Quality control of radiomic features using 3D-printed CT phantoms

Usman Mahmood,[a,*] **Aditya Apte**,[a] **Christopher Kanan**,[b] **David D. B. Bates**,[c] **Giuseppe Corrias**,[c] **Lorenzo Manneli**,[d] **Jung Hun Oh**,[a] **Yusuf Emre Erdi**,[a] **John Nguyen**,[e] **Joseph O. Deasy**,[a] **and Amita Shukla-Dave**[a,c]

[a]Memorial Sloan Kettering Cancer Center, Department of Medical Physics, New York, United States

[b]Rochester Institute of Technology, Department of Imaging Science, Rochester, New York, United States

[c]Memorial Sloan Kettering Cancer Center, Department of Radiology, New York, United States

[d]IRCCS SDN, Naples, Italy

[e]The Cooper Union, New York, United States

Abstract

Purpose: The lack of standardization in quantitative radiomic measures of tumors seen on computed tomography (CT) scans is generally recognized as an unresolved issue. To develop reliable clinical applications, radiomics must be robust across different CT scan modes, protocols, software, and systems. We demonstrate how custom-designed phantoms, imprinted with human-derived patterns, can provide a straightforward approach to validating longitudinally stable radiomic signature values in a clinical setting.

Approach: Described herein is a prototype process to design an anatomically informed 3D-printed radiomic phantom. We used a multimaterial, ultra-high-resolution 3D printer with voxel printing capabilities. Multiple tissue regions of interest (ROIs), from four pancreas tumors, one lung tumor, and a liver background, were extracted from digital imaging and communication in medicine (DICOM) CT exam files and were merged together to develop a multipurpose, circular radiomic phantom (18 cm diameter and 4 cm width). The phantom was scanned 30 times using standard clinical CT protocols to test repeatability. Features that have been found to be prognostic for various diseases were then investigated for their repeatability and reproducibility across different CT scan modes.

Results: The structural similarity index between the segment used from the patients' DICOM image and the phantom CT scan was 0.71. The coefficient variation for all assessed radiomic features was <1.0% across 30 repeat scans of the phantom. The percent deviation (pDV) from the baseline value, which was the mean feature value determined from repeat scans, increased with the application of the lung convolution kernel, changes to the voxel size, and increases in the image noise. Gray level co-occurrence features, contrast, dissimilarity, and entropy were particularly affected by different scan modes, presenting with pDV $> \pm 15\%$.

Conclusions: Previously discovered prognostic and popular radiomic features are variable in practice and need to be interpreted with caution or excluded from clinical implementation. Voxel-based 3D printing can reproduce tissue morphology seen on CT exams. We believe that this is a flexible, yet practical, way to design custom phantoms to validate and compare radiomic metrics longitudinally, over time, and across systems.

© *The Authors. Published by SPIE under a Creative Commons Attribution 4.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.* [DOI: 10.1117/1.JMI.8.3.033505]

Keywords: additive manufacturing; computed tomography; radiomics; quantitative imaging.

Paper 20227R received Aug. 27, 2020; accepted for publication Jun. 4, 2021; published online Jun. 29, 2021.

*Address all correspondence to Usman Mahmood, mahmoodu@mskcc.org

1 Introduction

There is a growing body of literature about the role of quantitative radiomics (QR) metrics as cancer imaging biomarkers for predicting lesion malignancy and the efficacy of treatments.[1–3] Although promising, a general lack of standardization and inconsistent performance of QR metrics across different computed tomography (CT) scan modes is well established.[4–10] A solution to improve standardization and quality control (QC) of the QR pipeline should include phantoms that can filter unreliable metrics.[11–13]

Historically, CT equipment operators have used QC phantoms to monitor the imaging performance of clinical scanners.[14] However, CT QC phantoms are engineered with homogeneous materials that lack the texture or shapes of tumors seen on CT exams. There is evidence that a textured QC phantom may reveal problems across the imaging pipeline that were not apparent with the routinely used[15] homogeneous QC phantom.[15] Another study[16] showed that the local noise and resolution properties of a lesion depend on the background tissue texture when iterative reconstruction is used to reconstruct the image. As CT scanner hardware and software become more technologically sophisticated, the phantom components will need to take on more realistic properties.

Several phantom types have been proposed to study QR feature variability.[9,17–20] In most literature reports, the phantoms are uniform, consist of patterns that may not be found in patient images, or are shaped in a way that is not characteristic of patient anatomy.[21] For example, an updated version of the credence cartridge radiomic phantom[17] consists of six oval cartridges encased in a high-density polystyrene buildup material. The cylindrical shape and encasing are modeled after the size or shape of human anatomy, but the six-round cartridges are placed within a uniform surrounding with each cartridge consisting of a single textured pattern. Evaluating the impact of iterative reconstruction schemes is challenging when the background textures are uniform.[6,22] Further, the homogeneous shapes will not be able to evaluate interobserver segmentation variability, which is known to contribute to QR feature instability.[23]

More recently, 3D-printed imaging phantoms have been used to evaluate QR feature robustness.[24–26] In one study,[26] a realistic liver phantom was constructed by first converting the patient images into surface models using stereolithography (STL) file formats. However, the STL format does not capture the internal structure or texture of images and only represents the shell or surface of the modeled object.[27] In another study,[24] simulated lung nodules were 3D printed and inserted into a chest anthropomorphic phantom to evaluate QR feature robustness. Although the approach demonstrates the possibility of voxel-based 3D printing, some of the methods are complex, requiring simulations that may be limited by the extent to which models of anatomy and the imaging system are realistic enough.[28] Other approaches include using standard desktop inkjet printers with ink cartridges that are filled with aqueous potassium iodide solutions to generate realistic 3D prints.[29,30] The doped ink is deposited either on standard or specialized paper. Our proposed method has the distinct advantage of going directly from a digital imaging and communication in medicine (DICOM) CT scan to a 3D printer, using commercially available technology. Consequently, we can overcome key issues, such as the lack of adhesion between layers, coarser resolution, and requirement for extensive simulations that are seen with other methods.[31]

To overcome the limitations, we evaluate the feasibility of translating anatomy seen on a CT scan to a physical phantom using a multimaterial 3D printer with commercially available voxel-printing software (PolyJet Objet 260 Connex 3, Stratasys, Eden Prairie, Minnesota). The proposed method uses voxel printing technology to (i) develop fit-for-use, custom-designed 3D-printed phantom's, imprinted with actual tumor patterns seen on CT exams, to QC and validate QR feature robustness and (ii) evaluate the repeatability and reproducibility of derived QR features.

2 Materials and Methods

2.1 *Study Participants*

Institutional review board approval was obtained, and the requirement for informed consent was waived.

We modeled the 3D-printed radiomic phantom after diseased tissues seen on CT scans of six unique patients. Four patients had pancreatic adenocarcinoma (PDAC), one patient had nonsmall cell carcinoma (NSCLC), and one patient presented with advanced hepatic cirrhosis. We chose these patient scans because of the heterogeneous appearance of the diseased tissues. The patients with PDAC and advanced hepatic cirrhosis received contrast-enhanced abdominal CT exams using a 64 slice CT scanner (Discovery CT750 HD; GE Medical Systems, Milwaukee, Wisconsin) with the following scan parameters: tube voltage of 120 kVp, noise index of 14, tube current modulation ranging from 220 to 380 mA, 0.7 s rotation time, and pitch of 0.984. The images were reconstructed using a 512×512 matrix, the filtered back projection reconstruction algorithm, and a standard convolution kernel. The reconstructed slice thickness was 2.5 mm, with an interval of 2.5 mm. Intravenous contrast administration included 150 mL of iodinated contrast material at 4 mL/s (Iohexol 300 mgI/mL, Omnipaque 300, GE Healthcare, Cork, Ireland), respectively. The tumors were manually outlined by radiologists on the axial scans [window/level: 400/40 Hounsfield unit (HU)] using Volume Viewer on Advantage volume share 7 (GE Medical Systems, Milwaukee, Wisconsin). The CT scan of the patient with NSCLC was from a publicly available dataset hosted by the Cancer Imaging Archive.[20,32,33] The patient was imaged on a 16-slice CT scanner (Lightspeed, General Electric, Madison, Wisconsin). Images were reconstructed using a standard and lung convolution kernel. The scan data from the standard convolution kernel were used in this study. Further details of image acquisition parameters are available in the associated publication.[20] Across all patient scans, the in-plane pixel size ranged from 0.695 to 0.977 mm (mean = 0.851 mm). The tumors' maximum diameter ranged from 21 to 61 mm, with a mean of 46 mm. The largest tumor diameters were manually measured on a transverse image plane viewed with Volume Viewer on Advantage volume share 7 (GE Medical Systems, Milwaukee, Wisconsin) by an experienced radiologist. The placement of the tumors within the background cirrhotic liver was arbitrary.

2.2 *Phantom Model Fabrication*

Figures 1(a)–1(e) show a graphical overview of the workflow used to 3D print the radiomic phantom. The multimaterial 3D printer used in this study can simultaneously deposit up to three different photopolymer resins.[34] The resolution of a single droplet of resin in the x-y direction and the layer thickness (the z-direction) is on the order of $48 \times 84 \times 30$ μm, which is smaller than the resolution of a typical CT scanner ($\sim 0.5 \times 0.5 \times 0.6$ mm).[34] The selection of printing material was determined by scanning several solid samples of available resin materials using the patient abdominal CT protocol described in Sec. 2.1 and measuring the HU values. The two materials that had the highest and lowest HU values were selected.

At first, an experienced radiologist segmented each tumor and a circular portion of the cirrhotic liver from the scans at their original resolution. The segments were then individually

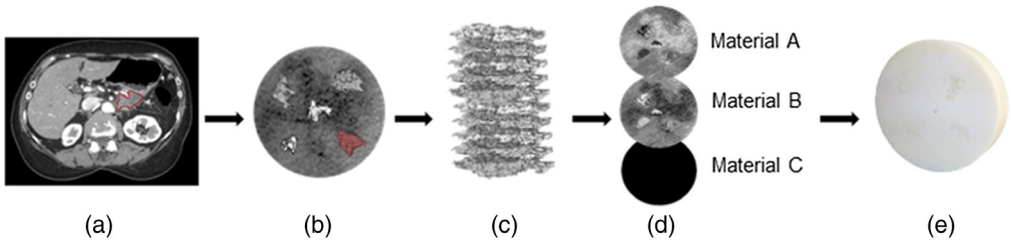


Fig. 1 Workflow to generate 3D-printed phantom. (a) The tumor was segmented from the patient CT exam. A cross-sectional slice from a single patient's CT image shows the contoured PDAC. (b) Binary masks of the tumors were generated and then used to replace the background voxel values with the tumor voxel intensity values. (c) The combined volume was then supersampled to the resolution of the 3D printer and stacked into slices. Each slice from (c) was then dithered using the Floyd–Steinberg dithering algorithm into binary raster files. (d) Three sets of raster files were generated, one for each resin material. These files define the spatial location of each resin material. (e) Resultant 3D print of the combined volume. Due to the material used, visualizing the internal structure is not possible with the naked eye.

normalized to have voxel intensity values between 0 and 1. The normalized ratios were used to determine the proportion of resin material deposited into a single voxel.[35] Binary images of the individual tumor volumes were then used to mask an area over the cirrhotic liver images [Fig. 1(b)]. Next, we merged the volumes without modifying the original HU values or the normalized ratios of each scan. The combined slices, as shown in Fig. 1(b), were then supersampled using the Whittaker–Shannon (SINC) interpolation method to the resolution of the 3D printer [Fig. 1(c)]. Finally, each slice was dithered using the Floyd–Steinberg dithering algorithm into binary raster files.[36] The raster files encode the spatial location over which each material is deposited [Fig. 1(d)]. The disc shape of the phantom was designed to have a diameter of 18 cm and thickness of 4 cm. These dimensions were selected so that the disc would fit within a tissue equivalent enclosure that was originally used with a commercially available low-contrast helical CT QC phantom (model 061, CIRS).

Since the printer can simultaneously print with three different resin materials, three sets of bitmap files were generated, one set for each resin material. Within the first raster file, a value of 1 indicates the deposition of material A, and a value of 0 indicates that material A will not be deposited. The second set of raster files (material B) was generated by inverting material A files so that a value of 0 now had a value of 1. The third set of bitmaps consisted of all zeros since two materials with opposing densities were enough to generate the desired contrast differences. The resulting 3D print is shown in Fig. 1(e).

2.3 *Computed Tomography Scan Modes*

A 64 slice CT scanner (HD750, General Electric, Madison, Wisconsin) was used to acquire 30 repeat scans of the radiomic phantom. The scanning parameters were as follows: 120 kVp, 280 mA, 0.7 s, pitch of 0.984, filtered back projection algorithm with a standard kernel, total collimation of 40 mm, display field of view (DFOV) 250 mm, reconstructed slice thickness, and interval of 1.25 mm. The phantom was centered in the gantry using the system onboard laser alignment lights. The associated volume CT dose index was 15.96 mGy. The average radiomic feature values determined from this protocol were considered the reference in percent deviation (pDV) calculations.[37] The deviation from the reference was determined by rescanning the phantom, sequentially, five times, without movement between scans using the scan modes listed in Table 1. All parameters of the reference protocol remained fixed while each scan mode was implemented. Figure 2 shows cross-sectional axial slices of the radiomic phantom scanned using each additional scan mode. In addition to commonplace scan modes, such as different tube potentials and currents, we evaluated QR feature robustness with adaptive statistical iterative reconstruction (ASiR), and the phantom positioned vertically off-center by 30 mm in the inferior

Table 1 Additional scan modes included in the study. The options were chosen due to their frequent use in the clinic.

Additional scanning modes	
Convolution kernel	Standard, lung and bone kernels[a]
ASiR	ASiR 10%, 20%,[b] and 30%
Peak tube potential (kVp)	100
Tube current (mA)	100
Phantom off-center in the *y*-axis inferior direction (mm)	30
DFOV	350 mm (0.684 mm pixel size)
Dual energy CT	Virtual monochromatic energy: 60 keV

[a]Reference kernel.
[b]Also performed as a combination of the lung kernel and ASiR 20%.

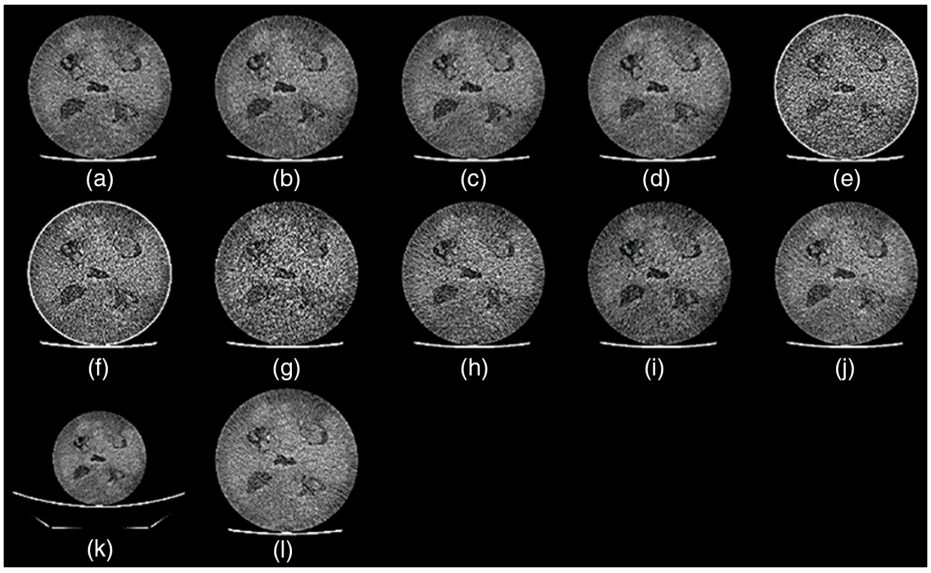


Fig. 2 Cross-sectional CT images of the radiomic phantom from each scan mode evaluated in this study. For each image, a window width of 40 and a level of 100 were applied. The differences between each scan mode are (a) baseline image reconstructed with standard kernel. (b) ASiR of 10%. (c) ASiR 20% and (d) ASiR 30%. (e) Lung kernel. (f) Lung kernel with ASiR 40%. (g) Bone kernel. (h) Reduced tube current of 100 mA. (i) Reduced tube potential of 100 kVp. (j) Dual-energy CT image reconstructed at monochromatic 70 keV. (k) Enlarged DFOV of 350 mm with a pixel size of 0.689 mm. (l) Phantom placed off-center by 30 mm. The off-center image was electronically centered within the field of view.

direction. The latter[38] is a practice commonly observed in the clinic. We evaluate its impact on QR features in this study because the off-center placement of the patient within the CT gantry misplaces the thickest portion of the bow-tie filter relative to the patient's anatomy, which leads to increased beam hardening artifacts and, consequently, increased noise or variability of CT HU values.[39] ASiR is a feature that reduces the pixel noise standard deviation while preserving structural detail and is available in 10 different strengths. As the strength or percentage of ASiR increases, the noise magnitude decreases, noise texture becomes coarser and more uniform, and images generally appear smoother.[40,41] The strengths and combinations that we use in this study are based on what we use in our clinic. The DFOV dictates the pixel size in the *x*-*y* direction. As DFOV increases, the size of the pixel increases, and the resolution in the *x*-*y* direction decreases. The rapid switching dual-energy CT variant used in this study acquires two projections nearly simultaneously while operating at a low and high peak tube potential of 80 and 140 kVp. With the two projections, reconstructing many image types is possible; these include virtual monochromatic images (VMI) that depict anatomy from the viewpoint of a monochromatic x-ray source ranging in energy from 40 to 140 keV or material density images. For this study, we chose to evaluate radiomic feature robustness on DECT scans reconstructed with a VMI of 60 keV. The scan modes were chosen due to their use in the clinic.

2.4 *Radiomic Feature Extraction*

The computational environment for radiobiological research (CERR)[42] was used to extract the prognostic QR features listed in Table 2. We chose these features because previous literature reports illustrated their potential prognostic capabilities for NSCLC and PDAC. They were extracted from the original images without any preprocessing, such as image smoothing or interpolation of voxel sizes, and the settings used for feature calculation were as follows: (1) the images were discretized using a fixed bin width of 25. (2) The average value of each texture feature was computed over all 13 directions to obtain rotational invariance. (3) For the prognostic PDAC QR features, images were discretized using a bin width of 25 and a patchwise volume of

Table 2 Prognostic radiomic features analyzed for repeatability and deviation relative to reference values.

Radiomic features	
NSCLC[43]	First-order energy[a]
	GLRLM: gray-level non-uniformity (GLN)[b]
	HLH wavelet preprocessed GLRLM GLN[c]
PDAC[2]	First-order energy[a]
	First-order entropy[a]
	GLCM-contrast[d]
	GLCM-dissimilarity[d]

Note: GLRLM, gray-level run length matrix; NSCLC, nonsmall cell lung carcinoma; PDAC, pancreatic ductal adenocarcinoma.
[a]First-order statistical features.
[b]Gray-level run length matrix feature.
[c]Wavelet-based feature.
[d]Gray-level co-occurrence matrix texture features.

$2 \times 2 \times 2$ mm voxels. Detailed descriptions of the feature definitions can be found in Refs. 42–44.

2.5 *Statistical Analysis*

The structural similarity index (SSIM)[45] was used to calculate the similarity between the original cirrhotic liver and the resulting 3D print. Repeatability (i.e., precision) of radiomic features was evaluated using the within-subject coefficient of variation (wCV, %):[46]

$$\mathrm{wCV\%} = \frac{\sigma_w}{\mu} \times 100, \tag{1}$$

where σ_w is the within-subject standard deviation and μ is the mean of individual radiomic features. A wCV% of <10% was considered to be repeatable. The 95% confidence interval (CI) was calculated using chi-squared (x^2) as the pivotal statistic as follows:

$$\mathrm{CI}(95\%) = \sqrt{\frac{Nxw(\mathrm{wCV}^2)}{x^2_{n,\propto}}}, \tag{2}$$

where N is the number of tumors and $x^2_{n,\propto}$ is the percentile of the distribution with n degrees of freedom. The lower bound α is 0.975, and the upper bound α is 0.025.

The pDV (%) of radiomic feature derived from the additional scan modes was calculated as follows:

$$\mathrm{pDV}(\%) = \left(\frac{\overline{f_n} - \hat{f}_o}{\hat{f}_o} \pm \delta\mathrm{pDV}\right) \times 100, \tag{3}$$

where $\overline{f_n}$ is the average value of the radiomic feature extracted from images of each tumor across the different scanning parameters and $\hat{f}_o$ is the average of the reference value, as described above.

The one-sample Wilcoxon signed-rank test was used to determine equality between the reference feature and the median feature value derived from the additional scan modes. A p-value of <0.05 was considered significant. The effect size was calculated as

$$r = \frac{Z}{\sqrt{N_{\text{obs}}}}, \tag{4}$$

where Z is the z-score, N_{obs} is the number of observations, and r ranges from -1 to 1. The 95th percentile CIs for the effect size estimate were determined using 100 bootstrap samples. All statistical analyses were completed using RStudio.[47]

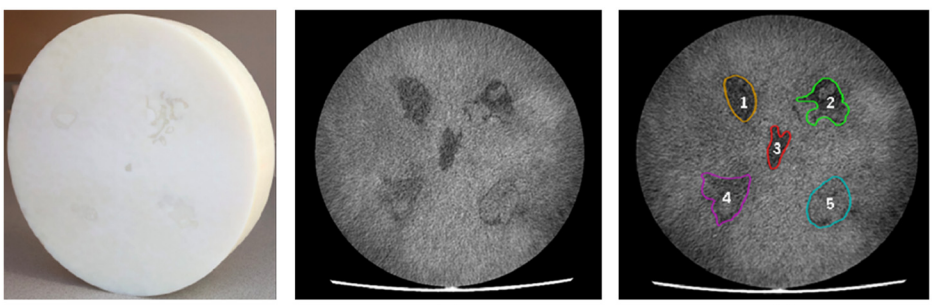
3 Results

3.1 *Hounsfield Unit of Printing Materials and Structural Similarity*

Figure 3 shows the resulting 3D-printed phantom, a cross-sectional CT scan, and the physician drawn contours overlaid onto each tumor. The overall time taken to 3D print the phantom was ~8 h. The final radiomic phantom was circular, with a measured diameter of 176 ± 0.2 mm and an axial length of 42 ± 0.2 mm. The two-resin materials with the lowest (65 ± 5 HU) and highest (125 ± 5 HU) CT numbers were VeroWhite (material A) and TangoPlus (material B). The SSIM between the cirrhotic liver background [Fig. 4(b)] and resultant 3D print [Fig. 4(c)] was 0.71. An SSIM value closer to 1 suggests more similarity between images.

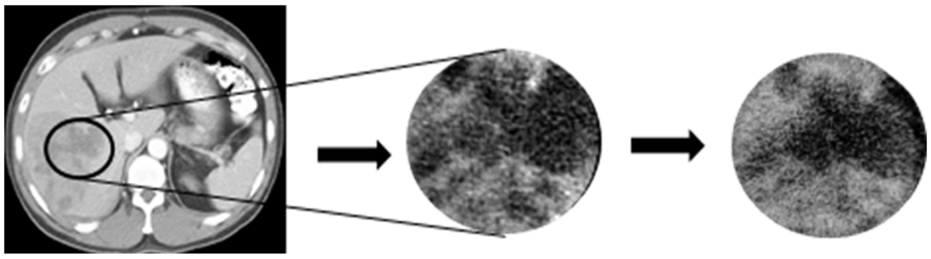
3.2 *Repeatability and Percent Deviation*

The repeatability of the prognostic NSCLC and PDAC radiomic features is shown in Fig. 5, where wCV (%) <1.0% across features. Figure 6 shows the pDV for the NSCLC radiomic features. The average pDV of first-order energy was 0.01% (range: −0.49% to 0.89%, $p = 0.290$) across all scan modes. The average pDV of GLRLM gray-level non-uniformity (GLN) was



(a) (b) (c)

Fig. 3 (a) The 3D-printed radiomic phantom. (b) Axial slice generated from a CT scan shows the embedded tumors within the background tissue. (c) The tumor contours were generated, and the tumor types were labeled as 1–Non-small cell lung carcinoma; 2 to 5–Pancreatic ductal adenocarcinoma.



(a) (b) (c)

Fig. 4 (a) An axial slice from a patient CT showing the region of interest (ROI) around the heterogeneous hepatic tissue. (b) A cropped and expanded view of the portion of the patient's liver on which the background of the 3D print was modeled. (c) An axial slice of the resulting 3D print CT scan.

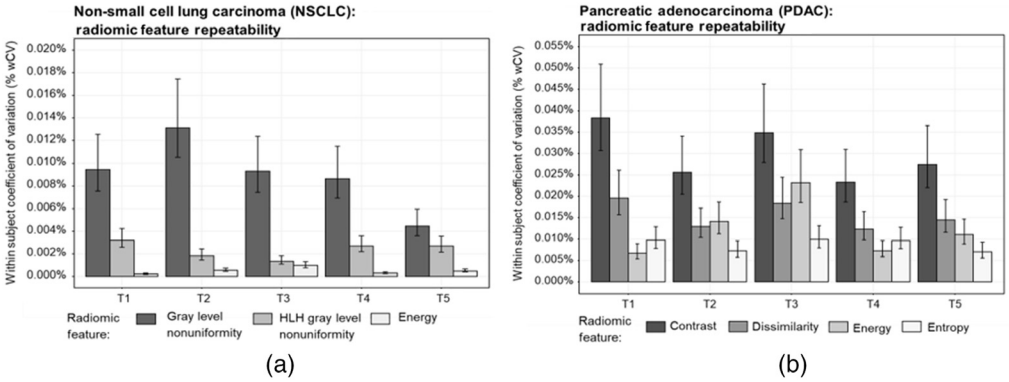


Fig. 5 The within-subject coefficient of variation (wCV, %) for prognostic non-small cell lung carcinoma radiomic features extracted from each tumor. The wCV was computed from the 30 repeated CT scans acquired with the reference protocol. The 95th percentile confidence intervals are displayed for each feature value.

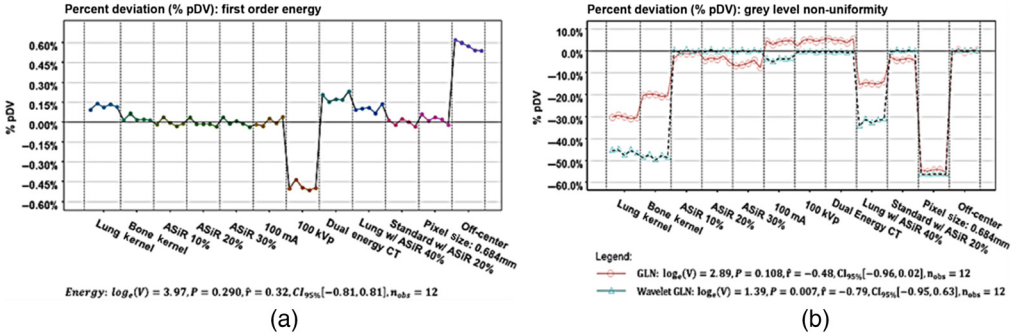


Fig. 6 The percent deviation (pDV, %) and one-sample Wilcoxon signed-rank test compare the prognostic NSCLC features. Comparisons are being made between the average feature value derived from the 30 repeat scans and the additional scan modes. (a) First-order energy, (b) gray-level non-uniformity (GLN), and HLH wavelet GLN.

10.2% (range: −55.2% to 5.57%, $p = 0.108$) [Fig. 6(b)]. The application of ASiR 40% to images reconstructed with the lung kernel resulted in the pDV for GLRLM GLN decreasing by a factor of 2, from −30% to −15%. The average pDV for HLH GLN pDV was 15.7% (range: −56.3% to 0.52%, $p = 0.007$). Similar to GLRLM GLN, pDV for HLH GLN decreased when ASiR 40% was applied to images reconstructed with the lung kernel. Figure 7 shows the pDV of prognostic PDAC radiomic features across scan modes. With the application of ASiR 10% to 30%, the pDV for contrast, dissimilarity, and entropy increased in the negative direction for all tumors, but overall, the deviation remained below 30%. Across all tumor types, the pDV for GLCM-contrast and GLCM-dissimilarity exceeded 40% and 20% when the phantom was scanned with the reduced dose scan modes and with the application of DECT. In addition, the pDV for GLCM-contrast was ≤10% when the phantom was scanned with a larger pixel size of 0.689 mm and the different strengths of ASiR.

4 Discussion

We have devised a method to 3D print what is seen on a CT scan to a physical imaging phantom using commercially available, multimaterial 3D-printing technology and software Voxel print. The method provides a straightforward approach to producing fit-for-use phantoms that validate longitudinally stable QR feature values in a clinical setting.

The results in this study also illustrate that previously discovered prognostic QR features were repeatable but sensitive to different scan modes. In contrast to previous works in which

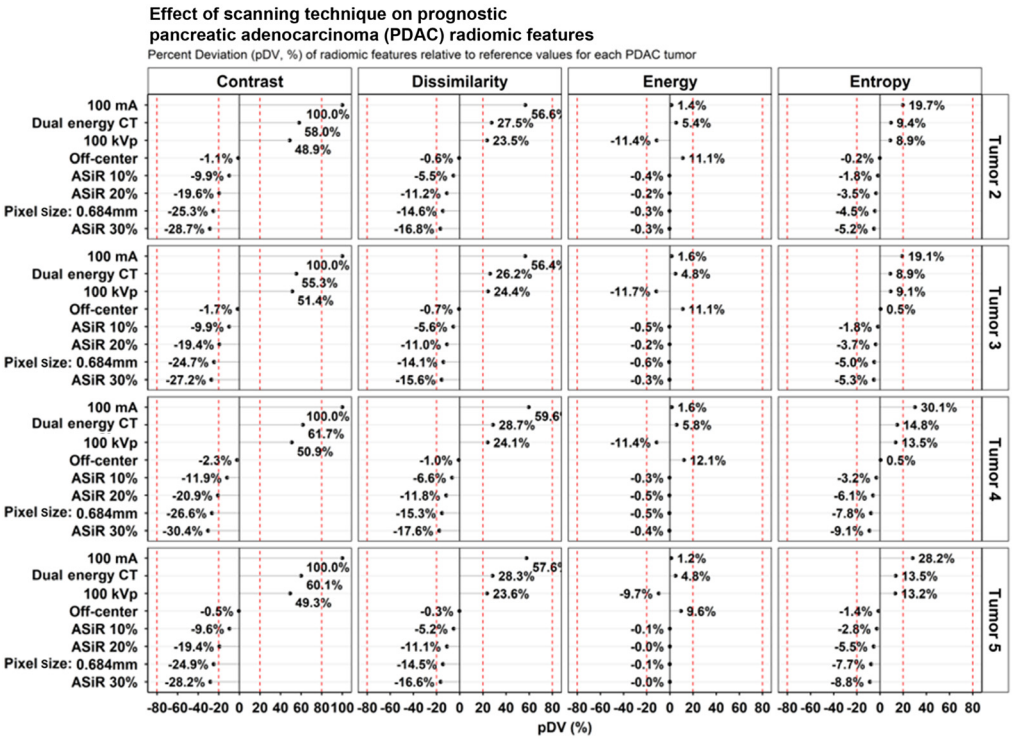


Fig. 7 Dot plots of the change in radiomic feature values as a function of each scanning technique and tumor type.

uniform phantoms were used,[48] we observe that, across all scan modes, first-order energy was not significantly affected by the reconstruction algorithm or other scan modes (pDV $= \pm 1.0\%$, $P > 0.290$). The deviation for GLN and HLH GLN increased as image noise increased, but the pDV reduced with the application of ASiR 40%. The findings suggest that increased scrutiny or exclusion of GLN and HLH GLN is warranted for future studies. The lack of reproducibility in some QR features, as noted by the high pDV, demonstrates the need for improved validation approaches. A correlation between QR features derived from tumors seen on abdominal CT scans and the underlying tumor microenvironment needs to be interpreted with caution.

The concept of quantifying diseased tissue seen on CT scans is not new. Efforts to quantify bone mineral density (BMD) from CT scans date to the mid-1970s.[49,50] However, similar to the issues that plague current QR efforts, BMD measurements were variable across scanning protocols and devices.[50] To address the lack of standardization and variability across CT scanners, the European spine phantom (ESP) was developed. It was designed to be a practical but effective tool that could standardize and cross calibrate BMD measurements made across CT scans. During its redesign in the 1990s, an international consortium published critical characteristics that a quantitative CT QC phantom should possess.[50] The characteristics included a phantom that is (1) geometrically defined with realistic dimensions, (2) fit-for-use across CT scanners, (3) closely anthropomorphic so that standard patient protocols can be used without alteration, and (4) composed of limited materials with a range of attenuation characteristics, so linearity can be assessed. Because of technological advances in CT scanning hardware and software, efforts to develop robust QR features would benefit from employing phantoms that meet the design criteria used to manufacture the ESP.

Additional caution is required when interpreting results derived from phantoms data. The applicability or translation of results from phantom studies to the clinic depends on the realism of the phantom components.[51] The radiomic phantom is crudely anthropomorphic (i.e., it lacks the fat planes or does not incorporate the influence of beam hardening artifacts from contrast or bone). However, the purpose of the phantom is to establish the minimum performance requirement for the QR features[51] and inform about the continued stability as CT scanning technology evolves. A distinct advantage of the radiomic phantom is that, by incorporating realistic shapes

and patterns, it offers tests for additional quantities, such as measuring tumor diameter, volume, or the comparative performance of manual, semiautomated, and automated segmentation techniques. In general, 3D printing offers the opportunity to generate the ground-truth value for parameters to be measured.[24] Although multimaterial 3D printing offers several advantages relative to previous approaches, some issues requiring further attention can be addressed with additional investigations. First, the density and HU value of available photopolymer resins is limited, but recent investigations into doping agents show that the HU value range of current resins could be increased.[52] Second, the size of the 3D print bed restricts the maximum anatomical area that could be printed. The Objet 260 printer used in this study can reproduce a volume with a maximum size of $255 \times 252 \times 200$ mm. However, the size of phantoms may be overcome by designing modular phantoms. Third, the phantom remained stationary during imaging and was not able to assess the impact of motion.

5 Conclusions

The strategy proposed here is to derive ROI tumor and normal tissue features from human CT scans and then to custom print a CT phantom capturing a facsimile of those features. Such a phantom can then be used to measure the variability of imaging features as well as the stability of the overall QR feature, and has the potential to be used as part of the QR QA process. In measurements using our prototype, we found that some previously reported prognostic radiomic features are noisy in practice and need to be used with caution or preferably excluded from clinical signature implementations. Personalized, custom-designed phantoms present a flexible, yet practical, way to validate and compare QR signatures over time and across systems.

Disclosures

One author in this study, CK, was employed at a commercial company, Paige, New York, during the preparation of this paper. This company played no role in the sponsorship, design, data collection and analysis, decision to publish, or preparation of the paper. The other authors have no relevant financial interests in the paper and no other potential conflicts of interest to disclose.

Acknowledgments

This research was funded in part through the National Institutes of Health/National Cancer Institute Cancer Center, Grant No. P30 CA008748.

References

1. D. Gu et al., "CT radiomics may predict the grade of pancreatic neuroendocrine tumors: a multicenter study," *Eur. Radiol.* **29**(12), 6880–6890 (2019).
2. L. Cozzi et al., "Computed tomography based radiomic signature as predictive of survival and local control after stereotactic body radiation therapy in pancreatic carcinoma," *PLoS One* **14**(1), e0210758 (2019).
3. S. H. Bak et al., "Prognostic impact of longitudinal monitoring of radiomic features in patients with advanced non-small cell lung cancer," *Sci. Rep.* **9**(1), 8730 (2019).
4. M. Shafiq-Ul-Hassan et al., "Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels," *Med. Phys.* **44**(3), 1050–1062 (2017).
5. A. Traverso et al., "Repeatability and reproducibility of radiomic features: a systematic review," *Int. J. Radiat. Oncol. Biol. Phys.* **102**(4), 1143–1158 (2018).
6. R. Yamashita et al., "Radiomic feature reproducibility in contrast-enhanced CT of the pancreas is affected by variabilities in scan parameters and manual segmentation," *Eur. Radiol.* **30**(1), 195–205 (2020).

7. M. Meyer et al., "Reproducibility of CT radiomic features within the same patient: influence of radiation dose and CT reconstruction settings," *Radiology* **293**(3), 583–591 (2019).
8. B. S. Erdal et al., "Are quantitative features of lung nodules reproducible at different CT acquisition and reconstruction parameters?" *PLoS One* **15**(10), e0240184 (2020).
9. B. A. Varghese et al., "Reliability of CT-based texture features: phantom study," *J. Appl. Clin. Med. Phys.* **20**(8), 155–163 (2019).
10. I. Fornacon-Wood et al., "Radiomics as a personalized medicine tool in lung cancer: separating the hope from the hype," *Lung Cancer* **146**, 197–208 (2020).
11. A. Ibrahim et al., "Radiomics for precision medicine: current challenges, future prospects, and the proposal of a new framework," *Methods* **188**, 20–29 (2021).
12. P. Lambin et al., "Radiomics: the bridge between medical imaging and personalized medicine," *Nat. Rev. Clin. Oncol.* **14**(12), 749–762 (2017).
13. J. E. Park et al., "Quality of science and reporting of radiomics in oncologic studies: room for improvement according to radiomics quality score and TRIPOD statement," *Eur. Radiol.* **30**(1), 523–536 (2020).
14. P. P. Lin et al., "Specification and acceptance testing of computed tomography scanners," *American Association of Physicists in Medicine Report no. 39*, American Institute of Physics, New York (1993).
15. L. R. F. Branco et al., "Technical note: proof of concept for radiomics-based quality assurance for computed tomography," *J. Appl. Clin. Med. Phys.* **20**(11), 199–205 (2019).
16. J. Solomon and E. Samei, "Quantum noise properties of CT images with anatomical textured backgrounds across reconstruction algorithms: FBP and SAFIRE," *Med. Phys.* **41**(9), 091908 (2014).
17. D. Mackin et al., "Measuring CT scanner variability of radiomics features," *Invest. Radiol.* **50**(11), 757 (2015).
18. A. Midya et al., "Influence of CT acquisition and reconstruction parameters on radiomic feature reproducibility," *J. Med. Imaging* **5**(1), 011020 (2018).
19. R. B. Ger et al., "Comprehensive investigation on controlling for CT imaging variabilities in radiomics studies," *Sci. Rep.* **8**(1), 13047 (2018).
20. B. Zhao, L. H. Schwartz, and M. G. Kris, "Data from RIDER_Lung CT," The Cancer Imaging Archive (2015).
21. U. Mahmood et al., "Investigating the robustness neighborhood gray tone difference matrix and gray level co-occurrence matrix radiomic features on clinical computed tomography systems using anthropomorphic phantoms: evidence from a multivendor study," *J. Comput. Assist. Tomogr.* **41**(6), 995–1001 (2017).
22. S. Rizzo et al., "Radiomics: the facts and the challenges of image analysis," *Eur. Radiol. Exp.* **2**(1), 36 (2018).
23. Q. Huang et al., "Interobserver variability in tumor contouring affects the use of radiomics to predict mutational status," *J. Med. Imaging* **5**(1), 011005 (2017).
24. E. Samei et al., "Design and fabrication of heterogeneous lung nodule phantoms for assessing the accuracy and variability of measured texture radiomics features in CT," *J. Med. Imaging* **6**(2), 021606 (2019).
25. J. I. Gear et al., "Abdo-Man: a 3D-printed anthropomorphic phantom for validating quantitative SIRT," *EJNMMI Phys.* **3**(1), 17 (2016).
26. S. Leng et al., "Construction of realistic liver phantoms from patient images using 3D printer and its application in CT image quality assessment," *Proc. SPIE* **9412**, 94124E (2015).
27. F. P. Melchels, J. Feijen, and D. W. J. B. Grijpma, "A review on stereolithography and its applications in biomedical engineering," *Biomaterials* **31**(24), 6121–6130 (2010).
28. L. G. Kessler et al., "The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions," *Stat. Methods Med. Res.* **24**(1), 9–26 (2015).
29. G. L. A. Pardo et al., "3D printing of anatomically realistic phantoms with detection tasks to assess the diagnostic performance of CT images," *Eur. Radiol.* **30**(8), 4557–4563 (2020).
30. P. Jahnke et al., "Paper-based 3D printing of anthropomorphic CT phantoms: feasibility of two construction techniques," *Eur. Radiol.* **29**(3), 1384–1390 (2019).

31. S. K. Eshkalak et al., "The role of three-dimensional printing in healthcare and medicine," *Mater. Des.* **194**, 108940 (2020).
32. K. Clark et al., "The cancer imaging archive (TCIA): maintaining and operating a public information repository," *J. Digital Imaging* **26**(6), 1045–1057 (2013).
33. B. Zhao et al., "Evaluating variability in tumor measurements from same-day repeat CT scans of patients with non-small cell lung cancer," *Radiology* **252**(1), 263–272 (2009).
34. E. Doubrovski et al., "Voxel-based fabrication through material property mapping: a design method for bitmap printing," *Comput.-Aided Des.* **60**, 3–13 (2015).
35. J. Solomon et al., "Comparison of low-contrast detectability between two CT reconstruction algorithms using voxel-based 3D printed textured phantoms," *Med. Phys.* **43**(12), 6497–6506 (2016).
36. R. W. Floyd, "An adaptive algorithm for spatial gray-scale," in *Proc. Soc. Inf. Disp.*, pp. 75–77 (1976).
37. J. D. Newell, Jr., J. Sieren, and E. A. Hoffman, "Development of quantitative CT lung protocols," *J. Thorac. Imaging* **28**(5), 266–271 (2013).
38. T. P. Szczykutowicz, A. DuPlissis, and P. J. Pickhardt, "Variation in CT number and Image noise uniformity according to patient positioning in MDCT," *AJR Am. J. Roentgenol.* **208**(5), 1064–1072 (2017).
39. T. Kaasalainen et al., "Effect of patient centering on patient dose and image noise in chest CT," *AJR Am. J. Roentgenol.* **203**(1), 123–130 (2014).
40. J. Hsieh, *Computed Tomography: Principles, Design, Artifacts, and Recent Advances*, SPIE Press, Bellingham, Washington, DC (2003).
41. S. Brady, B. Yee, and R. J. Kaufman, "Characterization of adaptive statistical iterative reconstruction algorithm for dose reduction in CT: a pediatric oncology perspective," *Med. Phys.* **39**(9), 5520–5531 (2012).
42. A. P. Apte et al., "Technical note: extension of CERR for computational radiomics: a comprehensive MATLAB platform for reproducible radiomics research," *Med. Phys.* **45**(8), 3713–3720 (2018).
43. H. J. Aerts et al., "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nat. Commun.* **5**, 4006 (2014).
44. A. Zwanenburg et al., "The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping," *Radiology* **295**(2), 328–338 (2016).
45. Z. Wang et al., "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).
46. D. L. Raunig et al., "Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment," *Stat Methods Med Res* **24**(1), 27–67 (2015).
47. R. J. R. Team Inc., "RStudio: integrated development for R," R. J. R. Team, Inc., Boston, MA, 2017, http://www.rstudio.com.
48. B. Zhao et al., "Exploring variability in CT characterization of tumors: a preliminary phantom study," *Transl. Oncol.* **7**(1), 88–93 (2014).
49. K. L. Boedeker et al., "Emphysema: effect of reconstruction algorithm on CT imaging measures," *Radiology* **232**(1), 295–301 (2004).
50. W. A. Kalender et al., "The European Spine Phantom: a tool for standardization and quality control in spinal bone mineral measurements by DXA and QCT," *Eur. J. Radiol.* **20**(2), 83–92 (1995).
51. N. A. Obuchowski et al., "Statistical issues in the comparison of quantitative imaging biomarker algorithms using pulmonary nodule volume as an example," *Stat. Methods Med. Res.* **24**(1), 107–140 (2015).
52. T. S. Yoo et al., "Toward quantitative x-ray CT phantoms of metastatic tumors using rapid prototyping technology," in *IEEE Int. Symp. Biomed. Imaging: From Nano to Macro*, pp. 1770–1773 (2011).

Usman Mahmood is a lead diagnostic medical physicist at Memorial Sloan Kettering Cancer. His research interests are in radiomics, machine learning, and 3D printing anatomically realistic phantoms for the evaluation of computed tomography systems.

Aditya P. Apte is a research faculty member at Memorial Sloan Kettering Cancer. He received his PhD in the Department of Mechanical Engineering at the University of Texas, Arlington, Texas, USA. His research interests include the development of radiation oncology informatics tools and systems to enable multi-institutional research and translation to clinic.

Christopher Kanan is an associate professor at the Rochester Institute of Technology (RIT), a Visiting Associate Professor at Cornell Tech, and the director of AI Science at Paige.ai, inc. At RIT, his lab conducts basic and applied research in deep learning. He received his PhD in computer science from the University of California at San Diego and an MS degree in computer science from the University of Southern California. He completed his postdoc at Caltech.

David D.B. Bates is a body radiologist and Director of CT at Memorial Sloan Kettering Cancer Center (MSKCC) and has interests in dual-energy CT, oncologic imaging and imaging biomarkers. He is a member of the MSKCC Colorectal Cancer Disease Management Team and an active member of the Society of Abdominal Radiology's Rectal and Anal Cancer Disease Focus Panel.

Jung Hun Oh is an assistant attending computer scientist at Memorial Sloan Kettering Cancer Center. He uses cutting-edge computational and statistical methods, informed by bioinformatics and machine learning techniques, to identify diagnostic biomarkers and imaging markers and to build models that predict radiation treatment outcomes.

Yusuf Emre Erdi is a clinical attending physicist at Memorial Sloan Kettering Cancer Center. He is the section head of Diagnostic Imaging Physics group that oversees QA/QC protocols for radiation emitting diagnostic imaging units. His past research interest was integrating PET images into the treatment planning systems, however, currently he is interested in performance characteristics of diagnostic monitors.

John Nguyen graduated from the Cooper Union in 2020 with a degree in Mechanical Engineering. He is currently a research technician at Weill Cornell in the Institute for Precision Medicine. He has various research experiences in 3D printing, imaging, wet lab techniques, and computational biology. He hopes to pursue a PhD in computational biology and later bring his experiences to innovative biotech startups that will turn research findings into actionable treatments for patients.

Joseph O. Deasy is the chair of the Department of Medical Physics at Memorial Sloan Kettering Cancer Center. He and his group apply statistical modeling to analyze large, complex datasets to understand the relationship between treatment, patient, and disease characteristics. He is also a chief of the Service for Predictive Informatics established to work with the Departments of Radiation Oncology, Radiology, and Pathology to conduct big medical data analyses, and to advance new clinical decision support tools.

Amita Shukla-Dave is an attending physicist at Memorial Sloan Kettering Cancer Center. Her group's research focuses on developing and implementing advanced quantitative imaging biomarkers derived from MR imaging physics techniques, including MR fingerprinting (MRF), diffusion-weighted MRI (DW-MRI), dynamic contrast-enhanced MRI (DCE-MRI), and MR spectroscopy (MRS) for clinical application in cancer.

Biographies of the other authors are not available.