Physics in Medicine & Biology

TOPICAL REVIEW

Applications and limitations of radiomics

To cite this article: Stephen S F Yip and Hugo J W L Aerts 2016 *Phys. Med. Biol.* **61** R150

IOP Publishing | Institute of Physics and Engineering in Medicine Physics in Medicine & Biology

Phys. Med. Biol. **61** (2016) R150–166 doi:10.1088/0031-9155/61/13/R150

Topical Review

Applications and limitations of radiomics

Stephen S F Yip[1] **and Hugo J W L Aerts**[1,2]

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

E-mail: Hugo_Aerts@dfci.harvard.edu

Received 21 July 2015, revised 29 February 2016
Accepted for publication 5 April 2016
Published 7 June 2016

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.

Keywords: PET, quantitative imaging, radiomics

(Some figures may appear in colour only in the online journal)

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 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, ^{18F}-fluorodeoxyglucose ([^{18F}]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* 2014b). Therefore, it is important to objectively and reproducibly quantify various imaging features that may reveal the underlying biology of tumours.

Radiomics 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). Radiomics 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 radiomics

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 (LoG) 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,

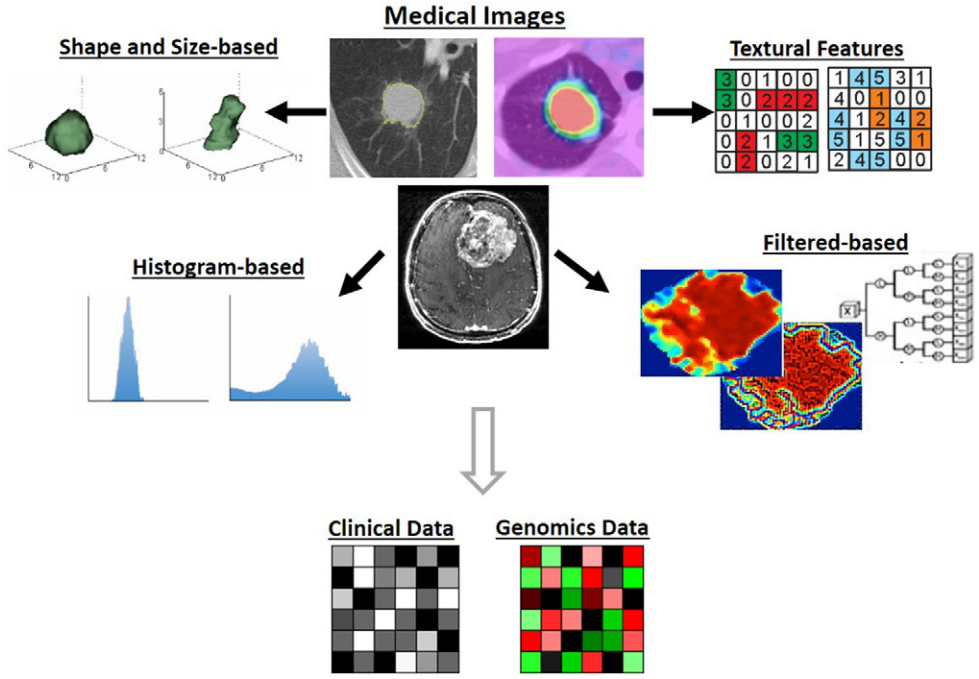

Figure 1. Radiomic workflow. (Top) Various radiomic features, such as shape/size-based, histogram-based, filtered-based, and textural features, can be extracted from the medical images within the tumours. (Bottom) The radiomic features are then compared with the clinical and genomics data.

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 CT 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* (2015b) 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 (7th 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 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 $\geqslant 0.80$) (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 CE-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 radiomics. 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 Platelet-derived growth factor receptor alpha (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, 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, 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 pathophysiological 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, 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, 2014b, van Velden *et al* 2014). Balagurunathan *et al* (2014a, 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 min apart. Of these features, only 66 of them were found to have ICC $\geqslant$ 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 d 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 d 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* 2014a, 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, 2015b). For instance, Leijenaar *et al* (2015b) 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 $= 2 \bullet \mathrm{IQR}/\mathrm{N}^{-1/3}$), 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 single-photon emission computed tomography (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, 2014, Cook *et al* 2013, Aerts *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, 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 H J W L *et al* 2014 Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach *Nat. Commun.* **5** 4006

Alic L, Niessen W J and Veenland J F 2014 Quantification of heterogeneity as a biomarker in tumor imaging: a systematic review *PLoS One* **9** e110300

Altman D G, Lausen B, Sauerbrei W and Schumacher M 1994 Dangers of using 'optimal' cutpoints in the evaluation of prognostic factors *J. Natl Cancer Inst.* **86** 829–35

Amadasun M and King R 1989 Textural features corresponding to textural properties *IEEE Trans. Syst. Man Cybern.* **19** 1264–74

Aristophanous M, Yong Y, Yap J T, Killoran J H, Allen A M, Berbeco R I and Chen A B 2012 Evaluating FDG uptake changes between pre and post therapy respiratory gated PET scans *Radiother. Oncol.* **102** 377–82

Baek H J, Kim H S, Kim N, Choi Y J and Kim Y J 2012 Percent change of perfusion skewness and kurtosis: a potential imaging biomarker for early treatment response in patients with newly diagnosed glioblastomas *Radiology* **264** 834–43

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

Balagurunathan Y *et al* 2014b Test–retest reproducibility analysis of lung CT image features *J. Digit. Imaging* **27** 805–23

Basu S, Kwee T, Gatenby R, Saboury B, Torigian D and Alavi A 2011 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 *Eur. J. Nucl. Med. Mol. Imaging* **38** 987–91

Boellaard R 2011 Need for standardization of 18F-FDG PET/CT for treatment response assessments *J. Nucl. Med.* **52** 93S–100S

Boellaard R *et al* 2015 FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0 *Eur. J. Nucl. Med. Mol. Imaging* **42** 328–54

Brooks F J and Grigsby P W 2014 The effect of small tumor volumes on studies of intratumoral heterogeneity of tracer uptake *J. Nucl. Med.* **55** 37–42

Buckler A J, Bresolin L, Dunnick N R, Sullivan D C and Group F T 2011 A collaborative enterprise for multi-stakeholder participation in the advancement of quantitative imaging *Radiology* **258** 906–14

Bundschuh R A *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** 891–7

Buvat I, Orlhac F and Soussan M 2015 Tumor texture analysis in PET: where do we stand? *J. Nucl. Med.* **56** 1642–4

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

Cheebsumon P, Boellaard R, de Ruysscher D, van Elmpt W, van Baardwijk A, Yaqub M, Hoekstra O S, Comans E F I, Lammertsma A A and van Velden F H P 2012 Assessment of tumour size in PET/CT lung cancer studies: PET- and CT-based methods compared to pathology *EJNMMI Res.* **2** 56–56

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 and Yen T-C 2013a Textural features of pretreatment 18F-FDG PET/CT images: prognostic significance in patients with advanced T-stage oropharyngeal squamous cell carcinoma *J. Nucl. Med.* **54** 1703–9

Cheng N-M, Fang Y-H and Yen T-C 2013b The promise and limits of PET texture analysis *Ann. Nucl. Med.* **27** 867–9

Cheng N-M *et al* 2014 Zone-size nonuniformity of 18F-FDG PET regional textural features predicts survival in patients with oropharyngeal cancer *Eur. J. Nucl. Med. Mol. Imaging* **42** 419–28

Chicklore S, Goh V, Siddique M, Roy A, Marsden P and Cook G R 2013 Quantifying tumour heterogeneity in 18F-FDG PET/CT imaging by texture analysis *Eur. J. Nucl. Med. Mol. Imaging* **40** 133–40

Cook G R, Siddique M, Taylor B, Yip C, Chicklore S and Goh V 2014 Radiomics in PET: principles and applications *Clin. Transl. Imaging* **2** 269–76

Cook G J R, Yip C, Siddique M, Goh V, Chicklore S, Roy A, Marsden P, Ahmad S and Landau D 2013 Are pretreatment 18F-FDG PET tumor textural features in non–small cell lung cancer associated with response and survival after chemoradiotherapy? *J. Nucl. Med.* **54** 19–26

Coroller T P, Grossmann P, Hou Y, Rios Velazquez E, Leijenaar R T H, Hermann G, Lambin P, Haibe-Kains B, Mak R H and Aerts H J W L 2015 CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma *Radiother. Oncol.* **114** 345–50

Davnall F, Yip C P, Ljungqvist G, Selmi M, Ng F, Sanghera B, Ganeshan B, Miles K, Cook G and Goh V 2012 Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice? *Insights Imaging* **3** 573–89

de Jong J S, van Diest P J and Baak J P 1995 Heterogeneity and reproducibility of microvessel counts in breast cancer *Lab. Invest.* **73** 922–6

Didierlaurent D, Ribes S, Batatia H, Jaudet C, Dierickx L O, Zerdoud S, Brillouet S, Caselles O and Courbon F 2012 The retrospective binning method improves the consistency of phase binning in respiratory-gated PET/CT *Phys. Med. Biol.* **57** 7829

Diehn M, Nardini C, Wang D S, McGovern S, Jayaraman M, Liang Y, Aldape K, Cha S and Kuo M D 2008 Identification of noninvasive imaging surrogates for brain tumor gene-expression modules *Proc. Natl Acad. Sci* **105** 5213–8

Dong X, Xing L, Wu P, Fu Z, Wan H, Li D, Yin Y, Sun X and Yu J 2013 Three-dimensional positron emission tomography image texture analysis of esophageal squamous cell carcinoma: relationship between tumor 18F-fluorodeoxyglucose uptake heterogeneity, maximum standardized uptake value, and tumor stage *Nucl. Med. Commun.* **34** 40–6

Doumou G, Siddique M, Tsoumpas C, Goh V and Cook G 2015 The precision of textural analysis in 18F-FDG-PET scans of oesophageal cancer *Eur. Radiol.* **25** 2805–12

Eary J F, O'Sullivan F, O'Sullivan J and Conrad E U 2008 Spatial heterogeneity in sarcoma 18F-FDG uptake as a predictor of patient outcome *J. Nucl. Med.* **49** 1973–9

El Naqa I *et al* 2009 Exploring feature-based approaches in PET images for predicting cancer treatment outcomes *Pattern Recognit.* **42** 1162–71

Ellingson B M *et al* 2013 Probabilistic radiographic atlas of glioblastoma phenotypes *Am. J. Neuroradiol.* **34** 533–40

Fanchon L M *et al* 2015 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 *J. Nucl. Med.* **56** 538–44

Fave X, Cook M, Frederick A, Zhang L, Yang J, Fried D, Stingo F and Court L 2015 Preliminary investigation into sources of uncertainty in quantitative imaging features *Comput. Med. Imaging Graph.* **44** 54–61

Fisher R, Pusztai L and Swanton C 2013 Cancer heterogeneity: implications for targeted therapeutics *Br. J. Cancer* **108** 479–85

Foroutan P, Kreahling J M, Morse D L, Grove O, Lloyd M C, Reed D, Raghavan M, Altiok S, Martinez G V and Gillies R J 2013 Diffusion MRI and novel texture analysis in osteosarcoma xenotransplants predicts response to anti-checkpoint therapy *PLoS One* **8** e82875

Galavis P E, Hollensen C, Jallow N, Paliwal B and Jeraj R 2010 Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters *Acta Oncol.* **49** 1012–6

Galloway M M 1975 Texture analysis using gray level run lengths *Comput. Graph. Image Process.* **4** 172–9

Ganeshan B, Abaleke S, Young R C D, Chatwin C R and Miles K A 2010 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* **10** 137–43

Ganeshan B, Skogen K, Pressney I, Coutroubis D and Miles K 2012 Tumour heterogeneity in oesophageal cancer assessed by CT texture analysis: preliminary evidence of an association with tumour metabolism, stage, and survival *Clin. Radiol.* **67** 157–64

García Vicente A M, Soriano Castrejón A M, Talavera Rubio M P, León Martín A A, Palomar Muñoz A M, Pilkington Woll J P and Poblete García V M 2010 18F-FDG PET-CT respiratory gating in characterization of pulmonary lesions: approximation towards clinical indications *Ann. Nucl. Med.* **24** 207–14

Gerlinger M *et al* 2012 Intratumor heterogeneity and branched evolution revealed by multiregion sequencing *New Engl. J. Med.* **366** 883–92

Gutman D, Dunn W Jr, Grossmann P, Cooper L D, Holder C, Ligon K, Alexander B and Aerts H W L 2015 Somatic mutations associated with MRI-derived volumetric features in glioblastoma *Neuroradiology* **57** 1227–37

Guyon I and André E 2003 An introduction to variable and feature selection *J. Mach. Learn. Res.* **3** 1157–82

Haberkorn U, Ziegler S I, Oberdorfer F, Trojan H, Haag D, Peschke P, Berger M R, Altmann A and Van Kaick G 1994 FDG uptake, tumor proliferation and expression of glycolysis associated genes in animal tumor models *Nucl. Med. Biol.* **21** 827–34

Haralick R M, Shanmugam K and Dinstein I H 1973 Textural features for image classification *IEEE Trans. Syst. Man Cybern.* **SMC-3** 610–21

Hatt M, Cheze-Lerest C, Turzo A, Roux C and Visvikis D 2009 A fuzzy locally adaptive Bayesian segmentation approach for volume determination in PET *IEEE Trans. Med. Imaging* **28** 881–93

Hatt M, Cheze-le Rest C, van Baardwijk A, Lambin P, Pradier O and Visvikis D 2011 Impact of tumor size and tracer uptake heterogeneity in 18F-FDG PET and CT non–small cell lung cancer tumor delineation *J. Nucl. Med.* **52** 1690–7

Hatt M, Tixier F, Cheze Le Rest C, Pradier O and Visvikis D 2013 Robustness of intratumour 18F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma *Eur. J. Nucl. Med. Mol. Imaging* **40** 1662–71

Hatt 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

Hayano K, Yoshida H, Zhu A and Sahani D 2014 Fractal analysis of contrast-enhanced CT images to predict survival of patients with hepatocellular carcinoma treated with sunitinib *Dig. Dis. Sci.* **59** 1996–2003

Henriksson E, Kjellen E, Wahlberg P, Ohlsson T, Wennerberg J and Brun E 2007 2-Deoxy-2-[18F] Fluoro-D-Glucose uptake and correlation to intratumoral heterogeneity *Anticancer Res.* **27** 2155–9

Higashi K, Clavo A C and Wahl R L 1993 Does FDG uptake measure proliferative activity of human cancer cells? *In vitro* comparison with DNA flow cytometry and tritiated thymidine uptake *J. Nucl. Med.* **34** 414–9

Higgins K A, Hoang J K, Roach M C, Chino J, Yoo D S, Turkington T G and Brizel D M 2012 Analysis of pretreatment FDG-PET SUV parameters in head-and-neck cancer: tumor suvmean has superior prognostic value *Int. J. Radiat. Oncol. Biol. Phys.* **82** 548–53

Hilsenbeck S, Clark G and McGuire W 1992 Why do so many prognostic factors fail to pan out? *Breast Cancer Res. Treat.* **22** 197–206

Huang T-C and Wang Y-C 2013 Deformation effect on SUVmax changes in thoracic tumors using 4-D PET/CT scan *PLoS One* **8** e58886

Johansen R, Jensen L R, Rydland J, Goa P E, Kvistad K A, Bathen T F, Axelson D E, Lundgren S and Gribbestad I S 2009 Predicting survival and early clinical response to primary chemotherapy for patients with locally advanced breast cancer using DCE-MRI *J. Magn. Reson. Imaging* **29** 1300–7

Karlo C A, Paolo P L D, Chaim J, Hakimi A A, Ostrovnaya I, Russo P, Hricak H, Motzer R, Hsieh J J and Akin O 2014 Radiogenomics of clear cell renal cell carcinoma: associations between CT imaging features and mutations *Radiol.* **270** 464–71

Kido S, Kuriyama K, Higashiyama M, Kasugai T and Kuroda C 2002 Fractal analysis of small peripheral pulmonary nodules in thin-section CT: evaluation of the lung-nodule interfaces *J. Comput. Assist. Tomogr.* **26** 573–8

Kim C K, Lim J H, Park C K, Choi D, Lim H K and Lee W J 2005 Neoangiogenesis and sinusoidal capillarization in hepatocellular carcinoma: correlation between dynamic CT and density of tumor microvessels *Radiology* **237** 529–34

King A D, Chow K-K, Yu K-H, Mo F K F, Yeung D K W, Yuan J, Bhatia K S, Vlantis A C and Ahuja A T 2013 Head and neck squamous cell carcinoma: diagnostic performance of diffusion-weighted MR imaging for the prediction of treatment response *Radiology* **266** 531–8

Kjær L, Ring P, Thomsen C and Henriksen O 1995 Texture analysis in quantitative MR imaging *Acta Radiol.* **36** 127–35

Kurland B F *et al* 2012 Promise and pitfalls of quantitative imaging in oncology clinical trials *Magn. Reson. Imaging* **30** 1301–12

Lambin P *et al* 2012 Radiomics: extracting more information from medical images using advanced feature analysis *Eur. J. Cancer* **48** 441–6

Lambin P *et al* 2013 'Rapid Learning health care in oncology'—an approach towards decision support systems enabling customised radiotherapy' *Radiother. Oncol.* **109** 159–64

Leijenaar R T H, Carvalho S, Hoebers F J P, Aerts H J W L, van Elmpt W J C, Huang S H, Chan B, Waldron J N, O'sullivan B and Lambin P 2015a External validation of a prognostic CT-based radiomic signature in oropharyngeal squamous cell carcinoma *Acta Oncol.* **54** 1423–9

Leijenaar R T H, Nalbantov G, Carvalho S, van Elmpt W J C, Troost E G C, Boellaard R, Aerts H J W L, Gillies R J and Lambin P 2015b The effect of SUV discretization in quantitative FDG-PET radiomics: the need for standardized methodology in tumor texture analysis *Sci. Rep.* **5** 11075

Leijenaar R T H *et al* 2013 Stability of FDG-PET Radiomics features: an integrated analysis of test-retest and inter-observer variability *Acta Oncol.* **52** 1391–7

Lerski R A, Straughan K, Schad L R, Boyce D, Blüml S and Zuna I 1993 VIII. MR image texture analysis—an approach to tissue characterization *Magn. Reson. Imaging* **11** 873–87

Mahmoud-Ghoneim D, Toussaint G, Constans J-M and de Certaines J D 2003 Three dimensional texture analysis in MRI: a preliminary evaluation in gliomas *Magn. Reson. Imaging* **21** 983–7

Maley C C *et al* 2006 Genetic clonal diversity predicts progression to esophageal adenocarcinoma *Nat. Genet.* **38** 468–73

Marusyk A, Almendro V and Polyak K 2012 Intra-tumour heterogeneity: a looking glass for cancer? *Nat. Rev. Cancer* **12** 323–34

McNitt-Gray M F, Hart E M, Wyckoff N, Sayre J W, Goldin J G and Aberle D R 1999 A pattern classification approach to characterizing solitary pulmonary nodules imaged on high resolution CT: preliminary results *Med. Phys.* **26** 880–8

Mu W, Chen Z, Liang Y, Shen W, Yang F, Dai R, Wu N and Tian J 2015 Staging of cervical cancer based on tumor heterogeneity characterized by texture features on 18 F-FDG PET images *Phys. Med. Biol.* **60** 5123

Mussurakis S, Buckley D L, Coady A M, Turnbull L W and Horsman A 1996 Observer variability in the interpretation of contrast enhanced MRI of the breast *Br. J. Radiol.* **69** 1009–16

Naeini K M *et al* 2013 Identifying the mesenchymal molecular subtype of glioblastoma using quantitative volumetric analysis of anatomic magnetic resonance images *Neuro-Oncology* **15** 626–34

Nair V S, Gevaert O, Davidzon G, Napel S, Graves E E, Hoang C D, Shrager J B, Quon A, Rubin D L and Plevritis S K 2012 Prognostic PET 18F-FDG uptake imaging features are associated with major oncogenomic alterations in patients with resected non–small cell lung cancer *Cancer Res.* **72** 3725–34

Nair V S, Gevaert O, Davidzon G, Plevritis S K and West R 2014 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* **83** 189–96

Nehmeh S A, Erdi Y E, Ling C C, Rosenzweig K E, Schoder H, Larson S M, Macapinlac H A, Squire O D and Humm J L 2002 Effect of respiratory gating on quantifying PET images of lung cancer *J. Nucl. Med.* **43** 876–81

Ng F, Kozarski R, Ganeshan B and 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** 342–8

Nie K, Chen J-H, Yu H J, Chu Y, Nalcioglu O and Su M-Y 2008 Quantitative analysis of lesion morphology and texture features for diagnostic prediction in breast MRI *Acad. Radiol.* **15** 1513–25

Nyflot M J, Yang F, Byrd D, Bowen S R, Sandison G A and Kinahan P E 2015 Quantitative radiomics: impact of stochastic effects on textural feature analysis implies the need for standards *J. Med. Imaging* **2** 041002

Orlhac F, Soussan M, Maisonobe J-A, Garcia C A, Vanderlinden B and 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** 414–22

Parmar C, Grossmann P, Bussink J, Lambin P and Aerts H J W L 2015a Machine learning methods for quantitative radiomic biomarkers *Sci. Rep.* **5** 13087

Parmar C, Leijenaar R T H, Grossmann P, Rios Velazquez E, Bussink J, Rietveld D, Rietbergen M M, Haibe-Kains B, Lambin P and Aerts H J W L 2015b Radiomic feature clusters and prognostic signatures specific for lung and head & neck cancer *Sci. Rep.* **5** 11044

Parmar C *et al* 2014 Robust radiomics feature quantification using semiautomatic volumetric segmentation *PLoS One* **9** e102107

Peng S-L, Chen C-F, Liu H-L, Lui C-C, Huang Y-J, Lee T-H, Chang C-C and Wang F-N 2013 Analysis of parametric histogram from dynamic contrast-enhanced MRI: application in evaluating brain tumor response to radiotherapy *NMR Biomed.* **26** 443–50

Pentland A P 1984 Fractal-based description of natural scenes *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-6** 661–74

Petkovska I, Shah S K, McNitt-Gray M F, Goldin J G, Brown M S, Kim H J, Brown K and Aberle D R 2006 Pulmonary nodule characterization: a comparison of conventional with quantitative and visual semi-quantitative analyses using contrast enhancement maps *Eur. J. Radiol.* **59** 244–52

Rahmim A, Schmidtlein C R, Jackson A, Sheikhbahaei S, Marcus C, Ashrafinia S, Soltani M and Subramaniam M R 2016 A novel metric for quantification of homogeneous and heterogeneous tumors in PET for enhanced clinical outcome prediction *Phys. Med. Biol.* **61** 227

Rajendran J G, Mankoff D A, O'Sullivan F, Peterson L M, Schwartz D L, Conrad E U, Spence A M, Muzi M, Farwell D G and Krohn K A 2004 Hypoxia and glucose metabolism in malignant tumors: evaluation by [18F]Fluoromisonidazole and [18F]Fluorodeoxyglucose positron emission tomography imaging *Clin. Cancer Res.* **10** 2245–52

Rizk N, Downey R J, Akhurst T, Gonen M, Bains M S, Larson S and Rusch V 2006 Preoperative 18[F]-Fluorodeoxyglucose positron emission tomography standardized uptake values predict survival after esophageal adenocarcinoma resection *Ann. Thorac. Surg.* **81** 1076–81

Shukla-Dave A *et al* 2012 Dynamic contrast-enhanced magnetic resonance imaging as a predictor of outcome in head-and-neck squamous cell carcinoma patients with nodal metastases *Int. J. Radiat. Oncol. Biol. Phys.* **82** 1837–44

Soret M, Bacharach S L and Buvat I 2007 Partial-volume effect in PET tumor imaging *J. Nucl. Med.* **48** 932–45

Steyerberg E W, Vickers A J, Cook N R, Gerds T, Gonen M, Obuchowski N, Pencina M J and Kattan M W 2010 Assessing the performance of prediction models: a framework for some traditional and novel measures *Epidemiology* **21** 128–38

Tan S, Kligerman S, Chen W, Lu M, Kim G, Feigenberg S, D'Souza W D, Suntharalingam M and Lu W 2013 Spatial-temporal [18F]FDG-PET features for predicting pathologic response of esophageal cancer to neoadjuvant chemoradiation therapy *Int. J. Radiat. Oncol. Biol. Phys.* **85** 1375–82

Tateishi U, Kusumoto M, Nishihara H, Nagashima K, Morikawa T and Moriyama N 2002 Contrast-enhanced dynamic computed tomography for the evaluation of tumor angiogenesis in patients with lung carcinoma *Cancer* **95** 835–42

Thibault G, Fertil B, Navarro C, Pereira S, Cau P, Levy N, Sequeira J and Mari J-L 2013 Shape and texture indexes application to cell nuclei classification *Int. J. Pattern Recognit. Artif. Intell.* **27** 1357002

Tixier F, Groves A M, Goh V, Hatt M, Ingrand P, Le Rest C C and Visvikis D 2014a Correlation of intra-tumor (18)F-FDG uptake heterogeneity indices with perfusion CT derived parameters in colorectal cancer *PLoS One* **9** e99567

Tixier F, Hatt M, Valla C, Fleury V, Lamour C, Ezzouhri S, Ingrand P, Perdrisot R, Visvikis D and Le Rest C C 2014b Visual versus quantitative assessment of intratumor 18F-FDG PET uptake heterogeneity: prognostic value in non-small cell lung cancer *J. Nucl. Med.* **55** 1235–41

Tixier F, Hatt M, Le Rest C C, Le Pogam A, Corcos L and Visvikis D 2012 Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in 18F-FDG PET *J. Nucl. Med.* **53** 693–700

Tixier F, Le Rest C C, Hatt M, Albarghach N, Pradier O, Metges J-P, Corcos L and Visvikis D 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** 369–78

Vallieres M, Freeman C R, Skamene S R and Naqa I E 2015 A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities *Phys. Med. Biol.* **60** 5471

van Velden F, Cheebsumon P, Yaqub M, Smit E, Hoekstra O, Lammertsma A and Boellaard R 2011 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* **38** 1636–47

van Velden F P, Nissen I, Jongsma F, Velasquez L, Hayes W, Lammertsma A, Hoekstra O and Boellaard R 2014 Test-retest variability of various quantitative measures to characterize tracer uptake and/or tracer uptake heterogeneity in metastasized liver for patients with colorectal carcinoma *Mol. Imaging Biol.* **16** 13–8

Velazquez E R *et al* 2013 Volumetric CT-based segmentation of NSCLC using 3D-Slicer *Sci. Rep.* **3** 3529

Way T W, Hadjiiski L M, Sahiner B, Chan H-P, Cascade P N, Kazerooni E A, Bogot N and Zhou C 2006 Computer-aided diagnosis of pulmonary nodules on CT scans: segmentation and classification using 3D active contours *Med. Phys.* **33** 2323–37

Wetzel S G, Cha S, Johnson G, Lee P, Law M, Kasow D L, Pierce S D and Xue X 2002 Relative cerebral blood volume measurements in intracranial mass lesions: interobserver and intraobserver reproducibility study *Radiology* **224** 797–803

Xu R, Kido S, Suga K, Hirano Y, Tachibana R, Muramatsu K, Chagawa K and Tanaka S 2014 Texture analysis on 18F-FDG PET/CT images to differentiate malignant and benign bone and soft-tissue lesions *Ann. Nucl. Med.* **28** 926–35

Yamamoto S, Korn R L, Oklu R, Migdal C, Gotway M B, Weiss G J, Iafrate A J, Kim D-W and Kuo M D 2014 ALK molecular phenotype in non-small cell lung cancer: CT radiogenomic characterization *Radiology* **272** 568–76

Yan J, Chu-Shern J L, Loi H Y, Khor L K, Sinha A K, Quek S T, Tham I W K and Townsend D 2015 Impact of image reconstruction settings on texture features in 18F-FDG PET *J. Nucl. Med.* **56** 1667–73

Yang X and Knopp M V 2011 Quantifying tumor vascular heterogeneity with dynamic contrast-enhanced magnetic resonance imaging: a review *J. Biomed. Biotechnol.* **2011** 12

Yang F, Thomas M, Dehdashti F and Grigsby P 2013 Temporal analysis of intratumoral metabolic heterogeneity characterized by textural features in cervical cancer *Eur. J. Nucl. Med. Mol. Imaging* **40** 716–27

Yip S, McCall K, Aristophanous M, Chen A B, Aerts H J W L and Berbeco R 2014 Comparison of texture features derived from static and respiratory-gated PET images in non-small cell lung cancer *PLoS One* **9** e115510

Yip S F S, Coroller P T, Sanford N N, Huynh E, Mamon H, Aerts J W L H and Berbeco I R 2016 Use of registration-based contour propagation in texture analysis for esophageal cancer pathologic response prediction *Phys. Med. Biol.* **61** 906

Yu H, Caldwell C, Mah K, Poon I, Balogh J, MacKenzie R, Khaouam N and Tirona R 2009 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** 618–25

Zhang H, Tan S, Chen W, Kligerman S, Kim G, D'Souza W D, Suntharalingam M and Lu W 2014 Modeling pathologic response of esophageal cancer to chemoradiation therapy using spatial-temporal 18F-FDG PET features, clinical parameters, and demographics *Int. J. Radiat. Oncol. Biol. Phys.* **88** 195–203

Zhang H-Q, Yu J-M, Meng X, Yue J-B, Feng R and Ma L 2011 Prognostic value of serial [18F] fluorodeoxyglucose PET-CT uptake in stage III patients with non-small cell lung cancer treated by concurrent chemoradiotherapy *Eur. J. Radiol.* **77** 92–6