Corresponding author mail id: mavanzo@cro.it

Machine and Deep Learning Methods for Radiomics

Michele Avanzo[1], Lise Wei[2], Joseph Stancanello[3], Martin Vallières[4,5], Arvind Rao[2,6], Olivier Morin[5], Sarah A. Mattonen[7], Issam El Naqa[2]

[1]Department of Medical Physics, Centro di Riferimento Oncologico di Aviano (CRO) IRCCS, 33081 Aviano, PN, Italy

[2]Department of Radiation Oncology, University of Michigan, Ann Arbor, MI 48103, USA

[3]Guerbet SA, Villepinte, France

[4]Medical Physics Unit, McGill University, Montreal, QC, Canada

[5]Department of Radiation Oncology, University of California, San Francisco, San Francisco, CA 94143, USA

[6]Department of Computational Medicine & Bioinformatics, University of Michigan, Ann Arbor, MI 48103, USA

[7]Department of Radiology, Stanford University, Stanford, California 94305, USA

Abstract

Radiomics is an emerging area in quantitative image analysis that aims to relate large-scale extracted imaging information to clinical and biological endpoints. The development of quantitative imaging methods along with machine learning has enabled the opportunity to move data science research towards translation for more personalized cancer treatments. Accumulating evidence has indeed demonstrated that non-invasive advanced imaging analytics, i.e., radiomics, can reveal key components of tumor phenotype for multiple three-dimensional lesions at multiple time points over and beyond the course of treatment. These developments in the use of CT, PET, US and MR

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/MP.13678

This article is protected by copyright. All rights reserved

Author Manuscript

imaging could augment patient stratification and prognostication buttressing emerging targeted therapeutic approaches. In recent years, deep learning architectures have demonstrated their tremendous potential for image segmentation, reconstruction, recognition, and classification. Many powerful open-source and commercial platforms are currently available to embark in new research areas of radiomics. Quantitative imaging research, however, is complex and key statistical principles should be followed to realize its full potential. The field of radiomics, in particular, require a renewed focus on optimal study design/reporting practices and standardization of image acquisition, feature calculation and rigorous statistical analysis for the field to move forward. In this article, the role of machine and deep learning as a major computational vehicle for advanced model building of radiomics-based signatures or classifiers, and diverse clinical applications, working principles, research opportunities and available computational platforms for radiomics will be reviewed with examples drawn primarily from oncology. We also address issues related to common applications in medical physics, such as standardization, feature extraction, model building, and validation.

Keywords: Quantitative image analysis, radiomics, machine learning, deep learning.

I. Introduction

Radiomics is an emerging area in quantitative image analysis that aims to relate large-scale data mining of images to clinical and biological endpoints[1]. The fundamental idea is that medical images are much richer in information than what the human eye can discern. Quantitative imaging features, called also "radiomic features" can provide richer information about intensity, shape, size or volume, and texture of tumor phenotypes using different imaging modalities (e.g., MRI, CT, PET, ultrasound, etc.)[2]. Tumor biopsy-based assays provide limited tumor characterization as the extracted sample may not always represent the heterogeneity of the whole patient's tumor, while radiomics can comprehensively assess the three-dimensional tumor landscape by means of extracting relevant imaging information[3]. It implies that, applying well-known machine learning methods to radiomic features extracted from medical images, it is possible to macroscopically decode the phenotype of many physio-pathological structures and, in theory, solve the inverse problem of inferring the genotype from the phenotype, providing valuable diagnostic, prognostic or predictive information[4,5].

The term radiomics originated from other –omics sciences (e.g., genomics and proteomics) and conveys the clear intent to invoke personalized medicine based on medical images. It traces its roots to computer-aided detection/diagnosis (CAD) of medical images[6,7]. However, with recent

This article is protected by copyright. All rights reserved

Author Manuscript

advances and the diversity of medical imaging acquisition technologies and processing, radiomics is establishing itself as an indispensable image analysis and understanding tool with applications that transcend diagnosis into prognosis and prediction approaches for personalizing patients' management and their treatment. One of the main differentiators from CAD consists of the link that radiomics has to establish between the current features of a physio-pathological structure at the time of investigation and its temporal evolution in order to personalize the therapeutical approach[8]. The recent availability of large databases of digital medical images and annotated information (e.g., evolution over time or response to treatment with a given prescription, clinical and survival information), the increase of computational power based on advanced hardware (e.g., GPU, cluster or cloud computing) as well as the tremendous mathematical and algorithmic development in areas like machine or deep learning have created favorable conditions to untap the potential of the enormous amount of imaging data wealth that is being generated.

Certainly, the complementarity of other information such as clinical or laboratory data as well as interaction measurements (e.g., *radiogenomics*[9], relating imaging to genomics, or *exposomics*, that is the complementary information from the interaction of the patient with environmental variables) will play a key role to drive future success of radiomics, such as accuracy and reproducibility, to levels that are acceptable for routine clinical practice.

Radiomics has been applied to many diseases including cancer and neurodegenerative diseases to name a few. Although the examples drawn here are from the cancer field, the principles presented here are generally universal across the medical imaging domain. The number of publications issued in the last years has grown almost exponentially. Although there are many review articles already about radiomics, its definition, technical details, and applications in different areas of medicine, the view of radiomics as an image mining tool lends itself naturally to application of machine/deep learning algorithms as computational instruments for advanced model building of radiomics-based signatures[9,10]. This will be the main subject of this article, addressing issues related to common applications in medical physics, standardization, feature extraction, model building, and validation.

II. Overview of Research and Clinical Applications of cancer Radiomics

In this section the applications of radiomics to tumor detection and characterization and prediction of outcome will be reviewed. All the studies described are retrospective and mono-institutional, except where noted.

This article is protected by copyright. All rights reserved

Author Manuscript

A. Radiomics in Diagnosis

a. Cancer detection and auto contouring

The radiomics approach of combining the extraction of radiomic features with machine learning, can be used either to detect/diagnose cancer or to automatically contour the tumor lesion. Methods for radiomics-driven automatic prostate tumor detection typically use a supervised method trained on a set of features calculated from multi-modality images[11]. For detection of prostate cancer, features were computed in a 3×3 pixels sliding window in multimodal MRI of prostate. The voxels were tagged as cancerous or non-cancerous using a support vector machine (SVM) classifier[12]. In Algohary *et al*.[13], the prostate was segmented into areas according to the aggressiveness between malignant and normal regions in the training groups. A voxel-wise random forest model (RF) with a conditional random field spatial regulation was used to classify the voxels in multimodal MRI (T1, Contrast – Enhanced (CE) T1, T2 and FLAIR) of the brain of glioblastoma multiforme (GBM) patients into five classes: non-tumor region and four tumor subregions including necrosis, edema, non-enhancing area, and enhancing area[14]. area[12]. Convolutional neural networks have also been applied to segment organs at risk in head and neck cancer radiotherapy [15] and in lung[16] and liver cancers[17] compared to traditional methods.

b. Prediction of histopathology and tumor stage

Radiomics holds the potential to revolutionize the conventional tumor characterization and replace classic approaches based on macroscopic variables and can be used to distinguish between malignant and benign lesions[3]. Breast cancer lesions, automatically detected using connected component labelling and adaptive fuzzy region growing algorithm, were classified using radiomic features as benign mass or malignant tumor on digital mammography[18], dynamic contrast enhanced (DCE) MRI, and ultrasound[19]. A radiomic model based on mean apparent diffusion coefficient (ADC), had better accuracy than radiologist assessment for characterization of prostate lesions as clinically significant cancer (Gleason grade group ≥ 2) during prospective MRI interpretation[20] . A deep learning multiparametric MRI transfer learning method has also shown the ability to classify prostate cancer high grade/low or grade [21]. Radiomic models based on CT images have been used to predict the histopathology (adenocarcinoma or squamous cell carcinoma)[22,23] and PET tumor stage[24] of lung cancer as well as micropapillary patterns in lung adercarcinomas[25].

c. Microenvironment and intra-tumor partitioning

This article is protected by copyright. All rights reserved

Author Manuscript

A radiomic signature combining features from CE-CT, and 18F-FDG PET was implemented for the presence of high level of hypoxia in head and neck cancer, defined in terms of maximum tumor-to-blood uptake ratio >1.4 in the 18F-FMISO PET[26]. Classification and clustering methods have been developed for tumor separation into subregions (habitat imaging), which contributes to the revelation of tumor heterogeneity, and potential selection of subregions to boost radiation dose[27]. A radiomics analysis focused on a characterization of GBM diversity, using various diversity indices to quantify habitat diversity of the tumor as well as to relate it to underlying molecular alterations and clinical outcomes[28].

d. Tumor genotype

Significant associations between the radiomic features and gene-expression patterns were found in lung cancer patients[3]. A radiogenomic study demonstrated the associations of radiomic phenotypes with breast cancer genomic features as mitochondrial DNA (miRNA) expressions, protein expressions, gene somatic mutations, and transcriptional activities. In particular, tumor size and enhancement texture had associations with transcriptional activities of pathways and miRNA expressions[29]. Radiomic models were implemented for identification of Epithelial Growth Factor Receptor (EGFR) mutant status from CT through multiple logistic regression and pairwise selection [30] and to decode ALK (anaplastic lymphoma kinase), ROS1 (c-ros oncogene 1), or RET (rearranged during transfection) fusions in lung adenocarcinoma[31].

Triple negative breast cancer (TNBC) is likely to be identified by considering heterogeneity of background parenchymal enhancement, characterized by quantitative texture features on DCE-MRI, adds value to such differentiation models as they are strongly associated with the TNBC subtype[32]. Furthermore, TNBC has been proven to be differentiated from fibroadenoma using ultrasound (US) radiomics. A radiomics score obtained by penalized logistic regression with a least absolute shrinkage and selection operator (LASSO) analysis showed significant difference between fibroadenoma and TNBC[33]. The extraction of radiomic features from MR of GBM was able to predict immunohistochemically identified protein expression patterns[34].

Despite large evidence of association among radiomics and genomics, few preclinical studies have demonstrated causal relationship between tumor genotype and radiomic. In one study, HCT116 colorectal carcinoma cells were grown as xenografts in the flanks of NMRI-nu mice. Then overexpression of GADD34 gene was induced by administration of HCT116 doxycycline (dox), or placebo was given. The radiomic analysis demonstrated that that gene overexpression causes

This article is protected by copyright. All rights reserved

Author Manuscript

change in radiomic features, as many features differed significantly between the dox-treated and placebo groups[4].

e. Clinical and macroscopic variables

Radiomic features, derived from T2-w and ADC MRI scan, correlate with clinical variables that are relevant for patient's prognosis. These include prostate specific antigen (PSA) level[35] in patients with prostate cancer, and Human Papilloma Virus (HPV) Status in head and neck squamous cell carcinoma[36,37]**.** Given the well-known behavior of HPV-positive head and neck cancer which is likely to respond at a lower dose of chemoradiation, this opens the way to a CT based patient stratification for a dose de-escalation.

B. Radiomics in therapy

Because radiomic features can describe histology[22] and genetic footprint [29-31] of the tumor, which are correlated with the tumor aggressiveness, they can be used to build models to predict the outcome, in terms of local/distant control or survival, of cancer therapy performed with various therapeutic options (radio-, chemotherapy, targeted molecular therapy, immunotherapy, non – ionizing radiation) or a combination of them.

a. Local control, response, recurrence

Radiomics predicts response to neoadjuvant chemoradiation assessed at time of surgery for Non-Small Cell Lung Cancer (NSCLC) and locally advanced rectal cancer[38]. Local control in patients treated with stereotactic radiotherapy for lung cancer was described using a PET and CT signature developed by using supervised principal component analysis was developed using features from PET and CT[39]. A Radiomic model was developed using first-order statistics, GLCM, and geometrical measurements computed in T2-w and ADC 3T MRI by RF approach for biochemical recurrence of prostate cancer after radiotherapy[35]. A total of 126 radiomic features were extracted using GLCM, GLGCM, Gabor transform, and GLSZM from contrast-enhanced 3T MRI using T1-w, T2-w, and DWI sequences to predict the therapeutic response of nasopharyngeal carcinoma (NPC) to chemoradiotherapy[40]. Deep learning methods with radiomics are also proposed to predict outcomes after liver[41] and lung cancers radiotherapy.

b. Distant metastases

Radiomic models to predict the development of distant metastases (DM) from NSCLC on patients treated with Stereotactic Body Radiotherapy (SBRT) patients for lung cancer were

This article is protected by copyright. All rights reserved

Author Manuscript

developed using features from CT[42] or from PET -CT[39]. Vallières *et al.* used texture-based model for the early evaluation of lung metastasis risk in soft-tissue sarcomas[43] from pre-treatment FDG-PET and MRI scans comprising T1-w and T2-w fat-suppressed sequences (T2FS). A radiomic signature was developed to predict DM after locally advanced adenocarcinoma[44]. Analysis of the peritumoral space can provide valuable information regarding the risk of distant failure, as more invasive tumors may have different morphologic patterns in the tumor periphery. An SVM classifier was trained to predict distant failure from radiomics analysis of the peritumoral space[45].

c. Survival

Aerts *et al.*[3] built a radiomic signature consisting of a combination of four features in a retrospective lung cancer cohort, which was predictive for survival in head and neck and NSCLC independent cohorts. One textural feature calculated from GLCM, SumMean[46], was identified using the LASSO procedure as an independent predictor of overall survival that complements metabolic tumor volume (MTV) in decision tree[47]. A radiomic signature was built from PET-CT for survival after SBRT for lung cancer [39]. Deep learning was also proposed to stratify NSCLC patients according to mortality risk using standard of care CT[48].

d. Molecular targeted therapy

Many tumors commonly overexpress oncogenes such as the EGFR and respond to molecular targeted therapies such as EGFR tyrosine kinase inhibitor. From the change in features between the CT acquisitions before and three weeks after therapy it was possible to identify NSCLC patients responding to treatment with gefitinib[49]. A radiomic prediction model was designed to stratify patients according to progression-free and overall survival after therapy with antiangiogenic for GBM [50].

e. Immunotherapy

Cancer immunotherapy by immune checkpoint blockade is a promising treatment modality that is currently under strong development, and there is a great need for models to select patients responding to immunotherapy. In a retrospective multicohort study, an eight-feature radiomic signature predictive of the presence of CD8 T cells, which is related to the tumor-immune phenotype, was developed from CE-CT images, using elastic-net regularized regression method[51]. The signature was successfully validated on external cohorts for discrimination of immune phenotype, and for the prediction of survival and response to anti-PD-1 or PD-L1 immunotherapy.

This article is protected by copyright. All rights reserved

Author Manuscript

f. Delta-radiomics

The longitudinal study of features and of their change during the treatment, with the goal of predicting response to therapy, is called delta-radiomics. Features calculated from pretreatment and weekly intra-treatment CT change significantly during radiation therapy (RT) for NSCLC[52]. Delta-radiomics could possibly be performed by the Cone Beam CT (CBCT) devices for image guidance of radiotherapy treatment, thus allowing large-scale study of tumor response to total dose, fractionation and fraction dose. It has been shown that reproducible features can be extracted from CBCT[53] predictive for overall survival in NSCLC patients as much as features from CT[54]. Nevertheless, the studies on CBCT delta-radiomics are still limited to assessment of feasibility and reproducibility[55].

g. Prediction of side effects

Radiomics-based models can help early identify the development of side effects such as radiation induced lung injury (RILI). The change from pre- to post-treatment (at 3, 6, and 9 months) CT features significantly correlates with lung-injury as scored by oncologist post-SBRT for lung cancer and was found to be correlated with dose and fractionation[56].

A logistic regression–based classifier was constructed to combine information from multiple features to identify patients that will develop grade ≥2 radiation pneumonitis among those who received RT for esophageal cancer[57]. The addition of normal lung image features produced superior model performance with respect to traditional dosimetric and clinical predictors of radiation pneumonitis (RP), suggesting that pre-treatment CT radiomic features should be considered in the context of RP prediction. CT radiomic features were extracted from the total lung volume defined using the treatment planning scan for RP [58].

h. Differentiation of recurrence from benign changes

The differentiation of tumor recurrence from benign radiation-induced changes in follow-up images can be a major challenge for the clinician. A radiomic signature consisting of 5 image-appearance features from CT demonstrated high discriminative capability to differentiate recurrence of lung tumor from consolidation and opacities in SBRT patients[59]. Similarly, a combination of five radiomic features from CE-T1w and T2w MR were found to be capable of distinguishing necrosis from progression in follow up MR images in patients treated with Gamma Knife radiosurgery for brain metastases[60].

This article is protected by copyright. All rights reserved

Author Manuscript

i. Non ionizing radiation and other therapies

Radiomic features in MRI respond differently when Laser interstitial thermal therapy (LITT), a highly promising focal strategy for low-grade, organ-confined prostate cancer, is performed on cancer or healthy prostate tissue. A radiomic signature then could allow to assess if prostate cancer is successfully ablated[61]. A radiomic model was predictive of complete response after transcatheter arterial chemoembolisation combined with high-intensity focused ultrasound treatment in hepatocellular carcinoma[62].

III. Radiomics Analysis with Machine and Deep Learning Methods

A. Preprocessing

Prior to radiomics analysis, preprocessing steps need to be applied to the images, which aim at reducing image noise, enhancing image quality, enabling the reproducible and comparable radiomic analysis. For some imaging modalities, such as PET, the images should be converted to a more meaningful representation (standardized uptake value, SUV). Image smoothing can be achieved by average or Gaussian filters[63]. Voxel size resampling is important for datasets that have variable voxel size[64]. Specifically, isotropic voxel size is required for some texture feature extraction. There are two main categories of interpolation algorithms: Polynomial and spline interpolation. Nearest neighbor is a zero-order polynomial method that assigns grey-level values of the nearest neighbor to the interpolated point. Bilinear or trilinear interpolation and bicubic or tricubic interpolation are often used for 2D in-plane interpolation or 3D cases. Cubic spline and convolution interpolation are third order polynomial method that interpolates smoother surface than linear method, while being slower in implementation. Linear interpolation is a rather commonly used algorithm, since it neither leads to the rough blocking artifacts images that are generated by nearest neighbors, nor will it cause out-of-range grey levels that might be produced by higher order interpolation[65].

In the context of feature-based radiomics analysis, as discussed below, the computation of textures would require discretization of the grey levels (intensity values). There are two ways to do the discretization: fixed bin number N and fixed bin width B. For fixed bin number, we first decide a fixed number of N bins, and the grey levels will be discretized into these bins using the formula below:

This article is protected by copyright. All rights reserved

Author Manuscript

$$X_{d,k} = \begin{cases} \left\lfloor N_g \frac{X_{gl,k} - X_{gl,min}}{X_{gl,max} - X_{gl,min}} \right\rfloor + 1 & X_{gl,k} < X_{gl,max} \\ N_g & X_{gl,k} = X_{gl,max} \end{cases}, \quad (1)$$

where $X_{gl,k}$ is the intensity of kth voxel.

For fixed bin width, starting at a minimum $X_{gl,min}$, a new bin will be assigned for every intensity interval of w_b. Discretized grey levels are calculated as follow:

$$X_{d,k} = \left\lfloor \frac{X_{gl,k} - X_{gl,min}}{w_b} \right\rfloor + 1. \quad (2)$$

The fixed bin number method is better when the modality used is not well calibrated. It maintains the contrast and makes the images of different patients comparable, but loses the relationship between image intensity, while fixed bin size method keeps the direct relationship with the original scale. Some investigations about the effect of both methods have shown that fixed bin size method offered better repeatability and thus may be suitable for intra- and inter- patient studies, however, this remains a subject of ongoing research[66,67]. In CT-radiomics the image pixel intensity maps to the Hounsfield Units (HU) and thus is much more directly comparable and interpretable. MRI-related modalities are more challenging since the pixel intensities are not directly interpretable, rather need to normalized relative to some standard reference (e.g., contralateral brain, or normal appearing white matter in neuroimaging, psoas muscle in abdominal imaging, etc.).

B. Machine and Deep Learning Algorithms for Radiomics

Machine and deep learning algorithms provide powerful modeling tools to mine the huge amount of image data available, reveal underlying complex biological mechanisms, and make personalized precision cancer diagnosis and treatment planning possible. Hereafter, two main types - feature-engineered (conventional radiomics) and non-engineered (deep learning-based) radiomics modeling methods – will be briefly introduced. Generally speaking, machine learning methods can also be divided into supervised, unsupervised and semi-supervised for both feature-based and featureless methods. Each of these categories will be briefly discussed in the following sections. A workflow diagram illustrating the radiomics analysis process after image acquisition is shown in Fig. 1.

a. Feature-engineered radiomics methods

Traditionally, the radiomic features being extracted are hand-crafted features that capture characteristic patterns in the imaging data, including shape-based, first-, second-, and higher order statistical determinants and model-based (e.g. fractal) features. Feature-based methods require a

This article is protected by copyright. All rights reserved

Author Manuscript

segmentation of the region of interest (ROI), either through a manual, semi-automated, or automatic methods. Shape-based features are external representations of a region, that characterize the shape, size and surface information of the ROIs[68]. Typical metrics include sphericity, and compactness[3,43,69,70]. First-order features (e.g. mean, median) describe the overall intensity and variation of the ROIs, while ignoring spatial relations[8,24]. Second-order (texture) features in contrast can provide inter-relationships among voxels. Textural features can be extracted from different matrices, e.g. grey-level co-occurrence matrix (GLCM), grey-level run-length matrix (GLRLM), etc[35,46,71]. Semantic features are another type of feature that can be extracted from medical images. These features describe qualitative features of the image typically used in the radiology workflow.

Hundreds or even thousands of radiomic features are not uncommon when we deal with outcome modeling. Feature selection and/or extraction thus is a crucial step that aims at obtaining the optimal feature subset or feature representation that correlates most with the endpoint and meanwhile correlates least between each other. After the feature subset is obtained, various machine learning algorithms can be applied based on them. 14 feature selection and 12 classification methods were evaluated in terms of their predictive performance on two independent lung cancer cohorts[72]. Sometimes, the feature selection and model construction can be implemented together, called the embedded method, such as least absolute shrinkage and selection operator (LASSO)[73]. In contrast, wrapper methods select the features based on the models' performance for different subsets of features, for which we need to rebuild the model again after features are selected, for instance, recursive feature elimination support vector machines (SVM-RFE). Filter method also separates the feature selection and model construction processes, whose uniqueness of it is its independence of the classifier being used for the subsequent model building, such as Pearson correlation-based feature ranking. In any feature selection method, it is essential to ensure that there is no "double dipping" into the training data for both feature selection, hyperparameter optimization and model selection. Rather the methods of "nested cross validation" should be used in order to prevent overfitting or incorrect estimates of generalization. According to whether or not the labels (ground truths) are used, feature selection and extraction can be divided into supervised, unsupervised and semi-supervised ways. The three feature selection methods discussed above are mostly supervised. Examples of unsupervised methods are principle component analysis (PCA)[74], clustering and t-Distributed Stochastic Neighbor Embedding (t-SNE)[75]. PCA uses an orthogonal linear transformation to convert the data into a new coordinate system so that large variances are projected to orthogonal coordinates. Clustering is another feature extraction algorithm which aims at finding relevant features and combining them by their cluster centroids based on some similarity measure, such as K-means and hierarchical clustering [76]. Unsupervised consensus clustering

This article is protected by copyright. All rights reserved

Author Manuscript

identified robust imaging subtypes using dynamic CE-MRI data for patients with breast cancer[77]. tSNE is a dimension reduction method capable of retaining the local structure (pairwise similarity) of data, while revealing some important global structure.

In the medical field, two types of questions are mainly investigated, binary problems (classification), such as whether or not a disease has recurred, the patient is alive beyond certain time threshold, etc; and survival analysis, that is able to show if a risk factor or treatment affects time to event. For the classification problem logistic regression fits the coefficients of the variables to predict a logit transformation of the probability of the presence of the event. SVM, frequently used in Computed Aided Diagnosis (CAD)[6] and radiomics[32,59,76,78], learns an optimal hyperplane that separates the classes as wide as possible, while trying to balance with misclassified cases. SVM can also perform non-linear classification using the "kernel trick" -- different basis functions (e.g. radial basis function), mapping to higher dimensional feature space. The hyperplane maximizes the margin between the two classes in a non-linear feature space. SVM also tolerates some points on the wrong side of the boundary, thus improving model robustness and generalization[79]. RF is based on decision trees, a popular concept in machine learning especially in the field of medicine, because their representation of hypotheses as sequential "if-then" resembles human reasoning[80]. RF applies bootstrap aggregating to decision trees and improve the performance by lowering the high variance of the trees[81]. Risk assessment models (classification and survival) were constructed via RFs and imbalance adjustment strategies for locoregional recurrences and distant metastases in head and neck cancer[82]

Neural networks, though usually used in the featureless context, can also be used in conventional feature selection and modeling[22,38,78]. These algorithms are mainly for supervised learning, while in particular in the medical field, there are a lot of data without labeling, in these cases, semi-supervised learning can be applied to make use of the unlabeled data combined with the small amount of labeled data. The self-training is bootstrapped with additional labelled data obtained from its predictions[83]. The transductive SVM (TSVM) tries to keep the unlabeled data as far away from the margin as possible[84]. Graph-based methods construct a graph connecting similar observations and enable the class information being transported through the graph[85].

For the survival analysis, Cox regression[86], random survival forests[87] and support vector survival[88] methods are also available to investigate the presence of a set of variables that may affect survival time. Due to the length limit, we will not go into the details. Interested readers can refer to the references to read more about these algorithms.

This article is protected by copyright. All rights reserved

Author Manuscript

b. Non feature-engineered radiomics methods

Though hand-crafted features introduced above provide prior knowledge, they also suffer from the tedious designing process and may not faithfully capture the underlying imaging information. Alternatively, with the development of deep learning technologies based on multi-layer neural networks, especially the convolutional neural networks (CNN), the extraction of machine learnt features is becoming widely applicable recently. In deep learning, the processes of data representation and prediction (e.g, classification or regression) are performed jointly[89]. In such a case, multi-stack neural layers of varying modules (e.g., convolution or pooling) with linear/non-linear activation functions perform the task of learning the representations of data with multiple levels of abstraction and subsequent fully connected layers are tasked with classification, for instance. A typical scenario to get such features is to use the data representation CNN layers as feature extractor. Each hidden layer module within the network transforms the representation at one level. For example, the first level may represent edges in an image oriented in a particular direction, the second may detect motifs in the observed edges, the third could recognize objects from ensembles of motifs[89]. Patch-/pixel-based machine learning (PML) methods use pixel/voxel values in images directly instead of features calculated from segmented objects as in other approaches [90,89]. Thus PML removes the need for segmentation, one of the major sources of variability of radiomic features. Moreover, the data representation removes the feature selection portion eliminating associated statistical bias in the process. For the CNN network, either self-designed (from scratch) or existing structures, e.g. VGG[91], Resnet[92], can be used. Depending on the data size, we can choose to fix the parameters or fine tune the network using our data, also called transfer learning. Instead of using deep networks as feature extractors, we can use them directly for the whole modeling process. Similarly to the conventional machine learning methods, there are also supervised, unsupervised and semi-supervised methods. CNN are similar to regular neural networks, but the architecture is modified to fit to the specific input of large-scale images. Inspired by the Hubel and Wiesel's work on the animal visual cortex[93], local filters are used to slide over the input space in CNNs, which not only exploit the strong local correlation in natural images, but also reduce the number of weights significantly by sharing weights for each filter. Recurrent neural networks (RNN) can use their internal memory to process sequence inputs and take the previous output as inputs. There are two popular types of RNN – Long short-term memory (LSTM)[94] and Gated recurrent units (GRU)[95]. They were invented to solve the problem of vanishing gradient for long sequences by internal gates that are able to learn which data in the sequence is important to keep or discard. Deep autoencoders (AE), which are unsupervised learning algorithms, have been applied to medical imaging for latent representative feature extraction. There are variations to the AEs, such as variational autoencoders

This article is protected by copyright. All rights reserved

Author Manuscript

that resemble the original AE and variational Bayesian methods to learn a probability distribution that represents the data[96], convolutional autoencoders that preserve spatial locality[97], etc. Another unsupervised method is the restricted Boltzmann machine (RBM), which is consists of visible and hidden layers[98]. The forward pass learns the probability of activations given the inputs, while the backward pass tries to estimate the probability of inputs given activations. Thus, the RBMs lead to the joint probability distribution of inputs and activations. Deep belief networks can be regarded as a stack of RBMs, where each RBM communicates with previous and subsequent layers. RBMs are quite similar with AEs, however, instead of using deterministic units, like RELU, RBMs use stochastic units with certain distribution. As mentioned above, labeled data is limited, especially in the medical field. Neural network based semi-supervised approaches combine unsupervised and supervised learning by training the supervised network with an additional loss component from the unsupervised generative models (e.g. AEs, RBMs)[99].

Machine learning methods are highly effective with large number of samples; however, they suffer from overfitting pitfalls with limited training samples. For deep learning, data augmentation (e.g. by affine transformation of the images) during training is commonly implemented. Transfer learning is another way to reduce the difficulty in training. Using deep models trained on other dataset (natural images) and then fine-tune on the target dataset. The structures of the networks can also be modified to reduce overfitting, such as, by adding dropout and batch normalization layers. Dropout randomly deactivates a fraction of the units during training and can be viewed as a regularization technique that adds noise to the hidden units[100]. Batch normalization reduces the internal covariate shift by normalizing for each training mini-batch[101].

Comparing with feature-based methods, deep learning methods are more flexible and can be used with some modifications in various tasks. In addition to classification, segmentation, registration, and lesion detection are widely explored by deep learning techniques. Fully CNN (FCN), trained end-to-end, merge features learnt from different stages in the encoders and then upsampling low resolution feature maps by deconvolutions[102]. Unet, built upon FCN, with the pooling layers being replaced by upsampling layers, resulted in a nearly symmetric U-shaped network[103]. Skipping structures combines the context information with the unsampled feature maps to achieve higher resolution. CNN, trained end-to-end from clinical images were directly used for binary classification of skin cancer and achieved performance on par with experts[104]. Chang *et al* proposed a multi-scale convolutional sparse coding method that provides an unsupervised solution for learning transferable base knowledge and fine-tuning it towards target tasks[105].

This article is protected by copyright. All rights reserved

Author Manuscript

Fig. 1. Workflow for radiomics analysis with feature-based (conventional machine learning) and featureless (deep learning) approaches.

C. Validation and Benchmarking of Radiomics Models

Once models are developed using the selected predictors, quantifying the predictive ability of the models (validation) is necessary. Based on the TRIPOD criteria[106], there are 4 types of validation: 1a. Developing and validating on the same data, which gives apparent performance. This evaluation is usually optimistic estimation of the true performance. 1b. Developing the models using all the data, then using resampling techniques to evaluate the performance. 2a. Randomly split the data into 2 groups for development and validation separately. 2b. Split the data non-randomly (e.g. by location or time), which is stronger than 2a. 3 & 4. Develop the model using one data set and validate on separate data. It is ideal if there is a separate data set for external validation, however, in the frequent case that only a single data set is available, internal validation (1b) is required. Two popular resampling methods are bootstrapping and cross-validation. Feature selection, which is required before machine learning, should precede cross-validation, or it will lead to a selection bias due to the leak of information by the pre-filtering of the features[107].

Radiomic classifiers output a score that indicates the likelihood of one event to happen, and a threshold, to generate positive or negative predictions according to the task at hand. For example, fewer false positives would be required if we are implementing a conservative experiment, thus larger threshold will be preferred. Classifiers are evaluated using either a numeric metric (e.g. accuracy), or the so-called confusion matrix, or a graphical representation of performance, such as a receiver operating characteristic curve (ROC), a two-dimensional graph with true positive rate being the Y axis, and false positive rate the X axis. It has the advantage that they show classifier performance without regard to threshold and class distribution, thus widely used in model evaluation. The area under an ROC curve (AUC) is more convenient when comparing, and is equivalent to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance[108]. For survival analysis, Harrell's C index[109] is commonly used to measure discrimination ability of the model, which is motivated by Kendall's tau correlation. Harrell defines the overall C index as the proportion of all usable pairs in which the predicted risk probabilities and outcomes are concordant (Usable pairs are two cases that at least one of them is event)[110].

Kaplan-Meier (KM) curves are used to estimate the survival function from lifetime data, and also used to compare different risk groups. The risk groups can be patients that are treated with certain plan and the control group, or they can be the outputs from a survival model (e.g. Cox model)

This article is protected by copyright. All rights reserved

Author Manuscript

that divides the patients into high and low risk groups. It is highly recommended to visualize confidence intervals of the curves. The log rank test gives a quantitative evaluation of the statistical significance of the difference for different curves, which is also widely provided for KM curves[111].

IV. Implementation in medical physics practice

A. Software tools for radiomics

In most published research studies in radiomics, in-house developed methods are used. However, some research groups developed image analysis/radiomic software tools, both commercial or open source, available to the scientific community. The main goals of these tools are: 1) to speed up the development of competences based on more recent skills on radiomics; 2) to allow reproducibility and comparability of results from different research groups, and 3) to standardize both feature definitions and computation methods to guarantee the reliability of radiomic results [112,113].

Table 1 shows a list of the software, web platforms, and toolkits available free of charge for the extraction of radiomics features, along with some of their main functionalities and relevant information. Given the high pace of radiomic developments, the list is not exhaustive and does not intend to cover all possible solutions. Furthermore, considering recent and increased interest in the radiomic field, many other dedicated tools are under development. All the open source solutions shown in this overview have been implemented by research teams (MaZda[114], LifeX [115], ePAD[116], HeterogeneityCAD[3], PyRadiomics/Radiomics [117], QuantImage [118], the Texture Analysis Toolbox[43], QIFE[119], IBEX [120], and MedomicsLab) and are capable of analyzing CT, MRI, and PET, some of them can process also other medical images, such as mammography, radiography, or ultrasound.

Four software programs (MaZda, LifeX, ePAD, IBEX) offer the possibility of manually or automatically segmenting medical images. Three toolkits (HeterogeneityCAD, PyRadiomics/Radiomics, QIFE) are designed exclusively for the extraction of features. They can be embedded in more complete solutions (e.g. 3D Slicer [121]). Morphological, first, second and third order statistical features can be extracted by all software solutions, except for ePAD. Four of them (TexRAD, MaZda, PyRadiomics/Radiomics, IBEX) offer also the possibility of extracting features from filtered images. Of note, MEDomicsLab is an open-source software currently being developed by a consortium of research institutions, which will be available in the second half of 2019.

B. Commercial Programs for radiomics

This article is protected by copyright. All rights reserved

Author Manuscript

Commercial software programs are also becoming increasingly available due to the interest of many medical device incumbents as well as newcomers such as commercial spin-off of research groups or de novo start-up companies. Such software programs can be divided into:

a. Research platforms

These platforms enable the discovery of new signatures by linking quantitative imaging biomarkers, clinical and –omics data to clinical endpoints. They are usually considered non-medical devices in that they do not affect the clinical routine, run usually on independent workstations, and are not used to drive clinical decisions. Their main differentiator from open access software consists of workflow optimization and efficiency improvements, enabling an automatic, end-to-end seamless processing pipeline. TexRad®, QIDS®, RadiomiX, iBiopsy® and EVIDENS offer research capabilities at a different level, ranging from simple features extraction to image filter application and machine learning modules. In the research mode, these software programs are usually open to process any 3D image, DICOM or not, up to 2D digital pathology images (histomics or pathomics).

b. Clinically validated software programs,

In order to use decision support systems (DSSs), based on an already discovered signatures and thoroughly validated on large independent datasets, also known as clinical grade DSS, in clinical practice, a regulatory clearance is usually needed, as they fall within the definition of medical devices in many regulatory systems, e.g., class I or II medical device as a function of their intended use (e.g. mere support to decision versus a computer aided diagnosis/prognosis). DSSs are usually limited to a specific modality, mostly CT, and to a specific disease in a specific body district: these constraints come primarily from the intended use definition to which these DSSs are subjected to be compliantly marketed.

Research tools or clinical grade DSSs can be embedded into more comprehensive platforms such as Picture Archive and Communication Systems (PACS), Hospital Information Systems (HIS), Oncology Information Systems (OIS) or Treatment Planning Systems (TPS), or being stand-alone. Usually, large medical device incumbents tend to embed DSSs into their research or clinical solutions, while newcomers often offer their solution as a standalone system.

It is not unusual that large medical device players embed open access or commercial software programs to provide their customers with the possibility of exploring or exploiting radiomic potential: examples are IntelliSPace Discovery (Philips, the Netherlands) which interfaces to Pyradiomics, Advantage Workstation (GE, Buc, France) which interfaces through a plugin to

This article is protected by copyright. All rights reserved

Author Manuscript

QuantibTM Brain or Syngo.via Frontier (Siemens, Erlangen, Germany) which interfaces to RadiomiX. It is also beneficial to mention the platform (www.envoyai.com) which offers the possibility of sharing applications and, once solutions reached the product maturity, to commercialize them.

V. Current challenges and recommendations

A. Interpretability issues

It is recognized that machine learning algorithms tend to generally trade interpretability for better prediction. Hence, clinicians are still reluctant to embrace these methods as part of their clinical practice, because they have long been perceived them as "black boxes", meaning that it is difficult to determine how they arrive at their predictions. For example, it is difficult to understand deep neural networks due to the large number of interacting, non-linear parts [122],[123]. In order to improve interpretability of radiomics for the clinician, methods based on graph approaches can be utilized[124], and in the context of deep learning better visualization tools are being developed such as maps highlighting regions of the tumor that impact the prediction of the deep learning classifier are also being proposed [123].

B. Repeatability and Reproducibility issues

In radiomics, repeatability is measured by extraction of features from repeated acquisition of images under identical or near-identical conditions and acquisition parameters[125], whereas reproducibility or robustness, is assessed when measuring system or parameters differ. These can be assessed by use of digital or physical phantoms. Physical phantoms usually contain inserts of different with different density, shape or texture properties in order to produce a wide range of radiomics feature values. These phantoms allow to assess the reproducibility or robustness of the entire workflow, from image acquisition to extraction of radiomic features. Their major drawback is that they do not reflect the variability of human anatomy in the clinical scenario.

A phantom for radiomics was created for use with CT [113] or CBCT[126] called Credence Cartridge Radiomics (CCR) Phantom. This consisted of 10 cartridges with different density and texture properties in order to produce a wide range of radiomics feature values: wood, rubber, cork, acrylic, and plaster. Phantoms for PET with heterogeneous lesions have been also proposed, e.g. with different 3D printed inserts reflecting different heterogeneities in FDG uptake[127].

This article is protected by copyright. All rights reserved

Author Manuscript

Digital phantoms are usually scans of patients acquired under controlled conditions. They are therefore realistic, but cannot be used for studying radiomic features' sensitivity to the image acquisition and its parameters. A dataset consisting of 31 sets of repeated CT scans acquired approximately 15 minutes apart is now publicly accessible through The Reference Image Database to Evaluate Therapy Response (RIDER). This dataset allows "test-retest" analysis, a comparison of the results from images acquired within a short time on the same patient[128].

C. Factors affecting stability

For CT, inter-scanner variability of image features produces differences in extracted features that are comparable to the variability in patient images acquired by the same scanner[113]. The choice of methods of reconstruction, such as filtered back projection or iterative algorithm, also affect radiomic feature[129]. Smoothing of the image and reducing the slice thicknesses can improve reproducibility of CT-extracted features[128,130]. In PET imaging, textural features are sensitive to different acquisition modes [131,132], reconstruction algorithms, and their user-defined parameters such as the number of iterations, the post-filtering level, input data noise, matrix size, and discretization bin size[133,134].

Radiomic features extracted from MRI scans depend on the field of view, field strength, reconstruction algorithm and slice thickness. Results of the DCE MRI depend on the contrast agent dose, method of administration, and the pulse sequence used. The radiomic features extracted from DW-MRI depend on acquisition parameters and conditions as k-space trajectory, gradient strengths and b-values. The repeatability of MR-based radiomic features has been investigated[135] using a ground truth digital phantom of brain glioma patients and an MRI simulator capable of generating images according to different acquisition (field strength, pulse sequence, arrangement of field coils) and reconstruction methods. It was found that some features are subject to small changes, compared with clinical effect size.

In presence of significant respiratory tumor motion as in the case of lung cancer, conventional PET images are influenced by motion as, because of their relatively long acquisition times, the counts measured are averaged over multiple breathing cycles. Respiratory-gated PET accounts for respiratory motion and textural features from gated PET have been found robust[136].

Segmentation affects the radiomics workflow, regardless of the imaging technique, because many extracted features depend on the segmented region[2,5]. Semiautomatic segmentation

This article is protected by copyright. All rights reserved

Author Manuscript

algorithms may improve the stability of radiomic features[137], and recently available fully automatic segmentation tools may be as accurate as manual segmentation by medical experts[138].

The studies on the comparisons of the performance of many classifier and feature selection methods indicate that the choice of classification method is the most dominant source of models' predictive performance variability [72]. Fourteen feature selection algorithms were compared on a set of 464 lung cancer patients considering 440 radiomic variables [76]. The feature selection method based on the Wilcoxon signed-rank (WLCX) test had the highest prognostic performance with high stability against data perturbation. Interestingly, WLCX is a simple univariate method based on ranks, which does not take into account the redundancy of selected features during feature ranking. In a comparison of performance of 24 feature selection methods for radiomic signature building for lung cancer histology it was shown that RELIEF with its variants were the best performing methods[22].

D. Quality, Radiomics quality score

The workflow for radiomic studies involves several steps, from data acquisition, selection, and curation, to feature extraction, feature selection, and modelling. There is an important need that radiomics studies are properly designed and reported to ensure the field can continue to develop and produce clinically useful tools and techniques. A number of issues can arise providing misleading information, including imaging artifacts, poor study design, overfitting of data, and incomplete reporting of results[8,139]. Although imaging artifacts are inevitable in medical imaging, consistent imaging parameters may help reduce variability in radiomic features[126]. To minimize the potential of overfitting of radiomic models, 10 patients are needed for each feature in the final model[140]. Ideally, an independent external validation dataset is also used to confirm the prognostic ability of any radiomic model. The radiomics quality score (RQS) has recently been developed to assess all areas of a radiomic study and determine whether it is compliant with best practice procedures[139], emulated from the TRIPOD initiative previously described.

E. Standardization and harmonization

Although research in the field of radiomics has drastically increased over the past several years, there still remains a lack of reproducibility and validation of current radiomic models. There are currently no guidelines and standard definitions for radiomic features and for constructing these features into clinical models. Current initiatives are underway to improve standardization and harmonization in radiomic studies.

This article is protected by copyright. All rights reserved

Author Manuscript

As a part of radiomic signature validation, there are efforts to explore distributed feature sharing and model development across contributing institutions[141]. A key component in this exercise is the assessment and redressal of batch effects[142] and confounding variables across contributing sites, so as to ameliorate systematic yet unmeasured sources of variation. Another key component is the use of methods to harmonize data as well as model parameters across study sites, with the intent of meaningful comparisons across clinical population[143]. Such efforts are necessary to enable the widespread and generalizable development of models that are transportable across institutions. In addition to the careful calibration and stability analysis of radiomic features within predictive models, there is also a need for ensuring model robustness through approaches like noise injection[144]. Adversarial training approaches from neural networks can have value in the modern deep learning modeling area by incorporating not only positive examples but negative ones too[145]. The workflow for computing features is complex and involves many steps, often leading to incomplete reporting of methodological information (e.g., texture matrix design choices and gray-level discretization methods). As a consequence, few radiomics studies in the current literature can be reproduced from start to end.

To accelerate the translation of radiomics methods to the clinical environment, the Image Biomarker Standardization Initiative (IBSI)[65] has the goal to provide standard definitions and nomenclature for radiomic features, reporting guidelines, and to provide benchmark datasets and values to verify image processing and radiomic feature calculations. Figure 2 presents the standardized radiomics workflow defined by the IBSI. The IBSI aims at standardizing both the computation of features and the image processing steps required before feature extraction. For this purpose, a simple digital phantom was designed and used in Phase 1 of the IBSI to standardize the computation of 172 features from 11 categories. In Phase 2 of the IBSI, a set of CT images from a lung cancer patient was used to standardize radiomics image processing steps using 5 different combinations of parameters including volumetric approaches (2D vs 3D), image interpolation, re-segmentation and discretization methods. The initiative is now reaching completion and a consensus on image processing and computation of features has been reached over time.

Overall, the use of standardized computation methods would greatly enhance the reproducibility of radiomics studies, and it may also lead to standardized software solutions available to the community. It would also be desirable that the code of existing software be updated to conform with standards established by the IBSI. Furthermore, it is essential to include in radiomics studies the comprehensive description of feature computation details as defined by the IBSI[65] and Vallières *et al*[146], as shown in Table 2. Ultimately, we envision the use of dedicated

This article is protected by copyright. All rights reserved

Author Manuscript

ontologies to improve the interoperability of radiomics analyses via consistent tagging of features, image processing parameters and filters. The Radiomics Ontology (www.bioportal.bioontology.org/ontologies/RO) could provide a standardized means of reporting radiomics data and methods, and would more concisely summarize the implementation details of a given radiomics workflow.

Finally, some guiding principles already exist to help radiomics scientists further implement the responsible research paradigm into their current practice. A concise set of principles for better scientific data management and stewardship, the "FAIR guiding principles"[147], stating that all research objects should be findable, accessible, interoperable, and reusable. Implementation of the FAIR principles within the radiomics field could facilitate its faster clinical translation. First, all methodological details and clinical information must be clearly reported or described to facilitate reproducibility and comparison with other studies and meta-analyses. Second, models must be tested in sufficiently large patient datasets distinct from teaching (training and validation) sets to statistically demonstrate their efficacy over conventional models (e.g., existing biomarkers, tumor volume, cancer stage, etc.). To allow for optimal reproducibility potential and further independent testing, all data, final models and programming code related to a given study needs to be made available to the community. Table 3 provides guidelines that can help to evaluate the quality of radiomics studies[146]. More guidelines on reproducible prognostic modeling can also be found in the TRIPOD statement[106].

VI. Conclusions

The field of radiomics is constantly growing within the field of medical physics and is an exciting opportunity for the medical physics community to participate in novel research for the safe translation of quantitative imaging. Machine and deep learning-based models have the potential to provide clinicians with DSS to improve diagnosis, treatment selection, and response assessment in oncology. As the field expands, the need to associate radiomic features with other clinical and biological variables will become of increased importance. The field should also continue to strive for standardized data collection, evaluation criteria, and reporting guidelines in order to mature as a field. Data-sharing will be crucial to develop the large-scale datasets needed for proper validation of radiomic models and there will be a need for collaborations to validate models across multiple institutions. In order to move radiomic models into the clinical practice it is imperative to demonstrate improvements to the clinical workflow and decision making, through expert observer studies and eventually clinical trials. Future developments in the areas of machine and deep learning

This article is protected by copyright. All rights reserved

Author Manuscript

with their improved balance of interpretability and prediction will also continue to advance radiomic studies.

FIGURE AND TABLE CAPTIONS

Fig. 1. Workflow for radiomics analysis with feature-based (conventional machine learning) and featureless (deep learning) approaches.

Fig. 2. Radiomics computation workflow as defined by the IBSI.

Table 1. Open access software programs for radiomics analysis.

Table 2. Reporting guidelines on the computation of radiomics features (adapted from Refs.[65] and [146]).

Table 3. Quality factors in radiomics studies (adapted from Refs [139] and [146])

REFERENCES

1 M. Avanzo, J. Stancanello, I. El Naqa, "Beyond imaging: The promise of radiomics," Phys. Med. **38**, 122-139 (2017).

2 P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R.G. van Stiphout, P. Granton, C.M. Zegers, R. Gillies, R. Boellard, A. Dekker, H.J. Aerts, "Radiomics: extracting more information from medical images using advanced feature analysis," Eur. J. Cancer **48**, 441-446 (2012).

3 H.J. Aerts, E.R. Velazquez, R.T. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, F. Hoebers, M.M. Rietbergen, C.R. Leemans, A. Dekker, J. Quackenbush, R.J. Gillies, P. Lambin, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," Nat. Commun. **5**, 4006 (2014).

4 K.M. Panth, R.T. Leijenaar, S. Carvalho, N.G. Lieuwes, A. Yaromina, L. Dubois, P. Lambin, "Is there a causal relationship between genetic changes and radiomics-based image features? An in

This article is protected by copyright. All rights reserved

Author Manuscript

vivo preclinical experiment with doxycycline inducible GADD34 tumor cells," Radiother. Oncol. **116**, 462-466 (2015).

5 V. Kumar, Y. Gu, S. Basu, A. Berglund, S.A. Eschrich, M.B. Schabath, K. Forster, H.J. Aerts, A. Dekker, D. Fenstermacher, D.B. Goldgof, L.O. Hall, P. Lambin, Y. Balagurunathan, R.A. Gatenby, R.J. Gillies, "Radiomics: the process and the challenges," Magn. Reson. Imaging **30**, 1234-1248 (2012).

6 J. Tang, R.M. Rangayyan, J. Xu, I. El Naqa, Y. Yang, "Computer-aided detection and diagnosis of breast cancer with mammography: recent advances," IEEE Trans. Inf. Technol. Biomed. **13**, 236-251 (2009).

7 M.L. Giger, N. Karssemeijer, J.A. Schnabel, "Breast image analysis for risk assessment, detection, diagnosis, and treatment of cancer," Annu. Rev. Biomed. Eng. **15**, 327-357 (2013).

8 R.J. Gillies, P.E. Kinahan, H. Hricak, "Radiomics: Images Are More than Pictures, They Are Data," Radiology **278**, 563-577 (2016).

9 J. Wu, K.K. Tha, L. Xing, R. Li, "Radiomics and radiogenomics for precision radiotherapy," J. Radiat. Res. **59**, i25-i31 (2018).

10 S. Napel, W. Mu, B.V. Jardim-Perassi, H.J.W.L. Aerts, R.J. Gillies, "Quantitative imaging of cancer in the postgenomic era: Radio(geno)mics, deep learning, and habitats," Cancer **124**, 4633-4649 (2018).

11 R. Stoyanova, M. Takhar, Y. Tschudi, J.C. Ford, G. SolÃ³rzano, N. Erho, Y. Balagurunathan, S. Punnen, E. Davicioni, R.J. Gillies, A. Pollack, "Prostate cancer radiomics and the promise of radiogenomics," Translational Cancer Research **5**, 432-447 (2016).

12 F. Khalvati, A. Wong, M.A. Haider, "Automated prostate cancer detection via comprehensive multi-parametric magnetic resonance imaging texture feature models," BMC Med. Imaging **15**, 27-015-0069-9 (2015).

13 A. Algohary, S. Viswanath, R. Shiradkar, S. Ghose, S. Pahwa, D. Moses, I. Jambor, R. Shnier, M. Bohm, A.M. Haynes, P. Brenner, W. Delprado, J. Thompson, M. Pulbrock, A.S. Purysko, S. Verma, L. Ponsky, P. Stricker, A. Madabhushi, "Radiomic features on MRI enable risk

This article is protected by copyright. All rights reserved

Author Manuscript

categorization of prostate cancer patients on active surveillance: Preliminary findings," J. Magn. Reson. Imaging (2018).

14 Q. Li, H. Bai, Y. Chen, Q. Sun, L. Liu, S. Zhou, G. Wang, C. Liang, Z.C. Li, "A Fully-Automatic Multiparametric Radiomics Model: Towards Reproducible and Prognostic Imaging Signature for Prediction of Overall Survival in Glioblastoma Multiforme," Sci. Rep. **7**, 14331-017-14753-7 (2017).

15 B. Ibragimov and L. Xing, "Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks," Med. Phys. **44**, 547-557 (2017).

16 T. Lustberg, J. van Soest, M. Gooding, D. Peressutti, P. Aljabar, J. van der Stoep, W. van Elmpt, A. Dekker, "Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer," Radiother. Oncol. **126**, 312-317 (2018).

17 G. Chlebus, A. Schenk, J.H. Moltz, B. van Ginneken, H.K. Hahn, H. Meine, "Automatic liver tumor segmentation in CT with fully convolutional neural networks and object-based postprocessing," Sci. Rep. **8**, 15497-018-33860-7 (2018).

18 S.G. Sapate, A. Mahajan, S.N. Talbar, N. Sable, S. Desai, M. Thakur, "Radiomics based detection and characterization of suspicious lesions on full field digital mammograms," Comput. Methods Programs Biomed. **163**, 1-20 (2018).

19 N. Antropova, B.Q. Huynh, M.L. Giger, "A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets," Med. Phys. **44**, 5162-5171 (2017).

20 D. Bonekamp, S. Kohl, M. Wiesenfarth, P. Schelb, J.P. Radtke, M. Gotz, P. Kickingereder, K. Yaqubi, B. Hitthaler, N. Gahlert, T.A. Kuder, F. Deister, M. Freitag, M. Hohenfellner, B.A. Hadaschik, H.P. Schlemmer, K.H. Maier-Hein, "Radiomic Machine Learning for Characterization of Prostate Lesions with MRI: Comparison to ADC Values," Radiology **289**, 128-137 (2018).

21 Y. Yuan, W. Qin, M. Buyyounouski, B. Ibragimov, S. Hancock, B. Han, L. Xing, "Prostate cancer classification with multiparametric MRI transfer learning model," Med. Phys. **46**, 756-765 (2019).

This article is protected by copyright. All rights reserved

Author Manuscript

22 W. Wu, C. Parmar, P. Grossmann, J. Quackenbush, P. Lambin, J. Bussink, R. Mak, H.J. Aerts, "Exploratory Study to Identify Radiomics Classifiers for Lung Cancer Histology," Front. Oncol. **6**, 71 (2016).

23 J.R. Ferreira Junior, M. Koenigkam-Santos, F.E.G. Cipriano, A.T. Fabro, P.M. Azevedo-Marques, "Radiomics-based features for pattern recognition of lung cancer histopathology and metastases," Comput. Methods Programs Biomed. **159**, 23-30 (2018).

24 B. Ganeshan, S. Abaleke, R.C. Young, C.R. Chatwin, K.A. Miles, "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**, 137-143 (2010).

25 S.H. Song, H. Park, G. Lee, H.Y. Lee, I. Sohn, H.S. Kim, S.H. Lee, J.Y. Jeong, J. Kim, K.S. Lee, Y.M. Shim, "Imaging Phenotyping Using Radiomics to Predict Micropapillary Pattern within Lung Adenocarcinoma," J. Thorac. Oncol. **12**, 624-632 (2017).

26 M. Crispin-Ortuzar, A. Apte, M. Grkovski, J.H. Oh, N.Y. Lee, H. Schoder, J.L. Humm, J.O. Deasy, "Predicting hypoxia status using a combination of contrast-enhanced computed tomography and [(18)F]-Fluorodeoxyglucose positron emission tomography radiomics features," Radiother. Oncol. **127**, 36-42 (2018).

27 A.N. Parra, H. Lu, Q. Li, R. Stoyanova, A. Pollack, S. Punnen, J. Choi, M. Abdalah, C. Lopez, K. Gage, J.Y. Park, Y. Kosj, J.M. Pow-Sang, R.J. Gillies, Y. Balagurunathan, "Predicting clinically significant prostate cancer using DCE-MRI habitat descriptors," Oncotarget **9**, 37125-37136 (2018).

28 J. Lee, S. Narang, J.J. Martinez, G. Rao, A. Rao, "Associating spatial diversity features of radiologically defined tumor habitats with epidermal growth factor receptor driver status and 12-month survival in glioblastoma: methods and preliminary investigation," J. Med. Imaging (Bellingham) **2**, 041006 (2015).

29 Y. Zhu, H. Li, W. Guo, K. Drukker, L. Lan, M.L. Giger, Y. Ji, "Deciphering Genomic Underpinnings of Quantitative MRI-based Radiomic Phenotypes of Invasive Breast Carcinoma," Sci. Rep. **5**, 17787 (2015).

This article is protected by copyright. All rights reserved

Author Manuscript

30 Y. Liu, J. Kim, Y. Balagurunathan, Q. Li, A.L. Garcia, O. Stringfield, Z. Ye, R.J. Gillies, "Radiomic Features Are Associated With EGFR Mutation Status in Lung Adenocarcinomas," Clin.Lung Cancer, 17(5), 441–448 (2016).

31 H.J. Yoon, I. Sohn, J.H. Cho, H.Y. Lee, J.H. Kim, Y.L. Choi, H. Kim, G. Lee, K.S. Lee, J. Kim, "Decoding Tumor Phenotypes for ALK, ROS1, and RET Fusions in Lung Adenocarcinoma Using a Radiomics Approach," Medicine (Baltimore) **94**, e1753 (2015).

32 J. Wang, F. Kato, N. Oyama-Manabe, R. Li, Y. Cui, K.K. Tha, H. Yamashita, K. Kudo, H. Shirato, "Identifying Triple-Negative Breast Cancer Using Background Parenchymal Enhancement Heterogeneity on Dynamic Contrast-Enhanced MRI: A Pilot Radiomics Study," PLoS One **10**, e0143308 (2015).

33 S.E. Lee, K. Han, J.Y. Kwak, E. Lee, E.K. Kim, "Radiomics of US texture features in differential diagnosis between triple-negative breast cancer and fibroadenoma," Sci. Rep. **8**, 13546-018-31906-4 (2018).

34 M. Diehn, C. Nardini, D.S. Wang, S. McGovern, M. Jayaraman, Y. Liang, K. Aldape, S. Cha, M.D. Kuo, "Identification of noninvasive imaging surrogates for brain tumor gene-expression modules," Proc. Natl. Acad. Sci. U. S. A. **105**, 5213-5218 (2007).

35 K. Gnep, A. Fargeas, R.E. Gutierrez-Carvajal, F. Commandeur, R. Mathieu, J.D. Ospina, Y. Rolland, T. Rohou, S. Vincendeau, M. Hatt, O. Acosta, R. de Crevoisier, "Haralick textural features on T2 -weighted MRI are associated with biochemical recurrence following radiotherapy for peripheral zone prostate cancer," J. Magn. Reson. Imaging , 45(1), 103-117 (2016).

36 M. Bogowicz, O. Riesterer, K. Ikenberg, S. Stieb, H. Moch, G. Studer, M. Guckenberger, S. Tanadini-Lang, "Computed Tomography Radiomics Predicts HPV Status and Local Tumor Control After Definitive Radiochemotherapy in Head and Neck Squamous Cell Carcinoma," Int. J. Radiat. Oncol. Biol. Phys. **99**, 921-928 (2017).

37 R.T. Leijenaar, M. Bogowicz, A. Jochems, F.J. Hoebers, F.W. Wesseling, S.H. Huang, B. Chan, J.N. Waldron, B. O'Sullivan, D. Rietveld, C.R. Leemans, R.H. Brakenhoff, O. Riesterer, S. Tanadini-Lang, M. Guckenberger, K. Ikenberg, P. Lambin, "Development and validation of a radiomic signature to predict HPV (p16) status from standard CT imaging: a multicenter study," Br. J. Radiol. **91**, 20170498 (2018).

This article is protected by copyright. All rights reserved

Author Manuscript

38 K. Nie, L. Shi, Q. Chen, X. Hu, S. Jabbour, N. Yue, T. Niu, X. Sun, "Rectal Cancer: Assessment of Neoadjuvant Chemo-Radiation Outcome Based on Radiomics of Multi-Parametric MRI," Clin. Cancer Res. 22(21), 5256-5263 (2016).

39 A. Oikonomou, F. Khalvati, P.N. Tyrrell, M.A. Haider, U. Tarique, L. Jimenez-Juan, M.C. Tjong, I. Poon, A. Eilaghi, L. Ehrlich, P. Cheung, "Radiomics analysis at PET/CT contributes to prognosis of recurrence and survival in lung cancer treated with stereotactic body radiotherapy," Sci. Rep. **8**, 4003-018-22357-y (2018).

40 J. Liu, Y. Mao, Z. Li, D. Zhang, Z. Zhang, S. Hao, B. Li, "Use of texture analysis based on contrast-enhanced MRI to predict treatment response to chemoradiotherapy in nasopharyngeal carcinoma " J. Magn. Reson. Imaging **44**, 445-455 (2016).

41 B. Ibragimov, D. Toesca, Y. Yuan, A. Koong, C. Daniel, L. Xing, "Neural networks for deep radiotherapy dose analysis and prediction of liver SBRT outcomes," IEEE J. Biomed. Health. Inform. IEEE J.Biomed.Health.Inform. doi: 10.1109/JBHI.2019.2904078, (2019).

42 E. Huynh, T.P. Coroller, V. Narayan, V. Agrawal, Y. Hou, J. Romano, I. Franco, R.H. Mak, H.J. Aerts, "CT-based radiomic analysis of stereotactic body radiation therapy patients with lung cancer," Radiother. Oncol. 120(2), 258-66 (2016).

43 M. Vallieres, C.R. Freeman, S.R. Skamene, I. El Naqa, "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**, 5471-5496 (2015).

44 T.P. Coroller, P. Grossmann, Y. Hou, E. Rios Velazquez, R.T. Leijenaar, G. Hermann, P. Lambin, B. Haibe-Kains, R.H. Mak, H.J. Aerts, "CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma," Radiother. Oncol. **114**, 345-350 (2015).

45 H. Hao, Z. Zhou, S. Li, G. Maquilan, M.R. Folkert, P. Iyengar, K.D. Westover, K. Albuquerque, F. Liu, H. Choy, R. Timmerman, L. Yang, J. Wang, "Shell feature: a new radiomics descriptor for predicting distant failure after radiotherapy in non-small cell lung cancer and cervix cancer," Phys. Med. Biol. **63**, 095007-6560/aabb5e (2018).

46 R.M. Haralick, K. Shanmugam, I. Dinstein, "Textural Features for Image Classification " IEEE Trans. Syst. Man Cybern. **3**, 610 - 621 (1973).

This article is protected by copyright. All rights reserved

Author Manuscript

841 47 N. Ohri, F. Duan, B.S. Snyder, B. Wei, M. Machtay, A. Alavi, B.A. Siegel, D.W. Johnson, J.D.
842 Bradley, A. DeNittis, M. Werner-Wasik, I. El Naqa, "Pretreatment 18F-FDG PET Textural
843 Features in Locally Advanced Non-Small Cell Lung Cancer: Secondary Analysis of ACRIN
844 6668/RTOG 0235," J. Nucl. Med. **57**, 842-848 (2016).

845 48 A. Hosny, C. Parmar, T.P. Coroller, P. Grossmann, R. Zeleznik, A. Kumar, J. Bussink, R.J.
846 Gillies, R.H. Mak, H.J.W.L. Aerts, "Deep learning for lung cancer prognostication: A
847 retrospective multi-cohort radiomics study," PLoS Med. **15**, e1002711 (2018).

848 49 H.J. Aerts, "The Potential of Radiomic-Based Phenotyping in Precision Medicine: A Review,"
849 JAMA Oncol, 2(12):1636-1642 (2016).

850 50 P. Kickingereder, M. Gotz, J. Muschelli, A. Wick, U. Neuberger, R.T. Shinohara, M. Sill, M.
851 Nowosielski, H.P. Schlemmer, A. Radbruch, W. Wick, M. Bendszus, K.H. Maier-Hein, D.
852 Bonekamp, "Large-scale Radiomic Profiling of Recurrent Glioblastoma Identifies an Imaging
853 Predictor for Stratifying Anti-Angiogenic Treatment Response," Clin. Cancer Res. 22(23),
854 5765-5771 (2016).

855 51 R. Sun, E.J. Limkin, M. Vakalopoulou, L. Dercle, S. Champiat, S.R. Han, L. Verlingue, D.
856 Brandao, A. Lancia, S. Ammari, A. Hollebecque, J.Y. Scoazec, A. Marabelle, C. Massard, J.C.
857 Soria, C. Robert, N. Paragios, E. Deutsch, C. Ferte, "A radiomics approach to assess tumour-
858 infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging
859 biomarker, retrospective multicohort study," Lancet Oncol. **19**, 1180-1191 (2018).

860 52 X. Fave, L. Zhang, J. Yang, D. Mackin, P. Balter, D. Gomez, D. Followill, A.K. Jones, F.
861 Stingo, Z. Liao, R. Mohan, L. Court, "Delta-radiomics features for the prediction of patient
862 outcomes in non-small cell lung cancer," Sci. Rep. **7**, 588-017-00665-z (2017).

863 53 X. Fave, D. Mackin, J. Yang, J. Zhang, D. Fried, P. Balter, D. Followill, D. Gomez, A.K. Jones,
864 F. Stingo, J. Fontenot, L. Court, "Can radiomics features be reproducibly measured from
865 CBCT images for patients with non-small cell lung cancer?" Med. Phys. **42**, 6784-6797 (2015).

866 54 J.E. van Timmeren, R.T. Leijenaar, W. van Elmpt, P. Lambin, "Interchangeability of a Radiomic
867 Signature Between Conventional and Weekly Cone Beam Computed Tomography Allowing
868 Response Prediction in Non-Small Cell Lung Cancer," Int. J. Radiat. Oncol. Biol. Phys. **96**,
869 S193 (2016).

This article is protected by copyright. All rights reserved

Author Manuscript

870 55 J.E. van Timmeren, R.T.H. Leijenaar, W. van Elmpt, B. Reymen, P. Lambin, "Feature selection
871 methodology for longitudinal cone-beam CT radiomics," Acta Oncol. **56**, 1537-1543 (2017).

872 56 A. Moran, M.E. Daly, S.S.F. Yip, T. Yamamoto, "Radiomics-based Assessment of Radiation-
873 induced Lung Injury After Stereotactic Body Radiotherapy," Clin. Lung Cancer. **18**, e425-e431
874 (2017).

875 57 A. Cunliffe, S.G. Armato 3rd, R. Castillo, N. Pham, T. Guerrero, H.A. Al-Hallaq, "Lung texture
876 in serial thoracic computed tomography scans: correlation of radiomics-based features with
877 radiation therapy dose and radiation pneumonitis development," Int. J. Radiat. Oncol. Biol.
878 Phys. **91**, 1048-1056 (2015).

879 58 S.P. Krafft, A. Rao, F. Stingo, T.M. Briere, L.E. Court, Z. Liao, M.K. Martel, "The utility of
880 quantitative CT radiomics features for improved prediction of radiation pneumonitis," Med.
881 Phys. **45**, 5317-5324 (2018).

882 59 S.A. Mattonen, D.A. Palma, C. Johnson, A.V. Louie, M. Landis, G. Rodrigues, I. Chan, R.
883 Etemad-Rezai, T.P. Yeung, S. Senan, A.D. Ward, "Detection of Local Cancer Recurrence
884 After Stereotactic Ablative Radiation Therapy for Lung Cancer: Physician Performance Versus
885 Radiomic Assessment," Int. J. Radiat. Oncol. Biol. Phys. **94**, 1121-1128 (2016).

886 60 Z. Zhang, J. Yang, A. Ho, W. Jiang, J. Logan, X. Wang, P.D. Brown, S.L. McGovern, N. Guha-
887 Thakurta, S.D. Ferguson, X. Fave, L. Zhang, D. Mackin, L.E. Court, J. Li, "A predictive model
888 for distinguishing radiation necrosis from tumour progression after gamma knife radiosurgery
889 based on radiomic features from MR images," Eur. Radiol. **28**, 2255-2263 (2018).

890 61 S. Viswanath, R. Toth, M. Rusu, D. Sperling, H. Lepor, J. Futterer, A. Madabhushi, "Identifying
891 Quantitative In Vivo Multi-Parametric MRI Features For Treatment Related Changes after
892 Laser Interstitial Thermal Therapy of Prostate Cancer," Neurocomputing **144**, 13-23 (2014).

893 62 J.Y. Yu, H.P. Zhang, Z.Y. Tang, J. Zhou, X.J. He, Y.Y. Liu, X.J. Liu, D.J. Guo, "Value of
894 texture analysis based on enhanced MRI for predicting an early therapeutic response to
895 transcatheter arterial chemoembolisation combined with high-intensity focused ultrasound
896 treatment in hepatocellular carcinoma," Clin. Radiol. **73**, 758.e9-758.e18 (2018).

Author Manuscript

This article is protected by copyright. All rights reserved

897 63 H. Bagher-Ebadian, F. Siddiqui, C. Liu, B. Movsas, I.J. Chetty, "On the impact of smoothing
898 and noise on robustness of CT and CBCT radiomics features for patients with head and neck
899 cancers," Med. Phys. **44**, 1755-1770 (2017).

900 64 M. Shafiq-Ul-Hassan, G.G. Zhang, K. Latifi, G. Ullah, D.C. Hunt, Y. Balagurunathan, M.A.
901 Abdalah, M.B. Schabath, D.G. Goldgof, D. Mackin, L.E. Court, R.J. Gillies, E.G. Moros,
902 "Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels,"
903 Med. Phys. **44**, 1050-1062 (2017).

904 65 Z. Alex, L. Stefan, V. Martin, L. Steffen, the,Image Biomarker Standardisation Initiative for,
905 "Image biomarker standardisation initiative," arXiv e-prints, ar:1612.07003 (2016).

906 66 F.H. van Velden, G.M. Kramer, V. Frings, I.A. Nissen, E.R. Mulder, A.J. de Langen, O.S.
907 Hoekstra, E.F. Smit, R. Boellaard, "Repeatability of Radiomic Features in Non-Small-Cell
908 Lung Cancer [18F]FDG-PET/CT Studies: Impact of Reconstruction and Delineation," Mol.
909 Imaging Biol, 18(5):788-95 (2016).

910 67 R.T. Leijenaar, G. Nalbantov, S. Carvalho, W.J. van Elmpt, E.G. Troost, R. Boellaard, H.J.
911 Aerts, R.J. Gillies, P. Lambin, "The effect of SUV discretization in quantitative FDG-PET
912 Radiomics: the need for standardized methodology in tumor texture analysis," Sci. Rep. **5**,
913 11075 (2015).

914 68 B. Julesz, "Experiments in the visual perception of texture," Sci. Am. **232**, 34-43 (1975).

915 69 I. El Naqa, P. Grigsby, A. Apte, E. Kidd, E. Donnelly, D. Khullar, S. Chaudhari, D. Yang, M.
916 Schmitt, R. Laforest, W. Thorstad, J.O. Deasy, "Exploring feature-based approaches in PET
917 images for predicting cancer treatment outcomes," Pattern Recognit **42**, 1162-1171 (2009).

918 70 T.P. Coroller, V. Agrawal, V. Narayan, Y. Hou, P. Grossmann, S.W. Lee, R.H. Mak, H.J. Aerts,
919 "Radiomic phenotype features predict pathological response in non-small cell lung cancer,"
920 Radiother. Oncol., 119(3), 480-6 (2016).

921 71 F. Tixier, M. Hatt, C. Valla, V. Fleury, C. Lamour, S. Ezzouhri, P. Ingrand, R. Perdrisot, D.
922 Visvikis, C.C. Le Rest, "Visual versus quantitative assessment of intratumor 18F-FDG PET
923 uptake heterogeneity: Prognostic value in non-small cell lung cancer," Journal of Nuclear
924 Medicine **55**, 1235-1241 (2014).

This article is protected by copyright. All rights reserved

Author Manuscript

925 72 C. Parmar, P. Grossmann, J. Bussink, P. Lambin, H.J. Aerts, "Machine Learning methods for
926 Quantitative Radiomic Biomarkers," Sci. Rep. **5**, 13087 (2015).

927 73 R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," Journal of the Royal
928 Statistical Society.Series B (Methodological) **58**, 267-288 (1996).

929 74 S. Wold, K. Esbensen, P. Geladi, "Principal component analysis," Chemometrics and Intelligent
930 Laboratory Systems **2**, 37-52 (1987).

931 75 Laurens van der Maaten and Geoffrey E. Hinton, "Visualizing Data using t-SNE," Journal of
932 Machine Learning Research, 9, 2579-2605, (2008).

933 76 C. Parmar, R.T.H. Leijenaar, P. Grossmann, E.R. Velazquez, J. Bussink, D. Rietveld, M.M.
934 Rietbergen, B. Haibe-Kains, P. Lambin, H.J.W.L. Aerts, "Radiomic feature clusters and
935 Prognostic Signatures specific for Lung and Head & Neck cancer," Scientific Reports 5, 11044
936 (2015).

937 77 J. Wu, Y. Cui, X. Sun, G. Cao, B. Li, D.M. Ikeda, A.W. Kurian, R. Li, "Unsupervised Clustering
938 of Quantitative Image Phenotypes Reveals Breast Cancer Subtypes with Distinct Prognoses
939 and Molecular Pathways," Clin. Cancer Res. **23**, 3334-3342 (2017).

940 78 P.P. Ypsilantis, M. Siddique, H.M. Sohn, A. Davies, G. Cook, V. Goh, G. Montana, "Predicting
941 Response to Neoadjuvant Chemotherapy with PET Imaging Using Convolutional Neural
942 Networks," PLoS One **10**, e0137036 (2015).

943 79 S. Chen, S. Zhou, F.F. Yin, L.B. Marks, S.K. Das, "Investigation of the support vector machine
944 algorithm to predict lung radiation-induced pneumonitis," Med. Phys. **34**, 3808-3814 (2007).

945 80 I. El Naqa, R. Li, M.J. Murphy, eds. "*Machine Learning in Radiation Oncology: Theory and*
946 *Application"*. Switzerland: Springer International
947 Publishing; 2015. https://doi.org/10.1007/978-3-319-18305-3, (2015)

948 81 D. Yang, G. Rao, J. Martinez, A. Veeraraghavan, A. Rao, "Evaluation of tumor-derived MRI-
949 texture features for discrimination of molecular subtypes and prediction of 12-month survival
950 status in glioblastoma " Med. Phys. **42**, 6725-6735 (2015).

This article is protected by copyright. All rights reserved

Author Manuscript

951 82 M. Vallieres, E. Kay-Rivest, L. Perrin, X. Liem, C. Furstoss, H.J.W.L. Aerts, N. Khaouam, P.
952 Nguyen-Tan, C. Wang, K. Sultanem, J. Seuntjens, I. El Naqa, "Radiomics strategies for risk
953 assessment of tumour failure in head-and-neck cancer," Scientific Reports **7**, 10117 (2017).

954 83 Charles Rosenberg, Martial Hebert and Henry Schneiderman, "Semi-Supervised Self-Training of
955 Object Detection Models," in: WACV-MOTION '05 Proceedings of the Seventh IEEE
956 Workshops on Application of Computer Vision (WACV/MOTION'05), Volume 1, 29-36
957 (2005). doi:10.1109/ACVMOT.2005.107

958 84 Thorsten Joachims, "Transductive Inference for Text Classification Using Support Vector
959 Machines," in: ICML '99 Proceedings of the Sixteenth International Conference on Machine
960 Learning, 200-209 (1999) doi:10.1109/ACVMOT.2005.107.

961 85 Avrim Blum, John Lafferty, Mugizi Robert Rwebangira and Rajashekar Reddy, "Semi-
962 supervised Learning Using Randomized Mincuts" in ICML '04 Proceedings of the twenty-first
963 international conference on Machine learning, 13 (2004) DOI: 10.1145/1015330.1015429.

964 86 D.R. Cox, "Regression Models and Life-Tables," Journal of the Royal Statistical Society.Series
965 B (Methodological) **34**, 187-220 (1972).

966 87 H. Ishwaran, U.B. Kogalur, E.H. Blackstone, M.S. Lauer, "Random survival forests " The
967 Annals of Applied Statistics **2**, 841 - 860 (2008).

968 88 V. Van Belle, K. Pelckmans, S. Van Huffel, J.A.K. Suykens, "Support vector methods for
969 survival analysis: a comparison between ranking and regression approaches," Artificial
970 Intelligence in Medicine **53**, 107-118 (2011).

971 89 Y. LeCun, Y. Bengio, G. Hinton, "Deep learning," Nature **521**, 436-444 (2015).

972 90 K. Suzuki, "Pixel-based machine learning in medical imaging," Int. J. Biomed. Imaging **2012**,
973 792079 (2012).

974 91 K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image
975 Recognition," arXiv:1409.1556 (2014).

976 92 K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition," presented at
977 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016. 770-778
978 (2016)

This article is protected by copyright. All rights reserved

Author Manuscript

979 93 D.H. Hubel and T.N. Wiesel, "Receptive fields and functional architecture of monkey striate
980 cortex," J. Physiol. (Lond.) **195**, 215-243 (1968).

981 94 S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Comput. **9**, 1735-1780
982 (1997).

983 95 K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio,
984 "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine
985 Translation," in Proceedings of the 2014 Conference on Empirical Methods in Natural
986 Language Processing (EMNLP), 1724-1734 (2014) DOI: 10.3115/v1/D14-1179.

987 96 D.P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," arXiv:1312.6114 (2013).

988 97 F. Li, H. Qiao, B. Zhang, "Discriminatively boosted image clustering with fully convolutional
989 auto-encoders," Pattern Recognition **83**, 161-173 (2018).

990 98 G.E. Hinton and R.R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural
991 Networks," Science **313**, 504-507 (2006).

992 99 D.P. Kingma, D.J. Rezende, S. Mohamed, M. Welling, "Semi-Supervised Learning with Deep
993 Generative Models," eprint arXiv:1406.5298 (2014).

994 100 H.G. E., S. Nitish, K. Alex, S. Ilya, S.R. R., "Improving neural networks by preventing co-
995 adaptation of feature detectors," arXiv e-prints, ar:1207.0580 (2012).

996 101 I. Sergey and S. Christian, "Batch Normalization: Accelerating Deep Network Training by
997 Reducing Internal Covariate Shift," arXiv e-prints, ar:1502.03167 (2015).

998 102 J. Long, E. Shelhamer, T. Darrell, "Fully convolutional networks for semantic segmentation,"
999 presented at 2015 IEEE Conference on Computer Vision and Pattern Recognition. Published in
1000 EEE Transactions on Pattern Analysis and Machine Intelligence 39(4): 640-651 doi:
1001 10.1109/TPAMI.2016.2572683.

1002 103 O. Ronneberger, P. Fischer, T. Brox, "U-Net: Convolutional Networks for Biomedical Image
1003 Segmentation," presented at Medical Image Computing and Computer-Assisted Intervention (
1004 MICCAI) 2015 DOI:10.1007/978-3-319-24574-4_28.

Author Manuscript

This article is protected by copyright. All rights reserved

104 A. Esteva, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau, S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," Nature **542**, 115-118 (2017).

105 H. Chang, J. Han, C. Zhong, A.M. Snijders, J.H. Mao, "Unsupervised Transfer Learning via Multi-Scale Convolutional Sparse Coding for Biomedical Applications," IEEE Trans. Pattern Anal. Mach. Intell. **40**, 1182-1194 (2018).

106 G.S. Collins, J.B. Reitsma, D.G. Altman, K.G. Moons, "Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement," BMC Med. **13**, 1-014-0241-z (2015).

107 T. Hastie, R. Tibshirani, J.H. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction (Springer, New York, NY, 2009).

108 J. Ren, J. Zhou, W. Ding, B. Zhong, J. Zhou, "Clinicopathological characteristics and imaging features of pulmonary adenocarcinoma with micropapillary pattern," Zhonghua Zhong Liu Za Zhi **36**, 282-286 (2014).

109 F.E. Harrell Jr, K.L. Lee, D.B. Mark, "Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors," Stat. Med. **15**, 361-387 (1996).

110 M.J. Pencina and R.B. D'Agostino, "Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation," Statist. Med. **23**, 2109-2123 (2004).

111 P. Royston and D.G. Altman, "External validation of a Cox prognostic model: principles and methods," BMC Med. Res. Methodol. **13**, 33-2288-13-33 (2013).

112 R. Boellaard, "Standards for PET image acquisition and quantitative data analysis," J. Nucl. Med. **50 Suppl 1**, 11S-20S (2009).

113 D. Mackin, X. Fave, L. Zhang, D. Fried, J. Yang, B. Taylor, E. Rodriguez-Rivera, C. Dodge, A.K. Jones, L. Court, "Measuring Computed Tomography Scanner Variability of Radiomics Features," Invest. Radiol. **50**, 757-765 (2015).

Author Manuscript

This article is protected by copyright. All rights reserved

114 Szczypinski PM, Strzelecki M FAU - Materka,Andrzej, Materka A FAU - Klepaczko,Artur, Klepaczko A, "MaZda--a software package for image texture analysis." Computer methods and programs in biomedicine 94(1):66-76. DOI: 10.1016/j.cmpb.2008.08.005.

115 C. Nioche, F. Orlhac, S. Boughdad, S. Reuze, J. Goya-Outi, C. Robert, C. Pellot-Barakat, M. Soussan, F. Frouin, I. Buvat, "LIFEx: A Freeware for Radiomic Feature Calculation in Multimodality Imaging to Accelerate Advances in the Characterization of Tumor Heterogeneity," Cancer Res. **78**, 4786-4789 (2018).

116 D.L. Rubin, D. Willrett, M.J. O'Connor, C. Hage, C. Kurtz, D.A. Moreira, "Automated tracking of quantitative assessments of tumor burden in clinical trials," Transl. Oncol. **7**, 23-35 (2014).

117 J.J.M. van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R.G.H. Beets-Tan, J.C. Fillion-Robin, S. Pieper, H.J.W.L. Aerts, "Computational Radiomics System to Decode the Radiographic Phenotype," Cancer Res. **77**, e104-e107 (2017).

118 Y. Dicente Cid, J. Castelli, R. Schaer, N. Scher, A. Pomoni, J.O. Prior, A. Depeursinge, "Chapter 12 - QuantImage: An Online Tool for High-Throughput 3D Radiomics Feature Extraction in PET-CT," in Biomedical Texture Analysis, 349-377 (2017). DOI: 10.1016/B978-0-12-812133-7.00012-0

119 S. Echegaray, S. Bakr, D.L. Rubin, S. Napel, "Quantitative Image Feature Engine (QIFE): an Open-Source, Modular Engine for 3D Quantitative Feature Extraction from Volumetric Medical Images," J. Digit. Imaging **31**, 403-414 (2018).

120 L. Zhang, D.V. Fried, X.J. Fave, L.A. Hunter, J. Yang, L.E. Court, "IBEX: an open infrastructure software platform to facilitate collaborative work in radiomics," Med. Phys. **42**, 1341-1353 (2015).

121 A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka, J. Buatti, S. Aylward, J.V. Miller, S. Pieper, R. Kikinis, "3D Slicer as an image computing platform for the Quantitative Imaging Network," Magn. Reson. Imaging **30**, 1323-1341 (2012).

122 J. Yosinski, J. Clune, A. Nguyen, T. Fuchs, H. Lipson, "Understanding Neural Networks Through Deep Visualization," eprint arXiv:1506.06579 (2015).

This article is protected by copyright. All rights reserved

Author Manuscript

123 V. Sankar, D. Kumar, D.A. Clausi, G.W. Taylor, A. Wong, "SISC: End-to-end Interpretable Discovery Radiomics-Driven Lung Cancer Prediction via Stacked Interpretable Sequencing Cells," eprint arXiv:1901.04641 (2019).

124 Y. Luo, D. McShan, D. Ray, M. Matuszak, S. Jolly, T. Lawrence, F. Kong, R. Ten Haken, I. El Naqa, "Development of a Fully Cross-Validated Bayesian Network Approach for Local Control Prediction in Lung Cancer," IEEE Transactions on Radiation and Plasma Medical Sciences **3**, 232-241 (2019).

125 M. Hatt, F. Tixier, L. Pierce, P.E. Kinahan, C.C. Le Rest, D. Visvikis, "Characterization of PET/CT images using texture analysis: the past, the present... any future?" Eur. J. Nucl. Med. Mol. Imaging **44**, 151-165 (2017).

126 X. Fave, M. Cook, A. Frederick, L. Zhang, J. Yang, D. Fried, F. Stingo, L. Court, "Preliminary investigation into sources of uncertainty in quantitative imaging features," Comput. Med. Imaging Graphics **44**, 54-61 (2015).

127 E. Pfaehler, R.J. Beukinga, J.R. de Jong, R.H.J.A. Slart, C.H. Slump, R.A.J.O. Dierckx, R. Boellaard, "Repeatability of (18) F-FDG PET radiomic features: A phantom study to explore sensitivity to image reconstruction settings, noise, and delineation method," Med. Phys. **46**, 665-678 (2019).

128 R.T. Leijenaar, S. Carvalho, E.R. Velazquez, W.J. van Elmpt, C. Parmar, O.S. Hoekstra, C.J. Hoekstra, R. Boellaard, A.L. Dekker, R.J. Gillies, H.J. Aerts, P. Lambin, "Stability of FDG-PET Radiomics features: an integrated analysis of test-retest and inter-observer variability," Acta Oncol. **52**, 1391-1397 (2013).

129 H. Kim, C.M. Park, M. Lee, S.J. Park, Y.S. Song, J.H. Lee, E.J. Hwang, J.M. Goo, "Impact of Reconstruction Algorithms on CT Radiomic Features of Pulmonary Tumors: Analysis of Intra- and Inter-Reader Variability and Inter-Reconstruction Algorithm Variability," PLoS One **11**, e0164924 (2016).

130 B. Zhao, L.P. James, C.S. Moskowitz, P. Guo, M.S. Ginsberg, R.A. Lefkowitz, Y. Qin, G.J. Riely, M.G. Kris, L.H. Schwartz, "Evaluating Variability in Tumor Measurements from Same-day Repeat CT Scans of Patients with Nonâ€“Small Cell Lung Cancer," Radiology **252**, 263-272 (2009).

This article is protected by copyright. All rights reserved

Author Manuscript

131 M.C. Desseroit, F. Tixier, W.A. Weber, B.A. Siegel, C. Cheze Le Rest, D. Visvikis, M. Hatt, "Reliability of PET/CT shape and heterogeneity features in functional and morphological components of Non-Small Cell Lung Cancer tumors: a repeatability analysis in a prospective multi-center cohort," J. Nucl. Med. 58(3), 406-411 (2016).

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

133 L. Lu, W. Lv, J. Jiang, J. Ma, Q. Feng, A. Rahmim, W. Chen, "Robustness of Radiomic Features in [11C]Choline and [18F]FDG PET/CT Imaging of Nasopharyngeal Carcinoma: Impact of Segmentation and Discretization," Mol. Imaging Biol. 18(6), 935-945 (2016).

134 C. Bailly, C. Bodet-Milin, S. Couespel, H. Necib, F. Kraeber-Bodéré, C. Ansquer, T. Carlier, "Revisiting the robustness of PET-based textural features in the context of multi-centric trials," PLoS ONE **11(7)**, e0159984 (2016) DOI: 10.1371/journal.pone.0159984.

135 F. Yang, N. Dogan, R. Stoyanova, J.C. Ford, "Evaluation of radiomic texture feature error due to MRI acquisition and reconstruction: A simulation study utilizing ground truth," Phys. Med. **50**, 26-36 (2018).

136 W. Grootjans, F. Tixier, C.S. van der Vos, D. Vriens, C.C. Le Rest, J. Bussink, W.J. Oyen, L.F. de Geus-Oei, D. Visvikis, E.P. Visser, "The impact of optimal respiratory gating and image noise on evaluation of intra-tumor heterogeneity in 18F-FDG positron emission tomography imaging of lung cancer," J. Nucl. Med. 57(11), 1692-1698 (2016).

137 C. Parmar, E. Rios Velazquez, R. Leijenaar, M. Jermoumi, S. Carvalho, R.H. Mak, S. Mitra, B.U. Shankar, R. Kikinis, B. Haibe-Kains, P. Lambin, H.J. Aerts, "Robust Radiomics feature quantification using semiautomatic volumetric segmentation," PLoS One **9**, e102107 (2014).

138 S. Echegaray, O. Gevaert, R. Shah, A. Kamaya, J. Louie, N. Kothary, S. Napel, "Core samples for radiomics features that are insensitive to tumor segmentation: method and pilot study using CT images of hepatocellular carcinoma," J. Med. Imaging (Bellingham) **2**, 041011 (2015).

139 P. Lambin, R.T.H. Leijenaar, T.M. Deist, J. Peerlings, E.E.C. de Jong, J. van Timmeren, S. Sanduleanu, R.T.H.M. Larue, A.J.G. Even, A. Jochems, Y. van Wijk, H. Woodruff, J. van Soest, T. Lustberg, E. Roelofs, W. van Elmpt, A. Dekker, F.M. Mottaghy, J.E. Wildberger, S.

This article is protected by copyright. All rights reserved

Author Manuscript

Walsh, "Radiomics: the bridge between medical imaging and personalized medicine," Nat. Rev. Clin. Oncol. **14**, 749-762 (2017).

140 A. Chalkidou, M.J. O'Doherty, P.K. Marsden, "False discovery rates in PET and CT studies with texture features: A systematic review," PLoS ONE **10**(5), e0124165 (2015).

141 E.A. AlBadawy, A. Saha, M.A. Mazurowski, "Deep learning for segmentation of brain tumors: Impact of cross-institutional training and testing," Med. Phys. **45**, 1150-1158 (2018).

142 W.E. Johnson, C. Li, A. Rabinovic, "Adjusting batch effects in microarray expression data using empirical Bayes methods," Biostatistics **8**, 118-127 (2007).

143 K. Chang, N. Balachandar, C. Lam, D. Yi, J. Brown, A. Beers, B. Rosen, D.L. Rubin, J. Kalpathy-Cramer, "Distributed deep learning networks among institutions for medical imaging," J. Am. Med. Inform. Assoc. **25**, 945-954 (2018).

144 R.M. Zur, Y. Jiang, L.L. Pesce, K. Drukker, "Noise injection for training artificial neural networks: a comparison with weight decay and early stopping," Med. Phys. **36**, 4810-4818 (2009).

145 S. Li, Y. Chen, Y. Peng, L. Bai, "Learning More Robust Features with Adversarial Training," arXiv:1804.07757 (2018).

146 M. Vallieres, A. Zwanenburg, B. Badic, C. Cheze Le Rest, D. Visvikis, M. Hatt, "Responsible Radiomics Research for Faster Clinical Translation," J. Nucl. Med. **59**, 189-193 (2018).

147 M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.W. Boiten, L.B. da Silva Santos, P.E. Bourne, J. Bouwman, A.J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, "The FAIR Guiding Principles for scientific data management and stewardship," Sci. Data **3**, 160018 (2016).

This article is protected by copyright. All rights reserved

Author Manuscript

Table 1. Open access software programs for radiomics analysis.

Software/ Toolbox	**MaZda**[114]	**lifeX**[115]	**ePAD**[116]	**QIFE**[119]	**HeterogeneityCAD**[3]	**PyRadiomics / Radiomics**[117]	**QuantImage**[118]	**Texture Analysis Toolbox**[43]	**IBEX**[120]	**MEDomicsLab**
Research group	Institute of Electronics, Technical University of Lodz, Poland	IMIV, CEA, Inserm, CNRS, Univ. Paris-Sud, Université Paris Saclay	Rubin Lab, Stanford University	Sandy Napel, Stanford University	V.Narayan, J. Jagadeesan	Dana-Farber Cancer Institute, Brigham Women's Hospital Harvard Medical School, Boston	University of Applied Science and Arts, Western Switzerland	M. Vallières	The University of Texas MD Anderson Cancer Center, Houston, Texas	MEDomics consortium
Image modalities	CT, MRI, PET	CT, MRI, PET, ultrasound	CT, MRI, radiography	CT, MRI, PET	CT, MRI, PET	CT, MRI, PET		CT, MRI, PET	CT, MRI, PET	CT, MRI, PET
Segmentation	YES	YES	YES	NO	NO	NO	NO	NO	YES	NO
Segmentation methods	manual, automatic (threshold, flood-filling)	manual, automatic (threshold, snake)	Manual	/	/	/	/	/	manual, automatic (threshold)	/
Radiomic features: morphology	YES	YES	NO	YES	YES	YES	YES	YES	YES	YES
statistical 1° order	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
statistical 2° order	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
statistical 3° order	YES	YES	NO	YES	YES	YES	YES	YES	YES	YES
Filtering	YES	NO	NO	NO	NO	YES	NO	NO	YES	YES
Feature selection	YES	NO	NO	NO	NO	NO	NO	YES	NO	YES
Feature selection methods	Fisher score, classification error, corr. coeff, **m**utual informat., minimal classification error	/	/	/	/	/	/	Maximal information coefficient	/	False discovery avoidance, Elastic Net, minimum Redundancy Maximum Relevance
Stratification	NO	NO	NO	NO	NO	NO	NO	NO	NO	YES

Author Manuscript

This article is protected by copyright. All rights reserved

Table 2. Reporting guidelines on the computation of radiomics features (adapted from Refs.[65] and [146]).

GENERAL	
Image acquisition	Acquisition protocols and scanner parameters: equipment vendor, reconstruction algorithms and filters, field of view and acquisition matrix dimensions, MRI sequence parameters, PET acquisition time and injected dose, CT x-ray energy (kVp) and exposure (mAs), etc.
Volumetric analysis	Imaging volumes are analyzed as separate images (2D) or as fully-connected volumes (3D).
Workflow structure	Sequence of processing steps leading to the extraction of features.
Software	Software type and version of code used for the computation of features.
IMAGE PRE-PROCESSING	
Conversion	How data were converted from input images: e.g, conversion of PET activity counts to SUV, calculation of ADC maps from raw DW-MRI signal, etc.
Processing	Image processing steps taken after image acquisition: e.g., noise filtering, intensity non-uniformity correction in MRI, partial-volume effect corrections, etc.
ROI SEGMENTATION[a;b]	How regions of interests (ROIs) were delineated in the images: software and/or algorithms used, how many different persons and what expertise (specialty, experience), how a consensus was obtained if several persons carried out the segmentation, in automatic or semi-automatic mode, etc.
INTERPOLATION	
Voxel dimensions	Original and interpolated voxel dimensions.
Image interpolation method	Method used to interpolate voxels values (e.g, linear, cubic, spline, etc.) as well as how original and interpolated grids were aligned.
Intensity rounding	Rounding procedures for non-integer interpolated gray levels (if applicable), e.g.,

Author Manuscript

This article is protected by copyright. All rights reserved

	rounding of Hounsfield units in CT imaging following interpolation.
ROI interpolation method	Method used to interpolate ROI masks. Definition of how original and interpolated grids were aligned.
ROI partial volume	Minimum partial volume fraction required to include an interpolated ROI mask voxel in the interpolated ROI (if applicable): e.g., a minimum partial volume fraction of 0.5 when using linear interpolation.
ROI RE-SEGMENTATION	
Inclusion/exclusion criteria	Criteria for inclusion and/or exclusion of voxels from the ROI intensity mask (if applicable), e.g., the exclusion of voxels with Hounsfield units values outside a pre-defined range inside the ROI intensity mask in CT imaging.
IMAGE DISCRETIZATION	
Discretization method	Method used for discretizing image intensities prior to feature extraction: e.g., fixed bin number, fixed bin width, histogram equalization, etc.
Discretization parameters	Parameters used for image discretization: the number of bins, the bin width and minimal value of discretization range, etc.
FEATURE CALCULATION	
Features set	Description and formulas of all calculated features.
Features parameters	Settings used for the calculation of features: voxel connectivity, with or without merging by slice, with or without merging directional texture matrices, etc.
CALIBRATION	
Image processing steps	Specifying which image processing steps match the benchmarks of the IBSI.
Features calculation	Specifying which feature calculations match the benchmarks of the IBSI.

[a]In order to reduce inter-observer variability, automatic and semi-automatic methods are favored.

[b]In multimodal applications (e.g., PET/CT, PET/MRI, etc.) ROI definition may involve the propagation of contours between modalities via co-registration. In that case, the technical details of the registration should also be provided.

This article is protected by copyright. All rights reserved

Author Manuscript

Table 3. Quality factors in radiomics studies (adapted from Refs [139] and [146]).

IMAGING	
Standardized imaging protocols	Imaging acquisition protocols are well described and ideally similar across patients. Alternatively, methodological steps are taken towards standardizing them.
Imaging quality assurance	Methodological steps are taken to only incorporate acquired images of sufficient quality.
Calibration	Computation of radiomics features and image processing steps match the benchmarks of the IBSI.
EXPERIMENTAL SETUP	
Multi-institutional/external datasets	Model construction and/or performance evaluation is carried out using cohorts from different institutions, ideally from different parts of the world.
Registration of prospective study	Prospective studies provide the highest level of evidence supporting the clinical validity and usefulness of radiomics models.
FEATURE SELECTION	
Feature robustness	The robustness of features against segmentation variations and varying imaging settings (e.g., noise fluctuations, inter-scanner differences, etc.) is evaluated. Unreliable features are discarded.
Feature complementarity	The inter-correlation of features is evaluated. Redundant features are discarded.
MODEL ASSESSMENT	
False discovery corrections	Correction for multiple testing comparisons (e.g., Bonferroni or Benjamini-Hochberg) is applied in univariate analysis.
Estimation of model performance	The teaching dataset is separated into training and validation set(s) to estimate optimal model parameters. Example methods include bootstrapping, cross-validation, random sub-sampling, etc.
Independent testing	A testing set distinct from the teaching set is used to evaluate the performance of complete models (i.e., without retraining and without adaptation of cut- off values). The evaluation of the performance is unbiased and not used to optimize model parameters.
Performance results consistency	Model performance obtained in the training, validation and testing sets is reported. Consistency checks of performance measures across the different sets are performed.

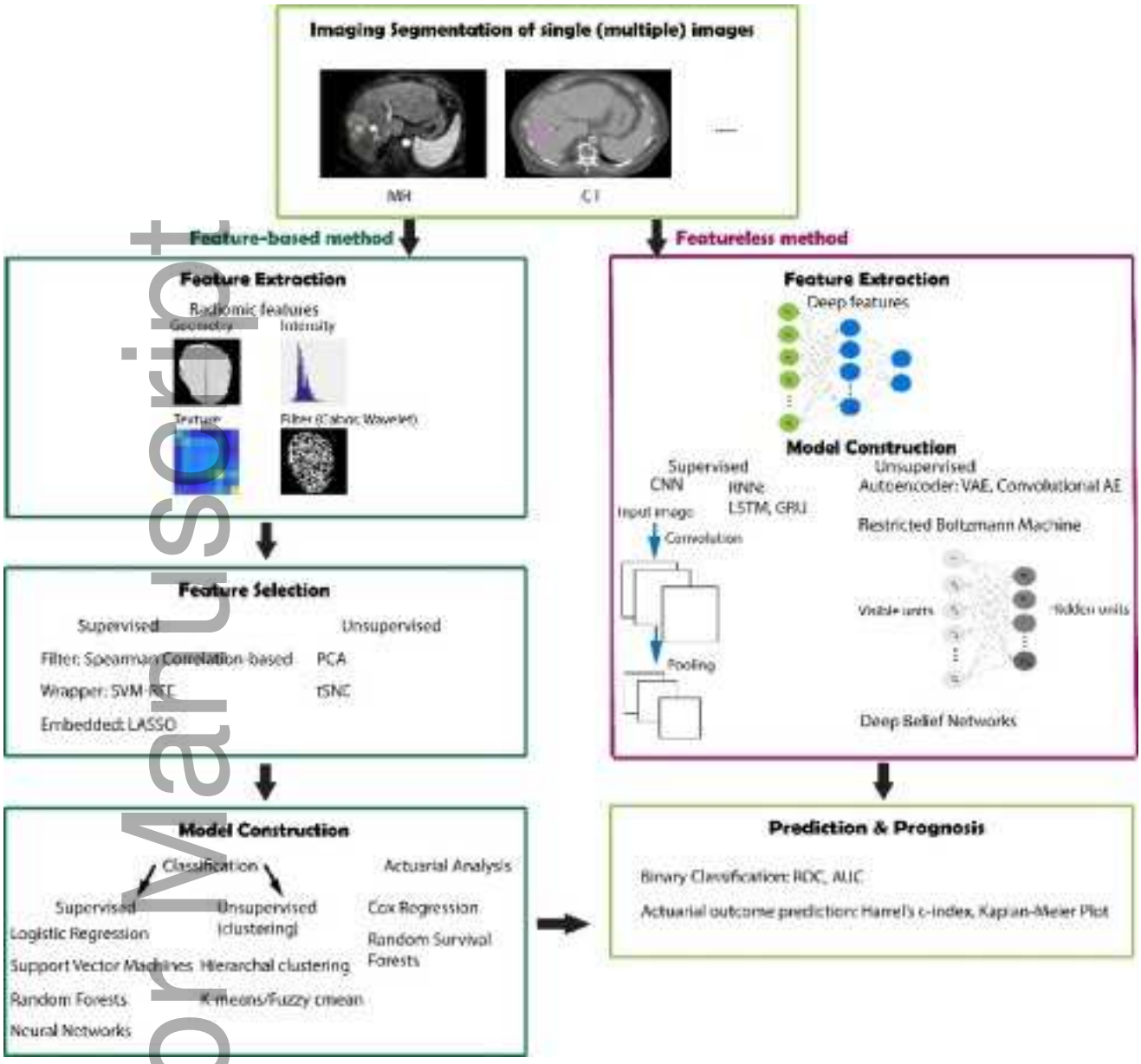
Author Manuscript

This article is protected by copyright. All rights reserved

Comparison to conventional metrics	Performance of radiomics-based models is compared against conventional metrics such as tumor volume and clinical variables (e.g., staging) in order to evaluate the added value of radiomics (e.g., by assessing the significance of AUC increase calculated with the DeLong test).
Multivariable analysis with non- radiomics variables	Multivariable analysis integrates variables other than radiomics features (e.g., clinical information, demographic data, panomics, etc.).
CLINICAL IMPLICATIONS	
Biological correlate	Assessment of the relationship between macroscopic tumor phenotype(s) described with radiomics and the underlying microscopic tumor biology.
Potential clinical application	The study discusses the current and potential application(s) of proposed radiomics-based models in the clinical setting.
MATERIAL AVAILABILITY	
Open data	Imaging data, tumor ROI and clinical information are made available.
Open code	All software code related to computation of features, statistical analysis and machine learning, and allowing to exactly reproduce results, is open source. This code package is ideally shared in the form of easy-to-run organized scripts pointing to other relevant pieces of code, along with useful sets of instructions.
Open models	Complete models are available, including model parameters and cut-off values.

Author Manuscript

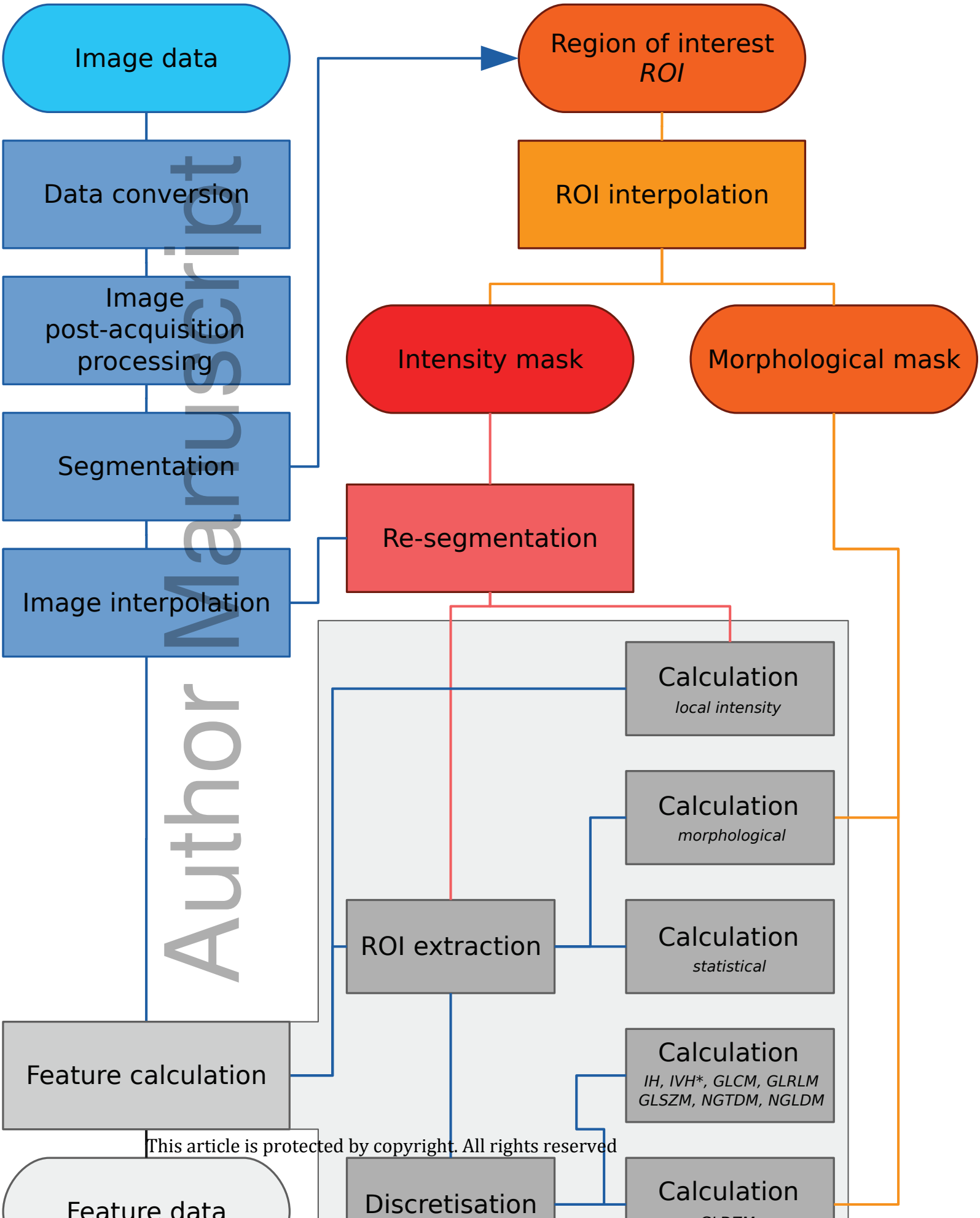
This article is protected by copyright. All rights reserved



mp_13678_f1.tif

Author Manuscript

This article is protected by copyright. All rights reserved



Author Manuscript

This article is protected by copyright. All rights reserved