracted from medical images, the studies mentioned in this section play an important role in identifying only a subset of features that might be most relevant to the underlying tumor biology and genetics. However, how the tumor patho-physiological processes give rise to imaging phenotypes that can be quantified by radiomic features remain unclear. Future studies would need to investigate these associations to further elucidate the biological meaning of the radiomic features.

## 3. Factors that affect radiomic features quantification

### 3.1. Acquisition modes, reconstruction parameters, smoothing, and segmentation thresholds

Despite the wide range of potential applications, radiomic feature quantification may be sensitive to a number of technical factors. For example, Galavis *et al* (2010) assessed the variability of 50 PET radiomic features due to different acquisition modes, matrix sizes, post-filtering widths, reconstruction algorithms and iteration numbers (Galavis *et al* 2010). Of these features, forty were shown to have substantial variability with a relative difference of >30%. Only four features, including intensity-histogram derived entropy and energy, GLCM-maximal correlation coefficient, RLM-low gray level run emphasis, were found to have variability <5%. The textures that are sensitive to acquisition modes and reconstruction

parameters are thus not recommended for radiomic applications, such as malignant and benign tissue differentiation (Galavis *et al* 2010). Yan *et al* (2015) identified features, including histogram-based entropy, normalized GLCM-inverse difference moment and inverse difference, RLM-low gray run emphasis and high gray run emphasis, and SZM-low gray zone emphasis, were robust to different PET image reconstruction settings (Yan *et al* 2015). These features may thus be useful for radiomic studies. However, (Galavis *et al* 2010) and (Yan *et al* 2015) did not investigate or elaborate on why certain radiomic features were more sensitive to the others. This may need to be further investigated.

On the other hand, (Doumou et al 2015) investigated the effect of PET image post-filtering width (noise smoothing) on feature quantification. They found that the radiomic features were generally insensitive to variations in filter width.

Accurate delineation of tumour volumes is crucial for the computation of radiomic features. Manual delineation of tumour volume is not only time-consuming, but can also be affected by inter-observer variability. Radiomic studies often recommend using automatic and semi-automatic methods for tumour volume delineation over manual contouring (Hatt *et al* 2009, Velazquez *et al* 2013, Parmar *et al* 2014, Yip *et al* 2016). For example, Velazquez *et al* (2013) compared the accuracy of manual and semiautomatic region growing tumour contouring methods on CT images. They found that the semiautomatic contouring method was better associated with gold-standard tumour sizes that were measured by surgical resection. Moreover, Parmar *et al* (2014) found that CT-based radiomic features were more stable when computed from a semiautomatic contouring method than from manual contours (Parmar *et al* 2014). A recent study investigated metabolic tumour volume auto-segmentation thresholds (45–60% of the maximum SUV) on the precision of PET-based radiomic texture quantification (Doumou *et al* 2015). The authors concluded that the variation in image segmentation thresholds only have small effects on the quantification, suggesting metabolic tumour volume may be precisely defined by thresholding.

(Hatt *et al* 2011) and (Cheebsumon *et al* 2012) found that lung tumour size computed with PET-based tumor delineation methods, such as fixed and adaptive thresholds, are in better agreement with surgical resection while manual contouring on CT images significantly overestimated the pathological tumour size. However, more advanced delineation algorithms, such as fuzzy locally adaptive Bayesian (FLAB), are recommended for larger lung tumors as simple threshold-based methods may result in underestimation of the metabolically active tumor region (Hatt *et al* 2011).

### 3.2. Reproducibility of radiomic features

While tumour heterogeneity can be potentially quantified using numerous radiomic features extracted from medical images, many features are often found to be unstable between imaging scans acquired within weeks–even minutes of each other (Tixier *et al* 2012, Leijenaar *et al* 2013, Balagurunathan *et al* 2014a, Balagurunathan *et al* 2014b, van Velden *et al* 2014). Balagurunathan *et al* (2014a and 2014b) assessed the intra-class correlation coefficient (ICC) of 219 radiomic features extracted from a test and retest CT scan in lung cancer patients which were acquired 15 minutes apart. Of these features, only 66 of them were found to have ICC≥0.90 across the test and retest scans, suggesting that a large number

of features may be unreliable. Tixier *et al* (2012) studied the reproducibility of 25 radiomic features between repeated PET scans acquired within 4 days of each other. GLCM-entropy and homogeneity, SZM-zone and intensity variability were not only found to be significant predictors of treatment response (Tixier *et al* 2011), but also exhibited the highest reproducibility (Tixier *et al* 2012). Leijenaar *et al* (2013) studied the stability of nearly 100 radiomic features to repeated PET images (1 day apart) and inter-observer variability in tumour delineation in lung cancer patients. They found that the PET-based features that are stable between repeated scans were also more robust to inter-observer variability, suggesting that features with poor reproducibility may also be sensitive to other factors and are thus not recommended.

However, to our knowledge, the repeatability of MR-based radiomic features has not been investigated. Understanding the stability of MR-based radiomic features between test and re-test scans can help identifying reliable features for radiomic applications, and thus would be a valuable future study.

### 3.3. Image discretization (resampling) schemes

Prior to radiomic feature computation, voxel intensities within tumour volumes need to be discretized to a limited range of intensity values in order to efficiently and practically compute the radiomic features (Cheng *et al* 2013b, Leijenaar *et al* 2013). For instance, an image with 1024 discrete intensity values will yield a $2^{1024} \times 2^{1024} \times 2^{1024}$ GLCM (Haralick *et al* 1973), which can be computationally intensive. The range of intensity values of an image thus needs to be reduced and limited for efficient radiomic feature computation. In PET radiomic studies, the most commonly used discretization scheme is to first normalize the medical image by the relative difference of the maximum and minimum intensity values within tumours, and then resample the voxel intensities to $2^N$ values (i.e. $2^N$ number of bins), where N ranges from 3 to 8 in literature (Orlhac *et al* 2014, Tixier *et al* 2014b, Doumou *et al* 2015, Mu *et al* 2015). Studies have shown that both the quantity and prognostic value of radiomic features, particularly GLCM-entropy, SZM-size zone high gray emphasis and SZM-size zone non-uniformity derived from PET images, can be substantially influenced by the number of discrete values ($2^N$) (Cheng *et al* 2013a, Orlhac *et al* 2014, Doumou *et al* 2015). At least $2^5$=32 discrete values is recommended to properly quantify tumour heterogeneity with PET-based radiomic features (Orlhac *et al* 2014). However, the Spearman correlation coefficient of tumour volume and GLCM-entropy was observed to be >0.85 for resampling values over 64, suggesting that textural features computed with resampling values over 64 may not provide additional prognostic information compared with the tumour volume (Hatt *et al* 2015). Therefore, Hatt *et al* (2015) limited the number of discrete bins to 64 for PET-based texture computation.

Alternatively, the voxel intensity range can be discretized into equally spaced bins with a fixed bin width (Leijenaar *et al* 2013, Leijenaar *et al* 2015b). For instance, Leijenaar *et al* (2015) compared two resampling strategies (i.e. fixed number of bins and fixed bin width) and found that quantification of radiomic features were more robust to a change in bin size than to a change in the number of bins (Leijenaar *et al* 2015b). They concluded that the resampled PET image voxel intensity (or SUV) using a fixed bin width may be more

appropriate for clinical case studies. This is because resampled PET voxel intensity using the alternative (i.e. a fixed number of bins) implicitly assumes that the tumour images of all patients have the same SUV range, which is usually not the case.

Another resampling strategy is to determine the bin size for each tumour image according to the Freedman–Diaconis rule (bin size= $2 \bullet IQR/N^{-1/3}$), where IQR is the interquartile SUV range and N is total number of voxels that the tumours are composed of (Brooks and Grigsby 2014). Comparison of different resampling strategies may be important to understand their effect on radiomic features in treatment outcome prediction.

### 3.4. Computation of radiomic features

The computations of radiomic features, even with the same features names, may be implemented differently in radiomic studies. For example, GLCM can be calculated either by averaging the values of the matrices computed for 13 distinct directions or a single matrix that accounts for tumour co-occurrence information in all 13 directions (Hatt *et al* 2015). Textural features can be extracted from the largest cross-sectional (axial) slice of the tumour boundary (2D textures) or extracted from the entire tumour volume (3D textures) (Ng *et al* 2013, Fave *et al* 2015). The impact of different feature implementation/computation methods on the predictive values of radiomic features needs to be carefully studied.

### 3.5. Respiratory motion

The accurate quantification of radiomic features can be hindered by respiratory motion in lung cancer patients (Yip *et al* 2014). Lung motion can lead to a reduction in the measured activity in tumor and other tissues due to insufficient data acquisition and limited reconstruction techniques in static PET images (3D PET) (Nehmeh *et al* 2002, Aristophanous *et al* 2012, Huang and Wang 2013). 4D PET imaging gates PET image acquisition with respiratory motion to improve PET image quality (Nehmeh *et al* 2002, García Vicente *et al* 2010, Didierlaurent *et al* 2012). Yip *et al* (2014) investigated the influence of the lung tumour motion on radiomic textures (Yip *et al* 2014). They observed that the radiomic textural features, blurred out by respiratory motion during 3D-PET acquisition, can be better resolved by 4D-PET imaging. 4D-PET textures may have better prognostic value as they are less susceptible to tumour motion although the hypothesis needs to be investigated in the future.

### 3.6. Tumour size and intratumoural heterogeneity

Intratumoural heterogeneity for small tumour volumes may not be accurately quantified due to the partial volume effect resulting from limited PET resolution (Soret *et al* 2007, Hatt *et al* 2013). Therefore, many studies often exclude tumours with volumes <3–5$cm^3$ from radiomic analysis (Orlhac *et al* 2014, Hatt *et al* 2015). To estimate the minimum tumour volume needed for texture computation, Brooks and Grigsby (2014) extracted GLCM-entropy from PET images in 70 cervical cancer tumours (Brooks and Grigsby 2014). Using probability theory, they found that GLCM-entropy computed for tumours <45$cm^3$ were strongly correlated with tumour size, and therefore may not accurately measure intratumoural heterogeneity. However, their conclusion was based on theoretical analysis, one radiomic texture, and a single tumour type.

Hatt *et al* (2015) computed four prognostic radiomic textures features on 555 PET images acquired from multiple cancer centers consisting of breast, cervical, NSCLC, esophageal, and head-and-neck tumours (Hatt *et al* 2015). The added prognostic value of the textures and their correlation to tumour volume were investigated. Both radiomic textures and tumour volume were observed to be independent predictors of survival for patients with bigger tumours, whereas the added value of textures in predicting survival was minimal for small tumours. They observed a strong correlation between textural features and tumour size for tumours with volumes less than 10 $cm^3$. The results of Hatt *et al* (2015)'s study suggest that radiomic textures have no added value in outcome prediction for tumours <10$cm^3$. However, instead of excluding tumours with volume <10$cm^3$ in the future radiomic studies, they recommended that the correlation of the radiomic features and tumour volume should be always reported to highlight if the features provide independent or redundant information (Hatt *et al* 2015).

### 3.7. The silver lining and the need for standardization

Besides the aforementioned factors, there are other factors, such as metal artifacts in CT images (Leijenaar *et al* 2015a), CT x-ray tube peak voltage and current (Fave *et al* 2015), that may also affect radiomic feature quantification. As CT images are often employed for attenuation correction of PET and SPECT images, factors that affect the quality of the CT images can also impact the quantification of features extracted from the PET and SPECT images. Despite the potential impact of these factors on quantification, strong prognostic signals of the features could still be found (Cheng *et al* 2013a, Cook *et al* 2013, Aerts *et al* 2014, Cheng *et al* 2014, Coroller *et al* 2015, Leijenaar *et al* 2015a, Parmar *et al* 2015b). While harmonization and standardization for imaging acquisition and feature computation may lead to more consistent findings in radiomic studies across institutions, the technical factors that affect the radiomic feature quantification may not be reduced (Boellaard 2011, Nyflot et al 2015). For example, in harmonization, as some PET systems fail to fully resolve small objects due to limited resolution (partial volume effect), additional smoothing steps are thus required for images acquired by certain PET systems, even with high resolution and sensitivity (Boellaard et al 2015). Thus, the impact of harmonization and standardization on the quantification and predictive values of radiomic features would be an important topic of future investigations for the field of radiomics. Of equal importance, standardization for proper statistical practice and study designs for current radiomic studies also need to be considered.

## 4. False positive discovery rate and proper study design

Many studies examined the prognostic value of radiomic features based on retrospective analysis of small patient datasets (<50 patients) (Tixier *et al* 2011, Dong *et al* 2013, Tan *et al* 2013, Bundschuh *et al* 2014, Zhang *et al* 2014). These retrospective studies are important for providing rationale (or proof-of-concept) for further investigation of radiomic features as imaging biomarkers and surrogates for intratumoural heterogeneity. However, it is not uncommon that the number of examined radiomic features is much greater than the number of patients, which can lead to feature selection bias and false positive results (Alic *et al* 2014, Chalkidou *et al* 2015). To demonstrate this bias, Chalkidou *et al* (2015) randomly

generated 100 features and assessed the association between the features and survival data extracted from a study by Ganeshan *et al* (2012) consisted of only 21 esophageal cancer patients (Ganeshan *et al* 2012). Ten random features were found to accurately identify patients surviving a follow-up period of over 2 years with the area under the receiver-operating-characteristics curves (AUC) of 0.68–0.80.

Ideally, an external validation dataset is required to confirm the prognostic value of the radiomic features to avoid optimism based on false positive results (Steyerberg *et al* 2010, Lambin *et al* 2013, Aerts *et al* 2014, Chalkidou *et al* 2015). However, acquiring a validation dataset is not always feasible due to high cost, requirement of excessive effort, differences in data collection practice and privacy issues between institutes (Lambin *et al* 2013).

As a rule of thumb, to reduce the false discovery rate, 10–15 patients are needed for each examined radiomic feature (Chalkidou *et al* 2015). As many of the radiomic features are highly correlated, radiomic studies should avoid including strongly correlated features that may provide redundant information about tumour characteristics (Orlhac *et al* 2014, Mu *et al* 2015). For analyses where large numbers of radiomic features are studied, the significant values (p-values) should be corrected for multiple hypotheses testing using the Holm-Bonferroni method or a false-discovery rate (FDR) controlling procedure, such as the Benjamini-Hochberg method (Alic *et al* 2014, Chalkidou *et al* 2015). For example, the Benjamini-Hochberg procedure has been used for multiple testing correction in the work of (Aerts *et al* 2014), (Hatt *et al* 2015), and (Yip *et al* 2016).

The optimal cutoff values of radiomic features are often used to stratify patients into two risk groups for Kaplan-Meier survival analysis (Cheng *et al* 2013a, Cook *et al* 2013). However, searching for the optimal cutoff values through testing multiple cutoffs can increase the likelihood of obtaining spurious significant results (Hilsenbeck *et al* 1992). Moreover, as the optimal cutoff value can vary in different datasets, the results may not be reproducible in different studies. Selection of an optimal cutoff for survival analysis is not recommended or must be accompanied by properly corrected significant values (p-values) (Altman *et al* 1994, Chalkidou *et al* 2015).

Numerous methods can be applied to reduce the number of radiomic features (Guyon *et al* 2003). The selected features can then be combined using various multivariate (classification) models to predict treatment outcome, tumour genetics, prognosis, metastatic potential, etc. In a study, Parmar *et al* (2014) investigated the prognostic values of 440 radiomic features using fourteen feature selection methods and twelve classification models in >460 lung cancer patients (Parmar *et al* 2015a). They found that the choice of the classification model could lead to variations in the predictive values of the radiomic features up to >30%, while choosing different feature selection methods only led to variations of about 6%. Furthermore, they identified feature selection methods and classification models that were stable to data perturbation while maintaining a decent performance for prediction of outcomes.

## 5. Summary

Here, we have reviewed applications and challenges of radiomics. Researchers have proposed to use radiomic features, which aim to quantify various tumour phenotypes on medical images, to describe this heterogeneity and furthermore, utilize these features as predictors of genetics and clinical outcomes. Despite the promising clinical potential of radiomics, there are precautions that must be taken in designing radiomics studies. For example, not all radiomics features are recommended for use due to their sensitivity to acquisition modes and reconstruction parameters. To examine the prognostic power of radiomic features, datasets consisting of ten to fifteen patients per feature evaluated has been recommended. Furthermore, the correlation of tumour volume and radiomic features should be reported to indicate the potential complementary value of the measures. Ideally, independent validation datasets are needed to confirm the prognostic value of the same radiomic features.

## Acknowledgments

The authors would like to acknowledge the support from National Institute of Health (Award Number U01CA190234 and U24CA194354) and research seed funding grant from the American Association of Physicists in Medicine. The authors would also like to thank Dr. Elizabeth Huynh for editorial assistance.

## References

Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D, Hoebers F, Rietbergen MM, Leemans CR, Dekker A, Quackenbush J, Gillies RJ, Lambin P. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014; 5:4006. [PubMed: 24892406]

Alic L, Niessen WJ, Veenland JF. Quantification of Heterogeneity as a Biomarker in Tumor Imaging: A Systematic Review. PLoS ONE. 2014; 9:e110300. [PubMed: 25330171]

Altman DG, Lausen B, Sauerbrei W, Schumacher M. Dangers of Using “Optimal” Cutpoints in the Evaluation of Prognostic Factors. Journal of the National Cancer Institute. 1994; 86:829–835. [PubMed: 8182763]

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

Aristophanous M, Yong Y, Yap JT, Killoran JH, Allen AM, Berbeco RI, Chen AB. Evaluating FDG uptake changes between pre and post therapy respiratory gated PET scans. Radiotherapy and Oncology. 2012; 102:377–382. [PubMed: 22265731]

Baek HJ, Kim HS, Kim N, Choi YJ, Kim YJ. Percent Change of Perfusion Skewness and Kurtosis: A Potential Imaging Biomarker for Early Treatment Response in Patients with Newly Diagnosed Glioblastomas. Radiology. 2012; 264:834–843. [PubMed: 22771885]

Balagurunathan Y, Gu Y, Wang H, Kumar V, Grove O, Hawkins S, Kim J, Goldgof DB, Hall LO, Gatenby RA, Gillies RJ. Reproducibility and Prognosis of Quantitative Features Extracted from CT Images. Translational Oncology. 2014a; 7:72–87. [PubMed: 24772210]

Balagurunathan Y, Kumar V, Gu Y, Kim J, Wang H, Liu Y, Goldgof D, Hall L, Korn R, Zhao B, Schwartz L, Basu S, Eschrich S, Gatenby R, Gillies R. Test–Retest Reproducibility Analysis of Lung CT Image Features. Journal of Digital Imaging. 2014b; 27:805–823. [PubMed: 24990346]

Basu S, Kwee T, Gatenby R, Saboury B, Torigian D, Alavi A. Evolving role of molecular imaging with PET in detecting and characterizing heterogeneity of cancer tissue at the primary and metastatic sites, a plausible explanation for failed attempts to cure malignant disorders. European Journal of Nuclear Medicine and Molecular Imaging. 2011; 38:987–991. [PubMed: 21451997]

Brooks FJ, Grigsby PW. The Effect of Small Tumor Volumes on Studies of Intratumoral Heterogeneity of Tracer Uptake. Journal of Nuclear Medicine. 2014; 55:37–42. [PubMed: 24263086]

Buckler AJ, Bresolin L, Dunnick NR, Sullivan DC, Group Ft. A Collaborative Enterprise for Multi-Stakeholder Participation in the Advancement of Quantitative Imaging. Radiology. 2011; 258:906–914. [PubMed: 21339352]

Bundschuh RA, Dinges J, Neumann L, Seyfried M, Zsótér N, Papp L, Rosenberg R, Becker K, Astner ST, Henninger M, Herrmann K, Ziegler SI, Schwaiger M, Essler M. Textural Parameters of Tumor Heterogeneity in 18F-FDG PET/CT for Therapy Response Assessment and Prognosis in Patients with Locally Advanced Rectal Cancer. Journal of Nuclear Medicine. 2014; 55:891–897. [PubMed: 24752672]

Buvat I, Orlhac F, Soussan M. Tumor Texture Analysis in PET: Where Do We Stand? Journal of Nuclear Medicine. 2015; 56:1642–1644. [PubMed: 26294296]

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. [PubMed: 25938522]

Cheebsumon P, Boellaard R, de Ruysscher D, van Elmpt W, van Baardwijk A, Yaqub M, Hoekstra OS, Comans EFI, Lammertsma AA, van Velden FHP. Assessment of tumour size in PET/CT lung cancer studies: PET- and CT-based methods compared to pathology. EJNMMI Research. 2012; 2:56–56. [PubMed: 23034289]

Cheng N-M, Dean Fang Y-H, Tung-Chieh Chang J, Huang C-G, Tsan D-L, Ng S-H, Wang H-M, Lin C-Y, Liao C-T, Yen T-C. Textural Features of Pretreatment 18F-FDG PET/CT Images: Prognostic Significance in Patients with Advanced T-Stage Oropharyngeal Squamous Cell Carcinoma. Journal of Nuclear Medicine. 2013a; 54:1703–1709. [PubMed: 24042030]

Cheng N-M, Fang Y-H, Lee L-y, Chang J-C, Tsan D-L, Ng S-H, Wang H-M, Liao C-T, Yang L-Y, Hsu C-H, Yen T-C. Zone-size nonuniformity of 18F-FDG PET regional textural features predicts survival in patients with oropharyngeal cancer. European Journal of Nuclear Medicine and Molecular Imaging. 2014:1–10. [PubMed: 24196913]

Cheng N-M, Fang Y-H, Yen T-C. The promise and limits of PET texture analysis. Annals of Nuclear Medicine. 2013b; 27:867–869. [PubMed: 23943197]

Chicklore S, Goh V, Siddique M, Roy A, Marsden P, Cook GR. Quantifying tumour heterogeneity in 18F-FDG PET/CT imaging by texture analysis. European Journal of Nuclear Medicine and Molecular Imaging. 2013; 40:133–140. [PubMed: 23064544]

Cook GJR, Yip C, Siddique M, Goh V, Chicklore S, Roy A, Marsden P, Ahmad S, Landau D. Are Pretreatment 18F-FDG PET Tumor Textural Features in Non–Small Cell Lung Cancer Associated with Response and Survival After Chemoradiotherapy? Journal of Nuclear Medicine. 2013; 54:19–26. [PubMed: 23204495]

Cook GR, Siddique M, Taylor B, Yip C, Chicklore S, Goh V. Radiomics in PET: principles and applications. Clinical and Translational Imaging. 2014; 2:269–276.

Coroller TP, Grossmann P, Hou Y, Rios Velazquez E, Leijenaar RTH, Hermann G, Lambin P, Haibe-Kains B, Mak RH, Aerts HJWL. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. Radiotherapy and Oncology. 2015; 114:345–350. [PubMed: 25746350]

Davnall F, Yip CP, Ljungqvist G, Selmi M, Ng F, Sanghera B, Ganeshan B, Miles K, Cook G, Goh V. Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice? Insights into Imaging. 2012; 3:573–589. [PubMed: 23093486]

de Jong JS, van Diest PJ, Baak JP. Heterogeneity and reproducibility of microvessel counts in breast cancer. Laboratory investigation; a journal of technical methods and pathology. 1995; 73:922–926.

Didierlaurent D, Ribes S, Batatia H, Jaudet C, Dierickx LO, Zerdoud S, Brillouet S, Caselles O, Courbon F. The retrospective binning method improves the consistency of phase binning in respiratory-gated PET/CT. Physics in Medicine and Biology. 2012; 57:7829. [PubMed: 23135238]

Diehn M, Nardini C, Wang DS, McGovern S, Jayaraman M, Liang Y, Aldape K, Cha S, Kuo MD. Identification of noninvasive imaging surrogates for brain tumor gene-expression modules. Proceedings of the National Academy of Sciences. 2008; 105:5213–5218.

Dong X, Xing L, Wu P, Fu Z, Wan H, Li D, Yin Y, Sun X, Yu J. 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. Nuclear Medicine Communications. 2013; 34:40–46. [PubMed: 23111378]

Doumou G, Siddique M, Tsoumpas C, Goh V, Cook G. The precision of textural analysis in 18F-FDG-PET scans of oesophageal cancer. European Radiology. 2015; 25:2805–2812. [PubMed: 25994189]

Eary JF, O'Sullivan F, O'Sullivan J, Conrad EU. Spatial Heterogeneity in Sarcoma 18F-FDG Uptake as a Predictor of Patient Outcome. Journal of Nuclear Medicine. 2008; 49:1973–1979. [PubMed: 18997052]

El Naqa I, Grigsby PW, Apte A, Kidd E, Donnelly E, Khullar D, Chaudhari S, Yang D, Schmitt M, Laforest R, Thorstad WL, Deasy JO. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. Pattern Recognition. 2009; 42:1162–1171. [PubMed: 20161266]

Ellingson BM, Lai A, Harris RJ, Selfridge JM, Yong WH, Das K, Pope WB, Nghiemphu PL, Vinters HV, Liau LM, Mischel PS, Cloughesy TF. Probabilistic Radiographic Atlas of Glioblastoma Phenotypes. American Journal of Neuroradiology. 2013; 34:533–540. [PubMed: 22997168]

Fanchon LM, Dogan S, Moreira AL, Carlin SA, Schmidtlein CR, Yorke E, Apte AP, Burger IA, Durack JC, Erinjeri JP, Maybody M, Schöder H, Siegelbaum RH, Sofocleous CT, Deasy JO, Solomon SB, Humm JL, Kirov AS. Feasibility of In Situ, High-Resolution Correlation of Tracer Uptake with Histopathology by Quantitative Autoradiography of Biopsy Specimens Obtained Under 18F-FDG PET/CT Guidance. Journal of Nuclear Medicine. 2015; 56:538–544. [PubMed: 25722446]

Fave X, Cook M, Frederick A, Zhang L, Yang J, Fried D, Stingo F, Court L. Preliminary investigation into sources of uncertainty in quantitative imaging features. Computerized Medical Imaging and Graphics. 2015; 44:54–61. [PubMed: 26004695]

Fisher R, Pusztai L, Swanton C. Cancer heterogeneity: implications for targeted therapeutics. Br J Cancer. 2013; 108:479–485. [PubMed: 23299535]

Foroutan P, Kreahling JM, Morse DL, Grove O, Lloyd MC, Reed D, Raghavan M, Altiok S, Martinez GV, Gillies RJ. Diffusion MRI and Novel Texture Analysis in Osteosarcoma Xenotransplants Predicts Response to Anti-Checkpoint Therapy. PLoS ONE. 2013; 8:e82875. [PubMed: 24358232]

Galavis PE, Hollensen C, Jallow N, Paliwal B, Jeraj R. Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters. Acta Oncologica. 2010; 49:1012–1016. [PubMed: 20831489]

Galloway MM. Texture analysis using gray level run lengths. Computer Graphics and Image Processing. 1975; 4:172–179.

Ganeshan B, Abaleke S, Young RCD, Chatwin CR, Miles KA. 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; 10:137–143. [PubMed: 20605762]

Ganeshan B, Skogen K, Pressney I, Coutroubis D, Miles K. Tumour heterogeneity in oesophageal cancer assessed by CT texture analysis: Preliminary evidence of an association with tumour metabolism, stage, and survival. Clinical Radiology. 2012; 67:157–164. [PubMed: 21943720]

García Vicente AM, Soriano Castrejón AM, Talavera Rubio MP, León Martín AA, Palomar Muñoz AM, Pilkington Woll JP, Poblete García VM. 18F-FDG PET-CT respiratory gating in characterization of pulmonary lesions: approximation towards clinical indications. Annals of Nuclear Medicine. 2010; 24:207–214. [PubMed: 20177834]

Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, Martinez P, Matthews N, Stewart A, Tarpey P, Varela I, Phillimore B, Begum S, McDonald NQ, Butler A, Jones D, Raine K, Latimer C, Santos CR, Nohadani M, Eklund AC, Spencer-Dene B, Clark G, Pickering L, Stamp G, Gore M, Szallasi Z, Downward J, Futreal PA, Swanton C. Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. New England Journal of Medicine. 2012; 366:883–892. [PubMed: 22397650]

Gutman D, Dunn W Jr, Grossmann P, Cooper LD, Holder C, Ligon K, Alexander B, Aerts HWL. Somatic mutations associated with MRI-derived volumetric features in glioblastoma. Neuroradiology. 2015; 57:1227–1237. [PubMed: 26337765]

Guyon I, Elisseeff André. An introduction to variable and feature selection. J Mach Learn Res. 2003; 3:1157–1182.

Haberkorn U, Ziegler SI, Oberdorfer F, Trojan H, Haag D, Peschke P, Berger MR, Altmann A, Van Kaick G. FDG uptake, tumor proliferation and expression of glycolysis associated genes in animal tumor models. Nuclear Medicine and Biology. 1994; 21:827–834. [PubMed: 9234332]

Haralick RM, Shanmugam K, Dinstein IH. Textural Features for Image Classification. Systems, Man and Cybernetics, IEEE Transactions on. 1973; SMC-3:610–621.

Hatt M, Cheze-le Rest C, van Baardwijk A, Lambin P, Pradier O, Visvikis D. Impact of Tumor Size and Tracer Uptake Heterogeneity in 18F-FDG PET and CT Non–Small Cell Lung Cancer Tumor Delineation. Journal of Nuclear Medicine. 2011; 52:1690–1697. [PubMed: 21990577]

Hatt M, Cheze-Lerest C, Turzo A, Roux C, Visvikis D. A fuzzy locally adaptive Bayesian segmentation approach for volume determination in PET. Ieee Transactions on Medical Imaging. 2009; 28:881–893. [PubMed: 19150782]

Hatt M, Majdoub M, Vallières M, Tixier F, Le Rest CC, Groheux D, Hindié E, Martineau A, Pradier O, Hustinx R, Perdrisot R, Guillevin R, El Naqa I, Visvikis D. 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. Journal of Nuclear Medicine. 2015; 56:38–44. [PubMed: 25500829]

Hatt M, Tixier F, Cheze Le, Rest C, Pradier O, Visvikis D. Robustness of intratumour 18F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma. European Journal of Nuclear Medicine and Molecular Imaging. 2013; 40:1662–1671. [PubMed: 23857457]

Hayano K, Yoshida H, Zhu A, Sahani D. Fractal Analysis of Contrast-Enhanced CT Images to Predict Survival of Patients with Hepatocellular Carcinoma Treated with Sunitinib. Digestive Diseases and Sciences. 2014; 59:1996–2003. [PubMed: 24563237]

Henriksson E, Kjellen E, Wahlberg P, Ohlsson T, Wennerberg J, Brun E. 2-Deoxy-2-[18F]Fluoro-D-Glucose Uptake and Correlation to Intratumoral Heterogeneity. Anticancer Research. 2007; 27:2155–2159. [PubMed: 17695498]

Higashi K, Clavo AC, Wahl RL. Does FDG Uptake Measure Proliferative Activity of Human Cancer Cells? In Vitro Comparison with DNA Flow Cytometry and Tritiated Thymidine Uptake. Journal of Nuclear Medicine. 1993; 34:414–419. [PubMed: 8441032]

Higgins KA, Hoang JK, Roach MC, Chino J, Yoo DS, Turkington TG, Brizel DM. Analysis of Pretreatment FDG-PET SUV Parameters in Head-and-Neck Cancer: Tumor SUVmean Has Superior Prognostic Value. International Journal of Radiation Oncology*Biology*Physics. 2012; 82:548–553.

Hilsenbeck S, Clark G, McGuire W. Why do so many prognostic factors fail to pan out? Breast Cancer Research and Treatment. 1992; 22:197–206. [PubMed: 1391986]

Huang T-C, Wang Y-C. Deformation Effect on SUVmax Changes in Thoracic Tumors Using 4-D PET/CT Scan. PLoS ONE. 2013; 8:e58886. [PubMed: 23516568]

Johansen R, Jensen LR, Rydland J, Goa PE, Kvistad KA, Bathen TF, Axelson DE, Lundgren S, Gribbestad IS. Predicting survival and early clinical response to primary chemotherapy for patients with locally advanced breast cancer using DCE-MRI. Journal of Magnetic Resonance Imaging. 2009; 29:1300–1307. [PubMed: 19472387]

Karlo CA, Paolo PLD, Chaim J, Hakimi AA, Ostrovnaya I, Russo P, Hricak H, Motzer R, Hsieh JJ, Akin O. Radiogenomics of Clear Cell Renal Cell Carcinoma: Associations between CT Imaging Features and Mutations. Radiology. 2014; 270:464–471. [PubMed: 24029645]

Kido S, Kuriyama K, Higashiyama M, Kasugai T, Kuroda C. Fractal Analysis of Small Peripheral Pulmonary Nodules in Thin-section CT: Evaluation of the Lung-nodule Interfaces. Journal of Computer Assisted Tomography. 2002; 26

Kim CK, Lim JH, Park CK, Choi D, Lim HK, Lee WJ. Neoangiogenesis and Sinusoidal Capillarization in Hepatocellular Carcinoma: Correlation between Dynamic CT and Density of Tumor Microvessels. Radiology. 2005; 237:529–534. [PubMed: 16244261]

King AD, Chow K-K, Yu K-H, Mo FKF, Yeung DKW, Yuan J, Bhatia KS, Vlantis AC, Ahuja AT. Head and Neck Squamous Cell Carcinoma: Diagnostic Performance of Diffusion-weighted MR Imaging for the Prediction of Treatment Response. Radiology. 2013; 266:531–538. [PubMed: 23151830]

Kjær L, Ring P, Thomsen C, Henriksen O. Texture Analysis in Quantitative MR Imaging. Acta Radiologica. 1995; 36:127–135. [PubMed: 7710790]

Kurland BF, Gerstner ER, Mountz JM, Schwartz LH, Ryan CW, Graham MM, Buatti JM, Fennessy FM, Eikman EA, Kumar V, Forster KM, Wahl RL, Lieberman FS. Promise and pitfalls of quantitative imaging in oncology clinical trials. Magnetic Resonance Imaging. 2012; 30:1301–1312. [PubMed: 22898682]

Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, Zegers CML, Gillies R, Boellard R, Dekker A, Aerts HJWL. Radiomics: Extracting more information from medical images using advanced feature analysis. European Journal of Cancer. 2012; 48:441–446. [PubMed: 22257792]

Lambin P, Roelofs E, Reymen B, Velazquez ER, Buijsen J, Zegers CML, Carvalho S, Leijenaar RTH, Nalbantov G, Oberije C, Scott Marshall M, Hoebers F, Troost EGC, van Stiphout RGPM, van Elmpt W, van der Weijden T, Boersma L, Valentini V, Dekker A. 'Rapid Learning health care in oncology' – An approach towards decision support systems enabling customised radiotherapy'. Radiotherapy and Oncology. 2013; 109:159–164. [PubMed: 23993399]

Leijenaar RTH, Carvalho S, Hoebers FJP, Aerts HJWL, van Elmpt WJC, Huang SH, Chan B, Waldron JN, O'sullivan B, Lambin P. External validation of a prognostic CT-based radiomic signature in oropharyngeal squamous cell carcinoma. Acta Oncologica. 2015a; 54:1423–1429. [PubMed: 26264429]

Leijenaar RTH, Carvalho S, Velazquez ER, van Elmpt WJC, Parmar C, Hoekstra OS, Hoekstra CJ, Boellaard R, Dekker ALAJ, Gillies RJ, Aerts HJWL, Lambin P. Stability of FDG-PET Radiomics features: An integrated analysis of test-retest and inter-observer variability. Acta Oncologica. 2013; 52:1391–1397. [PubMed: 24047337]

Leijenaar RTH, Nalbantov G, Carvalho S, van Elmpt WJC, Troost EGC, Boellaard R, Aerts HJWL, Gillies RJ, Lambin P. The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis. Scientific Reports. 2015b; 5:11075. [PubMed: 26242464]

Lerski RA, Straughan K, Schad LR, Boyce D, Blüml S, Zuna I. VIII. MR image texture analysis—An approach to tissue characterization. Magnetic Resonance Imaging. 1993; 11:873–887. [PubMed: 8371643]

Mahmoud-Ghoneim D, Toussaint G, Constans J-M, de Certaines JD. Three dimensional texture analysis in MRI: a preliminary evaluation in gliomas. Magnetic Resonance Imaging. 2003; 21:983–987. [PubMed: 14684200]

Maley CC, Galipeau PC, Finley JC, Wongsurawat VJ, Li X, Sanchez CA, Paulson TG, Blount PL, Risques R-A, Rabinovitch PS, Reid BJ. Genetic clonal diversity predicts progression to esophageal adenocarcinoma. Nat Genet. 2006; 38:468–473. [PubMed: 16565718]

Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? Nat Rev Cancer. 2012; 12:323–334. [PubMed: 22513401]

McNitt-Gray MF, Hart EM, Wyckoff N, Sayre JW, Goldin JG, Aberle DR. A pattern classification approach to characterizing solitary pulmonary nodules imaged on high resolution CT: Preliminary results. Medical Physics. 1999; 26:880–888. [PubMed: 10436888]

Mu W, Chen Z, Liang Y, Shen W, Yang F, Dai R, Wu N, Tian J. Staging of cervical cancer based on tumor heterogeneity characterized by texture features on 18 F-FDG PET images. Physics in Medicine and Biology. 2015; 60:5123. [PubMed: 26083460]

Mussurakis S, Buckley DL, Coady AM, Turnbull LW, Horsman A. Observer variability in the interpretation of contrast enhanced MRI of the breast. The British Journal of Radiology. 1996; 69:1009–1016. [PubMed: 8958017]

Naeini KM, Pope WB, Cloughesy TF, Harris RJ, Lai A, Eskin A, Chowdhury R, Phillips HS, Nghiemphu PL, Behbahanian Y, Ellingson BM. Identifying the mesenchymal molecular subtype of glioblastoma using quantitative volumetric analysis of anatomic magnetic resonance images. Neuro-Oncology. 2013; 15:626–634. [PubMed: 23444259]

Nair VS, Gevaert O, Davidzon G, Napel S, Graves EE, Hoang CD, Shrager JB, Quon A, Rubin DL, Plevritis SK. Prognostic PET 18F-FDG Uptake Imaging Features Are Associated with Major Oncogenomic Alterations in Patients with Resected Non–Small Cell Lung Cancer. Cancer Research. 2012; 72:3725–3734. [PubMed: 22710433]

Nair VS, Gevaert O, Davidzon G, Plevritis SK, West R. NF-κB protein expression associates with 18F-FDG PET tumor uptake in non-small cell lung cancer: A radiogenomics validation study to understand tumor metabolism. Lung Cancer. 2014; 83:189–196. [PubMed: 24355259]

Nehmeh SA, Erdi YE, Ling CC, Rosenzweig KE, Schoder H, Larson SM, Macapinlac HA, Squire OD, Humm JL. Effect of Respiratory Gating on Quantifying PET Images of Lung Cancer. Journal of Nuclear Medicine. 2002; 43:876–881. [PubMed: 12097456]

Ng F, Kozarski R, Ganeshan B, Goh V. Assessment of tumor heterogeneity by CT texture analysis: Can the largest cross-sectional area be used as an alternative to whole tumor analysis? European Journal of Radiology. 2013; 82:342–348. [PubMed: 23194641]

Nie K, Chen J-H, Yu HJ, Chu Y, Nalcioglu O, Su M-Y. Quantitative Analysis of Lesion Morphology and Texture Features for Diagnostic Prediction in Breast MRI. Academic Radiology. 2008; 15:1513–1525. [PubMed: 19000868]

Orlhac F, Soussan M, Maisonobe J-A, Garcia CA, Vanderlinden B, Buvat I. Tumor Texture Analysis in 18F-FDG PET: Relationships Between Texture Parameters, Histogram Indices, Standardized Uptake Values, Metabolic Volumes, and Total Lesion Glycolysis. Journal of Nuclear Medicine. 2014; 55:414–422. [PubMed: 24549286]

Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL. Machine Learning methods for Quantitative Radiomic Biomarkers. Scientific Reports. 2015a; 5:13087. [PubMed: 26278466]

Parmar C, Leijenaar RTH, Grossmann P, Rios Velazquez E, Bussink J, Rietveld D, Rietbergen MM, Haibe-Kains B, Lambin P, Aerts HJWL. Radiomic feature clusters and Prognostic Signatures specific for Lung and Head & Neck cancer. Scientific Reports. 2015b; 5:11044. [PubMed: 26251068]

Parmar C, Rios Velazquez E, Leijenaar R, Jermoumi M, Carvalho S, Mak RH, Mitra S, Shankar BU, Kikinis R, Haibe-Kains B, Lambin P, Aerts HJWL. Robust Radiomics Feature Quantification Using Semiautomatic Volumetric Segmentation. PLoS ONE. 2014; 9:e102107. [PubMed: 25025374]

Peng S-L, Chen C-F, Liu H-L, Lui C-C, Huang Y-J, Lee T-H, Chang C-C, Wang F-N. Analysis of parametric histogram from dynamic contrast-enhanced MRI: application in evaluating brain tumor response to radiotherapy. NMR in Biomedicine. 2013; 26:443–450. [PubMed: 23073840]

Pentland AP. Fractal-Based Description of Natural Scenes. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 1984; PAMI-6:661–674.

Petkovska I, Shah SK, McNitt-Gray MF, Goldin JG, Brown MS, Kim HJ, Brown K, Aberle DR. Pulmonary nodule characterization: A comparison of conventional with quantitative and visual semi-quantitative analyses using contrast enhancement maps. European Journal of Radiology. 2006; 59:244–252. [PubMed: 16616822]

Rahmim A, Schmidtlein CR, Jackson A, Sheikhbahaei S, Marcus C, Ashrafinia S, Soltani M, Subramaniam MR. A novel metric for quantification of homogeneous and heterogeneous tumors in PET for enhanced clinical outcome prediction. Physics in Medicine and Biology. 2016; 61:227. [PubMed: 26639024]

Rajendran JG, Mankoff DA, O'Sullivan F, Peterson LM, Schwartz DL, Conrad EU, Spence AM, Muzi M, Farwell DG, Krohn KA. Hypoxia and Glucose Metabolism in Malignant Tumors: Evaluation by [18F]Fluoromisonidazole and [18F]Fluorodeoxyglucose Positron Emission Tomography Imaging. Clinical Cancer Research. 2004; 10:2245–2252. [PubMed: 15073099]

Rizk N, Downey RJ, Akhurst T, Gonen M, Bains MS, Larson S, Rusch V. Preoperative 18[F]-Fluorodeoxyglucose Positron Emission Tomography Standardized Uptake Values Predict Survival After Esophageal Adenocarcinoma Resection. The Annals of Thoracic Surgery. 2006; 81:1076–1081. [PubMed: 16488726]

Rouschop KM, Dubois LJ, Keulers TG, van den Beucken T, Lambin P, Bussink J, van der Kogel AJ, Koritzinsky M, Wouters BG. PERK/eIF2α signaling protects therapy resistant hypoxic cells through induction of glutathione synthesis and protection against ROS. Proceedings of the National Academy of Sciences. 2013; 110:4622–4627.

Shukla-Dave A, Lee NY, Jansen JFA, Thaler HT, Stambuk HE, Fury MG, Patel SG, Moreira AL, Sherman E, Karimi S, Wang Y, Kraus D, Shah JP, Pfister DG, Koutcher JA. Dynamic Contrast-Enhanced Magnetic Resonance Imaging as a Predictor of Outcome in Head-and-Neck Squamous

Cell Carcinoma Patients With Nodal Metastases. International Journal of Radiation Oncology*Biology*Physics. 2012; 82:1837–1844.

Soret M, Bacharach SL, Buvat I. Partial-Volume Effect in PET Tumor Imaging. Journal of Nuclear Medicine. 2007; 48:932–945. [PubMed: 17504879]

Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW. Assessing the performance of prediction models: a framework for some traditional and novel measures. Epidemiology (Cambridge, Mass). 2010; 21:128–138.

Tan S, Kligerman S, Chen W, Lu M, Kim G, Feigenberg S, D'Souza WD, Suntharalingam M, Lu W. Spatial-Temporal [18F]FDG-PET Features for Predicting Pathologic Response of Esophageal Cancer to Neoadjuvant Chemoradiation Therapy. International Journal of Radiation Oncology*Biology*Physics. 2013; 85:1375–1382.

Tateishi U, Kusumoto M, Nishihara H, Nagashima K, Morikawa T, Moriyama N. Contrast-enhanced dynamic computed tomography for the evaluation of tumor angiogenesis in patients with lung carcinoma. Cancer. 2002; 95:835–842. [PubMed: 12209728]

Thibault G, FERTIL B, NAVARRO C, PEREIRA S, CAU P, LEVY N, SEQUEIRA J, MARI J-L. SHAPE AND TEXTURE INDEXES APPLICATION TO CELL NUCLEI CLASSIFICATION. International Journal of Pattern Recognition and Artificial Intelligence. 2013; 27:1357002.

Tixier F, Groves AM, Goh V, Hatt M, Ingrand P, Le Rest CC, Visvikis D. Correlation of Intra-Tumor (18)F-FDG Uptake Heterogeneity Indices with Perfusion CT Derived Parameters in Colorectal Cancer. PLoS ONE. 2014a; 9:e99567. [PubMed: 24926986]

Tixier F, Groves AM, Goh V, Hatt M, Ingrand P, Le Rest CC, Visvikis D. Correlation of Intra-Tumor <sup>18</sup>F-FDG Uptake Heterogeneity Indices with Perfusion CT Derived Parameters in Colorectal Cancer. PLoS ONE. 2014b; 9:e99567. [PubMed: 24926986]

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. Journal of Nuclear Medicine. 2012; 53:693–700. [PubMed: 22454484]

Tixier F, Hatt M, Valla C, Fleury V, Lamour C, Ezzouhri S, Ingrand P, Perdrisot R, Visvikis D, Le Rest CC. Visual Versus Quantitative Assessment of Intratumor 18F-FDG PET Uptake Heterogeneity: Prognostic Value in Non–Small Cell Lung Cancer. Journal of Nuclear Medicine. 2014c; 55:1235–1241. [PubMed: 24904113]

Tixier F, Le Rest CC, Hatt M, Albarghach N, Pradier O, Metges J-P, Corcos L, Visvikis D. Intratumor Heterogeneity Characterized by Textural Features on Baseline 18F-FDG PET Images Predicts Response to Concomitant Radiochemotherapy in Esophageal Cancer. Journal of Nuclear Medicine. 2011; 52:369–378. [PubMed: 21321270]

Vallieres M, Freeman CR, Skamene SR, Naqa IE. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. Physics in Medicine and Biology. 2015; 60:5471. [PubMed: 26119045]

van Velden F, Cheebsumon P, Yaqub M, Smit E, Hoekstra O, Lammertsma A, Boellaard R. 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; 38:1636–1647. [PubMed: 21617975]

van Velden FP, Nissen I, Jongsma F, Velasquez L, Hayes W, Lammertsma A, Hoekstra O, Boellaard R. Test-Retest Variability of Various Quantitative Measures to Characterize Tracer Uptake and/or Tracer Uptake Heterogeneity in Metastasized Liver for Patients with Colorectal Carcinoma. Molecular Imaging and Biology. 2014; 16:13–18. [PubMed: 23807457]

Velazquez ER, Parmar C, Jermoumi M, Mak RH, van Baardwijk A, Fennessy FM, Lewis JH, De Ruysscher D, Kikinis R, Lambin P, Aerts HJWL. Volumetric CT-based segmentation of NSCLC using 3D-Slicer. Sci Rep. 2013; 3:3529. [PubMed: 24346241]

Way TW, Hadjiiski LM, Sahiner B, Chan H-P, Cascade PN, Kazerooni EA, Bogot N, Zhou C. Computer-aided diagnosis of pulmonary nodules on CT scans: Segmentation and classification using 3D active contours. Medical physics. 2006; 33:2323–2337. [PubMed: 16898434]

Wetzel SG, Cha S, Johnson G, Lee P, Law M, Kasow DL, Pierce SD, Xue X. Relative Cerebral Blood Volume Measurements in Intracranial Mass Lesions: Interobserver and Intraobserver Reproducibility Study. Radiology. 2002; 224:797–803. [PubMed: 12202717]

Xu R, Kido S, Suga K, Hirano Y, Tachibana R, Muramatsu K, Chagawa K, Tanaka S. Texture analysis on 18F-FDG PET/CT images to differentiate malignant and benign bone and soft-tissue lesions. Annals of Nuclear Medicine. 2014; 28:926–935. [PubMed: 25107363]

Yamamoto S, Korn RL, Oklu R, Migdal C, Gotway MB, Weiss GJ, Iafrate AJ, Kim D-W, Kuo MD. ALK Molecular Phenotype in Non–Small Cell Lung Cancer: CT Radiogenomic Characterization. Radiology. 2014; 272:568–576. [PubMed: 24885982]

Yan J, Chu-Shern JL, Loi HY, Khor LK, Sinha AK, Quek ST, Tham IWK, Townsend D. Impact of Image Reconstruction Settings on Texture Features in 18F-FDG PET. Journal of Nuclear Medicine. 2015; 56:1667–1673. [PubMed: 26229145]

Yang F, Thomas M, Dehdashti F, Grigsby P. Temporal analysis of intratumoral metabolic heterogeneity characterized by textural features in cervical cancer. European Journal of Nuclear Medicine and Molecular Imaging. 2013; 40:716–727. [PubMed: 23340594]

Yang X, Knopp M. Quantifying Tumor Vascular Heterogeneity with Dynamic Contrast-Enhanced Magnetic Resonance Imaging: A Review. Journal of Biomedicine and Biotechnology. 2011; 2011:12.

Yip S, McCall K, Aristophanous M, Chen AB, Aerts HJWL, Berbeco R. Comparison of Texture Features Derived from Static and Respiratory-Gated PET Images in Non-Small Cell Lung Cancer. PLoS ONE. 2014; 9:e115510. [PubMed: 25517987]

Yip SFS, Coroller PT, Sanford NN, Huynh E, Mamon H, Aerts JWLH, Berbeco IR. Use of registration-based contour propagation in texture analysis for esophageal cancer pathologic response prediction. Physics in Medicine and Biology. 2016; 61:906. [PubMed: 26738433]

Yu H, Caldwell C, Mah K, Poon I, Balogh J, MacKenzie R, Khaouam N, Tirona R. Automated Radiation Targeting in Head-and-Neck Cancer Using Region-Based Texture Analysis of PET and CT Images. International Journal of Radiation Oncology*Biology*Physics. 2009; 75:618–625.

Zhang, H-q; Yu, J-m; Meng, X.; Yue, J-b; Feng, R.; Ma, L. Prognostic value of serial [18F]fluorodeoxyglucose PET-CT uptake in stage III patients with non-small cell lung cancer treated by concurrent chemoradiotherapy. European Journal of Radiology. 2011; 77:92–96. [PubMed: 19695804]

Zhang H, Tan S, Chen W, Kligerman S, Kim G, D'Souza WD, Suntharalingam M, Lu W. Modeling Pathologic Response of Esophageal Cancer to Chemoradiation Therapy Using Spatial-Temporal 18F-FDG PET Features, Clinical Parameters, and Demographics. International Journal of Radiation Oncology*Biology*Physics. 2014; 88:195–203.

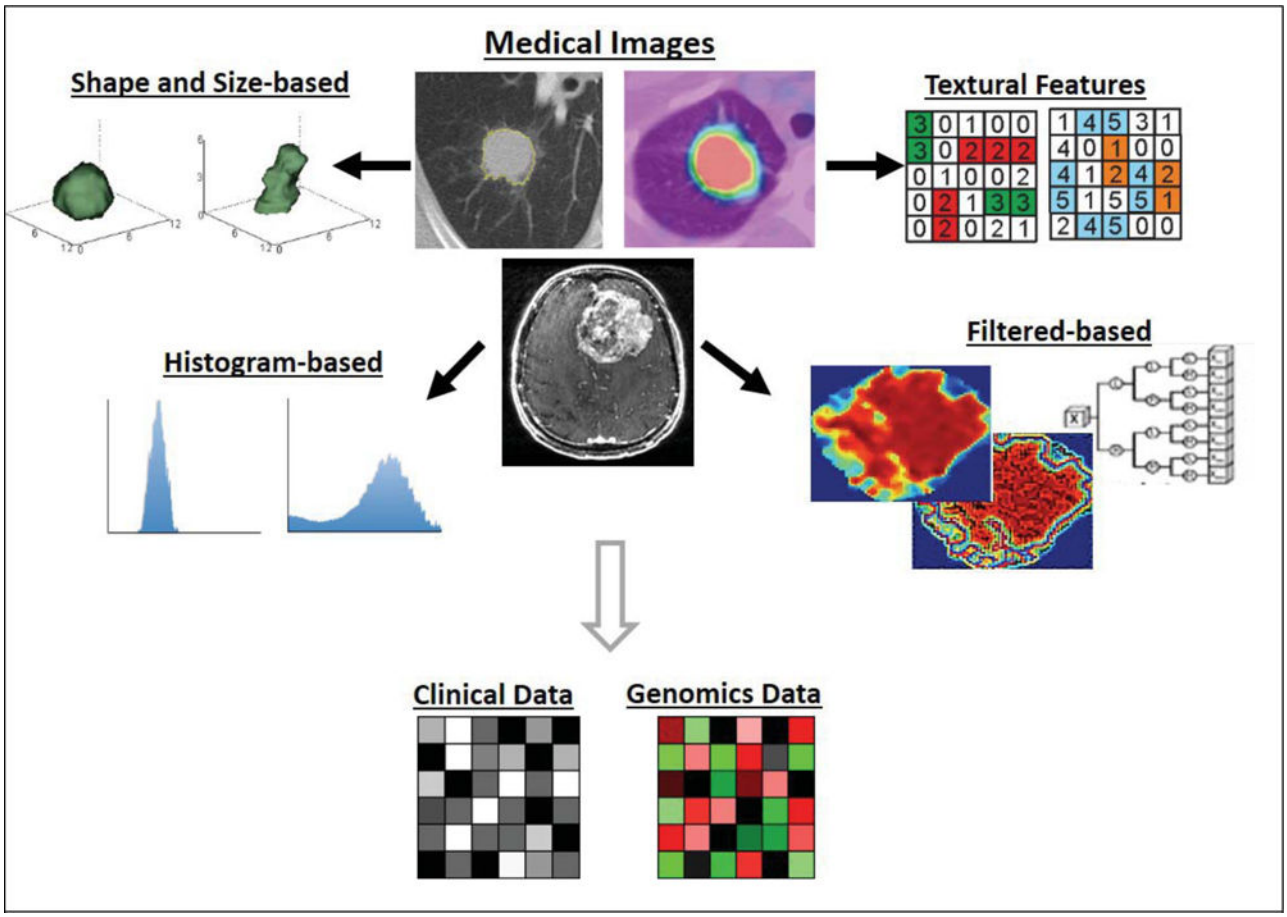


**Figure 1.**

Radiomic workflow. **(Top)** Various radiomic features, such as shape/size-based, Histogram-based, filtered-based, and textural features, can be extracted from the medical images within the tumours. **(Bottom)** The radiomic features are then compared with the clinical and genomics data.