Physica Medica 58 (2019) 141–148

Contents lists available at ScienceDirect

Physica Medica

journal homepage: www.elsevier.com/locate/ejmp

Original paper

Investigation of thoracic four-dimensional CT-based dimension reduction technique for extracting the robust radiomic features

Shohei Tanaka[a], Noriyuki Kadoya[a,*], Tomohiro Kajikawa[a], Shohei Matsuda[a], Suguru Dobashi[b], Ken Takeda[b], Keiichi Jingu[a]

[a] Department of Radiation Oncology, Tohoku University Graduate School of Medicine, Sendai, Japan
[b] Department of Radiological Technology, School of Health Sciences, Faculty of Medicine, Tohoku University, Sendai, Japan

ARTICLE INFO

Keywords:
Radiotherapy
Radiomics
4D-CT
Lung cancer

ABSTRACT

Robust feature selection in radiomic analysis is often implemented using the RIDER test-retest datasets. However, the CT Protocol between the facility and test-retest datasets are different. Therefore, we investigated possibility to select robust features using thoracic four-dimensional CT (4D-CT) scans that are available from patients receiving radiation therapy.

In 4D-CT datasets of 14 lung cancer patients who underwent stereotactic body radiotherapy (SBRT) and 14 test-retest datasets of non-small cell lung cancer (NSCLC), 1170 radiomic features (shape: n = 16, statistics: n = 32, texture: n = 1122) were extracted. A concordance correlation coefficient (CCC) > 0.85 was used to select robust features. We compared the robust features in various 4D-CT group with those in test-retest.

The total number of robust features was a range between 846/1170 (72%) and 970/1170 (83%) in all 4D-CT groups with three breathing phases (40%–60%); however, that was a range between 44/1170 (4%) and 476/1170 (41%) in all 4D-CT groups with 10 breathing phases. In test-retest, the total number of robust features was 967/1170 (83%); thus, the number of robust features in 4D-CT was almost equal to that in test-retest by using 40–60% breathing phases.

In 4D-CT, respiratory motion is a factor that greatly affects the robustness of features, thus by using only 40–60% breathing phases, excessive dimension reduction will be able to be prevented in any 4D-CT datasets, and select robust features suitable for CT protocol of your own facility.

1. Introduction

Radiomics is a new method for extracting quantitative data from medical images created using magnetic resonance imaging (MRI), computed tomography (CT), ultrasonography (US), and positron emission tomography (PET). Quantitative analysis of medical images using various softwares reportedly offers more and better information than a physician alone does [1]. Moreover, the prognostic power of radiomics has been recently established for various cancer types [2–9]. Nevertheless, radiomic features are susceptible to various factors; for example, CT protocols include scanning methods and parameter settings [10–13], multiple segmentation with several persons [14,15], and respiratory motion [16–18]. Thus, radiomics analyses may not always be accurate. Therefore, it is essential to focus only on robust features for precise radiomics analyses. In daily clinical practice, it is not feasible to obtain patients' data from two scans on one day. Nevertheless, National Cancer Institute (NCI)–sponsored Reference Imaging Database to Evaluate Response (RIDER) test–retest datasets from 32 patients with non–small-cell lung cancer (NSCLC) scanned twice at 15-min intervals are publicly available [19–21]. In addition, robust feature selection in radiomics analyses is often implemented using test–retest datasets [2,4,22]. However, it is imperative to select the robust features suitable for the CT protocol of your facility. If the CT protocol of the test–retest datasets differs from that used at your facility, the robust features selected using test–retest datasets could differ from feature selection using datasets of your facility, thereby introducing a risk that the features selected as robust using test–retest datasets would be non-robust for subsequent analyses at your facility. To overcome this problem, this study focuses on thoracic four-dimensional CT (4D-CT) scans available from patients undergoing radiotherapy. Although 4D-CT scans are a great alternative to test–retest data, little research has been conducted in this regard [23], thereby necessitating further investigation.

* Corresponding author at: Department of Radiation Oncology, Tohoku University Graduate School of Medicine, 1-1 Seiryo-machi, Aoba-ku, Sendai 980-8574, Japan.
E-mail address: kadoya.n@rad.med.tohoku.ac.jp (N. Kadoya).

https://doi.org/10.1016/j.ejmp.2019.02.009
Received 14 September 2018; Received in revised form 29 January 2019; Accepted 12 February 2019
Available online 19 February 2019

Hence, this study aims to investigate whether thoracic 4D-CT scans could be used for feature selection of reproducible features suitable for your facility in CT-based radiomics.

2. Materials and methods

2.1. Datasets

We examined 14 patients with lung cancer who were treated with stereotactic body radiotherapy (SBRT) at our hospital from 2014 to 2018. In all patients, thoracic 4D-CT images were obtained to ascertain the internal margin of the clinical target volume; a medical physicist and radiological technologist checked all CT scans. 4D-CT scanning was performed using a LightSpeed RT16 unit (GE Medical Systems, Waukesha, WI) in the cine mode with the Varian Real-Time Position Management System (RPM; pixel size, 0.98 × 0.98 mm; slice thickness, 2.5 mm; Varian Medical Systems, Palo Alto, CA). The X-ray tube current and peak tube voltage were 120 mA and 120 kV, respectively, and the reconstruction kernel was standard kernel; the matrix size in the axial plane was 512 × 512 matrix. We used a phase-based sorting pattern for 4D-CT images. For phase-based sorting, we used the GE Advantage 4D software to create a 4D-CT image set by sorting CT slices based on the RPM phase information into 10 breathing phases. Table 1 summarizes the patients' characteristics.

We used the publicly available RIDER test–retest datasets comprising 32 patients with NSCLC as comparison datasets to evaluate the feasibility of 4D-CT scans based on robust feature selection [19–21]. All datasets were downloaded from the Cancer Imaging Archive [20]. We randomly selected 14 of the 32 patients with NSCLC to match the number of people with 4D-CT; these 14 test–retest datasets were used for analysis. Each patient was scanned twice at 15-min intervals. Images were then acquired during breath hold with the X-ray tube current and peak voltage at 185–439 mA and 120 kV, respectively. The reconstruction kernel was lung kernel. Furthermore, the matrix size in the axial plane was 512 × 512 matrix, pixel size was 0.58 × 0.58–0.78 × 0.78 mm, slice thickness was 1.25 mm, and acquisition type was helical.

2.2. Volume of interest segmentation

Radiomic features are reportedly susceptible to segmentation methods [14]. We used the GrowCut algorithm, an emerging algorithm, implemented in 3D-slicer for 4D-CT (open software platform: https://www.slicer.org/) to decrease the uncertainty of segmentation. Chintan et al. reported that the extracted radiomic features from GrowCut segmentations were more reproducible than the features from manual delineation [14]. The process began by loading patient data of the breathing phase; the users manually scribbled inside and outside the tumor region. The algorithm then attempted to automatically label the remaining pixels inside and outside the tumor region in the volume of interest (VOI). We implemented pixel labeling by evaluating a weighted degree of similarity. Next, we split the merge volume to eliminate the label outside the tumor and manually edited the remaining label to delineate the appropriate shape of gross tumor volumes (GTV); this semiautomatic process was similarly applied to other breathing phases of patients.

Furthermore, some test–retest datasets were already semi-automatically delineated using semi-automated 3D segmentation techniques to isolate target lesions from surrounding anatomical structures [19,21]. In other test–retest datasets without contouring data, we semi-automatically delineated the GTV in the same way as for 4D-CT.

Table 1
The characteristics of 14 patients of four-dimentsional computed tomography (4D-CT) whose images were used in this study.

Patient Number	Age	Sex	Tumor motion [cm]	Tumor diameter [cm]	CT value range
1	67	F	0.15	1.8	285
2	76	M	0.46	1.8	293
3	65	M	0.15	1.2	193
4	60	F	0.21	1.1	160
5	74	M	0.25	2.4	217
6	75	M	0.4	2.6	255
7	83	F	0.2	3.7	448
8	46	F	0.58	1.3	194
9	80	M	0.7	1.7	582
10	79	F	1.2	2	268
11	83	M	2.4	3.1	252
12	82	M	1.6	4	280
13	70	M	1.3	3.2	240
14	75	M	1	3.2	306

2.3. Feature extraction and qualification

We examined 1170 radiomic features in this study. The features were divided into 16 shape features (shape, $n = 16$); 32 first-order statistics features (statistics, $n = 32$), including local intensity ($n = 15$) and intensity-based statistics ($n = 17$); and 1122 textural features including the Neighborhood Gray-Tone Difference Matrix 2.5 (NGTDM2.5, $n = 5$), Gray-Level Run-Length Matrix 2.5 (GLRLM, $n = 33$), Gray-Level Co-occurrence Matrix 2.5 (GLCM2.5, $n = 308$), and Gray-Level Co-occurrence Matrix 3 (GLCM3, $n = 776$). Of note, some statistical features needed setting of the number of bins—the values were set at 256. NGTDM explained the average difference between voxels and their neighboring voxels [24]. We used a 5 × 5 neighborhood region for NGTDM calculations; the number of bins was 2000. GLRLM explained the number of consecutive voxels with the same intensity [25]; in this, two directions (0° and 90°) were calculated and the number of bins was 115. GLCM defined a pixel of intensity co-occurring with another pixel of intensity with a predefined correlation between a certain direction and distance [26]. We calculated 4 directions (0°, 45°, 90°, and 135°) for GLCM2.5, 13 directions for GLCM3, and 3 different offsets (1, 4, and 7) for GLCM2.5 and GLCM3; the number of bins was 115, and the co-occurrence matrix was symmetrical. Moreover, 4D-CT and test–retest datasets were loaded and analyzed in the open-source software Imaging Biomarker Explorer (IBEX) [27]. In IBEX, the number 2.5 following the texture feature implies that the features for each two-dimensional (2D) slice are calculated and then combined. Likewise, the number 3 signifies that the features are calculated for 3D. In this study, we set a threshold to exclude air and bones for statistical and textural features. Of note, values less than −150 Hounsfield units (HU) or > 1000 HU were excluded for patients' images.

We used not only unfiltered CT images but also Butterworth low-pass filtered CT images for 4D-CT datasets to affect the usage of image filter in image preprocessing. In addition, order parameter 2 and cutoff parameter 100 of the Butterworth filter were used.

For standardizing radiomics analysis, the imaging biomarker standardization initiative (IBSI) that provided the validated definition and benchmark of features was established [28,29]. However, to the best of our knowledge, IBSI does not support IBEX [30,31]; thus, the features used in this study were not standardized.

2.4. Comparing radiomic features of different datasets: 4D-CT including all 10 phases, 4D-CT including only three phases, and test–retest datasets

Fig. 1 outlines the workflow of this study. We assessed the robustness of the radiomic features using 4D-CT and test–retest datasets. For 4D-CT, two different datasets were used: all 10 breathing phases and only 3 breathing phases (40–60%).

In 4D-CT, because the respiratory motion of 30% and 70% phases is swift, motion blurring of these phases is large [32]. In addition, the motion of the diaphragm during the exhalable phases is slower and

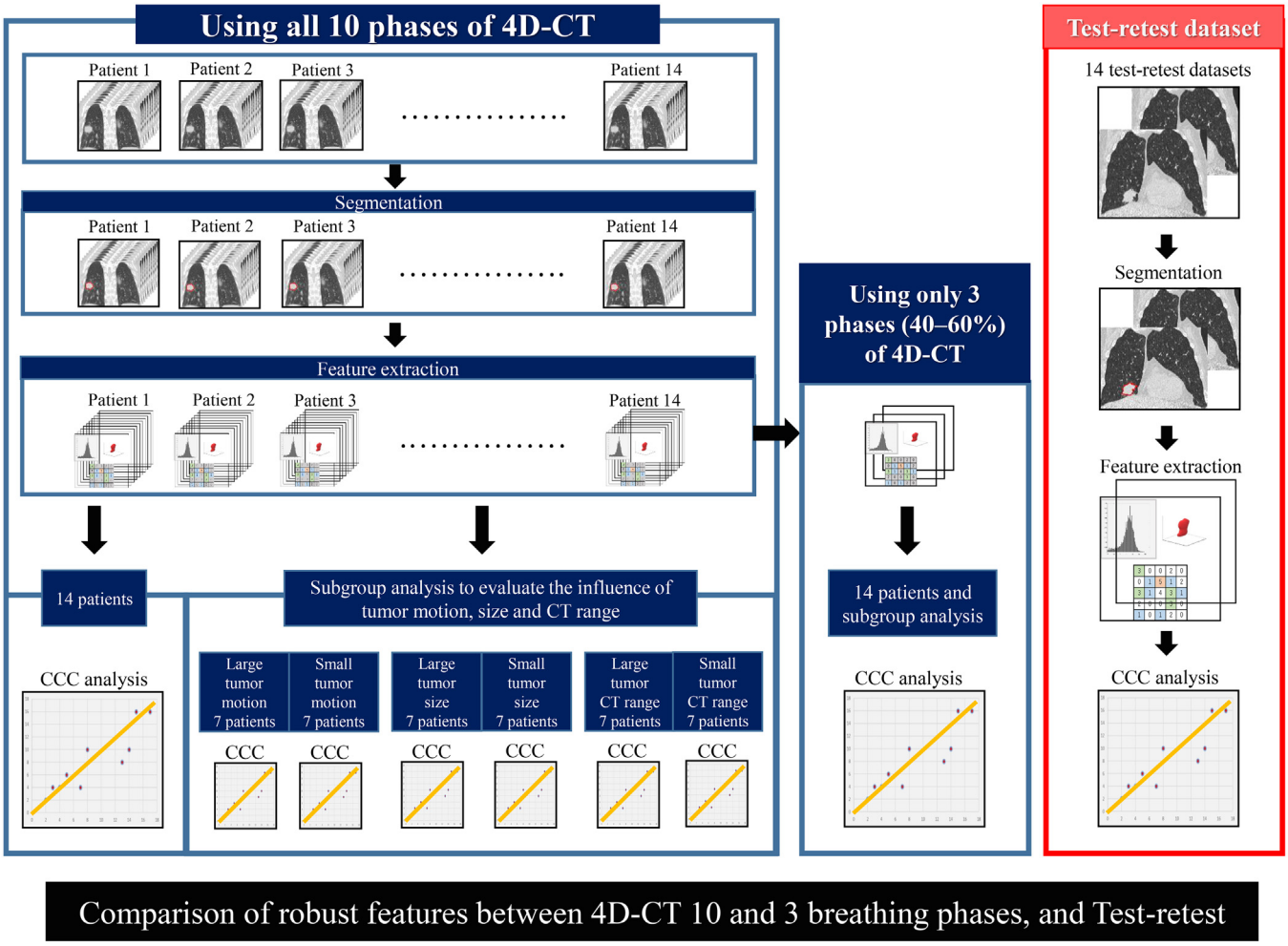

Fig. 1. The study workflow. We segmented 14 four-dimentsional computed tomography (4D-CT) and test–retest datasets for the analysis region with three-dimensional (3D)-slicer; radiomic features were extracted from this analysis region. In 4D-CT with 10 and 3 breathing phases, we implemented the concordance correlation coefficient (CCC) analysis for selecting robust features using all 14 patients. In addition, 14 patients were divided into two groups, and the CCC analysis was implemented with each group to assess the impact of the tumor motion, size, and CT range (subgroup analysis). In test–retest, all 14 patients were used for the CCC analysis. Finally, we compared robust features between 4D-CT, 10 and 3 breathing phases, and test–retest.

more reproducible than that during the inspiratory phase [33]; hence, the respiratory motion around the end-exhale phases (40–60%) of 4D-CT is relatively smaller than that around other phases. Assumedly, the features extracted from these phases were not markedly affected by the respiratory motion.

We used the concordance correlation coefficient (CCC) defined by Lin et al. [34] to evaluate the robustness of the radiomic features. Based on Larue et al. [23], we arbitrarily finalized 0.85 as the threshold of the minimum CCC value. In addition, we evaluated the histogram of the total percentage of robust radiomic features and CCC threshold value to evaluate the impact of threshold value for CCC. The CCC analysis was performed in MATLAB (Math Works, Natick, MA).

2.5. Effect of the tumor motion, tumor size, and tumor CT range value in 4D-CT

We classified 14 4D-CT datasets into two groups using the tumor motion, tumor diameter, and tumor CT value range, which could affect the image quality of each phase (Fig. 1); this subgroup analysis was implemented on the basis of unfiltered features. In addition, the tumor motion was manually measured with the coronal plane from the top of the tumor at the end-inhale phase (0%) to the top of the tumor at the end-exhale phase (50%). We compared robust features in the tumor motion ≥ 0.5 cm (7 datasets) with the tumor motion < 0.5 cm (7 datasets), referred to as the large tumor motion and small tumor motion, respectively, in this study. For the tumor diameter, we measured the equivalent spherical diameter using the Eclipse Treatment Planning System (Varian Medical Systems). In addition, robust features in tumor diameters > 2.0 cm (7 datasets) were compared with tumor diameters ≤ 2.0 cm (7 datasets), referred to as the large tumor size and small tumor size, respectively, in this study. We measured the tumor CT value range by IBEX from the minimum CT value (-150 HU) to the maximum CT value into the tumor with 50% breathing phase of each patient. Furthermore, robust features in the tumor CT value range > 260 (7 datasets) were compared with the tumor CT value range < 260 (7 datasets), referred to as the large tumor CT range and small tumor CT range, respectively, in this study.

3. Results

Fig. 2 summarizes the number of robust features in 4D-CT with 10 breathing phases and test–retest. The total number of robust features for unfiltered 4D-CT, Butterworth filtered 4D-CT, and unfiltered test–retest data were 397/1170 (34%) [shape: 9/16 (56%), statistics: 12/32 (38%), texture: 376/1122 (34%)], 468/1170 (40%) [shape: 9/16 (56%), statistics: 13/32 (41%), texture: 446/1122 (40%)], and 967/1170 (83%) [shape: 14/16 (88%), statistics: 21/32 (66%), texture: 932/1122 (83%)], respectively. Furthermore, we observed large

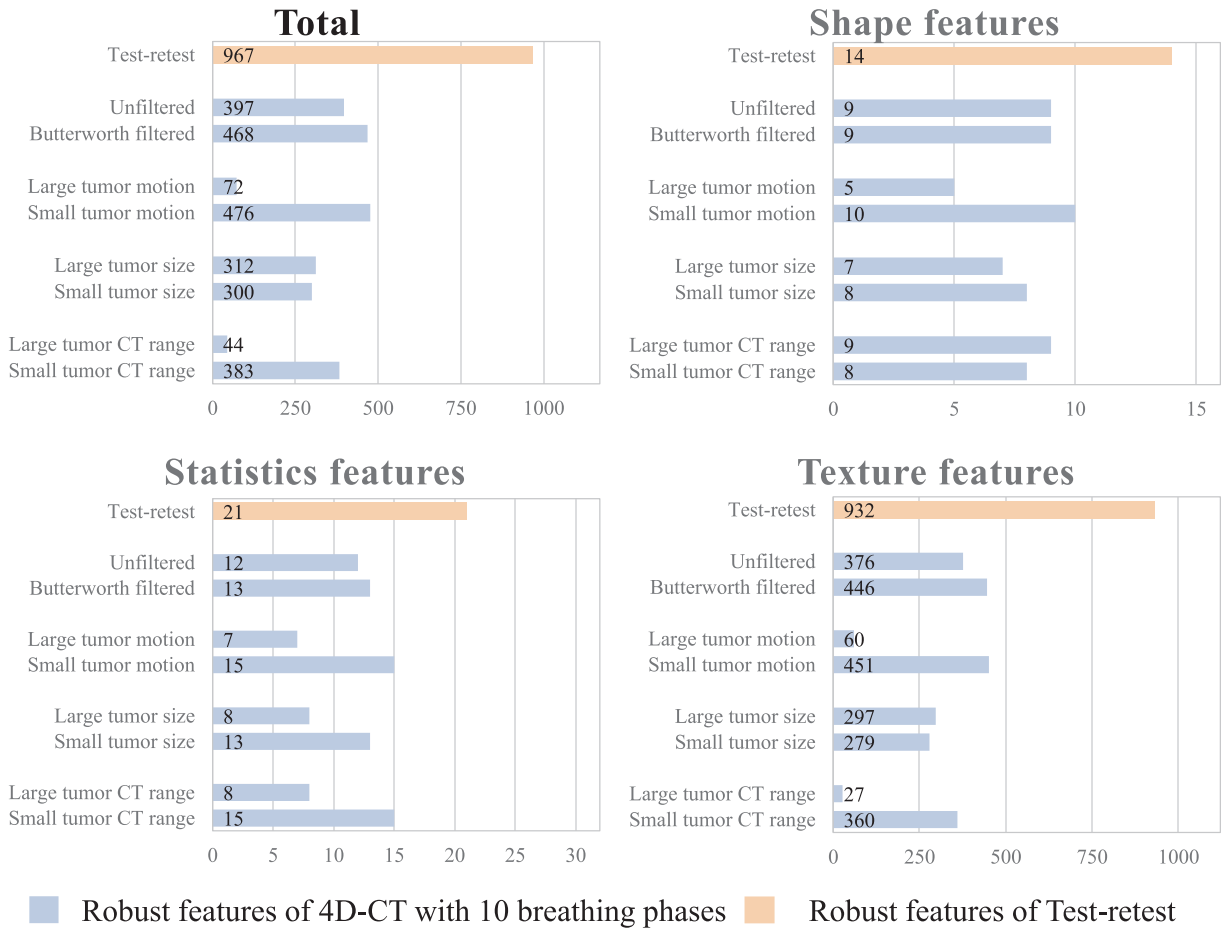

Fig. 2. The number of robust features of test–retest and all groups of four-dimentsional computed tomography (4D-CT) with 10 breathing phases. Top, the result of all features (n = 1170; left) and shape features (n = 16; right). Bottom, the result of statistics features (n = 32; left) and texture features (n = 1122; right).

differences in the number of robust features between unfiltered 4D-CT and test–retest (e.g., total number: 397 vs. 967).

Next, Regarding the effect of tumor motion, as shown in Fig. 2, the total number of robust features for datasets with large and small tumor motions were 72/1170 and 476/1170, respectively, revealing that datasets with the small tumor motion exhibited higher reproducibility than those with the large tumor motion. In addition, we observed significant differences in the number of robust features between these two datasets for shape features [5 vs. 10 (50%)], statistics features [7 vs. 15 (53%)], and texture features [60 vs. 451 (87%)]. Fig. 3 shows the CT images at each phase of 4D-CT datasets in 2 patients: a patient with the large tumor motion (a) and a patient with the small tumor motion (b). Visually, the shape and intensity of tumors were more similar in all 10 breathing phases in patients with the small tumor motion than patients with the large tumor motion.

Second, regarding the tumor size, as shown in Fig. 2, the total number of robust features in datasets with large and small tumor sizes were 312/1170 and 300/1170, respectively, suggesting that the number of robust features did not significantly change with the tumor diameter.

Finally, as shown in Fig. 2, the total number of robust features in datasets with the large and small tumor CT ranges was 44/1170 and 383/1170, respectively, revealing that datasets with the small tumor CT range exhibited higher reproducibility than that with the large tumor CT range.

Fig. 4 summarizes robust features in 4D-CT with three breathing phases. The total number of robust features were 920/1170 (79%) for datasets without filter, 965/1170 (82%) for datasets with Butterworth filter, 970/1170 (83%) for datasets with the large tumor motion, 846/1170 (72%) for datasets with the small tumor motion, 887/1170 (76%) for datasets with the large tumor size, 862/1170 (74%) for datasets with the small tumor size, 864/1170 (74%) for datasets with the large tumor CT range, and 889/1170 (76%) for datasets with the small tumor CT range, suggesting that 4D-CT with only three phases exhibited higher reproducibility than that with all 10 phases. As robust features in test–retest datasets were 967/1170 (83%), the number of robust features in all groups of 4D-CT with only three phases was almost equal to those of test–retest. Regarding the dependency of robust features between 4D-CT groups of three breathing phases, the dispersion of the total number of robust features between all groups was small (range: 846–970), compared with 10 breathing phases (range: 44–476).

Fig. 5 shows the Venn chart of the total number of non-robust features in unfiltered 4D-CT with 10 and 3 breathing phases, and test–retest datasets. The non-robust features in 4D-CT with 10 phases (i.e., 773) covered all non-robust features in those with three phases and 80% (n = 163) of all non-robust features in test–retest, whereas the non-robust features in 4D-CT with three breathing phases covered only 20% (n = 40) of those in test–retest.

Fig. 6 shows the histogram of the total percentage of robust features based on the threshold values of CCC for 4D-CT with 10 breathing phases and for 4D-CT with three breathing phases. For 4D-CT with 10 breathing phases (Fig. 6a), the higher the CCC threshold values, the lower the percentage of robust features. All 4D-CT groups exhibited smaller robust features than test–retest at all CCC threshold values. In addition, 4D-CT groups with the large tumor motion and large tumor CT range exhibited markedly smaller robust features than other 4D-CT groups. For 4D-CT with three breathing phases (Fig. 6b), the percentage of robust features for all 4D-CT groups was closer to that for test–retest,

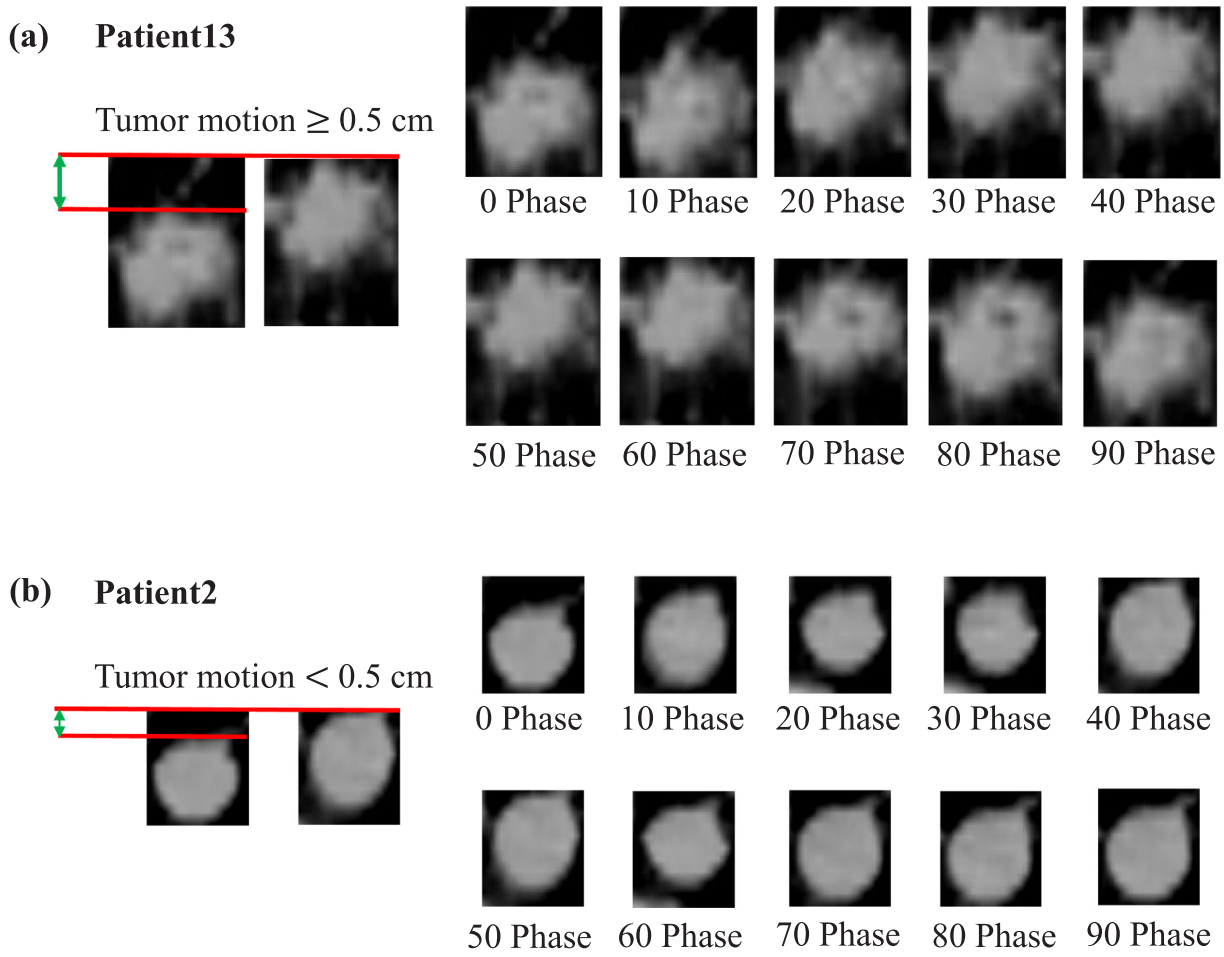

Fig. 3. The computed tomography (CT) images at each phase of 4D-CT datasets in 2 patients—a patient with the large tumor motion (a) and a patient with the small tumor motion (b).

compared with the result of 4D-CT with 10 breathing phases. In addition, the dispersion of the percentage of robust features among all 4D-CT groups was smaller than that in 10 breathing phases, except for the threshold value of about 0.95.

4. Discussion

This study investigated whether thoracic 4D-CT scans could be used for the feature selection of reproducible features suitable for your facility.

The total number of robust features was a range between 846/1170 (72%) and 970/1170 (83%) in all 4D-CT groups (including subgroups) with three breathing phases (40%–60%; Fig. 4); however, that was a range between 44/1170 (4%) and 476/1170 (41%) in all 4D-CT groups with 10 breathing phases (Fig. 2), suggesting the dispersion of the number of robust features among all 4D-CT groups with three breathing phases was smaller than that with 10 breathing phases. This finding could be attributed to the fact that effects influencing the robustness of features were smaller in any 4D-CT datasets with three breathing phases than that in 4D-CT datasets with 10 breathing phases. Of note, 4D-CT images include the residual phase error [35], residual motion artifacts around the tumor caused by partial projection artifacts [32,36], and density variations in the tumor caused by location changes over the breathing cycles [37] as factors affecting the robustness of the features. A slice of 4D-CT image comprises several phases closest to the target phase, in which phase error occurs. Mutaf et al. reported that the phase error resulted in deformations of the target shape and that more the target moved the more deformed the shape of the target would be [38]. Regarding motion artifacts, Castillo et al. reported that motion artifacts with 40–60% of breathing phases were smaller than other breathing phases [39]. Regarding the density variation, the mean difference of the tumor CT values between the end-inhale phase (0%) and the end-exhale phase (50%) in 14 patients was 13 HU. Nevertheless, the mean difference between three breathing phases (40% and 50%, 50% and 60%, and 40% and 60%) was 4 HU, suggesting that the density variation is smaller in three breathing phases than in 10 breathing phases. Hence, 4D-CT with three breathing phases was not susceptible to the impact of factors in any datasets compared with 4D-CT with 10 breathing phases, leading to the small dispersion of the total number of robust features in 4D-CT with three breathing phases. However, based on histograms of the percentage of the robust features (Fig. 6b), this dispersion in 4D-CT with three breathing phases was marginally large at the CCC threshold value of approximately 0.95 because the total percentage of robust features in the Butterworth filter was higher than that in other groups. Jinzhong et al. reported that using a smoothing filter, various features had CCC ≥ 0.9 than of those without smoothing filter [40]. Hence, differences in the percentage became larger at the CCC threshold value of approximately 0.95.

The total number of robust features in the unfiltered 4D-CT with three breathing phases was 920/1170 (79%), which was closer to that for test–retest [967/1170 (83%)] compared with the result of unfiltered 4D-CT with 10 phases [397/1170 (34%)]. However, as shown in the Venn chart (Fig. 5), the breakdown of features discarded as non-robust varied between 4D-CT and test–retest—the non-robust features in 4D-CT with 10 breathing phases covered 80% (n = 163) of those in test–retest and the non-robust features in 4D-CT with three breathing

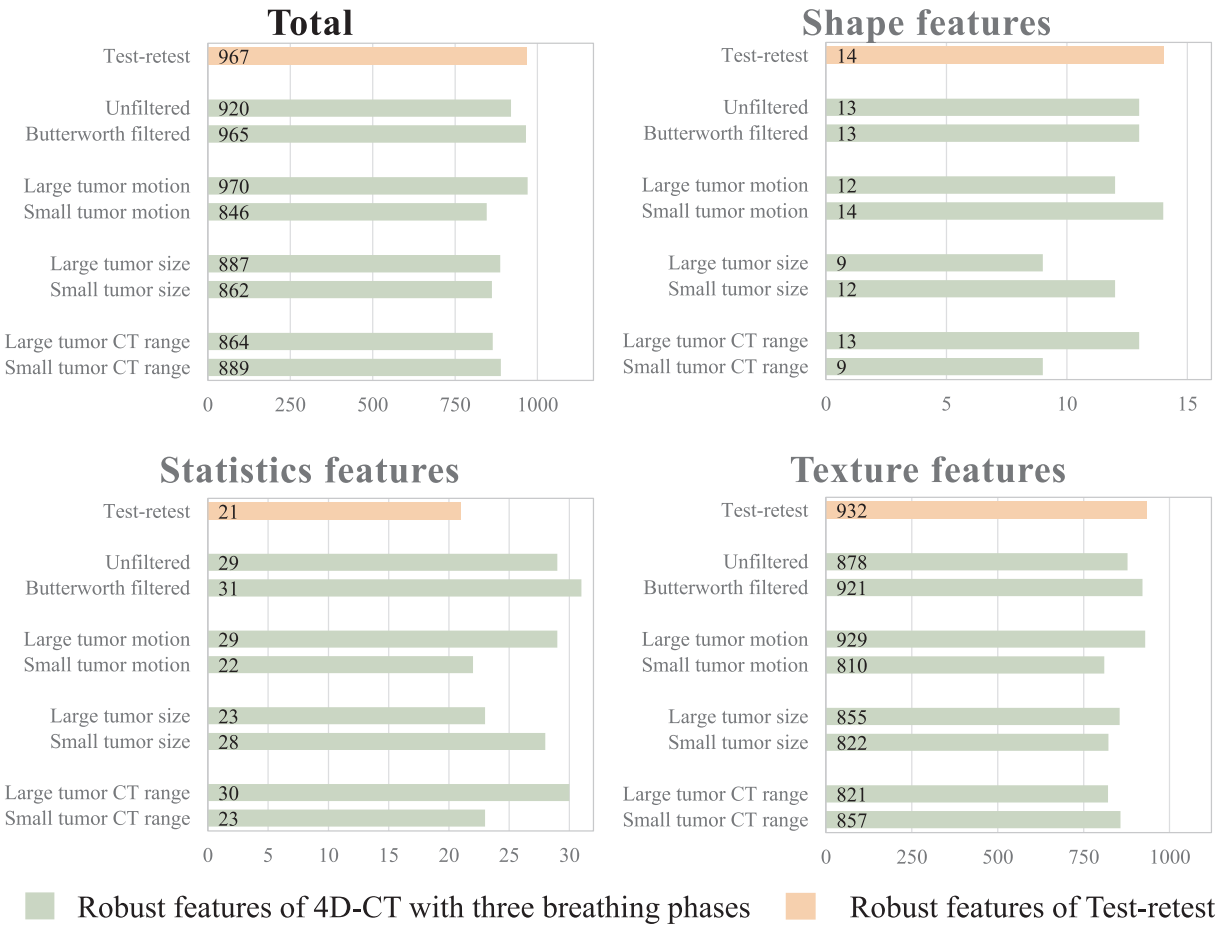

Fig. 4. The number of robust features of test–retest and all groups of four-dimensional computed tomography (4D-CT) with three breathing phases. Top, the result of all features (n = 1170; left) and shape features (n = 16; right). Bottom, the result of statistics features (n = 32; left) and texture features (n = 1122; right).

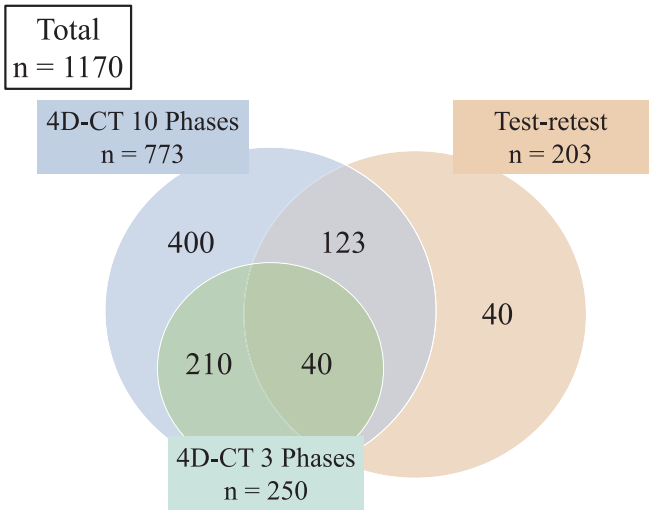

Fig. 5. The Venn chart of the total number of non-robust features [concordance correlation coefficient (CCC) threshold = 0.85] across three datasets [unfiltered four-dimensional computed tomography (4D-CT) 10 breathing phases, three breathing phases, and test–retest]

phases covered only 20% (n = 40) of those in test–retest—which could be attributed to differences in the imaging technique and scanner setting between 4D-CT and test–retest. Regarding the imaging technique, 4D-CT images included residual motion artifacts, phase error, and different breathing phase. Conversely, test–retest, which comprises two repeated breath hold CT images, included the differences in patients' position and breathing phases between two scans; thus, test–retest has different imaging techniques from 4D-CT, except for different breathing phase. Regarding the scanner setting, different parameters existed between 4D-CT and test–retest. For example, the tube currents, slice thicknesses, pixel sizes, and reconstruction kernels in 4D-CT and test–retest were 120 mA and 185–439 mA, 2.5 and 1.25 mm, 0.98 × 0.98, and 0.58 × 0.58–0.78 × 0.78 mm, and standard and lung kernel, respectively. Reportedly, the differences in the slice thickness, reconstruction algorithms, and pixel size markedly affected the accuracy of radiomic features [10,12,13]. These different errors between 4D-CT and test–retest caused a relatively small agreement in non-robust features between 4D-CT with three phases and test–retest. An ongoing study is investigating the difference in the number of robust features between 4D-CT and repeated breath hold CT using the same CT protocol with anthropomorphic phantom created by 3D printer.

In this study, the number of robust features was 397/1170 (34%) in the unfiltered 4D-CT with 10 phases and 967/1170 (83%) in test–retest, suggesting that the number of robust features in test–retest was markedly higher than that in 4D-CT with 10 breathing phases. In addition, based on histograms of the percentage of robust features (Fig. 6a), all 4D-CT groups displayed smaller robust features than test–retest at all CCC threshold values. Nevertheless, Larue et al. reported that the number of robust features based on the unfiltered data was 74/133 (56%) in 4D-CT and 80/133 (60%) in test–retest at the CCC threshold value of 0.85, indicating the number of robust features was almost same [23]; this result contradictory to our study, which could be attributed to differences in radiomic features, feature parameters, segmentation, or CT protocols, and patients between that and this study. In addition,

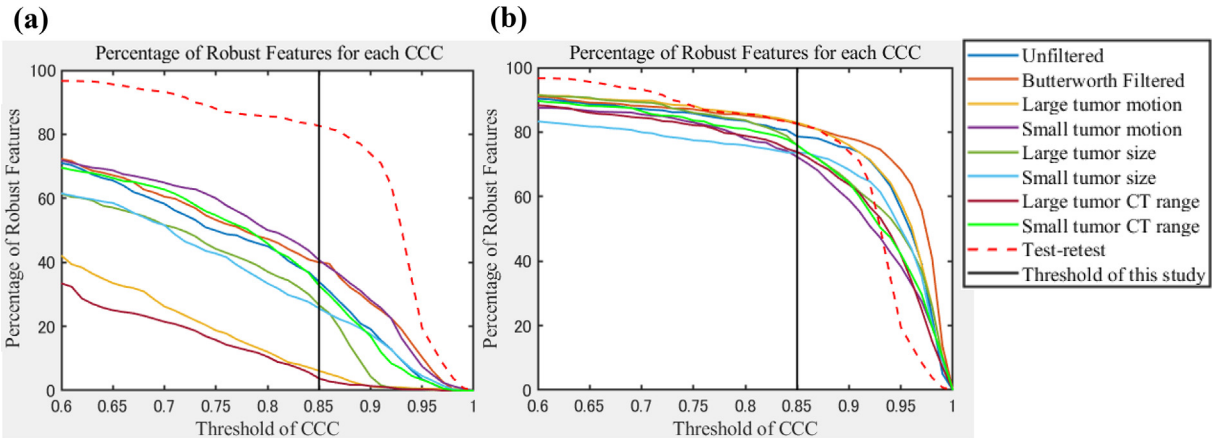

Fig. 6. The percentage of robust features according to the concordance correlation coefficient (CCC) threshold values for four-dimensional computed tomography (4D-CT) with 10 breathing phases (a) and 4D-CT with three breathing phases (b). Red dotted line, the result of test–retest; other lines, results of 4D-CT. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Larue et al. [23] used GLCM3 as the mean of the feature calculations for each of the 13 directions, and voxel distance was only one. Furthermore, they included the Gray-Level Distance Zone features (GLDZM) [41], Gray-Level Size Zone features (GLSZM) [42], and Neighboring Gray-Level Dependence features (NGLDM) [43], which were not used in this study. Nevertheless, ascertaining the reasons for these differences is difficult, thereby necessitating further research to clarify this issue.

5. Conclusions

This study reveals that if all 10 breathing phases of 4D-CT are used as the first step in a feature selection, excessive dimension reduction might be implemented on the basis of 4D-CT datasets and CCC threshold values, thereby introducing a risk that even useful features might be eliminated for subsequent analyses. Hence, using 4D-CT datasets around the end-exhale phase that has small respiratory motion might prevent excessive dimension reduction in any 4D-CT dataset, and non-robust features are eliminated as the first step of the feature selection. Furthermore, thoracic 4D-CT datasets are available in facilities conducting radiotherapy; thus, 4D-CT can select robust feature suitable for the CT protocol of your facility. However, the features discarded as non-robust differ between test–retest and 4D-CT; hence, further studies are warranted to investigate this difference.

Conflicts of interest

There is no conflict of interest with regard to this manuscript.

References

[1] Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer 2012;48:441–6.
[2] Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5:4006.
[3] Coroller TP, Agrawal V, Narayan V, Hou Y, Grossmann P, Lee SW, et al. Radiomic phenotype features predict pathological response in non-small cell lung cancer. Radiother Oncol 2016;119:480–6.
[4] Huynh E, Coroller TP, Narayan V, Agrawal V, Hou Y, Romano J, et al. CT-based radiomic analysis of stereotactic body radiation therapy patients with lung cancer. Radiother Oncol 2016;120:258–66.
[5] Jensen GL, Yost CM, Mackin DS, Fried DV, Zhou S, Court LE, et al. Prognostic value of combining a quantitative image feature from positron emission tomography with clinical factors in oligometastatic non-small cell lung cancer. Radiother Oncol 2017.
[6] Li Q, Kim J, Balagurunathan Y, Liu Y, Latifi K, Stringfield O, et al. Imaging features from pretreatment CT scans are associated with clinical outcomes in nonsmall-cell lung cancer patients treated with stereotactic body radiotherapy. Med Phys 2017;44:4341–9.
[7] Ouyang FS, Guo BL, Zhang B, Dong YH, Zhang L, Mo XK, et al. Exploration and validation of radiomics signature as an independent prognostic biomarker in stage III-IVb nasopharyngeal carcinoma. Oncotarget 2017;8:74869–79.
[8] van Timmeren JE, Leijenaar RTH, van Elmpt W, Reymen B, Oberije C, Monshouwer R, et al. Survival prediction of non-small cell lung cancer patients using radiomics analyses of cone-beam CT images. Radiother Oncol 2017;123:363–9.
[9] Zhang B, He X, Ouyang F, Gu D, Dong Y, Zhang L, et al. Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma. Cancer Lett 2017;403:21–7.
[10] Shafiq-Ul-Hassan M, Zhang GG, Latifi K, Ullah G, Hunt DC, Balagurunathan Y, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. Med Phys 2017;44:1050–62.
[11] Mackin D, Fave X, Zhang L, Fried D, Yang J, Taylor B, et al. Measuring computed tomography scanner variability of radiomics features. Invest Radiol 2015;50:757–65.
[12] Zhao B, Tan Y, Tsai WY, Qi J, Xie C, Lu L, et al. Reproducibility of radiomics for deciphering tumor phenotype with imaging. Sci Rep 2016;6:23428.
[13] 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.
[14] Parmar C, Rios Velazquez E, Leijenaar R, Jermoumi M, Carvalho S, Mak RH, et al. Robust Radiomics feature quantification using semiautomatic volumetric segmentation. PLoS One 2014;9:e102107.
[15] van Velden FH, Kramer GM, Frings V, Nissen IA, Mulder ER, de Langen AJ, et al. Repeatability of radiomic features in non-small-cell lung cancer [(18)F]FDG-PET/CT studies: impact of reconstruction and delineation. Mol Imag Biol 2016;18:788–95.
[16] Yip S, McCall K, Aristophanous M, Chen AB, Aerts HJ, Berbeco R. Comparison of texture features derived from static and respiratory-gated PET images in non-small cell lung cancer. PLoS One 2014;9:e115510.
[17] Oliver JA, Budzevich M, Zhang GG, Dilling TJ, Latifi K, Moros EG. Variability of image features computed from conventional and respiratory-gated PET/CT images of lung cancer. Transl Oncol 2015;8:524–34.
[18] Fave X, Mackin D, Yang J, Zhang J, Fried D, Balter P, et al. Can radiomics features be reproducibly measured from CBCT images for patients with non-small cell lung cancer? Med Phys 2015;42:6784–97.
[19] Zhao B, Schwartz LH, Kris MG. Data from RIDER_Lung CT. *The Cancer Imaging Archive* 2015. https://doi.org/10.7937/K9/TCIA.2015.U1X8A5NR.
[20] Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. J Digit Imag 2013;26:1045–57.
[21] Zhao B, James LP, Moskowitz CS, Guo P, Ginsberg MS, Lefkowitz RA, et al. Evaluating variability in tumor measurements from same-day repeat CT scans of patients with non-small cell lung cancer. Radiology 2009;252:263–72.
[22] Huynh E, Coroller TP, Narayan V, Agrawal V, Romano J, Franco I, et al. Associations of radiomic data extracted from static and respiratory-gated CT scans with disease recurrence in lung cancer patients treated with SBRT. PLoS One 2017;12:e0169172.
[23] Larue R, Van De Voorde L, van Timmeren JE, Leijenaar RTH, Berbee M, Sosef MN, et al. 4DCT imaging to assess radiomics feature stability: an investigation for thoracic cancers. Radiother Oncol J Eur Soc Therap Radiol Oncol 2017;125:147–53.
[24] Amadasun M, King R. Textural features corresponding to textural properties. IEEE Trans Syst Man Cyber 1989;19:1264–74.
[25] Galloway MM. Texture analysis using grey level run lengths. NASA STI/Recon Technical Report N. 1974;75.
[26] Haralick RM, Shanmugam K. Textural features for image classification. IEEE Trans Syst Man Cyber 1973:610–21.
[27] 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;42:1341–53.
[28] Zwanenburg A, Leger S, Vallières M, Löck S. Image Biomarker Standardisation

Initiative arXiv preprint arXiv:161207003 2016.
[29] Hatt M, Vallieres M, Visvikis D, Zwanenburg A. IBSI: an international community radiomics standardization initiative. J Nucl Med 2018;59:287-.
[30] Apte AP, Iyer A, Crispin-Ortuzar M, Pandya R, van Dijk LV, Spezi E, et al. Technical note: extension of CERR for computational radiomics: a comprehensive MATLAB platform for reproducible radiomics research. Med Phys 2018.
[31] Ferreira Junior JR, Koenigkam-Santos M, Cipriano FEG, Fabro AT, Azevedo-Marques PM. Radiomics-based features for pattern recognition of lung cancer histopathology and metastases. Comput Meth Prog Biomed 2018;159:23–30.
[32] Watkins WT, Li R, Lewis J, Park JC, Sandhu A, Jiang SB, et al. Patient-specific motion artifacts in 4DCT. Med Phys 2010;37:2855–61.
[33] Minohara S, Kanai T, Endo M, Noda K, Kanazawa M. Respiratory gated irradiation system for heavy-ion radiotherapy. Int J Radiat Oncol Biol Phys 2000;47:1097–103.
[34] Lin LI. A concordance correlation coefficient to evaluate reproducibility. Biometrics 1989;45:255–68.
[35] Pan T, Lee TY, Rietzel E, Chen GT. 4D-CT imaging of a volume influenced by respiratory motion on multi-slice CT. Med Phys 2004;31:333–40.
[36] Rietzel E, Pan T, Chen GT. Four-dimensional computed tomography: image formation and clinical protocol. Med Phys 2005;32:874–89.
[37] Mexner V, Wolthaus JW, van Herk M, Damen EM, Sonke J-J. Effects of respiration-induced density variations on dose distributions in radiotherapy of lung cancer. Int J Radiat Oncol Biol Phys 2009;74:1266–75.
[38] Mutaf Y, Antolak J, Brinkmann D. The impact of temporal inaccuracies on 4DCT image quality. Med Phys 2007;34:1615–22.
[39] Castillo SJ, Castillo R, Balter P, Pan T, Ibbott G, Hobbs B, et al. Assessment of a quantitative metric for 4D CT artifact evaluation by observer consensus. J Appl Clin Med Phys 2014;15:190–201.
[40] Yang J, Zhang L, Fave XJ, Fried DV, Stingo FC, Ng CS, et al. Uncertainty analysis of quantitative imaging features extracted from contrast-enhanced CT in lung tumors. Comput Med Imaging Graph 2016;48:1–8.
[41] Thibault G, Angulo J, Meyer F. Advanced statistical matrices for texture characterization: application to cell classification. IEEE Trans Bio-medical Eng 2014;61:630–7.
[42] Tixier F, Le Rest CC, Hatt M, Albarghach N, Pradier O, Metges JP, et al. Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. J Nucl Med Off Publ Soc Nucl Med 2011;52:369–78.
[43] Stoecker WV, Chiang CS, Moss RH. Texture in skin images: comparison of three methods to determine smoothness. Comput Med Imag Graph Off J Comput Med Imag Soc 1992;16:179–90.