



OPEN

Assessment of CT to CBCT contour mapping for radiomic feature analysis in prostate cancer

Ryder M. Schmidt^{1,2}, Rodrigo Delgadillo¹, John C. Ford^{1,2}, Kyle R. Padgett¹, Matthew Studenski¹, Matthew C. Abramowitz¹, Benjamin Spieler¹, Yihang Xu^{1,2}, Fei Yang¹ & Nesrin Dogan¹✉

This study provides a quantitative assessment of the accuracy of a commercially available deformable image registration (DIR) algorithm to automatically generate prostate contours and additionally investigates the robustness of radiomic features to differing contours. Twenty-eight prostate cancer patients enrolled on an institutional review board (IRB) approved protocol were selected. Planning CTs (pCTs) were deformably registered to daily cone-beam CTs (CBCTs) to generate prostate contours (auto contours). The prostate contours were also manually drawn by a physician. Quantitative assessment of deformed versus manually drawn prostate contours on daily CBCT images was performed using Dice similarity coefficient (DSC), mean distance-to-agreement (MDA), difference in center-of-mass position (Δ CM) and difference in volume (Δ Vol). Radiomic features from 6 classes were extracted from each contour. Lin's concordance correlation coefficient (CCC) and mean absolute percent difference in radiomic feature-derived data (mean $|\% \Delta|RF$) between auto and manual contours were calculated. The mean (\pm SD) DSC, MDA, Δ CM and Δ Vol between the auto and manual prostate contours were 0.90 ± 0.04 , 1.81 ± 0.47 mm, 2.17 ± 1.26 mm and $5.1 \pm 4.1\%$ respectively. Of the 1,010 fractions under consideration, 94.8% of DIRs were within TG-132 recommended tolerance. 30 radiomic features had a CCC > 0.90 and 21 had a mean $|\% \Delta|RF$ < 5%. Auto-propagation of prostate contours resulted in nearly 95% of DIRs within tolerance recommendations of TG-132, leading to the majority of features being regarded as acceptably robust. The use of auto contours for radiomic feature analysis is promising but must be done with caution.

Radiomics is a promising tool with potential diagnostic, prognostic and predictive powers. The extraction and analysis of quantitative radiological features provides valuable information before, during and after radiation therapy (RT)¹. Previous studies have linked several radiomic features directly to patient survival². Research has shown the power of radiomics for many disease sites; however, these studies also show variability with respect to imaging modality, reconstruction algorithms, feature selection, and volume of interest (VOI)³⁻⁹. Several groups have studied the robustness of radiomic features with respect to contouring variability^{3-5,10-12}. Contours are typically created by a trained radiation oncologist; however, inter-, and intra-observer contouring variation can still be significant when considering radiomics^{10,13}.

A recent study by Yang et al., investigated the impact of contouring variability on PET-based radiomic features for lung cancer¹⁴. The study demonstrated that the impact of contour uncertainty on PET-based radiomic features varied widely and cautioned predictive use in the context of contouring uncertainty for models involving PET-based radiomic features. A study by Pavic et al., examined intra-observer variation effects on radiomic features extracted from CT images¹². This study extracted a total of 137 radiomic features from planning CT images of head and neck cancer patients and warned that variation in delineation can significantly affect some radiomic features.

On-board imaging (OBI) utilizing megavoltage (MV) and kilovoltage (kV) cone beam CT (CBCT) is a widely used imaging technique for daily patient bony alignment and prostate marker alignment¹⁵. The prostate can deform and rotate daily due to differential bladder and rectal filling resulting in suboptimal dosimetry over the course of treatment¹⁶. By utilizing CBCT setup images, changes in anatomy can be accounted for at the time of treatment delivery¹⁷. Deformable image registration (DIR) can automatically propagate the contours drawn on

¹Department of Radiation Oncology, University of Miami Miller School of Medicine, 1475 NW 12th Ave, Miami, FL 33136, USA. ²Department of Biomedical Engineering, University of Miami, Coral Gables, FL, USA. ✉email: ndogan@med.miami.edu

the planning CT (pCT) to daily CBCT images, accounting for the anatomical changes and allowing adaptive radiotherapy (ART)^{18,19}. The American Association of Physicists in Medicine (AAPM) Task Group 132 (TG-132) provides recommendations on the use of image registration and fusion algorithms and provides quantitative methods of evaluating DIR accuracy²⁰.

Various studies have investigated the accuracy of the DIR algorithms for the automatic creation of contours (auto contours) for prostate cancer; however, these studies often used very small sample sizes^{18,21–23}. A study by Woerner et al.¹⁸, evaluated the DIR performance from pCT to CBCT that was acquired near the end of treatment for 6 prostate, 5 head and neck and 5 pancreas patients. The small sample size of their study limited their analysis to organs at risk (OAR). The authors cautioned the use of automatic DIR workflows to perform contour deformation to assess the changes during the course of treatment. Another study by Thor et al., evaluated DIR performance from pCT to CBCT through the course of prostate cancer treatment for 5 patients²⁴. The study concluded that more advanced imaging and/or DIR algorithms should be developed to confidently use DIR workflows for contour deformation.

Unlike CT, PET and MR images that are used for diagnostic and treatment planning purposes, CBCT images are used for daily patient setup prior to radiation delivery and are collected as the current standard of care on a daily basis. Most radiomic studies performed thus far have been evaluated in a pre-treatment setting and are lacking a knowledge of early tumor response to therapy, hindering the possibility for timely treatment adaptation. Hence, the day-to-day radiomic feature changes of the tumor obtained from CBCT may offer a possibility of treatment adaptation during the early course of treatment, distinct from radiomic predictions derived in the pre-treatment settings. These features can be examined for their use in early response assessment.

The automatic propagation of pCT contours to daily CBCT can also be used in the context of radiomics. However, to our knowledge, the robustness of radiomic features to varying prostate contours on daily CBCTs has not been previously examined. By updating radiomic feature derived data on a more frequent basis, as could be done through utilizing daily CBCT derived data, radiomic features can help influence clinical decision making. The goal of this study is to utilize a commercially available DIR algorithm to deform manual prostate contours to the daily CBCTs and determine the robustness of radiomic features to DIR-based contour propagation.

Methods and materials

Patient selection. Twenty-nine prostate cancer patients who were treated with volumetric modulated arc therapy (VMAT) and had daily CBCT images were considered. The ethical approval for this study was obtained from the University of Miami Institutional Review Board (IRB). Written informed consent was obtained from all patients in this study. The data was retrospectively collected and analyzed. All methods undertaken in this work were carried out in accordance with the relevant guidelines and regulations. One patient was excluded due to prosthetic implants causing extremely poor image quality. Each patient was treated using conventional fractionation, consisting of 28–40 fractions, totaling 1,010 total fractions. Prostate volumes ranged from 15 cm³ to 92 cm³.

Image acquisition and manual contouring. Planning CT (pCT) images were acquired using Somatom Definition AS and Sensation Open (Siemens Healthineers AG, Germany), and/or Gemini TF TOF 64 (Philips, Netherlands) for each patient. The mean field-of-view (FOV) was 670 mm with a range of 492–800 mm, reconstructed with dimensions of 512 × 512 pixels, a thickness of 2 mm, and an average in-slice pixel size of 1.3 mm with a range of 0.9–1.6 mm. On the day of treatment, each patient was imaged with the same FOV using kV CBCT with 465 slices, voxel sizes of (0.9 mm, 0.9 mm, 2.0 mm) and reconstructed with dimensions of (512, 512, 232). Each patient was imaged in supine position. All 28 patients had 4 gold fiducial markers implanted in the prostate prior to imaging. Prostate volumes were manually contoured on pCT and on daily CBCT setup images by the same expert radiation oncologist who has 4 years of experience in contouring prostate cases to eliminate interobserver variation. The pCT images and daily CBCT setup images, including manually drawn contours, were uploaded to a commercial image registration software (Velocity Advanced Imaging, ver. 4.1, Varian Medical Systems, Palo Alto, CA).

Deformable image registration and delineation propagation. The DIR of pCT to daily CBCT images were performed using an Adaptive Monitoring tool available in the commercial image registration software. Figure 1 shows the DIR and delineation propagation workflow. DIR creation utilized the ‘Adaptive Monitoring’ navigator, which is comprised of three steps: (1) manual alignment, (2) rigid registration, (3) deformable registration. During the manual alignment step, a manual rigid alignment between the CBCT image and the pCT image was made using bony anatomy and the implanted gold fiducials as a guide. These were done independently of clinical set-up shifts for consistency. In step 2 the region of interest was adjusted to include the prostate and, using the manual alignment created in step 1 as a starting point, a rigid registration was created by the software. Step 3 uses the rigid registrations of step 2 as a reference to create DIRs. After the rigid registration, the DIR algorithm was utilized to deformably register the pCT to the daily CBCTs. The DIR algorithm uses an intensity-based B-spline algorithm based on the Mattes formulation; the details have been described elsewhere^{14,25}. For poor quality deformations, the deformable image registration workflow was repeated with smaller ROIs (67 fractions). Poor quality deformations were defined as a DSC < 0.75 and MDA > 3.5 mm, just beyond the TG-132 tolerance recommendations. For the three fractions that continued to produce poor quality deformations, a structure-guided deformation (hybrid DIR) was employed.

Assessment of deformable image registration. The DIRs from pCT to daily CBCTs were done using both qualitative and quantitative assessment metrics. First, a visual assessment of deformation vector fields

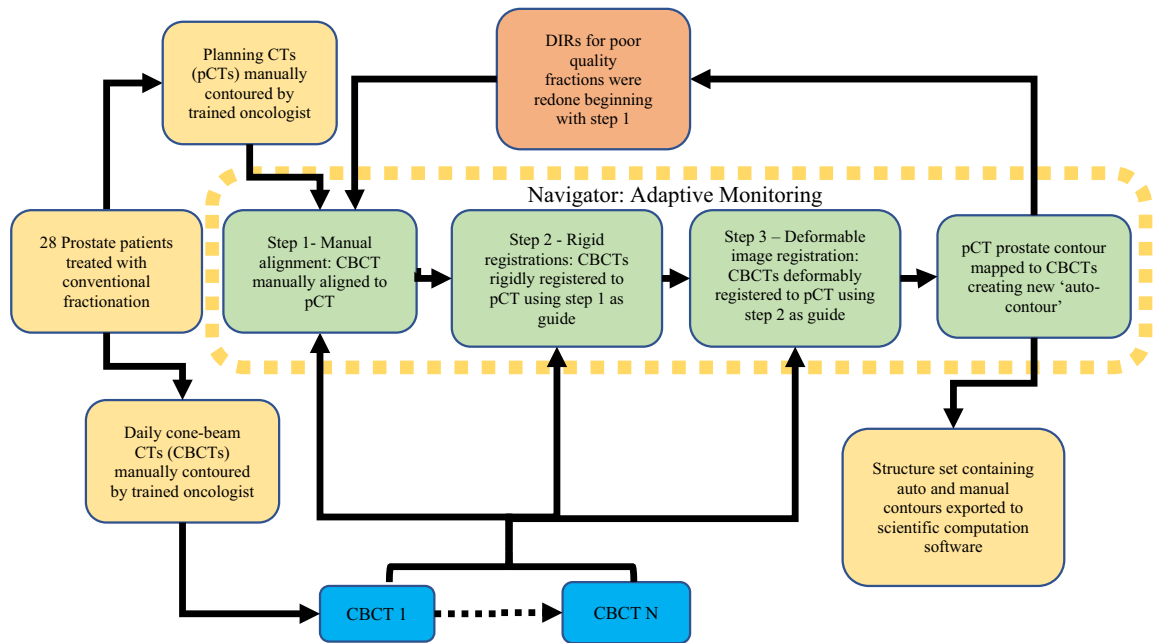


Figure 1. Deformable image registration and delineation propagation workflow. All steps included in the Adaptive Monitoring Navigator on Velocity are inside the dotted-line box. The re-evaluation of poor-quality fractions ($DSC < 0.75$, $MDA > 3.5$ mm) is shown in the light-red box. Exportation of data to MATLAB for data extraction and analysis is done in the final green box.

(DVF) was carried out to ensure that the DIR transformation was physically and anatomically reasonable. The locations of anatomical landmarks (e.g., bones) and fiducial markers were visually inspected on fused images of the pCT and CBCT to verify that they matched. DIRs were refined by adjusting the ROI around the prostate to improve DIR accuracy²⁶, or by applying a structure-guided deformation.

The quantitative assessment of the registered prostate contours was done using several metrics. Dice similarity coefficient (DSC) is a statistical measure of contour overlap with 0 being no overlap and 1 being a perfect match²⁷. Distance to agreement (DTA) between two contours, sometimes referred to as distance to conformity, is the shortest distance from a given point on the surface of one contour to the surface of the other contour. Mean distance to agreement (MDA) is the mean of all DTA distances²⁸. The geometric centers of both the auto and manual contours on the CBCTs were calculated and used to determine the difference in center of mass position (ΔCM). The difference in volume between the auto and manual contour (ΔVol) was also evaluated for all 1,010 fractions. For a smaller sub-sample of 10 patients, on a bi-weekly basis (totaling 47 fractions), the Jacobian determinant (JD) was computed. Jacobian determinant values corresponding to volume expansion, no volume change and volume reduction are > 1 , 1, and < 1 , respectively. JD values equal to or less than zero correspond to non-physical transformations which are indications of a poor DIR²⁰. DSC, MDA, ΔCM , and ΔVol were calculated between the auto and manual prostate contours for all 1,010 fractions. TG-132 defines a clear method for assessing DIR algorithms²⁰. In this work, the tolerances defined in the TG-132 were used for quantitative assessment of the DIRs.

Radiomic features. To study the impact of contouring variability due to the DIR on CBCT radiomic features in the prostate a total of 46 radiomic features derived from 6 different classes were analyzed for all 1,010 fractions. Additionally, a subpopulation consisting of 149 fractions was identified having $|\Delta Vol| > |10\%|$ to study the impact of larger contour variability on CBCT radiomic features. Radiomic features were extracted using the procedure described in Delgadillo et al.¹¹, including Gray-Level Co-occurrence Matrices (GLCM), Neighborhood Gray-Tone Difference Matrix (NGTDM), Gray-Level Run-Length Matrices (GLRLM), Gray-Level Size Zone Matrices (GLSZM), Morphological and statistical features^{29–31}. Each radiomic feature is distinguished by the image biomarker standardization initiative (IBSI) code³².

Percent difference in radiomic feature derived data (% ΔRF) was compared to DSC to assess radiomic feature dependency on contouring variability using Spearman's rank correlation coefficient (ρ)^{33,34}. These correlations were classified as weak if $|\rho| < 0.4$, moderate if $0.4 \leq |\rho| < 0.6$, relatively strong if $0.6 \leq |\rho| < 0.8$, and strong if $0.8 \leq |\rho|$ ³³.

Lin's concordance correlation coefficient (CCC) was computed for all 46 radiomic features between the auto and manual contours to find the strength of correlation³⁵, with a perfectly linear relationship equal to 1 and no relation being 0^{34,36}. Adapting the classification scheme proposed by McGraw et al.³⁷, radiomic features were classified as robust with $CCC > 0.90$, acceptable with $0.75 < CCC < 0.90$, and uncertain with $CCC < 0.75$.

Additionally, mean absolute percent difference in radiomic feature derived data (mean $|\% \Delta| RF$) was used to evaluate the stability of radiomic features to differences in prostate contours. Radiomic features were independently classified as robust with mean $|\% \Delta| RF < 5\%$, acceptable with $5\% < \text{mean } |\% \Delta| RF < 15\%$, and uncertain

	Mean	SD	Range	TG-132 recommendation
Dice similarity coefficient	0.90	0.04	(0.74, 0.98)	~ 0.8 to 0.9
Mean distance to agreement (mm)	1.80	0.50	(0.88, 4.16)	~ 2 to 3
Difference in center of mass position (mm)	2.17	1.38	(0.06, 8.18)	NA
Difference in volume (%)	5.10	4.10	(0.06, 22.8)	NA
Jacobian minimum	0.77	0.18	(0.25, 0.97)	> 0
Jacobian maximum	1.30	0.23	(1.02, 1.97)	NA

Table 1. Similarity metrics between auto and manual contours.

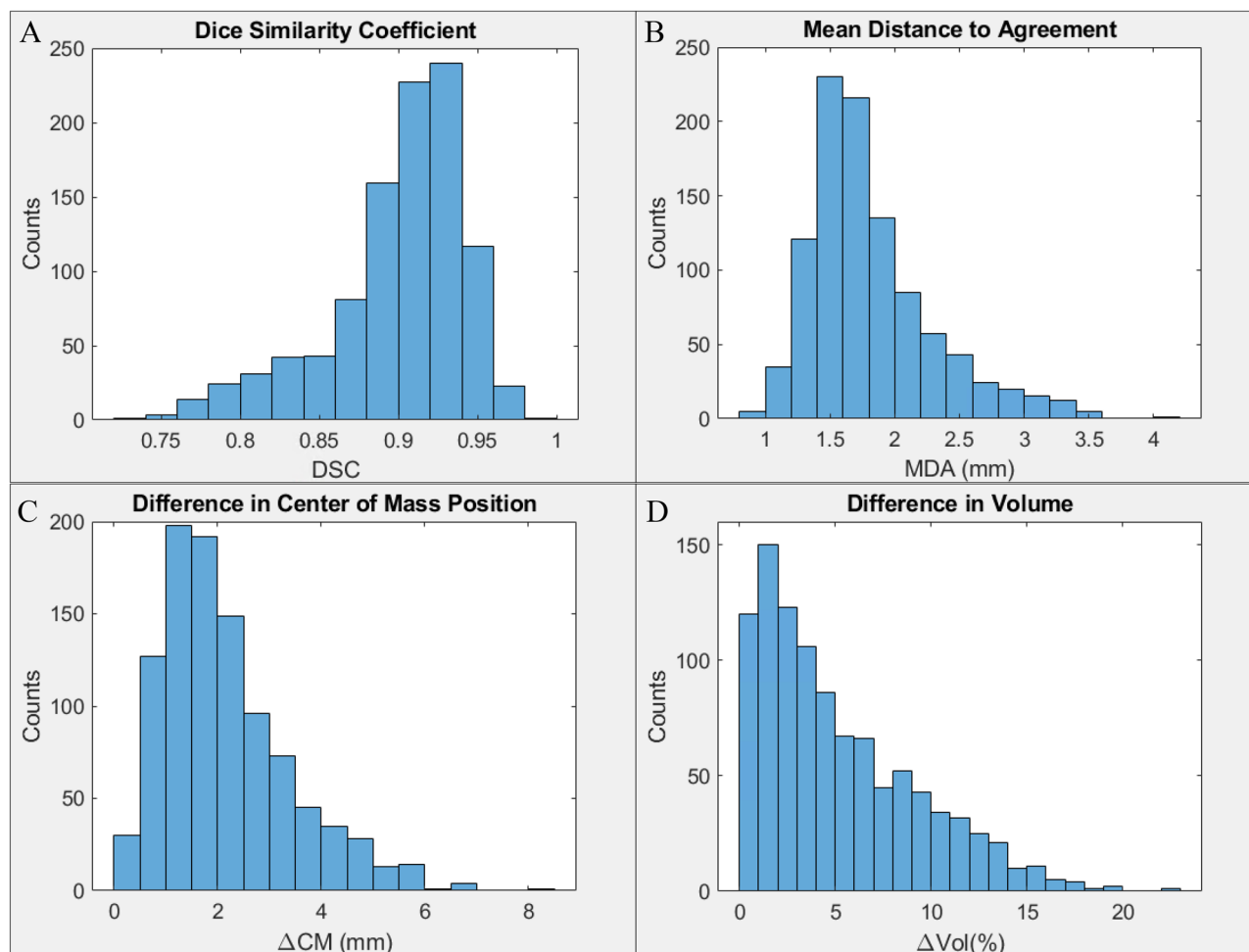


Figure 2. Histograms of (A) DSC, (B) MDA, (C) Δ CM and (D) % Δ volume between the auto and manual contours for all 1010 fractions are shown in (A)–(D) respectively.

with $15\% < \text{mean } |\% \Delta|_{\text{RF}} < 50\%$. Processing and analysis were performed using scientific computation software (MATLAB, ver. 2018b, Math-Works Inc., Natick, MA).

Results

Assessment of prostate contour accuracy. Table 1 summarizes the mean, standard deviation and range for DSC, MDA, Δ CM, Δ Vol, JD minimum and JD maximum between auto and manual prostate contours, and Fig. 2 shows the distribution of DSC, MDA, Δ CM and Δ Vol over 1010 fractions. The mean DSC of all 1,010 fractions, 0.90 ± 0.04 , is within the TG-132 recommended DSC value of ~0.8 to 0.9. A total of 42 fractions had DSC < 0.8, below the lower tolerance recommendation of TG-132.

The mean MDA of 1.81 ± 0.47 mm is well within the TG-132 recommended ~2 to 3 mm. A total of 33 fractions had MDA > 3 mm with a maximum MDA of 4 mm. The mean Δ CM found to be 2.17 ± 1.26 mm and mean Δ Vol of $5.1 \pm 4.1\%$. The mean minimum and mean maximum values of the JD were found to be 0.77 ± 0.18 and 1.31 ± 0.23 respectively, with no JD values ≤ 0 (Table 1).

Impact of contouring variability on radiomic features. Spearman rank correlation coefficient for % Δ RF versus DSC for each radiomic feature is shown in Fig. 3. As previously mentioned, a subpopulation of fractions with $|\Delta\text{Vol}| > |10\%|$ was also considered, results from these fractions are plotted in red (Fig. 3). All 46 radiomic features were classified as having a weak correlation between % Δ RF and DSC with $|\rho| < 0.4$ for both populations under consideration (all fractions and sub-population of fractions with $|\Delta\text{Vol}| > |10\%|$).

Table 2 displays Lin's concordance correlation coefficient (CCC) for all 46 radiomic features. Using the classification scheme mentioned earlier, 30 of 46 radiomic features were classified as robust, 8 radiomic features were classified as acceptable, and 8 radiomic features were classified as uncertain (Table 2). Neighborhood Gray-Tone Difference Matrix (NGTDM) and Gray-Level Co-occurrence Matrices (GLCM) had the highest mean CCC with values of 0.963 and 0.943, respectively.

Independent classification according to mean $|\% \Delta|$ RF can also be seen in Table 2 separated by class. In total 21 of 46 radiomic features were classified as robust and 7 of the 13 GLRLM radiomic features were classified as robust. 14 radiomic features were classified as robust according to both CCC and mean $|\% \Delta|$ RF. 24 radiomic features did not have matching stability classifications when comparing the classifications from CCC and mean $|\% \Delta|$ RF.

Discussion

DIR and contour accuracy. Mean DSC and mean MDA of the current study (Table 1) are both well within the TG-132 recommended tolerances, indicating that the DIR workflow followed here can produce accurate prostate contours. A study done by Forde et al., found that 37% of contours created through DIR had a DSC $< 0.7^5$. The DIR algorithm of our study, having no contours with DSC < 0.7 , outperformed the one used by Forde et al. which used an earlier version of Varian's auto contouring (SmartSegmentation version 15.5) which used an atlas-based algorithm. However, 4.1% of contours created in this study still had DSC below the TG-132 recommendation of 0.8—further inspection of these fractions found acceptable displacement vector fields and acceptable contours as shown in Fig. 4 for a poor performing fraction (DSC = 0.79, MDA = 3.02).

While ΔCM and ΔVol are not included in the TG-132 recommendation, both are useful indications of auto and manual contour matching. A previous study by Studenski et al., considered 16 patients on hypo-fractionated schemes and found percent difference in prostate volume from contours created through a DIR workflow to be $< 10\%^{38}$, and a previous study by Forde et al., which looked at inter-observer delineation variability and its impacts on radiomic feature robustness found the interobserver variation in contour sizes were as high as 80%⁵. In the study presented here, 149 of the 1,010 fractions had a difference in volume between the auto and manual contours ($\Delta\text{Vol}) > 10\%$ (Table 1). Large discrepancies in volume indicate a poor DIR, however, the mean DSC of these 149 fractions was 0.88, and visual inspections resulted in acceptable contours.

The radiomic features belonging to the morphological features class can also be used as a measure of contour accuracy. As can be seen in Table 2, the mean $|\% \Delta|$ RF for all 4 morphological features is less than 0.1%. These low mean absolute percent difference in radiomic feature derived data for the morphological feature class is another indication that the automatically generated contour and the manual contour are in good agreement.

Radiomic feature robustness. Spearman rank correlation coefficient for % Δ RF versus DSC for each radiomic feature as shown in Fig. 3, shows that all 46 radiomic features were classified as having a weak correlation (for both populations under consideration). A low Spearman rank correlation coefficient between % Δ RF versus DSC simply highlights the complex interplay between radiomic feature derived data of differing contours. That is, a large difference in contour does not necessarily translate to a large difference in radiomic feature derived data, and a small difference in contour does not necessarily translate to a small difference in radiomic feature derived data.

Higher Lin's concordance correlation coefficient (CCC) translates to robust radiomic features. As seen from Table 2, the NGTDM class radiomic features had the highest mean CCC of 0.963, while GLRLM class radiomic features had the lowest mean of 0.820. Based on this, NGTDM was the most robust class of radiomic features, while GLRLM was the least robust class of radiomic features when considering differences in prostate contours. In contrast to our study, Rizzetto et al., found that GLRLM was the most robust to differing contours, considering colorectal liver metastases contours¹⁰. These incongruous results may be due to the differing locations within the body (prostate versus liver) and/or the differing contour sizes, but only further the idea that the robustness of radiomic features should be evaluated as they can vary by location. Future radiomic studies should consider the location specific radiomic feature robustness, as the radiomic feature derived data has varying dependence on contour as seen in this study and others^{3–5,10–12}.

Similar to the results of this study, a different study done by Yang et al., evaluating radiomic feature robustness taken from PET images of lung cancer patients, found only weak or moderate correlations between % Δ RF and DSC¹⁴. A weak correlation implies a more complex interplay between the delineation of the volume and the contents of the volume. While two contours of a volume may have a high DSC, the percent difference in radiomic feature-derived data from the two contours may be very different.

Mean $|\% \Delta|$ RF alone was also used to evaluate the stability of radiomic features to differences in prostate contours. Forde et al., found that GLRLM had the smallest mean $|\% \Delta|$ RF considering parotid gland contours⁵. The results of Forde et al. agree with the work presented here, finding 7 of the 13 GLRLM radiomic features were classified as robust, the highest performing class.

Severe discrepancies between CCC and mean $|\% \Delta|$ RF do occur. For example, statistical features of 'Skewness' and 'Kurtosis' had CCC values of 0.879 and 0.946 respectively and mean \pm SD in $|\% \Delta|$ RF of 143.2 ± 694.1 and 85.7 ± 4786 , respectively. If the auto contour radiomic feature derived data and manual contour radiomic feature derived data have similar means and standard deviations over all fractions, CCC will be high and indicate high

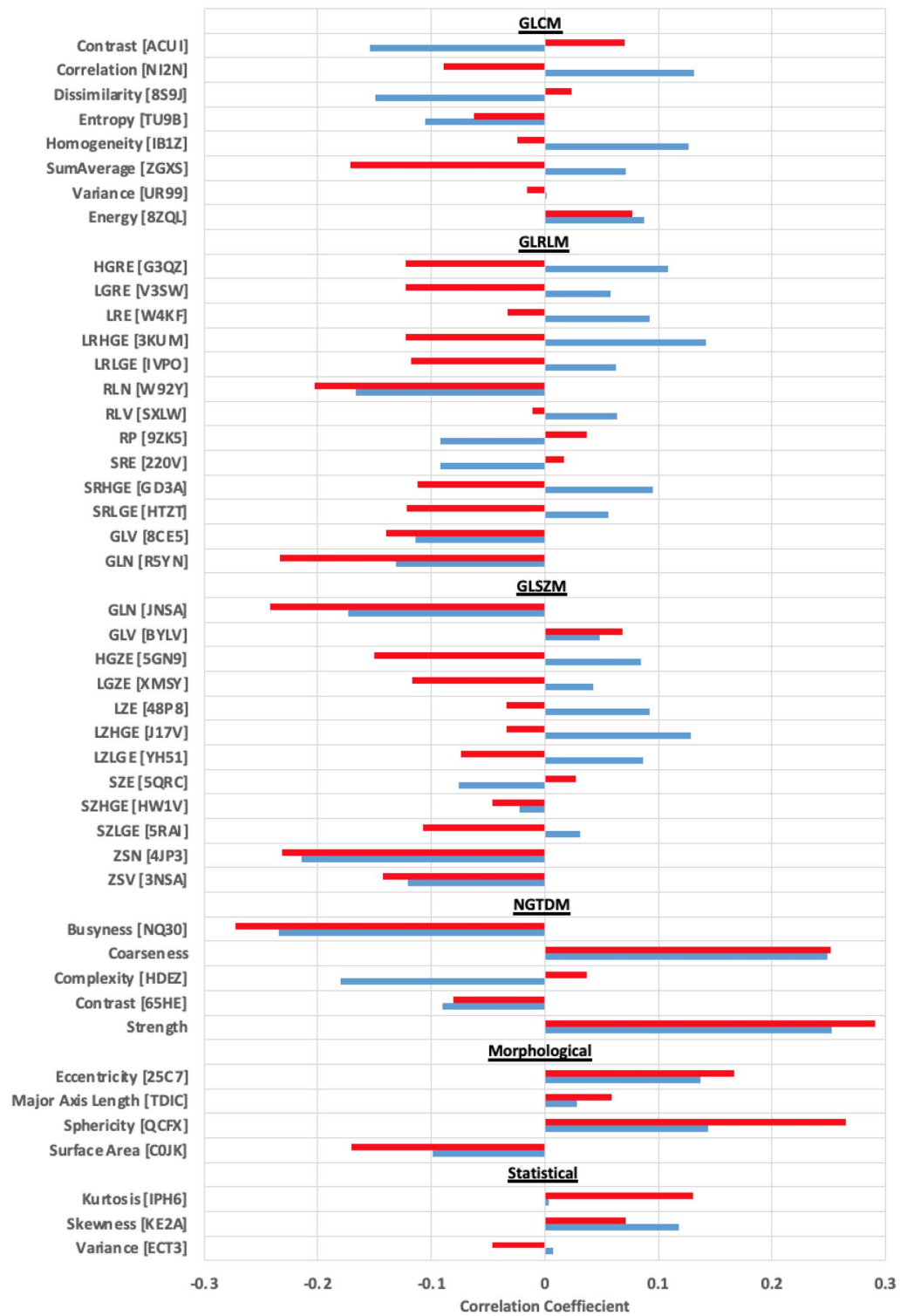


Figure 3. Spearman's rank correlation coefficient between the mean absolute percent difference in radiomic feature derived data (%ΔRF) between auto and manual contours plotted against Dice similarity coefficient (DSC), stratified by class. This was done for two populations, all fractions (blue) and for fractions with ΔVol > |10% (red).

	Radiomic feature [IBSI code]	CCC	Stability classification	Mean \pm SD % Δ RF	Stability classification 2	Stability classification match?
GLCM Mean CCC = 0.943	Energy [8ZQL]	0.959	Robust	7.4 \pm 14.5	Acceptable	No
	Contrast [ACUI]	0.959	Robust	6.7 \pm 9.1	Acceptable	No
	Correlation [NI2N]	0.957	Robust	3.0 \pm 3.0	Robust	Yes
	Dissimilarity [8S9J]	0.957	Robust	3.6 \pm 4.7	Robust	Yes
	Entropy [TU9B]	0.949	Robust	0.6 \pm 0.9	Robust	Yes
	Homogeneity [IB1Z]	0.968	Robust	2.8 \pm 3.8	Robust	Yes
	SumAverage [ZGXS]	0.861	Acceptable	0.6 \pm 0.8	Robust	No
	Variance [UR99]	0.936	Robust	5.6 \pm 8.6	Acceptable	No
GLRLM Mean CCC = 0.820	GLN [R5YN]	0.972	Robust	7.5 \pm 13.0	Acceptable	No
	GLV [BYLV]	0.580	Uncertain	12.6 \pm 11.8	Uncertain	Yes
	HGRE [G3QZ]	0.891	Acceptable	1.1 \pm 1.7	Robust	No
	LGRE [V3SW]	0.705	Uncertain	23.5 \pm 23.0	Uncertain	Yes
	LRE [W4KF]	0.969	Robust	0.3 \pm 0.5	Robust	Yes
	LRHGE [3KUM]	0.875	Acceptable	1.1 \pm 1.7	Robust	No
	LRLGE [IVPO]	0.699	Uncertain	23.6 \pm 23.2	Uncertain	Yes
	RLN [W92Y]	0.985	Robust	6.1 6.4	Robust	Yes
	RLV [SXLW]	0.447	Uncertain	25.8 \pm 31.2	Uncertain	Yes
	RP [9ZK5]	0.968	Robust	0.1 \pm 0.2	Robust	Yes
	SRE [220V]	0.968	Robust	0.1 \pm 0.1	Robust	Yes
	SRHGE [GD3A]	0.898	Acceptable	1.0 \pm 1.7	Robust	No
SRLGE [HTZT]	0.706	Uncertain	23.4 \pm 22.9	Uncertain	Yes	
GLN [JNSA]	0.982	Robust	6.6 \pm 10.7	Acceptable	No	

Table 2. (continued)

GLSZM Mean CCC = 0.884	GLV [8CE5]	0.958	Robust	21.1 ± 20.8	Uncertain	No
	HGZE [5GN9]	0.880	Acceptable	1.0 ± 1.6	Robust	No
	LGZE [XMSY]	0.697	Uncertain	23.6 ± 22.9	Uncertain	Yes
	LZE [48P8]	0.969	Robust	5.0 ± 11.0	Robust	Yes
	LZHGE [J17V]	0.968	Robust	5.4 ± 11.0	Acceptable	No
	LZLGE [YH51]	0.586	Uncertain	27.0 ± 28.5	Uncertain	Yes
	SRE [220V]	0.963	Robust	0.1 ± 2.2	Robust	Yes
	SZE [5QRC]	0.950	Robust	0.7 ± 0.9	Robust	Yes
	SZHGE [HW1V]	0.917	Robust	1.3 ± 2.0	Robust	Yes
	SZLGE [5RAI]	0.678	Uncertain	25.0 ± 24.6	Uncertain	Yes
	ZSN [4JP3]	0.985	Robust	6.2 ± 7.5	Acceptable	No
	ZSV [3NSA]	0.959	Robust	20.6 ± 22.1	Uncertain	No
NGTDM Mean CCC = 0.963	Busyness [NQ30]	0.986	Robust	6.2 ± 7.1	Acceptable	No
	Coarseness	0.966	Robust	5.8 ± 7.1	Acceptable	No
	Complexity [HDEZ]	0.956	Robust	4.0 ± 4.9	Acceptable	No
	Contrast [65HE]	0.937	Robust	7.4 ± 14.6	Acceptable	No
	Strength	0.968	Robust	6.4 ± 7.1	Acceptable	No
Morphological Mean CCC = 0.907	Eccentricity [25C7]	0.810	Acceptable	0.08 ± 0.08	Robust	No
	Major Axis Length [TDIC]	0.983	Robust	0.02 ± 0.01	Robust	Yes
	Sphericity [QCFX]	0.854	Acceptable	0.04 ± 0.03	Robust	No
	Surface Area [COJK]	0.980	Robust	0.04 ± 0.03	Robust	Yes
Statistical Mean CCC = 0.920	Kurtosis [IPH6]	0.946	Robust	85.7 ± 478.6	Unstable	No
	Skewness [KE2A]	0.879	Acceptable	143.2 ± 694.1	Unstable	No
	Variance [ECT3]	0.935	Robust	5.3 ± 8.3	Acceptable	No

Table 2. Lin's concordance correlation coefficient (CCC) and mean absolute percent difference in radiomic feature derived data ($|\% \Delta|_{RF}$) between auto and manual contours for all radiomic features. Corresponding stability classifications were given according to each independently and compared for consistency. CCC values classified as robust $CCC > 0.90$, acceptable with $0.75 < CCC < 0.90$, and uncertain with $CCC < 0.75$. Radiomic features were classified according to mean absolute percent difference in radiomic feature derived data ($|\% \Delta|_{RF}$) as robust with mean $|\% \Delta|_{RF} < 5\%$, acceptable with $5\% < \text{mean } |\% \Delta|_{RF} < 15\%$, and uncertain with $15\% < \text{mean } |\% \Delta|_{RF} < 50\%$.

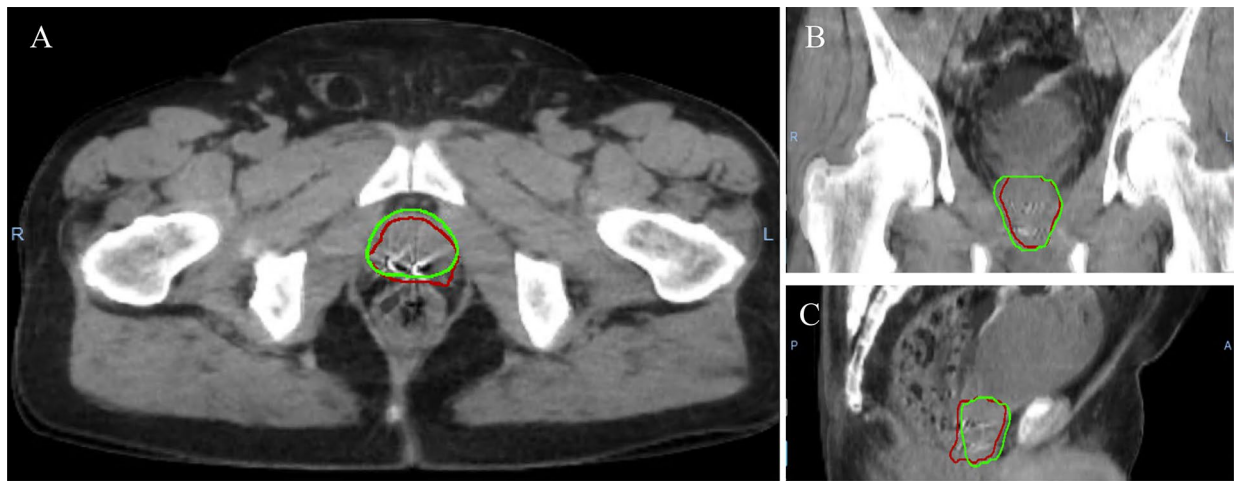


Figure 4. Automatically generated prostate contour (red) and manually drawn contour (green) for sample patient with poor match statistics ($DSC < 0.8$ and $MDA > 3$). Shown axial slice (A), coronal slice (B), and sagittal slice (C).

similarity. However, in the same situation, $|\% \Delta|_{RF}$ per auto and manual contour pair can be large and varying. Thus, leading to situations where a radiomic feature has both a desirable high CCC and an undesirable high mean $|\% \Delta|_{RF}$.

Conclusions

This study demonstrated that an intensity-based DIR algorithm applied to daily CBCTs is sufficiently robust and accurate to meet the recommendations of TG-132 for prostate cancer. The radiomic features derived from DIR-generated auto contours and manually drawn contours were acceptably similar for 22 of 46 radiomic features. However, there is a varying dependence on contours from one radiomics class to another and from one radiomic feature to another. Weak correlations between mean $|\% \Delta|_{RF}$ and DSC imply a complex interplay of volume and contents when considering radiomic feature data extraction from prostate contours that requires further insight.

Received: 12 June 2021; Accepted: 3 November 2021

Published online: 23 November 2021

References

1. Aerts, H. J. W. L. *et al.* Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach (vol 5, pg 4006, 2014). *Nat. Commun.* **5**, 1–9 (2014).
2. Zhang, H. *et al.* Locally advanced squamous cell carcinoma of the head and neck: CT texture and histogram analysis allow independent prediction of overall survival in patients treated with induction chemotherapy. *Radiology* **269**(3), 801–809 (2013).
3. Zhao, B. Understanding sources of variation to improve the reproducibility of radiomics. *Front. Oncol.* **11**, 633176 (2021).
4. Traverso, A. *et al.* Sensitivity of radiomic features to inter-observer variability and image pre-processing in Apparent Diffusion Coefficient (ADC) maps of cervix cancer patients. *Radiother. Oncol.* **143**, 88–94 (2020).
5. Forde, E. *et al.* Influence of inter-observer delineation variability on radiomic features of the parotid gland. *Phys. Med.* **82**, 240–248 (2021).
6. Ford, J. *et al.* Quantitative radiomics: Impact of pulse sequence parameter selection on MRI-based textural features of the brain. *Contrast Media Mol. Imaging* **2018**, 1729071 (2018).
7. Yang, F. *et al.* Evaluation of radiomic texture feature error due to MRI acquisition and reconstruction: A simulation study utilizing ground truth. *Physica Med.* **50**, 26–36 (2018).
8. Yang, F. *et al.* Magnetic resonance imaging (MRI)-based radiomics for prostate cancer radiotherapy. *Transl. Androl Urol.* **7**(3), 445–458 (2018).
9. Simpson, G. *et al.* Impact of quantization algorithm and number of gray level intensities on variability and repeatability of low field strength magnetic resonance image-based radiomics texture features. *Physica Med.* **80**, 209–220 (2020).
10. Rizzetto, F. *et al.* Impact of inter-reader contouring variability on textural radiomics of colorectal liver metastases. *Eur. Radiol. Exp.* **4**(1), 62 (2020).
11. Delgadillo, R. *et al.* Repeatability of CBCT radiomic features and their correlation with CT radiomic features for prostate cancer. *Med. Phys.* **48**, 2386–2399 (2021).
12. Pavic, M. *et al.* Influence of inter-observer delineation variability on radiomics stability in different tumor sites. *Acta Oncol.* **57**(8), 1070–1074 (2018).
13. Qiao, H. *et al.* Interobserver variability in tumor contouring affects the use of radiomics to predict mutational status. *J. Med. Imaging* **5**(1), 1–9 (2017).
14. Yang, F. *et al.* Impact of contouring variability on oncological PET radiomics features in the lung. *Sci. Rep.* **10**(1), 369 (2020).
15. Lawson, J. D. *et al.* Early clinical experience with kilovoltage image-guided radiation therapy for interfraction motion management. *Med. Dosim.* **33**(4), 268–274 (2008).
16. Xu, Y. *et al.* Assessment of daily dose accumulation for robustly optimized intensity modulated proton therapy treatment of prostate cancer. *Physica Med.* **81**, 77–85 (2021).
17. Yan, D. *et al.* Adaptive radiation therapy. *Phys. Med. Biol.* **42**(1), 123–132 (1997).
18. Woerner, A. J. *et al.* Evaluation of deformable image registration-based contour propagation from planning CT to cone-beam CT. *Technol. Cancer Res. Treat.* **16**(6), 801–810 (2017).

19. Simon, A. *et al.* Roles of deformable image registration in adaptive RT: From contour propagation to dose monitoring. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **2015**, 5215–5218 (2015).
20. Brock, K. K. *et al.* Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the AAPM Radiation therapy committee task group no. 132. *Med. Phys.* **44**(7), e43–e76 (2017).
21. Kim, J. *et al.* A novel approach for establishing benchmark CBCT/CT deformable image registrations in prostate cancer radiotherapy. *Phys. Med. Biol.* **58**(22), 8077–8097 (2013).
22. Motegi, K. *et al.* Usefulness of hybrid deformable image registration algorithms in prostate radiation therapy. *J. Appl. Clin. Med. Phys.* **20**(1), 229–236 (2019).
23. Nie, K. *et al.* Performance variations among clinically available deformable image registration tools in adaptive radiotherapy—How should we evaluate and interpret the result?. *J. Appl. Clin. Med. Phys.* **17**(2), 328–340 (2016).
24. Thor, M. *et al.* Deformable image registration for contour propagation from CT to cone-beam CT scans in radiotherapy of prostate cancer. *Acta Oncol.* **50**(6), 918–925 (2011).
25. Kadoya, N. *et al.* Evaluation of various deformable image registration algorithms for thoracic images. *J. Radiat. Res.* **55**(1), 175–182 (2014).
26. Lin, H. *et al.* SU-GG-I-109: A quantitative evaluation of velocity AI deformable image registration. *Med. Phys.* **37**(6Part4), 3126–3126 (2010).
27. Dice, L. R. Measures of the amount of ecologic association between species. *Ecology* **26**(3), 297–302 (1945).
28. Chalana, V. & Kim, Y. A methodology for evaluation of boundary detection algorithms on medical images. *IEEE Trans. Med. Imaging* **16**(5), 642–652 (1997).
29. Thibault, G. *et al.* Shape and texture indexes application to cell nuclei classification. *Int. J. Pattern Recogn. Artif. Intell.* **27**(01), 1357002 (2013).
30. Haralick, R. M. Statistical and structural approaches to texture. *Proc. IEEE* **67**(5), 786–804 (1979).
31. Stoecker, W. V., Chiang, C.-S. & Moss, R. H. Texture in skin images: Comparison of three methods to determine smoothness. *Comput. Med. Imaging Gr.* **16**(3), 179–190 (1992).
32. Zwanenburg, A. *et al.* The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* **295**(2), 328–338 (2020).
33. Kendall, M. G. A new measure of rank correlation. *Biometrika* **30**(1–2), 81–93 (1938).
34. Spearman, C. Demonstration of formulae for true measurement of correlation. *Am. J. Psychol.* **18**(2), 161–169 (1907).
35. Liu, J. *et al.* Correlation and agreement: Overview and clarification of competing concepts and measures. *Shanghai Arch. Psychiatry* **28**(2), 115–120 (2016).
36. Lawrence, I. K. L. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* **45**(1), 255–268 (1989).
37. McGraw, K. O. & Wong, S. P. “Forming inferences about some intraclass correlations coefficients”: Correction. *Psychol. Methods* **1**(4), 390–390 (1996).
38. Studenski, M. T. *et al.* Margin verification for hypofractionated prostate radiotherapy using a novel dose accumulation workflow and iterative CBCT. *Phys. Med.* **77**, 154–159 (2020).

Acknowledgements

This work was supported in part by a research grant from Varian Medical Systems, Palo Alto, CA. (GR013242).

Author contributions

R.S. wrote the manuscript text, prepared all figures/tables, and carried out data collection/analysis. B.S. completed manual contouring. Y.X. completed structure guided (hybrid) DIRs. ND provided guidance for the overall manuscript. All authors reviewed the data and revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to N.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021