Stanford DRO Toolkit: Digital Reference Objects for Standardization of Radiomic Features

Akshay Jaggi[1], Sarah A. Mattonen[1,2], Michael McNitt-Gray[3], and Sandy Napel[1]

[1]Department of Radiology, Stanford University, Stanford, CA; [2]Department of Medical Biophysics, The University of Western Ontario, London, Ontario, Canada; and [3]Department of Radiology, University of California, Los Angeles, Los Angeles, CA

Corresponding Author:
Sandy Napel, PhD
Department of Radiology, James H. Clark Center, 318 Campus Drive, Room S323, Stanford, CA 94305;
E-mail: snapel@stanford.edu

Key Words: Radiomics, radiology, quantitative imaging, phantoms, standardization
Abbreviations: digital reference object (DRO), DICOM segmentation objects (DSO), computed tomography (CT), Hounsfield unit (HU)

ABSTRACT

Several institutions have developed image feature extraction software to compute quantitative descriptors of medical images for radiomics analyses. With radiomics increasingly proposed for use in research and clinical contexts, new techniques are necessary for standardizing and replicating radiomics findings across software implementations. We have developed a software toolkit for the creation of 3D digital reference objects with customizable size, shape, intensity, texture, and margin sharpness values. Using user-supplied input parameters, these objects are defined mathematically as continuous functions, discretized, and then saved as DICOM objects. Here, we present the definition of these objects, parameterized derivations of a subset of their radiomics values, computer code for object generation, example use cases, and a user-downloadable sample collection used for the examples cited in this paper.

INTRODUCTION

Radiomics approaches provide quantitative image features computed from medical images and hold promise for improved computer-aided diagnosis, treatment selection, and response prediction (1–6). Radiomics features belong to 5 broad classes of descriptors, namely, size, shape, intensity, texture, and margin sharpness (7, 8). Recent findings strongly suggest that these image features may hold diagnostic and predictive information, some of which may not be visible to the human eye (3, 9, 10).

Several institutions have independently developed software packages for the generation of radiomics features (7, 11, 12). When different pipelines are run on the same imaging data, features may vary significantly across institutions and pipelines owing to differences in feature definition, software implementations, and/or parameter settings (13). This raises concerns about the reproducibility and repeatability of both the feature computation itself and the subsequent model building (2, 4). Phantoms with known characteristics should prove helpful for standardization across institutions.

Multiple physical phantoms have been developed for analyzing the effect of scanner variation on the reproducibility of quantitative image features (14–16) and for analyzing the computation of specific radiomics features (eg, shape phantoms) (17). However, physical phantoms must be designed for specific experimental questions, are difficult to share, and are subject to the variations introduced by the physical scanning and reconstruction process (eg, different intensity values across different devices).

Several radiomics standardization initiatives have contributed patient cohorts as digital radiomics "phantoms." Some project teams asked several institutions to calculate radiomics features on a small number of patient images and then compared the computed values across institutions (13, 18, 19). Although results could be compared across pipelines, the underlying ground truth values of these features were unknown. One study used a digital reference object (DRO) with a known standardized uptake value to analyze variations in PET standardized uptake value computation across institutions (20). This study showed the utility of a "ground-truth" value in standardization across institutions.

In addition, while recent work has already shown that radiomics features can be dependent on voxel size, scanner model and acquisition/reconstruction settings, image rotation, and translation (15, 21–25), some features might be dependent on other radiomics features themselves (like object size and shape). One recent phantom design allowed for testing the effect of controlled changes in object size on the calculation of shape features (17). This work revealed that a number of shape features were unstable with respect to changes in volume. Because many classes of radiomics features (intensity, texture, margin, etc.) may also show these interdependencies, phantoms are needed to address these relationships in a controlled, hypothesis-driven fashion.

This paper presents a toolkit for the creation of DROs and a sample collection of DROs made using it for radiomics experiments and illustration. The DROs are mathematically defined and

ISSN 2379-1381 https://doi.org/10.18383/j.tom.2019.00030

are output as DICOM image stacks with segmentations formatted as DICOM segmentation objects (DSOs). The DROs' mathematical definitions allow the derivation of "theoretical" radiomics values for some radiomics features, thereby allowing the accuracy of radiomics algorithms to be verified. They can be altered along 10 axes of variation to allow for investigations into the stability and robustness of extracted features to controlled variation of object construction. We present the calculation of several theoretical radiomics values for some DROs in our sample collection and compare them to the corresponding features extracted using a particular radiomics pipeline.

METHODOLOGY

Definition of Objects and Software

Definition of Objects. The toolkit creates DROs of chosen size, shape, intensity, texture, and margin sharpness, sampled and embedded in stacks of 300 512- × 512-Gy value images. As currently implemented, the images have a pixel spacing of 1 mm in the image (*x*.*y*) plane, a slice thickness of 1 mm, and a slice spacing of 1 mm. Therefore, voxels are 1 mm^3. These image stacks are then saved as DICOM images and segmentation objects (26). These DROs do not attempt to simulate computed tomography (CT), magnetic resonance, or any specific scanner or modality behavior. Instead, we define these objects with known, continuous functions and then sample these functions as images. Each DRO is a variation on a sphere as defined by 10 parameters divided into 5 categories, as described in the following sections.

Size (1 Parameter). **Mean Radius**: The average radius ($\tilde{\rho}$) of the object in millimeters.

Shape (5 Parameters). **XYZ Deformation**: The scaling of the object along each Cartesian dimension [*x*,*y*,*z*] defined as a scalar multiple of the radius. Therefore, [1,1,1] would be a perfect sphere, while [2,1,1] would be an ellipsoid with twice the radius in the *x* direction.

Surface Frequency and Amplitude: The frequency (ω) and amplitude (α) of a sinusoidal variation along the surface of the sphere, with frequency unitless and amplitude as a multiple of the radius. Together, in spherical coordinates, the shape parameters can be represented as:

$$\rho = f(\theta, \phi) = \tilde{\rho}(1 + \alpha)sin(\omega\theta)cos(\omega\phi)$$

where θ is the polar angle and ϕ is the azimuthal angle.

Intensity (1 Parameter). **Mean Intensity**: The mean internal intensity ($\tilde{\mu}$) of the object's gray values. Note that the intensity values can be converted to specific intensity values (eg, Hounsfield unit, or HU, for CT) when desired. For the provided objects, intensity values have been scaled to HU. That is, intensities are scaled linearly with air at −1024 and water at 0.

Texture (2 Parameters). We conceptualize the texture as a 3-dimensional sinusoidal variation of the voxel value. Therefore, 2 texture parameters control aspects of this sinusoidal model.

Texture Wavelength and Amplitude: The wavelength (λ) and amplitude (α) of the sinusoidal variation of the intensity of the image, wavelength in (mm), and amplitude in intensity units. In the case of these objects, these units are scaled to HU. Together, in 3D Cartesian coordinates, we model the texture variation as:

$$\mu(x, y, z) = \tilde{\mu} + \alpha cos\left(\frac{2\pi}{\lambda}x\right)cos\left(\frac{2\pi}{\lambda}y\right)cos\left(\frac{2\pi}{\lambda}z\right)$$

Where μ is the intensity at any given coordinate. Note, as currently implemented, the wavelength is identical in all 3 dimensions. In addition, when α is set to zero, the object will have a uniform internal value of $\tilde{\mu}$, the mean internal intensity.

Margin Sharpness (1 Parameter). We conceptualize margin sharpness as the transition from the internal intensity value to the external intensity value. To smoothly transition the internal intensity to the background, we apply a Gaussian blur, with a single parameter, to the object.

Gaussian Standard Deviation: The standard deviation (σ) of the uniform, 3-dimensional Gaussian image blur applied to the image. The larger the standard deviation, the greater the blurring effect. A standard deviation of zero defaults to no blurring. This blur can be modeled continuously as:

$$f(x, y, z) = \frac{1}{\sigma^3 2\pi^{\frac{3}{2}}}\exp -\frac{x^2 + y^2 + z^2}{2\sigma^2}$$

Because the function is applied as a filter to the image, it is centered on each pixel. Therefore, the filter blurs not only the edges but also all internal intensity values. To repair the now-blurred internal intensities, we replace all voxels within the original segmentation map with their original values. Therefore, the region outside the object is the filtered version of the image while the interior of the object has the prefiltered intensity values.

Description of Code. We have implemented a command-line-driven software package that allows the creation of objects exhibiting the parameters as described above. Using a YAML configuration file, all 10 parameters from the 5 categories described above can be set for user-driven generation of new DROs. Users can specify a range of values for each parameter, by providing a minimum, maximum, and number of values. The program will divide the range between the minimum and maximum into the number of requested values at equal intervals (eg, minimum radius of 20, maximum radius of 40, and 3 values produces: radii 20, 30, and 40). If the user desires just 1 value for a given parameter, the minimum and maximum should be equal and the number of values should be 1. If the user provides ranges for n different parameters, the program generates an n-dimensional matrix of DROs for each combination of parameter values. Each point in this matrix corresponds to a unique object produced by the code. For example, if the user requests 3 values for 3 parameters, (3^3), the toolkit will generate 27 DROs.

To provide unique and relevant names for every generated DRO, the DRO name (saved in the DICOM header "Patient Name" and "Patient ID" tags and as the enclosing folder name) is a "-" separated list of the values of all 10 settable parameters in the same order as listed in the "Definition of Objects" subsection. For example, if a DRO has a radius of 100, an *x*-, *y*-, and *z*-deformation of 1, a shape frequency of 9, a shape amplitude of 0.2, a mean intensity of 100, a texture wavelength of 10, a texture amplitude of 50.0, and a margin sharpness Gaussian standard deviation of 10, then the unique name of the generated object will be Phantom-100.0-1.0-1.0-1.0-9.0-0.2-100.0-10.0-50.0-10.0. Each number corresponds to each parameter in order.

Table 1. Parameters for the Sample Collection of DROs

Feature Class	Parameter Name	Unit Measure	First Value	Second Value
Size	Mean Radius	mm	20	100
Shape	X Deformation	Multiple of Radius	1	
Shape	Y Deformation	Multiple of Radius	1	
Shape	Z Deformation	Multiple of Radius	1	
Shape	Surface Frequency	Unitless	9	
Shape	Surface Amplitude	Multiple of Radius	0	0.2
Intensity	Mean Intensity	HU	−100	100
Texture	Texture Wavelength	mm	10	
Texture	Texture Amplitude	HU	0	50
Margin Sharpness	Gaussian Standard Deviation	Unitless	0	10

Parameters and units are taken from the "Definition of Objects" subsection. Five chosen parameters from each feature class have 2 values. All other parameters have 1 value. Because a 0 amplitude negates a sinusoidal variation, we specify only 1 value for texture wavelength and shape frequency.

Once the DICOM series has been generated, the software then produces a corresponding DSO file for each DICOM series. The DSO file is named with the unique DRO name. The software finally returns 2 zip files, namely, DICOMs and DSOs. The DICOM zip file is a folder of subfolders. Each subfolder is named after a specific DRO and contains the DICOM series for that DRO. The DSO zip file is a folder of files. Each file is the DSO for a specific DRO.

The command line tool is open-source and available for use from this GitHub repository: https://github.com/riipl/dro_cli.

Illustrative Use Cases

Generation of Sample Collection. To offer out-of-the-box, ready-to-use DROs for immediate studies, we generated a collection of DROs; each DRO consists of a DICOM image series with accompanying DSO. For distribution, we also provide this specific collection of DROs in Neuroimaging Informatics Technology Initiative (NIfTI) format segmentations (27).

For each of the 5 classes of image features (size, shape, intensity, texture, and margin sharpness), we have selected 1 parameter from each class to have 2 values. We chose mean radius, shape variation amplitude, mean intensity, texture variation amplitude, and Gaussian standard deviation. All other parameters are held constant. Table 1 specifies the values we used for all 10 parameters and highlights the 5 parameters that have 2 values. Because each chosen parameter has 2 values, we generated 2^5 or 32 unique objects. Note that, while all objects have a texture wavelength of 10 mm, Table 1 specifies that the texture amplitude is zero for half of the 32 objects, resulting in 16 objects with uniform intensity.

Calculation of Theoretical Radiomic Values. To show the use of the theoretical definitions of these phantoms, we generated the theoretical "ground truth" values for a subset of radiomics features. Acknowledging that there are many potential radiomic values to compute, we chose 8 features as defined by the Image Biomarker Standardisation Initiative (IBSI) (8, 28). These features are volume, surface area, 2D diameter, 3D diameter, sphericity, intensity mean, intensity standard deviation, and intensity kurtosis. Online supplemental Appendix 1 specifies their IBSI definitions. All of the theoretical values of the features in our DROs were computed by applying the IBSI definitions of the features to the mathematical definition of the continuous object. We provide the value of each of these features for 3 objects from the collection: uniform sphere with uniform intensity (Table 2 **first bold**, hereafter named "Uniform" for convenience), uniform sphere with texture variation (Table 2 **second bold**, hereafter named "Texture Variation"), and nonuniform sphere with uniform intensity (Table 2 **third bold**, hereafter named "Shape Variation").

Comparison of Theoretical Radiomics Values with Output of a Pipeline. To show the utility of the theoretical values in comparison with pipeline output, we compared the theoretical radiomics values defined above against radiomics values produced by the Stanford Quantitative Image Feature Engine (QIFE) on the 3 DROs described above (7). Online supplemental Appendix 2 gives the configuration file parameters we used for running the Stanford QIFE. See Echegaray et al. (7) for definitions and implementation of all QIFE features.

RESULTS

Description of Sample Collection

The 32 unique objects generated for the sample collection have every combination of the 2 values for the 5 chosen parameters (Table 2). As the colors indicate, there are 16 objects with each value for each parameter. This diversity of objects allows users to explore each parameter in isolation and in the context of other parameter changes. Figure 1 presents 8 objects sampled from the collection. The entire collection is available in zipped folders in the project GitHub repository: https://github.com/riipl/dro_cli. The collection is also available in The Cancer Imaging Archive (https://doi.org/10.7937/t062-8262).

Comparison of Theoretical Computation of "Ground Truth" Radiomics Features to QIFE Calculated Values

We derived the theoretical values for the 8 IBSI-defined radiomics features described above for the Uniform, Texture Variation, and Shape Variation DROs defined above. These

Table 2. Table of All Generated DROs with Unique Parameters

Name	Size	Shape	Intensity	Texture	Margin
	Mean Radius (mm)	Amplitude (multiple of radius)	Mean Intensity (HU)	Amplitude (HU)	Gaussian Standard Deviation (unitless)
Phantom-20.0-1.0-1.0-1.0-9.0-0.0--100.0-10.0-0.0-0.0	20	0	-100	0	0
Phantom-20.0-1.0-1.0-1.0-9.0-0.0--100.0-10.0-0.0-10.0	20	0	-100	0	10
Phantom-20.0-1.0-1.0-1.0-9.0-0.0--100.0-10.0-50.0-0.0	20	0	-100	50.0	0
Phantom-20.0-1.0-1.0-1.0-9.0-0.0--100.0-10.0-50.0-10.0	20	0	-100	50.0	10
Phantom-20.0-1.0-1.0-1.0-9.0-0.0-100.0-10.0-0.0-0.0	20	0	100	0	0
Phantom-20.0-1.0-1.0-1.0-9.0-0.0-100.0-10.0-0.0-10.0	20	0	100	0	10
Phantom-20.0-1.0-1.0-1.0-9.0-0.0-100.0-10.0-50.0-0.0	20	0	100	50.0	0
Phantom-20.0-1.0-1.0-1.0-9.0-0.0-100.0-10.0-50.0-10.0	20	0	100	50.0	10
Phantom-20.0-1.0-1.0-1.0-9.0-0.2--100.0-10.0-0.0-0.0	20	0.2	-100	0	0
Phantom-20.0-1.0-1.0-1.0-9.0-0.2--100.0-10.0-0.0-10.0	20	0.2	-100	0	10
Phantom-20.0-1.0-1.0-1.0-9.0-0.2--100.0-10.0-50.0-0.0	20	0.2	-100	50.0	0
Phantom-20.0-1.0-1.0-1.0-9.0-0.2--100.0-10.0-50.0-10.0	20	0.2	-100	50.0	10
Phantom-20.0-1.0-1.0-1.0-9.0-0.2-100.0-10.0-0.0-0.0	20	0.2	100	0	0
Phantom-20.0-1.0-1.0-1.0-9.0-0.2-100.0-10.0-0.0-10.0	20	0.2	100	0	10
Phantom-20.0-1.0-1.0-1.0-9.0-0.2-100.0-10.0-50.0-0.0	20	0.2	100	50.0	0
Phantom-20.0-1.0-1.0-1.0-9.0-0.2-100.0-10.0-50.0-10.0	20	0.2	100	50.0	10
Phantom-100.0-1.0-1.0-1.0-9.0-0.0--100.0-10.0-0.0-0.0	100	0	-100	0	0
Phantom-100.0-1.0-1.0-1.0-9.0-0.0--100.0-10.0-0.0-10.0	100	0	-100	0	10
Phantom-100.0-1.0-1.0-1.0-9.0-0.0--100.0-10.0-50.0-0.0	100	0	-100	50.0	0
Phantom-100.0-1.0-1.0-1.0-9.0-0.0--100.0-10.0-50.0-10.0	100	0	-100	50.0	10
Phantom-100.0-1.0-1.0-1.0-9.0-0.0-100.0-10.0-0.0-0.0 (1)	**100**	**0**	**100**	**0**	**0**
Phantom-100.0-1.0-1.0-1.0-9.0-0.0-100.0-10.0-0.0-10.0	100	0	100	0	10
Phantom-100.0-1.0-1.0-1.0-9.0-0.0-100.0-10.0-50.0-0.0 (2)	**100**	**0**	**100**	**50.0**	**0**
Phantom-100.0-1.0-1.0-1.0-9.0-0.0-100.0-10.0-50.0-10.0	100	0	100	50.0	10
Phantom-100.0-1.0-1.0-1.0-9.0-0.2--100.0-10.0-0.0-0.0	100	0.2	-100	0	0
Phantom-100.0-1.0-1.0-1.0-9.0-0.2--100.0-10.0-0.0-10.0	100	0.2	-100	0	10
Phantom-100.0-1.0-1.0-1.0-9.0-0.2--100.0-10.0-50.0-0.0	100	0.2	-100	50.0	0
Phantom-100.0-1.0-1.0-1.0-9.0-0.2--100.0-10.0-50.0-10.0	100	0.2	-100	50.0	10
Phantom-100.0-1.0-1.0-1.0-9.0-0.2-100.0-10.0-0.0-0.0 (3)	**100**	**0.2**	**100**	**0**	**0**
Phantom-100.0-1.0-1.0-1.0-9.0-0.2-100.0-10.0-0.0-10.0	100	0.2	100	0	10
Phantom-100.0-1.0-1.0-1.0-9.0-0.2-100.0-10.0-50.0-0.0	100	0.2	100	50.0	0
Phantom-100.0-1.0-1.0-1.0-9.0-0.2-100.0-10.0-50.0-10.0	100	0.2	100	50.0	10

While every DRO has 10 parameters, we present just the 5 parameters we varied between 2 values in the sample collection. The unique name of each DRO is generated as described in "Description of Code" subsection. Each parameter is colored by value. Bolded objects: (1) referred to as "Uniform," (2) Referred to as "Texture Variation," (3) referred to as "Shape Variation," in *Calculation of Theoretical Radiomics Values*.

derivations serve as a model to researchers interested in comparing their radiomics pipelines to "ground-truth" values. Table 3 compares the derived theoretical radiomics values to those produced by Stanford QIFE. Across all values, there is <10% difference. Excluding surface area and sphericity, there is a <1% difference. Note that for the DROs with uniform intensity, QIFE appropriately returns NaN for kurtosis, as it is undefined.

Furthermore, as the Uniform DRO only differs from the Texture Variation and Shape Variation DROs by 1 parameter each, Table 3 also presents a simple demonstration of how QIFE output reflects changes in individual DRO parameters. For example, increasing the texture amplitude from 0 to 50 HU produces a change in intensity standard deviation and kurtosis but not intensity mean or any shape feature (Table 3). Similarly, increasing the shape variation amplitude from 0% to 20% of the radius leads to changes in all size and shape features but no change in any intensity features (Table 3).

DISCUSSION

Customizable DROs allow for valuable comparisons of existing pipelines. As presented in McNitt-Gray et al. (19), DROs can be used for large-scale, multi-institutional studies as a tool for comparing radiomics features across many pipelines. We narrow this work by focusing on the benefits of the DRO's theoretical values as a benchmark for features when no additional pipelines are available for comparison. Unlike MD Anderson's Credence Cartridge Radiomics (CCR) phantom (15, 16) or the American College of Radiology CT Phantom (ACR CT) (14, 29), our DROs have theoretically defined values that match closely to computed values from an existing pipeline (Table 3).

Notably, the QIFE calculated surface area at 8% higher than the theoretical value (with sphericity 8% lower as a result) compared with the theoretical value. Limkin et al. analytically computed shape radiomics features from the surface of mathematically specified objects and compared these values to the radiomics features computed from the images of CT scans of their 3D-printed versions (17). In their experiments, the surface area of the images also showed between a 5% and 15% difference from the value computed from the mathematical description. However, many more factors can influence radiomics values for a scanned object, including noise, partial volume, reconstruction parameters including voxel size, etc., which could cause differences even in repeated

scans of the same object. Although these results are important, our DROs allow investigation of the accuracy of radiomic pipelines before considering these confounding effects.

One limitation of the theoretical values for our DROs is that they are computed from the continuous definitions of the objects and not the discretized version embedded in image space. For instance, the volume, surface area, and sphericity of a continuous sphere are different from the corresponding values computed for a discretized sphere with the same dimensions. Further work could investigate the impact of discretization (number and size of voxels) on the difference between the theoretical values and the pipeline-computed values. Nonetheless, the general agreement between the theoretical values and the QIFE values for the 8 features we studied confirms the theoretical values' accuracy and utility in checking the output of a radiomics pipeline.

Another limitation of this work is that, although we have scaled the image gray values in HU, these DROs do not attempt to simulate CT noise or any of its known artifacts. Therefore, these DROs could be argued as being "hyperidealized" objects with no relevance to feature computation in clinical contexts. However, their idealized nature is also their benefit because it allows for controlled studies in developing and/or fine-tuning radiomics pipelines and comparing their performance across institutions. In addition, not all radiomics features can be trivially computed from the mathematical definitions of the objects. Many second-order texture features are the main source of disagreement between radiomics implementations (13, 18, 19) but must be computed from the voxel-based embedding of the objects rather than their continuous definitions. Further work could consider theoretical estimations of these kinds of features from the object definitions.

This work presents the implementations of just 10 user-definable image features. We provide the code in this GitHub repository (https://github.com/riipl/dro_cli) with the hope that users will modify and/or implement new features to resolve specific questions of interest. There are many possible extensions of this feature set. For example, as implemented, the sinusoidal texture has equal wavelength in all 3 Cartesian dimensions. Future developers could easily extend the current code to implement different wavelengths in all dimensions allowing for a greater diversity in object textures. Similarly, the sinusoidal shape variation has equal frequency in both angular dimensions. This limitation could easily be removed to develop more complex object shapes. In addition, margin blurring is currently computed by blurring the object and then resetting all intensity values within the original segmentation. More sophisticated methods could attempt to create a smooth transition between the interior and exterior of the object.

We provide a collection of 32 DROs with a large combination of image features of interest. These objects allow for investigation of image features in isolation and in the context of other changes (eg, the impact of object shape on GLCM texture features). Although we only analyzed 3 DROs from this collection, we hope this collection of DROs can serve a diversity of experimental questions and inspire the generation of new DROs using the available command line tool.

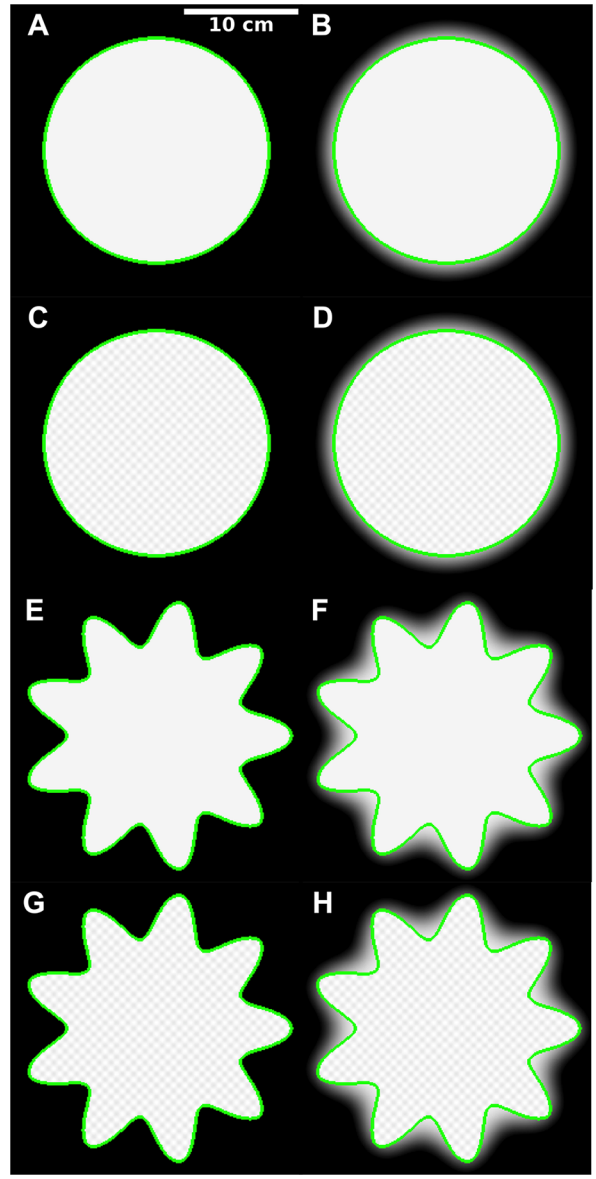

Figure 1. DRO Collection Subset. Representative images of the maximum area cross-section of 8 digital reference objects (DROs) from the provided collection. The window level is equivalent to −400 HU, and the window width is equivalent to 800 HU. The scale of the images indicated by the 10-cm scale bar in (A). DROs are aligned such that all characteristics between DROs in the same row are identical except margin sharpness. All DROs have an average radius of 100 mm and an average internal intensity of −100 HU. (A) and (B) are both uniform shape and uniform intensity. (C) and (D) are both uniform shape and varying intensity. (E) and (F) are both varying shape and uniform intensity. (G) and (H) are both varying shape and varying intensity.

Table 3. Theoretical and QIFE Radiomics Feature Values for 8 IBSI-Defined Features Computed on 3 DROs, with Percent Differences of Feature Value between the Theoretical and QIFE Calculations

DRO	Uniform			Texture Variation			Shape Variation		
Institution	Theoretical	QIFE	Percent Difference	Theoretical	QIFE	Percent Difference	Theoretical	QIFE	Percent Difference
Volume	4188790.00	4158712.00	−0.72	4188790.00	4158712.00	−0.72	4314060.00	4284044.00	−0.70
Surface Area	125664.00	135581.51	7.89	125664.00	135581.51	7.89	244451.00	262306.95	7.30
2D Longest Diameter	200.00	199.58	−0.21	200.00	199.58	−0.21	236.49	235.77	−0.30
3D Longest Diameter	200.00	199.61	−0.20	200.00	199.61	−0.20	236.49	238.82	0.99
Sphericity	1.00	0.92	−7.76	1.00	0.92	−7.76	0.52	0.49	−7.34
Intensity Mean	100.00	100.00	0.00	100.00	99.51	−0.49	100.00	100.00	0.00
Intensity Stdev	0.00	0.00	0.00	17.68	17.68	−0.01	0.00	0.00	0.00
Intensity Kurtosis	0.00	NaN	NaN	3.38	3.37	−0.01	0.00	NaN	NaN

The presented DROs address 3 major needs in the radiomics literature. These objects are the first digital, customizable reference objects with some mathematically derivable radiomics features. Not relying on physical phantoms or patient images democratizes experimental design and allows for faster image dissemination and project collaboration. Theoretical values for radiomics features provide pipeline-independent reference values. Finally, the multiparametric customizability of the objects allows for controlled studies of individual radiomics features and their stability to changes in other image features.

Supplemental Materials

Supplemental Appendix 1: https://doi.org/10.18383/j.tom.2019.00030.sup.01

Supplemental Appendix 2: https://doi.org/10.18383/j.tom.2019.00030.sup.02

ACKNOWLEDGMENTS

A.J. designed, conducted, and wrote the experiments, and wrote the initial draft of the manuscript. S.M. advised on results' interpretation, writing, and data presentation. M.M.G advised on writing and data presentation. S.N. advised and oversaw project conception, design, experimentation, and writing.

S.N., S.M., and A.J. were supported, in part, through funding from NIH/NCI (U01 CA187947); M.M.G. was supported, in part, through funding from NIH/NCI (U01 CA181156). We would also like to thank Dev Gude and Emel Alkim for essential technical support and Elizabeth Colvin for administrative support.

Disclosure: Dr. Sandy Napel is on the Medical Advisory Board for Fovia Inc., a Scientific Advisor for EchoPixel Inc., and a Scientific Advisor for RADLogics Inc.

REFERENCES

1. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, Zegers CML, Gillies R, Boellard R, Dekker A, Aerts HJWL. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer. 2012;48:441–446.
2. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, Forster K, Aerts HJWL, Dekker A, Fenstermacher D, Goldgof DB, Hall LO, Lambin P, Balagurunathan Y, Gatenby RA, Gillies RJ. Radiomics: the process and the challenges. Magn Reson Imaging. 2012;30:1234–1248.
3. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D, Hoebers F, Rietbergen MM, Leemans CR, Dekker A, Quackenbush J, Gillies RJ, Lambin P. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014;5:4006.
4. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, Sanduleanu S, Larue RTHM, Even AJG, Jochems A, van Wijk Y, Woodruff H, van Soest J, Lustberg T, Roelofs E, van Elmpt W, Dekker A, Mottaghy FM, Wildberger JE, Walsh S. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol. 2017;14:749–762.
5. Gillies R, Kinahan PE, Hricak H. Radiomics: images Are More than Pictures, They Are Data. Radiology [Internet]. 2019;278.
6. Napel S, Mu W, Jardim-Perassi BV, Aerts H, Gillies RJ. Quantitative imaging of cancer in the postgenomic era: radio(geno)mics, deep learning, and habitats. Cancer. 2018;124:4633–4649.
7. Echegaray S, Bakr S, Rubin DL, Napel S. Quantitative Image Feature Engine (QIFE): an open-source, modular engine for 3D quantitative feature extraction from volumetric medical images. J Digit Imaging. 2018;31:403–414.
8. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, Ashrafinia S, Bakas S, Beukinga RJ, Boellaard R, Bogowicz M, Boldrini L, Buvat I, Cook GJR, Davatzikos C, Depeursinge A, Desseroit MC, Dinapoli N, Dinh CV, Echegaray S, El Naqa I, Fedorov AY, Gatta R, Gillies RJ, Goh V, Götz M, Guckenberger M, Ha SM, Hatt M, Isensee F, Lambin P, Leger S, Leijenaar RTH, Lenkowicz J, Lippert F, Losnegård A, Maier-Hein KH, Morin O, Müller H, Napel S, Nioche C, Orlhac F, Pati S, Pfaehler EAG, Rahmim A, Rao AUK, Scherer J, Siddique MM, Sijtsema NM, Socarras Fernandez J, Spezi E, Steenbakkers RJHM, Tanadini-Lang S, Thorwarth D, Troost EGC, Upadhaya T, Valentini V, van Dijk LV, van Griethuysen J, van Velden FHP, Whybra P, Richter C, Löck S. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology. 2020 Mar;10:191145.
9. Fave X, Zhang L, Yang J, Mackin D, Balter P, Gomez D, Followill D, Jones AK, Stingo F, Liao Z, Mohan R, Court L. Delta-radiomics features for the prediction of patient outcomes in non–small cell lung cancer. Sci Rep. 2017;7:588.
10. Gevaert O, Mitchell LA, Achrol AS, Xu J, Echegaray S, Steinberg GK, Cheshier SH, Napel S, Zaharchuk G, Plevritis SK. Glioblastoma multiforme: exploratory radiogenomic analysis by using quantitative image features. Radiology. 2015;276:313–313.
11. 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.
12. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RGH, Fillion-Robin J-C, Pieper S, Aerts HJWL. Computational radiomics system to decode the radiographic phenotype. Cancer Res. 2017;77:e104–7.
13. Kalpathy-Cramer J, Mamomov A, Zhao B, Lu L, Cherezov D, Napel S, Echegaray S, Rubin D, McNitt-Gray M, Lo P, Sieren JC, Uthoff J, Dilger SKN, Driscoll B, Yeung I, Hadjiiski L, Cha K, Balagurunathan Y, Gillies R, Goldgof D. Radiomics of lung

nodules: a multi-institutional study of robustness and agreement of quantitative imaging features. Tomography. 2016;2:430–437.
14. Zhao B, Tan Y, Tsai WY, Schwartz LH, Lu L. Exploring variability in CT characterization of tumors: a preliminary phantom study. Transl Oncol. 2014;7:88–93.
15. Mackin D, Fave X, Zhang L, Fried D, Yang J, Taylor B, Rodriguez-Rivera E, Dodge C, Jones AK, Court L. Measuring computed tomography scanner variability of radiomics features. Invest Radiol. 2015;50:757.
16. Mackin D, Fave X, Zhang L, Fried D, Yang J, Taylor B, Rodriguez-Rivera E, Dodge C, Jones AK, Court L. Data from Credence Cartridge Radiomics Phantom CT Scans. The Cancer Imaging Archive. 2017; Available from: http://doi.org/10.7937/K9/TCIA.2017.zuzrml5b.
17. Limkin EJ, Reuzé S, Carré A, Sun R, Schernberg A, Alexis A, Deutsch E, Ferté C, Robert C. The complexity of tumor shape, spiculatedness, correlates with tumor radiomic shape features. Sci Rep. 2019;9:1–12.
18. Chenevert TL, Malyarenko DI, Newitt D, Li X, Jayatilake M, Tudorica A, Fedorov A, Kikinis R, Liu TT, Muzi M, Oborski MJ, Laymon CM, Li X, Thomas Y, Jayashree K-C, Mountz JM, Kinahan PE, Rubin DL, Fennessy F, Huang W, Hylton N, Ross BD. Errors in quantitative image analysis due to platform-dependent image scaling. Transl Oncol. 2014;7:65–71.
19. McNitt-Gray M, Napel S, Jaggi A, Mattonen SA, Hadjiiski L, Muzi M, Goldgof D, Balagurunathan Y, Pierce LA, Kinahan PE, Jones EF, Nguyen A, Virkud A, Chan HP, Emaminejad N, Wahi-Anwar M, Daly M, Abdalah M, Yang H, Lu L, Lv W, Rahmim A, Gastounioti A, Pati S, Bakas S, Kontos D, Zhao B, Kalpathy-Cramer J, Farahani K. Standardization in quantitative imaging: a multicenter comparison of radiomics features from different software packages on digital reference objects and patient datasets. Tomography 2020;6:118–128.
20. Pierce LA, Elston BF, Clunie DA, Nelson D, Kinahan PE. A digital reference object to analyze calculation accuracy of PET standardized uptake value. Radiology. 2015;277:538–45.
21. Zhao B, Tan Y, Tsai W-Y, Qi J, Xie C, Lu L, Schwartz LH. Reproducibility of radiomics for deciphering tumor phenotype with imaging. Sci Rep. 2016;6:1–7.
22. Zwanenburg A, Leger S, Agolli L, Pilz K, Troost EGC, Richter C, Löck S. Assessing robustness of radiomic features by image perturbation. Sci Rep. 2019;9:1–10.
23. 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;8:1–9.
24. Traverso A, Wee L, Dekker A, Gillies R. Repeatability and reproducibility of radiomic features: a systematic review. Int J Radiat Oncol. 2018;102:1143–1158.
25. 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 (Bellingham). 2015;2. [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4524811/].
26. NEMA PS3/ISO 12052. Digital Imaging and Communications in Medicine (DICOM) Standard [Internet]. National Electrical Manufacturers Association; Available from http://medical.nema.org/.
27. Cox RW, Ashburner J, Breman H, Fissell K, Haselgrove C, Holmes CJ, Lancaster JL, Rex DE, Smith SM, Woodward JB, Strother S. A (sort of) new image data format standard: NiFTI-1. In: 10th Annual Meeting of the Organization for Human Brain Mapping. 2004.
28. Vallières M, Zwanenburg A, Badic B, Rest CCL, Visvikis D, Hatt M. Responsible radiomics research for faster clinical translation. J Nucl Med. 2018;59:189–93.
29. Lu L, Liang Y, Schwartz LH, Zhao B. Reliability of radiomic features across multiple abdominal CT image acquisition settings: a Pilot Study using ACR CT phantom. Tomography. 2019;5:226–231.