International Journal of Radiation Oncology biology • physics

www.redjournal.org

Critical Review

NCTN Assessment on Current Applications of Radiomics in Oncology

Ke Nie, PhD,* Hania Al-Hallaq, PhD,† X. Allen Li, PhD,‡ Stanley H. Benedict, PhD,§ Jason W. Sohn, PhD,‖ Jean M. Moran, PhD,¶ Yong Fan, PhD,# Mi Huang, PhD,** Michael V. Knopp, MD,†† Jeff M. Michalski, MD,‡‡ James Monroe, PhD,§§ Ceferino Obcemea, PhD,‖‖ Christina I. Tsien, MD,‡‡ Timothy Solberg, PhD,¶¶ Jackie Wu, PhD,## Ping Xia, PhD,*** Ying Xiao, PhD,** and Issam El Naqa, PhD†

**Department of Radiation Oncology, Rutgers Cancer Institute of New Jersey, Rutgers—Robert Wood Johnson Medical School, New Brunswick, New Jersey; †Department of Radiation and Cellular Oncology, University of Chicago, Chicago, Illinois; ‡Department of Radiation Oncology, Medical College of Wisconsin, Milwaukee, Wisconsin; §Department of Radiation Oncology, University of California-Davis, Sacramento, California; ‖Department of Radiation Oncology, Allegheny Health Network, Pittsburgh, Pennsylvania; ¶Department of Radiation Oncology, University of Michigan, Ann Arbor, Michigan; #Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania; **Department of Radiation Oncology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania; ††Division of Imaging Science, Department of Radiology, Ohio State University, Columbus, Ohio; ‡‡Department of Radiation Oncology, Washington University School of Medicine in St. Louis, St. Louis, Missouri; §§Department of Radiation Oncology, St. Anthony's Cancer Center, St. Louis, Missouri; ‖‖Radiation Research Program, National Cancer Institute, Bethesda, Maryland; ¶¶Department of Radiation Oncology, University of California-San Francisco, San Francisco, California; ##Department of Radiation Oncology, Duke University, Durham, North Carolina; and ***Department of Radiation Oncology, Cleveland Clinic, Cleveland, Ohio*

Received Jul 19, 2018. Accepted for publication Jan 23, 2019.

Radiomics is a fast-growing research area based on converting standard-of-care imaging into quantitative minable data and building subsequent predictive models to personalize treatment. Radiomics has been proposed as a study objective in clinical trial concepts and a potential biomarker for stratifying patients across interventional treatment arms. In

Note—An online CME test for this article can be taken at https://academy.astro.org.

Reprint requests to: Ke Nie, PhD, Rutgers Cancer Institute of New Jersey, Rutgers—Robert Wood Johnson Medical School, 195 Little Albany St, New Brunswick, NJ, 08903. Tel: 732-824-3998; E-mail: Kn231@cinj.rutgers.edu

The materials presented represent the expert consensus among a group of NRG Oncology investigators developed specifically for use in NCTN or NCTN-endorsed clinical trials. Authors are not responsible for use in any other context.

Supported by National Cancer Institute grants U24CA180803 (Imaging and Radiation Oncology Core) and U10CA180868 (NRG).

Conflict of interest: none.

Int J Radiation Oncol Biol Phys, Vol. 104, No. 2, pp. 302—315, 2019
0360-3016/$ - see front matter
https://doi.org/10.1016/j.ijrobp.2019.01.087

recognizing the growing importance of radiomics in oncology, a group of medical physicists and clinicians from NRG Oncology reviewed the current status of the field and identified critical issues, providing a general assessment and early recommendations for incorporation in oncology studies. © 2019 Published by Elsevier Inc.

Radiomics and Applications

Radiomics introduction

The concept of relating imaging information to predicting prognosis and therapeutic response traces its roots to the early days of robotics and computer vision in the 1960s, but its systemic application to quantitative imaging analysis dates to the beginning of the 1980s in areas such as computer-aided detection or diagnosis.[1] The application of this approach to biologic markers and therapeutic endpoints started only in the past decade, when the concept of personalized medicine arose after the increasing use of genomics. Some early examples include the investigation of correlations between hepatocellular carcinoma imaging phenotypes with gene expression,[2] and between positron emission tomography (PET)-based features and radiation therapy response.[3]

When initial studies to investigate whether magnetic resonance imaging (MRI)-based measurements of breast cancer volume could accurately assess the response to neoadjuvant chemotherapy proved promising,[4] they were used as a springboard for a series of American College of Radiology Imaging Network (ACRIN) trials. The ACRIN trials investigated the use of serial MRI studies to predict therapeutic response to chemotherapy; this was called the I-SPY TRIAL (Investigation of Serial Studies to Predict Your Therapeutic Response with Imaging And moLecular Analysis).[5]

Since 2010 this field has been formalized with the term "radiomics."[6] The term originates from the word "radio," which refers to radiology, the science of acquiring medical images through the use of radiation (eg, computed tomography [CT], PET, MRI). The suffix "-omics" follows from the notion of the whole, which was first used in the term "genomics" to indicate the entire mapping of human genetics.[7] Currently this process of extraction of massive quantitative information from anatomic/molecular images and relating them to corresponding biologic information and clinical endpoints is an emerging field referred to as "radiomics."[8]

Radiomics applications

With the recent advances in imaging techniques, imaging has extended its role to the entire spectrum of cancer management—from detection and diagnosis to treatment response monitoring and further risk surveillance. The clinical application of radiomics is expected to play important roles in every aspect of cancer management.

Tumor detection and diagnosis

One of the earliest applications for the radiomics-driven method is in tumor detection, with the greatest success in lung and breast imaging. US Food and Drug Administration—approved systems are currently being used in the clinic. A recent study using commercially available clinical tools identified many lung cancers that were initially missed by radiologists in the International Early Lung Cancer Action Program trial.[9]

Currently the trend is moving in a direction called "discovery radiomics."[10,11] Instead of using a predefined radiomics feature set, the image data are directly fed into the discovery engine, where a customized radiomics sequencer is constructed using deep learning architecture such as convolutional neural networks (CNNs); the descriptive radiomics sequencer then can be applied to identify normal or abnormal tissues. In addition to tumor detection, radiomics features have been demonstrated to be helpful in identifying various types of lesions. For example, Li et al suggested that mammographic images contain computer-extractable information, which may distinguish between *BRCA1/2* gene mutation carriers and noncarriers.[12] Grimm et al also observed that imaging features from dynamic contrast-enhanced MRI were strongly associated with luminal A and luminal B hormone receptor—positive molecular subtypes.[13] Overall, early detection and identification of tumors could be useful for better stratification of patients and identification of subsequent treatment options.

Treatment outcome prediction for decision-making support

Significant interest in using radiomics for early prediction of treatment response has emerged recently. In predicting pathologic complete response after neoadjuvant chemoradiation for locally advanced rectal cancer, Nie et al showed improved prognostic values could be achieved using a voxelized radiomics analysis approach over conventional imaging metrics.[14] Zhang et al identified MRI-based radiomics as a prognostic factor for progression-free survival in patients with nasopharyngeal carcinoma. The prediction power significantly outweighed that of traditional TNM staging.[15] Although TNM staging is the cornerstone for treatment decision making, it is typically assessed based on gross anatomy information, not reflecting the intratumor heterogeneity, whereas the radiomics approach can characterize intratumor heterogeneity noninvasively and thus add incremental value to clinical information in assessing treatment outcome.

Radiomics features from CT, PET/CT, and cone beam CT have also shown predictive value for response to treatment.[16-24] Investigators analyzed the daily noncontrast

CT scans, acquired during routine image-guided radiation therapy using in-room CT, from patients with head and neck,[25] lung,[26] and pancreatic cancers.[27] They reported that radiation can induce patient-specific changes in CT texture features and that these changes can be detected in the early phase of radiation therapy. Ohri et al showed that pretreatment metabolic tumor volume and heterogeneity textural metrics on PET/CT can be good prognostic factors for patients with locally advanced non-small cell lung cancer treated with chemoradiation therapy based on data from the ACRIN 6668/RTOG 0235 study.[28] In addition, Buizza et al showed that the longitudinal temporal and spatial changes from PET/CT imaging could improve early survival prediction for chemoradiation treatment.[29] Radiomics has also been used to predict radiation-induced normal tissue toxicities, such as radiation pneumonitis[21] or xerostomia.[30,31] These results suggest that a radiomics-based signature may emerge as an accepted imaging biomarker for predicting therapeutic outcome and for improving decision support in cancer treatment.

Risk assessment

Radiomics has also been extended to risk surveillance in several cancers. Liu et al investigated the association between imaging features and low-grade glioma-related epilepsy and proposed a radiomics-based model for the prediction of associated risk.[32] Similarly, it is well known that mammographic density is an independent risk factor, and radiomics may provide much more information than breast density.[33-35] Li et al investigated breast parenchymal patterns in mammographic images in 456 patients (53 *BRCA1/2* gene carriers, 75 with unilateral cancer, and 328 with low-risk disease).[12] They demonstrated that women at high risk tend to have dense breasts with coarse and low-contrast texture patterns. Haberle et al performed a case-control study with 864 cases versus 418 controls.[36] Of the 470 radiomics features explored, 46 remained in the final risk model; the radiomics model outperformed the conventional risk model with mammographic density. These studies may promote future breast cancer prevention trials to investigate the role of radiomics to measure breast tissue composition in individual women for personalized risk management.

Radiogenomics

In radiogenomics, the radiomics phenotype is correlated with a genomic profile. The hypothesis is that imaging may provide insight into tumor phenotypes that are driven by the heterogeneity of the genetic evolution. Radiogenomics is a very young field owing to the lack of data consisting of both imaging and genomic measurements on the same set of tumors.[37-39]

Recent studies in brain tumors,[37,39,40] lung cancer,[28,41,42] and breast cancer[13,43] suggest value for radiogenomics. National shared databases, such as the Cancer Imaging Archive and the Cancer Genome Atlas (TCGA), provide researchers opportunities to explore this field. Using MRI scans from the Cancer Imaging Archive and clinical, histopathologic, and genomic data from the TCGA, Li et al investigated the relationship between MRI phenotypes and multigene assays, including Oncotype DX, (Genomic Health, Redwood City, CA), MammaPrint (Agendia, Irvine, CA), and PAM50 (NanoString Technologies, Seattle, WA).[43] Multiple linear regression analyses demonstrated significant associations between radiomics signatures and multigene assay recurrence scores. Zinn et al identified an association between high T2 fluid-attenuated inversion recovery volumes, upregulation of periostin (*POSTN*), and downregulation of miR-219 using data from the TCGA.[40] They noted high levels of *POSTN* were associated with mesenchymal tumors and shorter survival and further concluded that this approach may be valuable for identifying new targets for molecular inhibition or future therapies.

Although numerous radiomics features can be extracted from medical images, the method by which tumor pathophysiologic processes give rise to imaging phenotypes remains unclear. More studies are required to confirm these associations to further elucidate the biologic meaning of the radiomics features.

Radiomics Processes and Components

One goal of radiomics is to convert images into data that can be mined using high-throughput computing. The process follows several general steps, as shown in Figure 1: (1) image acquisition, (2) region of interest (ROI) identification and segmentation, (3) quantitative image feature extraction, and (4) data mining and informatics analysis.

Image acquisition

The first step of image-based radiomics phenotyping involves image acquisition. Different image acquisitions provide different, and often complementary, information. For example, size- or volume-based analysis can be obtained using anatomic MRI or CT scans. Measurements of perfusion can be determined from a series of dynamic contrast-enhanced MRI or CT acquisitions. Functional MRI, such as diffusion-weighted imaging, can be used for tissue microcirculation and cellularity evaluation.[14] Metabolic changes, such as rate of glucose metabolism, can be measured using fluorodeoxyglucose PET.[44] Emerging functional and molecular imaging methods that are increasingly used in clinical trials may offer additional biomarkers. Human experts can help guide the choice of the imaging modalities tailored to the disease site of interest, clinical endpoints, and potential treatment options.

Historically, imaging devices are designed for subjective interpretation of images, allowing clinicians to identify, for example, the presence and the location of a lesion. Subsequent technical innovation has largely focused on improving imaging quality, shortening scanning time, or integrating with treatment machines. Conversely, devices are not designed primarily to provide quantitative

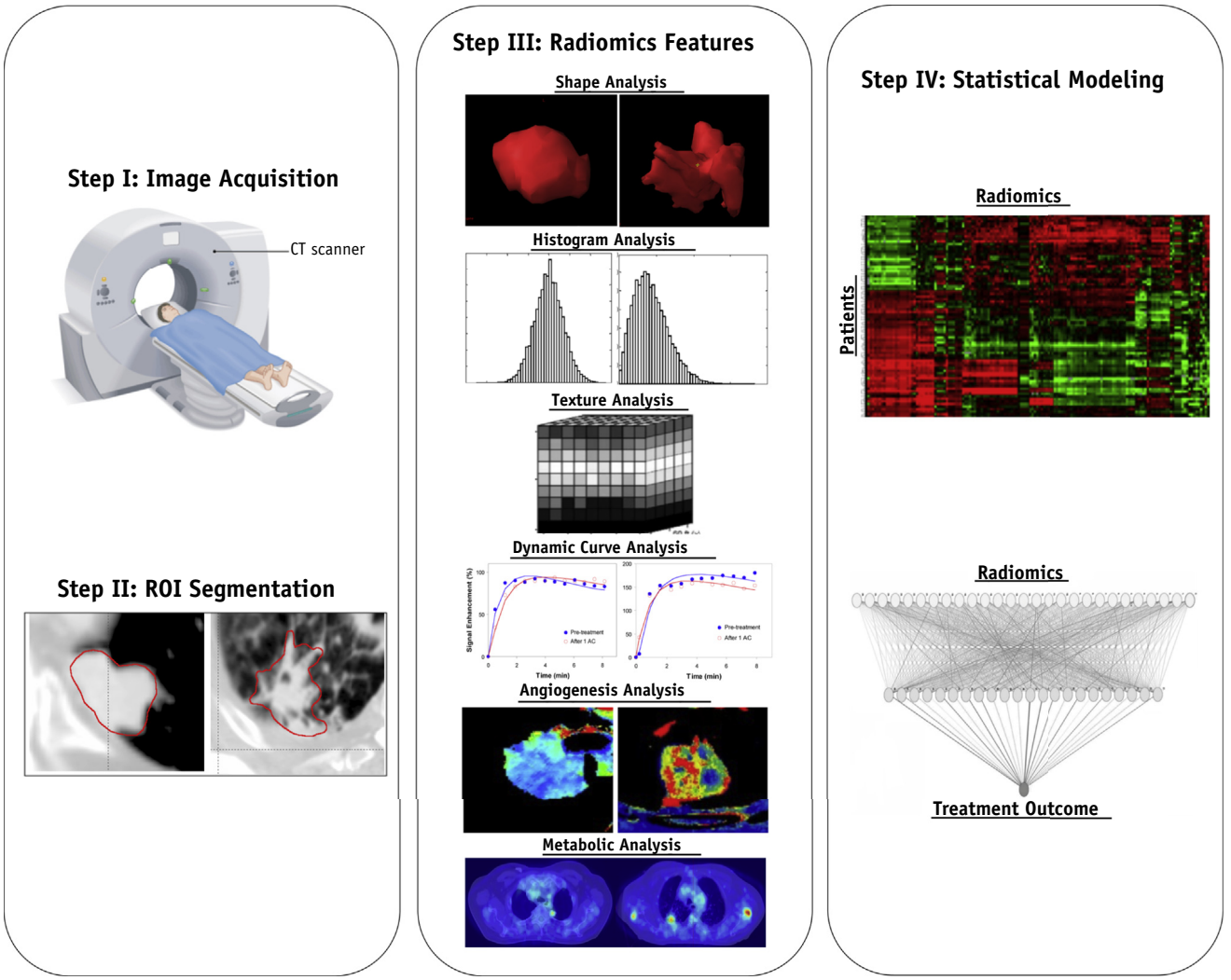

Fig. 1. The general radiomics study workflow. Step 2: image acquisition. Step 2: region of interest identification and segmentation. Step 3: quantitative image feature extraction. Step 4: data mining and informatics analysis.

measurements in a reproducible manner. The standardization of imaging acquisition protocols is typically lacking, and wide variations in reconstruction or acquisition parameters can exist. Zhao et al studied a repeat CT dataset from patients with lung cancer and concluded that smooth and sharp reconstruction algorithms should not be used interchangeably.[45] Galavis et al assessed the variability of radiomics features extracted from PET owing to different acquisition modes, reconstruction algorithms, postfiltering, and iteration numbers. Of 50 features, 40 were shown to have substantial variability: up to 30%.[46] Because of the gradient strength of the scanner, pulse sequence used, method of contrast agent administration, k-space trajectory sampling, and other factors, results from MRI can vary more significantly.[47-49] Because the quality of the radiomics data depends on the reliability of the acquisition protocols used in clinical centers, the impact of these variations on the stability of radiomics features needs to be thoroughly investigated and understood in future studies.

ROI identification and segmentation

Defining ROIs is a fundamental task within the practice of oncology. In radiology, human experts identify the presence, location, and size of the suspicious areas for diagnosis, staging, or response assessment. In radiation oncology, the human experts must identify the tumor extent for treatment and organs at risk for radiation sparing. Manual outlining by experienced radiologists or radiation oncologists is often treated as the gold standard, but it is labor intensive and has high interoperator or even intraoperator variabilities. ROIs may be contoured more consistently using semiautomated or fully automated methods, such as thresholding, region growing, classifiers, clustering, Markov random field models, artificial neural networks, deformable models, and atlas-guided approaches.[50,51] Although full automation may present a new opportunity for standardized segmentation methods, challenges related to complex anatomy or areas with low

soft-tissue contrast persist, and manual correction of contours by an experienced physician is often required. A rather new idea to avoid segmentation pitfalls is the use of a "digital biopsy," which samples rather than segments the ROI. This approach has been recently applied to sampling CT lung nodules.[52] More recently, advanced machine learning–based algorithms have been applied for image segmentation or sampling.[53,54] There are also several large initiatives aimed at developing automatic segmentation solutions using deep learning. These include Google's DeepMind,[55] Microsoft's Project InnerEye,[56] Mirada's DLCExpert,[57] and the Grand Challenges in Biomedical Image Analysis.[58] These automated segmentation tools have been shown to improve efficiency of structure set generation, particularly for organs at risk. In the near future, deep learning–based segmentation tools may be robust enough for routine radiomics applications.

Radiomics features

The core of radiomics is the extraction of high-dimensional feature sets to quantify images. The features extracted from images can be divided into static (ie, a snapshot of enhancement at one point in time) or dynamic (ie, time-variant) categories according to the acquisition protocol used at the time of scanning.[59]

Static image features

Several types of static image features can be applied to radiomics studies, including morphology- and intensity-based features. Morphology-based features are used to capture 3-dimensional (3D) shape characteristics, such as volume and surface area, and sphericity, which quantifies how closely a 3D volume resembles a sphere.[60] Higher sphericity indicates a round shape, whereas a lower value indicates an irregular or elongated shape. Intensity-based features are used to quantify the gray-level distribution inside the ROI. Examples of first-order intensity-based features include mean, standard deviation, percentiles, kurtosis, and skewness. They are used to characterize the overall intensity variability, whereas second-order intensity-based features, also referred to as texture features, look into the local distribution. Example metrics include the gray-level co-occurrence matrices (GLCMs),[61] gray-level run length matrices (GLRLMs),[62] and gray-level size zone matrices.[63] GLCMs indicate the probability of observing a pair of values in voxels at a given distance in a given direction.[61] GLRLMs measure the number of consecutive voxels with the same value aligned in a given direction,[62] and gray-level size zone matrices reflect the number of neighboring voxels with the same value.[63] Higher-order intensity values may be achieved using image transformations (eg, Laplacian or Gaussian)[64,65] with different filter grids (e.g. Laws filters),[66] which highlight edge structures, or by using wavelet composition,[67] which characterize sharp transitions in the intensity frequency spectrum.

Dynamic image features

Pharmacokinetic modeling is typically used to quantify the dynamic behavior of a contrast agent or other tracer within a region (which can be one or more voxels), In general, pharmacokinetic modeling considers tissue concentration as a convolution between the arterial input function and the residual function for the decay of contrast agent inside the ROI. The intravascular and interstitial space can be modeled under different assumptions. The most widely used kinetic model, the Toft model, for example, assumes instantaneous mixing of contrast upon arrival in the intravascular and interstitial space, whereas the extended Toft model considers a delay effect of the tissue concentration transferred from the artery.[68] The adiabatic tissue homogeneity model is motivated by the fact that the concentration of contrast agent in the extravascular distribution volume changes slowly relative to that in the intravascular space.[69] Thus, the model assumes that there exists a finite transit time for contrast solutes to travel from the arterial to the venous phase.

The Patlak model is a linearization of irreversible compartment models in an equilibrium state wherein the tracers flow into the tissue without leaving.[70] The Logan model is also a linearization of the reversible compartment model in the equilibrium states wherein the tracers can move freely back into the plasma.[70] Another approach is to directly fit the residual function using deconvolution without making any additional model assumptions.[71] Typically derived dynamic image features include regional blood flow, regional blood volume, mean transit time, extraction fraction, permeability surface area product, and most frequently volume transfer constant (Ktrans) and extravascular extracellular volume (v_e). With an increasing emphasis on imaging of the tumor microenvironment, dynamic contrast-enhanced CT/MRI and fluorodeoxyglucose PET have evolved as important functional techniques in this setting.

Overall, the current radiomics pipeline typically incorporates thousands of extracted radiomics features, and these are expected to further widen as the field continues to evolve.

Analytical tools

As in many other "-omics" fields, the number of input variables often far exceeds the number of patients. To reduce the probability of false-positive results, feature selection or dimension reduction is often needed, and filter-based score ranking approaches, such as Wilcoxon, χ^2, and principal component analysis, are typically used.[72] This can be carried out using either univariate methods, as the scoring criterion depends only on the feature relevancy, or multivariate methods using a weighted sum to maximize relevancy and minimize redundancy.[72,73] Feature selection can also be combined with feature classification into a single model; examples include least absolute shrinkage and election operator[74] and elastic net.[75]

Once a feature set is obtained, a data-driven model can be constructed. These models include supervised and

nonsupervised approaches.[76] Unsupervised analysis does not provide an outcome variable but rather summary information of the data. The most frequently used graphic display is a heat map, which simultaneously reveals cluster structures in a data matrix.[77] Supervised analysis, in contrast, creates models that attempt to separate the data with respect to a treatment outcome, such as responders versus nonresponders. Typical classification methods include conventional logistic regression or more advanced machine learning techniques.[78,79]

Outcome modeling by logistic regression

Logistic regression is a common tool for multimetric modeling. A logit transformation is used as follows:

$$f(x_i)=\frac{e^{g(x_i)}}{1+e^{g(x_i)}},\ i=1,\ldots,n, \qquad (1)$$

where n is the number of cases (patients), $\mathbf{x}_i$ is a vector of the input variable values (ie, image features) used to predict $f(\mathbf{x}_i)$ for outcome y_i (eg, tumor control or toxicity) of the i*th* patient,

$$g(x_i)=\beta_0+\sum_{j=1}^{d}\beta_j x_{ij}, i=1,\ldots,n, j=1,\ldots,d, \qquad (2)$$

where d is the number of model variables, and the βs are the set of model coefficients determined by maximizing the probability that the data gave rise to the observations. This gives a linear combination of selected features with coefficients of respective weightings to the outcome, providing an intuitive tool for clinicians to interpret the associations between selected variables and the outcome.

Outcome modeling by machine learning

A wide class of artificial intelligence techniques (eg, neural networks, decision trees, support vector machine) can provide a nonlinear association of input variables to the outcome.[80,81] Indeed, prognostic biomarkers developed using these machine learning methods have increased performance when compared with conventional statistical methods.[78,82,83] Recently deep learning algorithms such as CNNs have achieved breakthrough prediction power in a variety of medical studies, including detection of lung nodules on CT scans[84-86] and detection of breast cancer on mammograms.[87,88] A comparison in mortality prediction from chest CT between a deep learning framework and a standard framework with radiomics features showed increased accuracy with CNN-based classification.[89] Multitask learning is expected to help provide a degree of interpretation for deep learning approaches.[76,81-90] Given enough high-quality data (text and images), it is expected that the role of CNNs will continue to expand in medicine and quantitative imaging. Despite these advances, however, concerted efforts are needed to promote detailed understanding of these approaches, including the relationship between dataset sizes, possible confounders, and performance of outcome prediction.

Quality Assurance of the Images and Methodologies

Despite the promise that radiomics may hold for precision medicine, there are significant concerns regarding the lack of reproducibility in results within and across modalities and among multiple institutions. In this section, the challenges associated with the clinical translation of radiomics are highlighted, and recommendations are provided for application in National Clinical Trials Network (NCTN) clinical trials.

Standardization of image acquisition parameters

In recent years, the field has striven to improve standardization by defining standard acquisition protocols. National efforts have been led by the Quantitative Imaging Network initiated by the National Cancer Institute (NCI),[91,92] the Radiological Society of North of America, the Quantitative Imaging Biomarkers Alliance,[93,94] and others.[95,96] The National Cancer Institute—Centers for Quantitative Image Excellence project was initiated in 2010, and the NCTN is a key focus for this effort.[96,97] The Centers for Quantitative Image Excellence provide PET/CT and MRI phantoms and protocols for site qualification, and the Quantitative Imaging Biomarkers Alliance provides consensus "profiles" on the measurement accuracy of quantitative imaging biomarkers and the requirement/procedures needed to achieve this level of accuracy.[98]

Despite the progress made by these groups, there are still no universal acquisition protocols for any imaging modality in clinical practice. For studies involved with radiomics in NTCN trials, therefore, we recommend the following:

1. A comprehensive description of the image acquisition parameters should be documented, including manufacturer, model, types of images (eg, CT, MRI, contrast-enhanced); contrast agents; image acquisition parameters (eg, slice by slice or 3D acquisition, MRI magnetic field/repetition time/echo time/flip angle, CT tube current/voltage, axial or helical mode); reconstruction package; software version; image resolution; signal-to-noise ratio, and management of motion artifact.

2. If a clinical trial is being conducted in institutions with the same scanners, the same scanning protocols should be strictly followed. Comparison across institutions with different scanners may be difficult. It is suggested, when possible, to use each patient as his or her own control and to use the delta changes instead of the absolute value, or other corrections means for such variability including accounting for contrast.

3. Direct measurement from scans with contrast should be used cautiously because uptake and the time from injection to imaging will differ and cause large variability. A practical strategy is to control the normal tissue ROIs

as a baseline. For instance, on an individual basis, the average background of parenchyma or muscle can be used to normalize breast scans.[99,100]

4. Radiomics feature(s) that are less dependent on variations in image acquisition protocol and/or platform should be used in the final model.

Standardization of image preprocessing/postprocessing

After image acquisition, there can still be a large range of voxel intensities and image noise; therefore, filtering procedures may be needed to enhance the signal and reduce the unwanted noise.[101] Fave et al tested the effect of different image preprocessing filters, such as bit-depth resampling and smoothing filters, on radiomics features, and concluded that the correlation of extracted radiomics features with clinical outcome changes with different filters.[102] The impact of noise, which directly affects intensity and GLCM features, has also been studied.[103] In addition, images obtained on 2 different scanners may result in different pixel values because of different detector materials, image resolution, or acquisition techniques.[76] A very recent study by Reuzé et al showed GLCM entropy values were higher on a scanner equipped with time-of-flight capabilities, leading to very different cutoff values for predicting recurrence.[76] Therefore, some postimage acquisition processing such as filtering and normalization is necessary.

To reduce the impact of noise, the signal-to-noise ratio from all acquired images should be well controlled. A smoothing filtering procedure may be used so that all images come close to a target spatial resolution value.[104-106] A consequence of this approach is that images with the highest initial resolution will be degraded, which is especially adverse for analysis of image texture. CT images without contrast may be used directly if geometry distortions are calibrated properly and consistency in terms of Hounsfield units can be obtained. MRI scans in general are susceptible to different distortions. A typical correction procedure would follow normalization before radiomics analysis.[107] PET images tend to have more noise than other imaging modalities used in radiomics analysis; a typical preprocessing procedure to correction procedure has been to use Poisson-to-Gaussian conversion (a root square transform).[108]

Additional procedures include the discretization signal intensities into finite intervals for intensity-based feature analysis, using either absolute (fixed bin size) or relative (fixed number of bins) discretization.[109] The choice of method is important because the extracted features will vary. Absolute discretization shows better repeatability and lower sensitivities to changes and is not volume dependent.[76,109] Thus, absolute discretization with fixed bin size should be adopted for intensity-based radiomics analysis. Each discretization method, however, has advantages and drawbacks and can lead to substantially different results.[76]

Overall, effects of these pre-/postprocessing techniques on variations in radiomics analysis remain an open area. Although specific recommendations are being worked out by the Image Biomarker Standardisation initiative (IBSI) group[110] and others, methods to avoid more general pitfalls include the following:

- Inclusion of a detailed description of filtering technique, noise reduction technique, intensity correction, intensity discretization, and bitdepth resampling.

- Use of radiomics feature(s) that are less dependent on variations in image pre-/postprocessing techniques.

Reproducibility of radiomics features

Differences in segmentation methods are likely to bias the stability of shape metrics and intensity-based features. Parmar et al performed a stability analysis based on an interobserver study.[111] Fifty-six radiomics features, quantifying shape, intensity, and texture, were extracted from lung CT images. These authors showed that for manual delineation among 5 experienced operators, only 52% of these features had high reproducibility compared with 88% based on semiautomatic segmentation.

Moreover, contours of ROIs are typically stored in 2 image formats: directly as voxels (eg, NIFTI [Neuroimaging Informatics Technology Initiative]), or as (x, y, z) coordinates (eg, Digital Imaging and Communications in Medicine—Radiation Therapy Structure Set). We need to determine which voxel centers lie within the space enclosed by the contour polygon. Thus, different interpolation and partial-volume fraction threshold would also affect the contour-based feature calculation. Researchers showed contour data—rendered volumes exhibited large variations (up to 20%) across the commercialized stereotactic radiosurgery platforms using both the phantom and patient cases.[112]

The methodologies to calculate radiomics feature can vary as well. Radiomics features can be obtained from either a 2D image slice[13,113] or from a reconstructed 3D volume.[114] As another example, GLCM texture analysis can be calculated either by averaging the values of the matrices computed for 13 distinct directions or a single matrix that accounts for tumor co-occurrence information in all 13 directions.[102,115] In addition, features computed from different matrices can have the same name.[116] For example, entropy can be computed either from a histogram of intensities or from the GLCM matrix accounting for spatial similarity.[117] Thus the impact of different feature identification methods used in radiomics also needs to be carefully studied.

Furthermore, the variability of radiomics indices has been found to be highly feature dependent. Recent studies on mammography datasets reported that robust features were those that described spatial patterns rather than directionality or image intensity.[118] Deformable registration

of CT lung data was shown to alter underlying texture, but certain features that were robust to registration effects could still be identified.[21] Consensus has been reached that first-order and shape features in both CT and cone beam CT are generally more repeatable than texture or higher-order features.[119-121] First-order statistically derived features from a standard uptake value histogram are generally robust with respect to segmentation, whereas texture features consistently showed greater sensitivity to segmentation differences.[111,119,122]

Regarding the computation of radiomics features, we recommend the following:

- Unambiguous definitions of each radiomics feature should be provided and evaluated.
- If contouring is involved, describe how ROIs are delineated in the image. Specify if segmentation is performed manually, semiautomatically, or automatically and by how many users/experts, as well as how consensus has been reached. The reproducibility of the radiomics features based on segmentations by multiple observers needs to be assessed.
- The stability and accuracy of features should be confirmed in terms of calculation algorithms, such as interpolation criteria to include or exclude voxels from an ROI mask, 2D/3D calculations, or by use of features that do not require accurate segmentation.
- Additional suggestions include comparison of the means and standard deviations of these results and their correlations through the use of test−retest scans with physical phantom and/or patient studies during the developmental phase (ie, scanned on the same scanner but repositioned between scans). Within the radiation oncology workflow, there are situations that could mimic the test−retest scenario; one example is a 4D CT scan if each phase is treated as a separate CT scan. A good practice policy is to eliminate features that prove to be unreliable in the test−retest. To this end, several datasets are publically available. Of note, the Reference Image Database to Evaluate Therapy Response dataset allows validation of results in the same set of patients with 2 scans taken 15 minutes apart.[123] Investigators are encouraged to develop and share additional shared test−retest datasets.
- Again, radiomics feature(s) that are less dependent on the previous variabilities should be used to build the final model.

Radiomics phantoms

To help in addressing some of the aforementioned challenges, the design of radiomics-specific phantoms is now an active area of research.[124-126] One example is the Credence Cartridge Radiomics (CCR) phantom, which was designed for use in studies of texture feature robustness.[125] Using the CCR phantom, Mackin et al investigated the interscanner variability of the calculated texture features, looking for clustering effects related to the scanner manufacturer and CT acquisition parameters.[125] They subsequently developed a correction technique to reduce or eliminate the variability in radiomics features owing to differences in image pixel size based on CT images using the physical CCR phantom.[127]

Digital phantoms such as the Zobel and NCAT phantoms have long been used in image processing applications.[128] It is unclear, however, whether these phantoms are suited for radiomics, particularly for texture analysis. An alternative approach is to use standardized patterns such as the Brodatz textures,[129] although the accompanying signatures are not necessarily clinically relevant. Alternately, simulated images can be used to determine optimal parameters for feature analysis. McGurk et al augmented PET images from 30 patients with soft tissue sarcoma by varying the extent of axial data combined per slice ("span").[130] Simulated T1-weighted and T2-weighted MRI scans were acquired by varying the repetition time and echo time in a spin-echo pulse sequence, respectively. The impact of PET and MRI acquisition parameter variation on individual textures was investigated to assess the global response and the predictive properties of a texture-based model. The results suggested that such a process is feasible for identifying an optimal set of image acquisition parameters to improve prediction performance.[130]

Overall, these radiomics phantoms may be helpful in assessing the interscanner and intrascanner variabilities, and thus protocols for regular phantom quality assurance may be worth developing, similar to dosimetric study in current radiation therapy quality assurance programs, to monitor interscan and intervendor variability of image-derived features. In addition, phantom studies will also be useful in optimizing imaging protocols and image pre-/postprocessing techniques that allow for reliable radiomics characterization. Ultimately, specific acquisition protocols optimized to generate superior radiomics measurements for a given clinical problem should be developed and standardized and thereafter validated using clinical scanners.

Robustness of modeling with radiomics

In the current iteration of radiomics, image features have to be extracted with high throughput, putting a premium on statistical modeling and machine learning algorithm development. Currently there is no consensus on which feature selection or learning methods should be used. Parmar et al carried out a comparative study using 12 machine learning algorithms combined with 14 feature selection methods for radiomics-based prediction of 2-year survival of patients with lung cancer with radical radiation therapy alone or with chemotherapy.[78] They concluded the 30%

variation was observed for different classification methods but not the feature selection algorithm. Given the state of the art, an analysis of the impact of different learning approaches on the stability and robustness of the proposed models is needed.

The general understanding is that simpler models involving few radiomics features are more robust. To avoid overfitting, a reasonable rule of thumb with feature selection or dimension reduction is that 10 positive samples (patients) are needed for each feature selected for binary classification according to the Harrell guidelines.[131,132] For example, radiomics analysis can be performed for 100 positive events, which will result in no more than 10 features selected for the final prediction model.

Recently deep learning has emerged as a productive force across many health care disciplines, especially in diagnostic radiology and pathology. However, there is a significant mismatch between the perceived capabilities of artificial intelligence compared with actual capabilities in present imaging studies.[133,134] The scarce availability of high-quality data, as well as the lack of standardized processes among institutions, are the key barriers. In addition, interpretation of a nonlinear relationship between input variables versus outcome, as given by many current "black box" machine learning models, requires high-level expertise in this field. Additional complementary information, such as patient demographics, genomics, histology, and biomarkers, may be helpful in interpreting the results.

Nevertheless, robust models should address the following points:

- The need to assess the overfitting risk by using cross-validation methods during the classification step and controlled by dimension reduction methods.
- The need to be built on rigorous training, testing, and validation. Estimation of predictive performance in single-institution cohorts should include multiple-folded repeated cross-validation to minimize the risk of overfitting; validation with an external dataset is highly recommended.
- The possible need to consider accommodation of clinical information, and with covariates of genomic profiles, histology, biomarkers, patient histories, and so on, to generate clinically understandable and acceptable decisions.

Overview of Commercial and Open Source Radiomics Systems

Commercial systems

Several commercial applications providing radiomics capabilities applicable to cancer clinical trials have been created, including HealthMyne.com,[135] TexRad.com,[136] and Oncoradiomics.com.[137] The commercial entities offer platforms that automate image analysis and clinical interpretation for radiologists and oncologists. Radiomic feature extraction is included as part of image analysis functionality. In addition, intensity, shape, and texture metrics, along with conventional size measurements used in routine image interpretation as Response Evaluation Criteria In Solid Tumors, therapy response monitoring, and cancer screening are generated on approval of structure segmentation. Along with imaging data (CT, MRI, and PET), these systems can import and display electronic health record information via an HL7 interface (Corepoint Health); imported and generated data are stored in a minable patient database that can be examined.

Open source systems

Several open source software packages are capable of performing radiomics analysis. Examples include IBEX (ibex Solutions)[138]; MaZda (Technical Unviersity of Lodz, Institute of Electronics)[139]; CGITA (Lab for Biochemical Information Analysis and Integration, National Cheng Kung University)[140]; pyradiomics (GitHub, Inc.)[141,142]; and CERR (Memorial Sloan Kettering Cancer Center)[143] Because there is no official index of the various open source radiomics packages, this list is not necessarily exhaustive. These open source packages are typically capable of calculating first-order texture, GLCM, and GLRLM features, but not necessarily more sophisticated ones such as fractal features. Certain software packages can calculate texture features in 2D and 3D (IBEX, CGITA, pyradiomics), whereas others are limited to 2D datasets (MaZda). Furthermore, the ability to alter calculation parameters, such as the number of gray levels and directions for GLCMs, is not available in all open source packages (eg, MaZda). Some packages were tailored for certain imaging modalities. For example, IBEX was initially developed for CT data, whereas MaZda was initially developed for MRI studies. As a result, both IBEX and MaZda rescale pixel values in an image to remove any nonnegative pixel values.[144] Algorithmic implementation of features in each package could also vary. Some efforts have been made, such as IBSI,[110] to standardize feature definitions. Some packages such as pyradiomics attempt to adhere to IBSI definitions. Beyond open source software tools, a number of groups have also developed in-house tools for radiomics analyses. One such example is an open source Matlab (MathWorks) that calculates different texture matrices in addition to global metrics. It also supports multivariate model building using a logistic regression model with bootstrapping for order selection.[145] Overall, when selecting an open source package, users should verify that it meets their task-specific need and provides expected results.

The following points are recommended for investigators:

- Investigators should provide access to code and datasets to ensure results can be reproduced. The datasets should include images, segmentations, and clinical information

if available and should be compliant with Health Insurance Portability and Accounting Act regulations.
- Investigators should promote an open source approach to radiomics software as an important step toward independent validation and general dissemination of the approach.

Implementation in Clinical Trials and Recommendations

For radiomics to gain broad acceptance in clinical medicine, the value that the technology brings to clinical assessment and decision making must be affirmed in the context of clinical trials. Ideally, clinical trials that are conducted across diverse institutional environments and that are agnostic to the vendor or technologic platform will provide the most robust validation of the significance of radiomics. High-quality clinical research is resource intensive and consumes time, energy, and resources of physicians, research staff, and patients alike. The recently published roadmap for imaging biomarkers is a notable advancement, showcasing key recommendations for clinical translation of radiomics.[146] The application of these considerations to the role of genomics in clinical trials may also apply to radiomics. Because of the expense of clinical trials, it is crucial that any prospective trials involving radiomics be carefully designed.

- An important clinical question is what type of studies will be best suited to test the value of radiomics. Strategies must be defined and mitigate and/or quantify uncertainties, risk, and cost associated with any potential biomarker in making research or clinical decisions.

- Trials need to be carefully designed considering both the initial and any secondary analyses to ensure that the proper data have been collected to support the study questions. Whether there is value—and how much the value of radiomics would add to the current gold standard clinical measurement—needs to be verified. It is also prudent to incorporate radiomics studies as exploratory aims in the initial phase of the implementation in clinical trials. It should be clearly specified whether radiomics input will be used to stratify or determine patient eligibility or be integrated within the trial to assess the quality of the test.

- For clinical trials, a standard radiomics data format similar to the needs defined by the Advanced Technology Consortium to support clinical trials, one of the main proponents of the digital imaging and communications−radiation therapy standard, may be necessary. Such a standard would need to be developed by researchers, clinicians, and manufacturers together to address the competing needs for robust quantitative evaluation of data. In addition, a guideline for reporting results from radiomics studies should be established. The concept of a radiomics quality score has been proposed as a possible evaluation criterion for radiomics studies.[147] The radiomics quality score contains 16 key components intended to minimize bias and enhance the usefulness of the radiomics models. These recommendations may establish reporting guidelines for future radiomics studies.

- For a given trial, the ability to share a standard acquisition protocol in support of the trial across participating institutions would establish a strong foundation for meaningful radiomics data and analysis.

Summary

Radiomics is an evolving, rapidly growing field with great potential to affect the design of future clinical trials in oncology. As in a 2009 report, however, many genomics-based studies have been published that contained significant analytical errors,[110] and scientists reported that they failed to replicate 47 of 53 landmark studies.[148] These findings raise similar concerns regarding the accountability of the statistical analysis, the transparency of the raw data accessibility, and validation of results in radiomics studies. Because the ultimate goal of radiomics is to build reliable imaging biomarkers to assist clinical decision making, a prospective investigation must be trained, tested, and validated against a completely independent dataset, with a systematic study design with uniform treatment delivery, complete reporting of results, and robust statistical analysis. Protocols for standardizing image acquisition, feature extraction, and analysis will help to streamline the process for clinical trials. This can build on already existing expertise in the community in credentialing and evaluating other metrics (eg, treatment plan delivery, imaging quality assurance) to accelerate the process.

Nevertheless, the promise of radiomics is still quite positive. Medical imaging is redefining its role as a valuable data source for precision medicine in the guise of image-based phenotyping. Furthermore, it could characterize tumor heterogeneity at a macroscopic level, a critical limitation of biopsy-based or current genomic approaches, which could potentially provide important complementary information for precision medicine. The successful introduction of these methods into clinical care will require much additional research to determine how underlying driving biologic phenomena are related to the tumor imaging phenotypes. In the foreseeable future, we expect that the data gleaned from oncology examinations will be converted into quantitative data that will be interfaced with knowledge databases to improve diagnostic and prognostic power for clinical decision support.

References

1. Doi K. Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Comput Med Graph* 2007; 31:198-211.

2. Kuo MD, Gollub J, Sirlin CB, Ooi C, Chen X. Radiogenomic analysis to identify imaging phenotypes associated with drug response gene expression programs in hepatocellular carcinoma. *J Vasc Interv Radiol* 2007;18:821-831.
3. El Naqa I, Grigsby P, Apte A, et al. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern Recognit* 2009;42:1162-1171.
4. Partridge SC, Gibbs JE, Lu Y, et al. MRI measurements of breast tumor volume predict response to neoadjuvant chemotherapy and recurrence-free survival. *AJR Am J Roentgen* 2005;184:1774-1781.
5. Hylton NM, Gatsonis CA, Rosen MA, et al. Neoadjuvant chemotherapy for breast cancer: Functional tumor volume by MR imaging predicts recurrence-free survival—results from the ACRIN 6657/CALGB 150007 I-SPY 1 TRIAL. *Radiology* 2016;279:44-55.
6. Gillies RJ, Anderson AR, Gatenby RA, Morse DL. The biology underlying molecular imaging in oncology: From genome to anatome and back again. *Clin Radiol* 2010;65:517-521.
7. Avanzo M, Stancanello J, El Naqa I. Beyond imaging: The promise of radiomics. *Phys Med* 2017;38:122-139.
8. Lambin P, Rios-Velazquez E, Fau-Leijenaar R, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012;48:441-446.
9. Liang M, Tang W, Xu DM, et al. Low-dose CT screening for lung cancer: Computer-aided detection of missed lung cancers. *Radiology* 2016;281:279-288.
10. Kumar V, Gu Y, Basu S, et al. Radiomics: The process and the challenges. *Magn Reson Imaging* 2012;30:1234-1248.
11. Shafiee MJ, Chung AG, Khalvati F, Haider MA, Wong A. Discovery radiomics via evolutionary deep radiomic sequencer discovery for pathologically proven lung cancer detection. *J Med Imaging (Bellingham)* 2016;4:041305.
12. Li H, Giger ML, Lan L, Janardanan J, Sennett CA. Comparative analysis of image-based phenotypes of mammographic density and parenchymal patterns in distinguishing between BRCA1/2 cases, unilateral cancer cases, and controls. *J Med Imaging (Bellingham)* 2014;1:031009.
13. Grimm LJ, Zhang J, Mazurowski MA. Computational approach to radiogenomics of breast cancer: Luminal A and luminal B molecular subtypes are associated with imaging features on routine breast MRI extracted using computer vision algorithms. *J Magn Reson Imaging* 2015;42:902-907.
14. Nie K, Shi L, Chen Q, et al. Rectal cancer: Assessment of neoadjuvant chemoradiation outcome based on radiomics of multiparametric MRI. *Clin Cancer Res* 2016;22:5256-5264.
15. Zhang B, Tian J, Dong D, et al. Radiomics features of multiparametric MRI as novel prognostic factors in advanced nasopharyngeal carcinoma. *Clin Cancer Res* 2017;23:4259-4269.
16. Goh V, Ganeshan B, Nathan P, Juttla JK, Vinayan A, Miles KA. Assessment of response to tyrosine kinase inhibitors in metastatic renal cell cancer: CT texture as a predictive biomarker. *Radiology* 2011;261:165-171.
17. Yip C, Davnall F, Kozarski R, et al. Assessment of changes in tumor heterogeneity following neoadjuvant chemotherapy in primary esophageal cancer. *Dis Esophagus* 2015;28:172-179.
18. Aerts HJ, Grossmann P, Tan Y, et al. Defining a radiomic response phenotype: A pilot study using targeted therapy in NSCLC. *Sci Rep* 2016;6:33860.
19. Coroller TP, Grossmann P, Hou Y, et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. *Radiother Oncol* 2015;114:345-350.
20. Lopez CJ, Nagornaya N, Parra NA, et al. Association of radiomics and metabolic tumor volumes in radiation treatment of glioblastoma multiforme. *Int J Radiat Oncol Biol Phys* 2017;97:586-595.
21. Cunliffe AR, Armato SG 3rd, Fei XM, Tuohy RE, Al-Hallaq HA. Lung texture in serial thoracic CT scans: Registration-based methods to compare anatomically matched regions. *Med Phys* 2013;40:061906.
22. Palma DA, van Sornsen de Koste J, Verbakel WF, Vincent A, Senan S. Lung density changes after stereotactic radiotherapy: A quantitative analysis in 50 patients. *Int J Radiat Oncol Biol Phys* 2011;81:974-978.
23. Diot Q, Kavanagh B, Schefter T, Gaspar L, Stuhr K, Miften M. Regional normal lung tissue density changes in patients treated with stereotactic body radiation therapy for lung tumors. *Int J Radiat Oncol Biol Phys* 2012;84:1024-1030.
24. Li H, Galperin-Aizenberg M, Pryma D, Simone CB, Fan Y. Unsupervised machine learning of radiomic features for predicting treatment response and overall survival of early stage non-small cell lung cancer patients treated with stereotactic body radiation therapy. *Radiother Oncol* 2018;129:218-226.
25. Feng M, Yang C, Chen X, et al. Computed tomography number changes observed during computed tomography-guided radiation therapy for head and neck cancer. *Int J Radiat Oncol Biol Phys* 2015;91:1041-1047.
26. Paul J, Yang C, Wu H, et al. Early assessment of treatment responses during radiation therapy for lung cancer using quantitative analysis of daily computed tomography. *Int J Radiat Oncol Biol Phys* 2017;98:463-472.
27. Chen X, Oshima K, Schott D, et al. Assessment of treatment response during chemoradiation therapy for pancreatic cancer based on quantitative radiomic analysis of daily CTs: An exploratory study. *PLoS One* 2017;12:e0178961.
28. Ohri N, Duan F, Snyder BS, et al. Pretreatment 18F-FDG PET textural features in locally advanced non-small cell lung cancer: Secondary analysis of ACRIN 6668/RTOG 0235. *J Nucl Med* 2016;57:842-848.
29. Buizza G, Toma-Dasu I, Lazzeroni M, et al. Early tumor response prediction for lung cancer patients using novel longitudinal pattern features from sequential PET/CT image scans. *Phys Med* 2018;54:21-29.
30. Nakatsugawa M, Cheng Z, Goatman KA, et al. Radiomic analysis of salivary glands and its role for predicting xerostomia in irradiated head and neck cancer patients. *Int J Radiat Oncol Biol Phys* 2016;96(2 suppl):S217.
31. Nardone V, Tini P, Nioche C, et al. Texture analysis as a predictor of radiation-induced xerostomia in head and neck patients undergoing IMRT. *Radiol Med* 2018;123:415-423.
32. Liu Z, Wang Y, Liu X, et al. Radiomics analysis allows for precise prediction of epilepsy in patients with low-grade gliomas. *Neuroimage Clin* 2018;19:271-278.
33. Titus-Ernstoff L, Tosteson AN, Kasales C, et al. Breast cancer risk factors in relation to breast density (United States). *Cancer Causes Control* 2006;17:1281-1290.
34. Boyd NF, Guo H, Martin LJ, et al. Mammographic density and the risk and detection of breast cancer. *N Engl J Med* 2007;356:227-236.
35. Vachon CM, Brandt KR, Ghosh K, et al. Mammographic breast density as a general marker of breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 2007;16:43-49.
36. Häberle L, Wagner F, Fasching PA, et al. Characterizing mammographic images by using generic texture features. *Breast Cancer Res* 2012;14:R59.
37. Gutman DA, Cooper LA, Hwang SN, et al. MR imaging predictors of molecular profile and survival: Multi-institutional study of the TCGA glioblastoma data set. *Radiology* 2013;267:560-569.
38. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than pictures, they are data. *Radiology* 2016;278:563-577.
39. Zinn PO, Mahajan B, Sathyan P, et al. Radiogenomic mapping of edema/cellular invasion MRI-phenotypes in glioblastoma multiforme. *PLoS One* 2011;6:e25451.
40. Hong SJ, Kim TJ, Choi YW, Park JS, Chung JH, Lee KW. Radiogenomic correlation in lung adenocarcinoma with epidermal growth factor receptor mutations: Imaging features and histological subtypes. *Eur Radiol* 2016;26:3660-3668.
41. Hsu JS, Huang MS, Chen CY, et al. Correlation between EGFR mutation status and computed tomography features in patients with advanced pulmonary adenocarcinoma. *J Thorac Imaging* 2014;29:357-363.

42. Liu Y, Kim J, Balagurunathan Y, et al. Radiomic features are associated with EGFR mutation status in lung adenocarcinomas. *Clin Lung Cancer* 2016;17:441-448.e6.
43. Li H, Zhu Y, Burnside ES, et al. MR imaging radiomics signatures for predicting the risk of breast cancer recurrence as given by research versions of MammaPrint, Oncotype DX, and PAM50 gene assays. *Radiology* 2016;281:382-391.
44. Choi NC, Chun TT, Niemierko A, et al. Potential of (18)F-FDG PET toward personalized radiotherapy or chemoradiotherapy in lung cancer. *Eur J Nucl Med Mol Imaging* 2013;40:832-841.
45. Zhao B, Tan Y, Tsai W-Y, et al. Reproducibility of radiomics for deciphering tumor phenotype with imaging. *Sci Rep* 2016;6:23428.
46. Galavis PE, Hollensen C, Jallow N, Paliwal B, Jeraj R. Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters. *Acta Oncol* 2010;49:1012-1016.
47. Bologna M, Corino VDA, Montin E, et al. Assessment of stability and discrimination capacity of radiomic features on apparent diffusion coefficient images. *J Digit Imaging* 2018;31:879-894.
48. Ford J, Dogan N, Young L, Yang F. Quantitative radiomics: Impact of pulse sequence parameter selection on MRI-based textural features of the brain. *Contrast Media Mol Imaging* 2018;2018:1729071.
49. Molina D, Perez-Beteta J, Martinez-Gonzalez A, et al. Lack of robustness of textural measures obtained from 3D brain tumor MRIs impose a need for standardization. *PLoS One* 2017;12:e0178843.
50. Pham DL, Xu C, Prince JL. Current methods in medical image segmentation. *Annu Rev Biomed Eng* 2000;2:315-337.
51. El Naqa I, Yang D, Apte A, et al. Concurrent multimodality image segmentation by active contours for radiotherapy treatment planning. *Med Phys* 2007;34:4738-4749.
52. Echegaray S, Nair V, Kadoch M, et al. A rapid segmentation-insensitive "digital biopsy" method for radiomic feature extraction: Method and pilot study using CT images of non-small cell lung cancer. *Tomography* 2016;2:283-294.
53. Kamnitsas K, Ledig C, Newcombe VFJ, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal* 2017;36:61-78.
54. Birenbaum A, Greenspan H. Multi-view longitudinal CNN for multiple sclerosis lesion segmentation. *Eng Appl Artif Intell* 2017;65:111-118.
55. Google's DeepMind Health. Available at: https://deepmind.com/applied/deepmind-health/. Accessed March 12, 2019.
56. Project InnerEye—Medical Imaging AI to Empower Clinicians. Available at: https://www.microsoft.com/en-us/research/project/medical-image-analysis/. Accessed March 12, 2019.
57. Aljabar P, Gooding MJ. The cutting edge: Delineating contours with deep learning. Oxford, UK: Mirada Medical Ltd; 2017.
58. The Grand Challenges in Biomedical Image Analysis. Available at: https://grand-challenge.org/. Accessed March 12, 2019.
59. El Naqa I. The role of quantitative PET in predicting cancer treatment outcomes. *Clin Transl Imaging* 2014;2:305-320.
60. Wadell H. Volume, shape, and roundness of quartz particles. *J Geol* 1935;43:250-280.
61. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybern* 1973;SMC-3:610-621.
62. Xu D-H, Kurani AS, Furst JD, Raicu DS. Run-length encoding for volumetric texture. In: Proceedings of the 4th IASTED International Conference, Visualization and Imagine Processing, Marbella, Spain.
63. Thibault G, Fertil B, Navarro C, et al. Texture indexes and gray level size zone matrix. Application to cell nuclei classification. Paper presented at: 10th International Conference on Pattern Recognition and Information Processing. May 19-21, 2009; Minsk, Belarus.
64. Wu X, Wang K, Zhang D. Palmprint texture analysis using derivative of Gaussian filters. Paper presented at: 2006 International Conference on Computational Intelligence and Security. November 3-6, 2006; Guangzhou, China.
65. Du H, Jin X, Willis PJ. Two-level joint local laplacian texture filtering. *The Vis Comput* 2016;32:1537-1548.
66. Laws KI. Rapid texture identification. In: *SPIE Image Processing for Missile Guidance*; 1980;376-380.
67. Livens S, Scheunders, van de Wouwer G, Van Duck D. Wavelets for texture analysis, an overview. Papers presented at: 6th International Conference on Image Processing and Its Applications. July 14-17, 1997; Dublin, Ireland.
68. Tofts PS, Brix G, Buckley DL, et al. Estimating kinetic parameters from dynamic contrast-enhanced T(1)-weighted MRI of a diffusable tracer: Standardized quantities and symbols. *J Magn Reson Imaging* 1999;10:223-232.
69. St. Lawrence KS, Lee TY. An adiabatic approximation to the tissue homogeneity model for water exchange in the brain: I. Theoretical derivation. *J Cereb Blood Flow Metab* 1998;18:1365-1377.
70. Niu TY, Yang P, Sun X, et al. Variations of quantitative perfusion measurement on dynamic contrast enhanced CT for colorectal cancer: Implication of standardized image protocol. *Phys Med Biol* 2018;63:1650009.
71. Jerosch-Herold M, Swingen C, Seethamraju RT. Myocardial blood flow quantification with MRI by model-independent deconvolution. *Med Phys* 2002;29:886-897.
72. Chandrashekar G, Sahin F. A survey on feature selection methods. *Comput Electr Eng* 2014;40:16-28.
73. Kohavi R, John GH. Wrappers for feature subset selection. *Artif Intell* 1997;97:273-324.
74. Guyon I, Elisseeff André E. An introduction to variable and feature selection. *J Mach Learn Res* 2003;3:1157-1182.
75. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Series B Methodol* 2005;67:301-320.
76. Reuzé S, Schernberg A, Orlhac F, et al. Radiomics in nuclear medicine applied to radiation therapy: Methods, pitfalls, and challenges. *Int J Radiat Oncol Biol Phys* 2018;102:1117-1142.
77. Wilkinson L, Friendly M. The history of the cluster heat map. *Am Stat* 2009;63:179-184.
78. Parmar C, Grossmann P, Rietveld D, Rietbergen MM, Lambin P, Aerts HJ. Radiomic machine-learning classifiers for prognostic biomarkers of head and neck cancer. *Front Oncol* 2015;5:272.
79. Gao X, Chu C, Li Y, et al. The method and efficacy of support vector machine classifiers based on texture features and multi-resolution histogram from (18)F-FDG PET-CT images for the evaluation of mediastinal lymph nodes in patients with lung cancer. *Eur J Radiol* 2015;84:312-317.
80. Lao J, Chen Y, Li Z-C, et al. A Deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. *Sci Rep* 2017;7:10353.
81. Li Z, Wang Y, Yu J, Guo Y, Cao W. Deep learning based radiomics (DLF) and its usage in noninvasive IDH1 prediction for low grade glioma. *Sci Rep* 2017;7:5467.
82. Wu W, Parmar C, Grossmann P, et al. Exploratory study to identify radiomics classifiers for lung cancer histology. *Front Oncol* 2016;6:71.
83. McLaren CE, Chen WP, Nie K, Su MY. Prediction of malignant breast lesions from MRI features: A comparison of artificial neural network and logistic regression techniques. *Acad Radiol* 2009;16:842-851.
84. Huang X, Shan J, Vaidya V. Lung nodule detection in CT using 3D convolutional neural networks. Paper presented at: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). April 18-21, 2017; Melbourne, Australia.
85. da Silva GLF, Valente TLA, Silva AC, de Paiva AC, Gattass M. Convolutional neural network-based PSO for lung nodule false positive reduction on CT images. *Comput Methods Programs Biomed* 2018;162:109-118.
86. Jin H, Li Z, Tong R, Lin L. A deep 3D residual CNN for false-positive reduction in pulmonary nodule detection. *Med Phys* 2018;45:2097-2107.
87. Sampaio WB, Diniz EM, Silva AC, de Paiva AC, Gattass M. Detection of masses in mammogram images using CNN, geostatistic functions and SVM. *Comput Biol Med* 2011;41:653-664.

88. Charan S, Khan MJ, Khurshid K. Breast cancer detection in mammograms using convolutional neural network. Paper presented at: 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET). March 3-4, 2018; Sukkur, Sindh, Pakistan.
89. Carneiro G, Oakden-Rayner L, Bradley AP, Nascimento J, Palmer L. Automated 5-year mortality prediction using deep learning and radiomics features from chest computed tomography. Paper presented at: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). April 18-21, 2017; Melbourne, Australia.
90. Thompson RF, Valdes G, Fuller CD, et al. Artificial intelligence in radiation oncology: A specialty-wide disruptive transformation? *Radiother Oncol* 2018;129:421-426.
91. Clarke LP, Nordstrom RJ, Zhang H, et al. The Quantitative Imaging Network: NCI's historical perspective and planned goals. *Transl Oncol* 2014;7:1-4.
92. Shankar LK. The clinical evaluation of novel imaging methods for cancer management. *Nat Rev Clin Oncol* 2012;9:738-744.
93. Huang EP, Wang XF, Choudhury KR, et al. Meta-analysis of the technical performance of an imaging procedure: Guidelines and statistical methodology. *Stat Methods Med Res* 2015;24:141-174.
94. Sullivan DC, Obuchowski NA, Kessler LG, et al. Metrology standards for quantitative imaging biomarkers. *Radiology* 2015;277:813-825.
95. National Cancer Institute. Cancer Imaging Program, https://imaging.cancer.gov/. Accessed March 12, 2019.
96. Rosen M, Kinahan PE, Gimpel JF, et al. Performance observations of scanner qualification of NCI-designated cancer centers: Results from the Centers of Quantitative Imaging Excellence (CQIE) program. *Acad. Radiol* 2017;24:232-245.
97. Kurland BF, Gerstner ER, Mountz JM, et al. Promise and pitfalls of quantitative imaging in oncology clinical trials. *Magn Reson Imaging* 2012;30:1301-1312.
98. Quantitative Imaging Biomarkers Alliance. Available at: https://www.rsna.org/QIBA/. Accessed March 12, 2019.
99. Wang J, Kato F, Oyama-Manabe N, et al. Identifying triple-negative breast cancer using background parenchymal enhancement heterogeneity on dynamic contrast-enhanced MRI: A pilot radiomics study. *PLoS One* 2015;10:e0143308.
100. Limkin EJ, Sun R, Dercle L, et al. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. *Ann Oncol* 2017;28:1191-1206.
101. Wells WM. Efficient synthesis of Gaussian filters by cascaded uniform filters. *IEEE Trans Pattern Anal Mach Intell* 1986;8:234-239.
102. Fave X, Cook M, Frederick A, et al. Preliminary investigation into sources of uncertainty in quantitative imaging features. *Comput Medical Imaging Graph* 2015;44:54-61.
103. Fave X, Mackin D, Yang J, et al. Can radiomics features be reproducibly measured from CBCT images for patients with non-small cell lung cancer? *Med Phys* 2015;42:6784-6797.
104. Lasnon C, Majdoub M, Lavigne B, et al. ^{18}F-FDG PET/CT heterogeneity quantification through textural features in the era of harmonisation programs: A focus on lung cancer. *Eur J Nucl Med Mol Imaging* 2016;43:2324-2335.
105. Boellaard R. Standards for PET image acquisition and quantitative data analysis. *J Nucl Med* 2009;50(suppl 1):11S-20S.
106. Quak E, Le Roux P-Y, Hofman MS, et al. Harmonizing FDG PET quantification while maintaining optimal lesion detection: Prospective multicentre validation in 517 oncology patients. *Eur J Nucl Med Mol Imaging* 2015;42:2072-2082.
107. Collewet G, Strzelecki M, Mariette F. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. *Magn Reson Imaging* 2004;22:81-91.
108. Nyflot MJ, Yang F, Byrd D, Bowen SR, Sandison GA, Kinahan PE. Quantitative radiomics: Impact of stochastic effects on textural feature analysis implies the need for standards. *J Med Imaging* 2015;2:041002.
109. Leijenaar RT, Nalbantov G, Carvalho S, et al. The effect of SUV discretization in quantitative FDG-PET radiomics: The need for standardized methodology in tumor texture analysis. *Sci Rep* 2015;5:11075.
110. Zwanenburg A, Leger S, Vallières M, Löck S, and for the Image Biomarker Standardisation Intiative. Image Biomarker Standardisation Initiative. Available at: arXiv preprint arXiv:161207003; 2016.
111. Parmar C, Rios Velazquez E, Leijenaar R, et al. Robust radiomics feature quantification using semiautomatic volumetric segmentation. *PLoS One* 2014;9:e102107.
112. Ma L, Sahgal A, Nie K, et al. Reliability of contour-based volume calculation for radiosurgery. *J Neurosurg* 2012;117:203-210.
113. Nie K, Chen JH, Chan S, et al. Development of a quantitative method for analysis of breast density based on three-dimensional breast MRI. *Med Phys* 2008;35:5253-5262.
114. Baek HM, Chen JH, Nie K, et al. Predicting pathologic response to neoadjuvant chemotherapy in breast cancer by using MR imaging and quantitative 1H MR spectroscopy. *Radiology* 2009;251:653-662.
115. Hatt M, Majdoub M, Vallieres M, et al. 18F-FDG PET uptake characterization through texture analysis: Investigating the complementary nature of heterogeneity and functional tumor volume in a multi-cancer site patient cohort. *J Nucl Med* 2015;56:38-44.
116. Buvat I, Orlhac F, Soussan M. Tumor texture analysis in PET: Where do we stand? *J Nucl Med* 2015;56:1642-1644.
117. Dercle L, Ammari S, Bateson M, et al. Limits of radiomic-based entropy as a surrogate of tumor heterogeneity: ROI-area, acquisition protocol and tissue site exert substantial influence. *Sci Rep* 2017;7:7952.
118. Mendel KR, Li H, Lan L, et al. Quantitative texture analysis: Robustness of radiomics across two digital mammography manufacturers' systems. *J Med Imaging (Bellingham)* 2018;5:011002.
119. Traverso A, Wee L, Dekker A, Gillies R. Repeatability and reproducibility of radiomic features: A systematic review. *Int J Radiat Oncol Biol Phys* 2018;102:1143-1158.
120. van Timmeren JE, Leijenaar RTH, van Elmpt W, et al. Test-retest data for radiomics feature stability analysis: Generalizable or study-specific? *Tomography* 2016;2:361-365.
121. Bagher-Ebadian H, Siddiqui F, Liu C, Movsas B, Chetty IJ. On the impact of smoothing and noise on robustness of CT and CBCT radiomics features for patients with head and neck cancers. *Med Phys* 2017;44:1755-1770.
122. van Velden FH, Kramer GM, Frings V, et al. Repeatability of radiomic features in non-small-cell lung cancer [(18)F]FDG-PET/CT studies: Impact of reconstruction and delineation. *Mol Imaging Biol* 2016;18:788-795.
123. National Institutes of Health. NCI Wiki. RIDER. Available at: https://wiki.nci.nih.gov/display/CIP/RIDER. Accessed March 12, 2019.
124. Nordstrom RJ. The Quantitative Imaging Network in precision medicine. *Tomography* 2016;2:239-241.
125. Mackin D, Fave X, Zhang L, et al. Measuring computed tomography scanner variability of radiomics features. *Invest Radiol* 2015;50:757-765.
126. Shafiq-Ul-Hassan M, Latifi K, Zhang G, Ullah G, Gillies R, Moros E. Voxel size and gray level normalization of CT radiomic features in lung cancer. *Sci Rep* 2018;8:10545.
127. Mackin D, Fave X, Zhang L, et al. Harmonizing the pixel size in retrospective computed tomography radiomics studies. *PLoS One* 2017;12:e0178524.
128. McGurk R, Seco J, Riboldi M, Wolfgang J, Segars P, Paganetti H. Extension of the NCAT phantom for the investigation of intrafraction respiratory motion in IMRT using 4D Monte Carlo. *Phys Med Biol* 2010;55:1475-1490.
129. Brodatz textures. Available at: http://www.ux.uis.no/~tranden/brodatz.html. Accessed March 12, 2019.
130. Vallières M, Laberge S, Diamant A, El Naqa I. Enhancement of multimodality texture-based prediction models via optimization of PET and MR image acquisition protocols: A proof of concept. *Phys Med Biol* 2017;62:8536-8565.

131. Steyerberg EW, Eijkemans MJ, Harrell FE Jr., Habbema JD. Prognostic modeling with logistic regression analysis: In search of a sensible strategy in small data sets. *Med Decis Making* 2001;21:45-56.
132. Harrell FE Jr., Lee KL, Mark DB. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361-387.
133. Obermeyer Z, Lee TH. Lost in thought - the limits of the human mind and the future of medicine. *N Engl J Med* 2017;377:1209-1211.
134. Chen JH, Asch SM. Machine learning and prediction in medicine—beyond the peak of inflated expectations. *N Engl J Med* 2017;376:2507-2509.
135. HealthMyne. Available at: https://www.healthmyne.com/. Accessed March 12, 2019.
136. TexRAD. Availabe at: http://texrad.com/. Accessed March 12, 2019.
137. OncoRadiomics. Available at: https://www.oncoradiomics.com/. Accessed March 12, 2019.
138. Zhang L, Fried DV, Fave XJ, Hunter LA, Yang J, Court LE. IBEX: An open infrastructure software platform to facilitate collaborative work in radiomics. *Med Phys* 2015;42:1341-1353.
139. MaZda Web Site. Available at: http://www.eletel.p.lodz.pl/programy/mazda/. Accessed March 12, 2019.
140. Fang YH, Lin CY, Shih MJ, et al. Development and evaluation of an open-source software package "CGITA" for quantifying tumor heterogeneity with molecular images. *Biomed Res Int* 2014;2014:248505.
141. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res* 2017;77:e104-e107.
142. Arimura H, Soufi M, Kamezawa, Ninomiya K, Yamada M. Radiomics with artificial intelligence for precision medicine in radiation therapy. *J Radiat Res* 2019;60:150-157.
143. Apte AP, Iyer A, Crispin-Ortuzar M, et al. Technical note: Extension of CERR for computational radiomics: A comprehensive MATLAB platform for reproducible radiomics research. *Med Phys* 2018;45:3713-3720.
144. Foy JJ, Robinson KR, Li H, Giger ML, Al-Hallaq H, Armato SG. Variation in algorithm implementation across radiomics software. *J Med Imag* 2018;5:044505.
145. Vallieres M, Freeman CR, Skamene SR, El Naqa I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Phys Med Biol* 2015;60:5471-5496.
146. O'Connor JP, Aboagye EO, Adams JE, et al. Imaging biomarker roadmap for cancer studies. *Nat Rev Clin Oncol* 2017;14:169-186.
147. Sanduleanu S, Woodruff HC, de Jong EEC, et al. Tracking tumor biology with radiomics: A systematic review utilizing a radiomics quality score. *Radiother Oncol* 2018;127:349-360.
148. Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. *Nature* 2012;483:531-533.