



# Phantom-based radiomics feature test–retest stability analysis on photon-counting detector CT

Alexander Hertel<sup>1</sup> · Hishan Tharmaseelan<sup>1</sup> · Lukas T. Rotkopf<sup>1,2</sup> · Dominik Nörenberg<sup>1</sup> · Philipp Riffel<sup>1</sup> · Konstantin Nikolaou<sup>3</sup> · Jakob Weiss<sup>4</sup> · Fabian Bamberg<sup>4</sup> · Stefan O. Schoenberg<sup>1</sup> · Matthias F. Froelich<sup>1</sup> · Isabelle Ayx<sup>1</sup>

Received: 15 August 2022 / Revised: 2 January 2023 / Accepted: 22 January 2023 / Published online: 21 February 2023  
© The Author(s) 2023

## Abstract

**Objectives** Radiomics image data analysis offers promising approaches in research but has not been implemented in clinical practice yet, partly due to the instability of many parameters. The aim of this study is to evaluate the stability of radiomics analysis on phantom scans with photon-counting detector CT (PCCT).

**Methods** Photon-counting CT scans of organic phantoms consisting of 4 apples, kiwis, limes, and onions each were performed at 10 mAs, 50 mAs, and 100 mAs with 120-kV tube current. The phantoms were segmented semi-automatically and original radiomics parameters were extracted. This was followed by statistical analysis including concordance correlation coefficients (CCC), intraclass correlation coefficients (ICC), as well as random forest (RF) analysis, and cluster analysis to determine the stable and important parameters.

**Results** Seventy-three of the 104 (70%) extracted features showed excellent stability with a CCC value > 0.9 when compared in a test and retest analysis, and 68 features (65.4%) were stable compared to the original in a rescan after repositioning. Between the test scans with different mAs values, 78 (75%) features were rated with excellent stability. Eight radiomics features were identified that had an ICC value greater than 0.75 in at least 3 of 4 groups when comparing the different phantoms in a phantom group. In addition, the RF analysis identified many features that are important for distinguishing the phantom groups.

**Conclusion** Radiomics analysis using PCCT data provides high feature stability on organic phantoms, which may facilitate the implementation of radiomics analysis likewise in clinical routine.

## Key Points

- Radiomics analysis using photon-counting computed tomography provides high feature stability.
- Photon-counting computed tomography may pave the way for implementation of radiomics analysis in clinical routine.

**Keywords** Tomography, X-ray computed · Data science · Radiology

Alexander Hertel and Hishan Tharmaseelan contributed equally.

✉ Matthias F. Froelich  
Matthias.froelich@medma.uni-heidelberg.de

<sup>1</sup> Department of Radiology and Nuclear Medicine, University Medical Center Mannheim, Heidelberg University, Theodor-Kutzer-Ufer 1-3, 68167 Mannheim, Germany

<sup>2</sup> Department of Radiology, German Cancer Research Center, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany

<sup>3</sup> Department of Diagnostic and Interventional Radiology, University of Tübingen, Hoppe-Seyler-Straße 3, 72076 Tübingen, Germany

<sup>4</sup> Department of Diagnostic and Interventional Radiology, Medical Center-University of Freiburg, Hugstetter Str. 55, 79106 Freiburg Im Breisgau, Germany

## Abbreviations

CCC	Concordance correlation coefficient
CT	Computed tomography
EID	Energy-integrating detector
GLCM	Gray level co-occurrence matrix
GLDM	Gray level dependence matrix
GLRLM	Gray level run length matrix
GLSZM	Gray level size zone matrix
ICC	Intraclass correlation coefficient
MRI	Magnetic resonance imaging
NGTDM	Neighboring gray tone difference matrix
PCCT	Photon-counting computed tomography
PCD	Photon-counting detector
RF	Random forest
ROI	Region of interest
SD	Standard deviation

## Introduction

Radiomics analysis is widely used in clinical research due to the rising interest in transforming subjective and qualitative to objective and quantifiable medical image analysis. This pixel-based extraction of features from imaging data promises additional insights beyond the level of information visible to the human eye [1] using dedicated software packages [2]. The further information consists of a vast number of features of the region of interest (ROI), fostering the big data trends in healthcare and creating new possibilities and promises, especially in oncologic imaging [3, 4]. It is evident that radiomics is able to outperform traditional clinical scores [5, 6] and shows promising results in terms of outcome prediction [7, 8], and tumor classification [4, 9, 10]. Besides oncologic imaging, radiomics analysis paved the way for future developments also in cardiovascular imaging by outlining a better risk stratification, e.g., via coronary plaque analysis [11], and defining potential biomarkers for the development of arteriosclerosis [12]. Additionally, potential myocardial diffuse fibrosis could be detected by radiomics analysis [13]. The power of radiomics analyses has also been shown in lung fibrosis [14, 15], kidney stones [16], and COVID pneumonia [17, 18]. Even the ability of a CT-derived radiomics score to predict the added benefit of adjuvant chemotherapy following surgery in patients with non-small cell lung cancer could be demonstrated by using three intratumoral and 10 peritumoral CT radiomics features [19].

Despite these excellent capabilities, radiomics analysis has not been implemented in clinical practice. One of the most important limitations for clinical implementation is the lack of comparability of radiomics analyses. In the past, several studies have demonstrated the influence of various parameters on the stability of radiomics features, including multiple image acquisition parameter settings such as tube voltage, reconstruction kernel, choice of contrast agent and contrast media phases, slice thickness, and the choice of the scanner which impede the reproducibility [20, 21].

Even the choice of segmentation method [22] and feature extraction software can have an influence on radiomics feature stability [23]. For accurate texture analysis, optimal spatial resolution and signal-to-noise ratio have been defined as the two most important image quality factors [24–26]. In the past, these parameters may have been impacted by the indirect conversion of X-ray photons to an electrical signal with an intermediate scintillator-based step.

Through the implementation of photon-counting CT (PCCT), this hurdle could be possibly addressed. This new technology allows direct conversion of the X-ray photons into electric pulses without the need of converting into visible light, as in traditional energy-integrating CT (EICT). Hence, each photon that reaches the detector contributes to the final image. Additionally, a threshold for image noise

can be set for each photon, reducing the overall image noise and in total resulting in better spatial resolution, lower beam-hardening artifacts, and better signal-to-noise ratio [27, 28].

Primary studies have shown radiomics feature properties compared between EICT and PCCT [29, 30], but to our knowledge, an organic phantom-based analysis regarding scan and rescan as well as reposition feature stability with mAs variance on PCCT has not been studied. The aim of this study is to analyze feature stability within the organic phantoms as well as the possibility of radiomics-based differentiation between different phantom groups. We hypothesized that given its technical innovation, PCCT may provide improved radiomics feature stability.

## Materials and methods

### Study design

For this phantom-based single-center study, 16 organic phantoms were included consisting of four apples, four onions, four limes, and four kiwis. The methodological structure of the study is shown in Fig. 1.

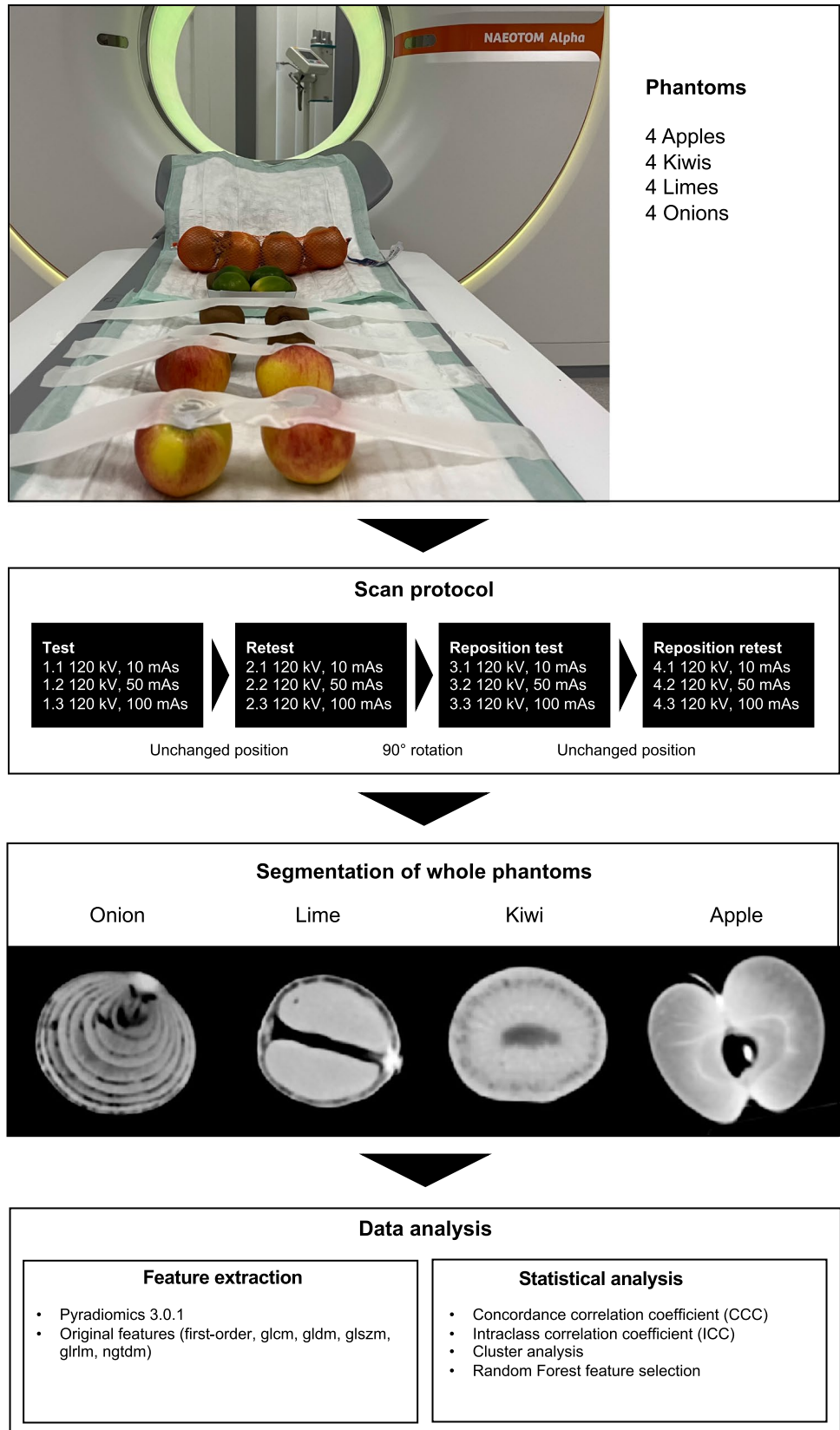
### CT imaging

All 16 phantoms were scanned in one session on a first-generation whole-body dual-source PCCT system (NAEOTOM Alpha; Siemens Healthcare GmbH). Underlying scan parameters were a tube voltage of 120 kV and gantry rotation time of 0.25 s. The in-plane resolution of the PCCT is 0.11 mm. A three-level variance was established for mAs, consisting of a scan with 10 mAs, 50 mAs, and 100 mAs. Each of these scans was repeated again immediately afterward without any change in the phantom position or the underlying scan parameters. In the next step, all phantoms were rotated 90° in the same direction (clockwise) and all scans including rescans and mAs variation were repeated (Fig. 1).

### CT imaging analysis

Axial images of all scans were reconstructed with a slice thickness of 1 mm (increment 1 mm) using a soft tissue kernel (Br40). This data was exported and stored in Digital Imaging and Communications in Medicine (DICOM) file format. DICOM files were converted to Neuroimaging Informatics Technology Initiative (NIFTI) file format for further processing with a dedicated segmentation tool (MITK workbench, version 2021.10) [31]. Segmentation for each organic phantom in each scan was done semi-automatically by a clinical radiologist with 3 years of experience in segmentation. Figure 2 shows an example segmentation as well as the original DICOM images.

Fig. 1 Study flowchart



## Radiomics feature extraction

Phantom segmentations were further analyzed by radiomics feature extraction using a dedicated imaging biomarker standardization initiative definition–based Python package (PyRadiomics, version 3.0.1.) [2]. For each phantom regarding each scan, first-order features, shape features, and second-order features (in total 104 features), namely gray level co-occurrence matrix (GLCM), gray level dependence matrix (GLDM), gray level size zone matrix (GLSZM), gray level run length matrix (GLRLM), and neighboring gray tone difference matrix (NGTDM), were extracted. The extraction was performed with voxel normalization, resampling to  $2 \times 2 \times 2$  mm and rebinning with a fixed bin width of 25 HU. The Chebyshev distance is 1.

## Statistical analysis

All statistical analyses were performed in R (R Core Team 2021) and RStudio (version 1.3.1093) [32].

The statistical analysis was performed in four steps (I to IV):

In the first step (I), concordance correlation coefficient (CCC) analysis was used to determine the feature stability between test and retest for the original and repositioned phantoms. In the CCC analysis, a default of  $\geq 0.9$  was defined as a threshold for feature stability in the test–retest setting. This threshold was defined in accordance to McBride et al (2005): “A proposal for strength-of-agreement criteria for Lin’s concordance” [33].

In the second step (II), the feature stability between the scans with different mAs values (10 mAs, 50 mAs, and 100 mAs) was determined by intraclass correlation coefficient (ICC) analysis, for which ICC (3,1) (a fixed set of raters, no average observations) was used as recommended in literature [34]. In the ICC analyses performed, in accordance

with Koo and Li, values below 0.5, between 0.5 and 0.75, between 0.75 and 0.9, and above 0.90 were considered suggestive of poor, moderate, good, and excellent stability [35]. In the next step (III), two different stability tests were performed: In the inter-phantom group stability test, the stability of the radiomics features was compared between two different groups (for example, the group of apples and onions), while in the intra-phantom group stability test, the stability of the individual phantoms within a phantom group (for example, between apples 1 and 4) was tested using ICC analysis. For this ICC analysis, ICC (3,1) was likewise used as explained in step II.

In step IV, a random forest (RF) analysis was performed with the Boruta algorithm in R [36] to determine the feature importance with respect to the discrimination of the different phantom groups. For each parameter variable, shadow variables are created by permutation by the algorithm. If the real variable is significantly more important than the shadow variable, high importance is assigned to the feature by the algorithm.

In addition, the feature correlations were calculated by Pearson’s correlation coefficients and were visualized in form of heatmaps using the “ComplexHeatmap” package in R.

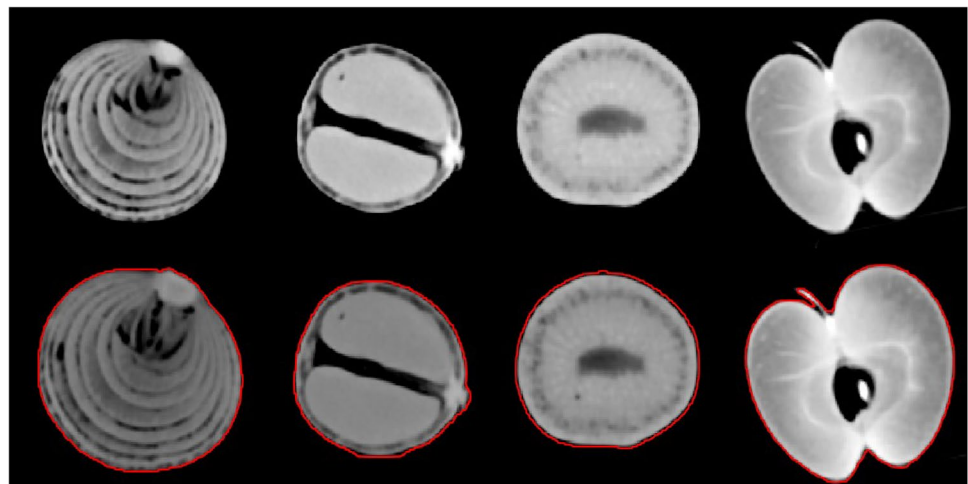
## Results

### Feature stability analysis

#### Test–retest stability and reposition stability (I)

In the CCC analysis performed (I), 73 features (70% of the extracted features) with excellent stability (CCC value  $> 0.9$ ) could be identified in the comparison between the test and retest scan (Supplementary Table 1). For non-shape features,

**Fig. 2** Example segmentation of each phantom group



66% of the features still showed excellent stability. All of the extracted shape features showed excellent stability. Percentage-wise, there were more stable first-order features than second-order features: 14 of the 18 extracted first-order features (78%) showed excellent stability with a CCC value above 0.9 while 46 of the 73 extracted second-order features (63%) showed excellent test–retest stability.

In comparison between the original test and the first reposition scan, 68 features showed a CCC value above 0.9 (65.4%).

### Stability with different mAs values (II)

The ICC analysis in the next evaluation step (II) of the test scans with different mAs values resulted in 78 (75%) features with excellent stability.

### Inter-phantom group stability (III)

The phantom group is defined by the four fruit/vegetable categories. To analyze the feature stability between the different phantom groups, all apples, onions, kiwis, and limes respectively were included in one group each. To compare the group of, e.g., apples against the group of, e.g., onions, an inter-phantom group stability test was performed. In step (III), the comparison of radiomics stability of individual phantoms within a phantom group resulted in a total of 8 features with an ICC value  $> 0.75$  across at least 3 of the total 4 phantom groups: *original\_firstorder\_90Percentile*, *original\_firstorder\_Median*, *original\_firstorder\_RobustMeanAbsoluteDeviation*, *original\_glcm\_Correlation*, *original\_glszm\_GrayLevelNonUniformityNormalized*, *original\_glszm\_SizeZoneNonUniformity*, *original\_glszm\_ZonePercentage*, and *original\_ngtdm\_Contrast* (see Table 1). The results of the ICC analysis are presented in Fig. 3.

### Intra-phantom group stability analysis

To investigate the comparative stability of radiomics features within each phantom group, ICC values were compared

(Fig. 3). For the intra-phantom group stability analysis, the phantom group was defined by the fruit/vegetable category, meaning one group consisted, for example, of four apples. For apple and onion phantoms, two features reached an ICC value of above 0.9. For the kiwi and lime phantom, this value was reached in 29 and 19 features, respectively.

Table 2 offers an overview of the distribution of stable features by test type and by category.

## Feature selection

### Feature importance (IV)

The performed Boruta random forest analysis (step IV) revealed that 88 of the 104 extracted radiomics features (84.6%) were important for the discrimination between different phantom types. The results are shown as a Boruta plot in Fig. 4.

## Cluster analysis

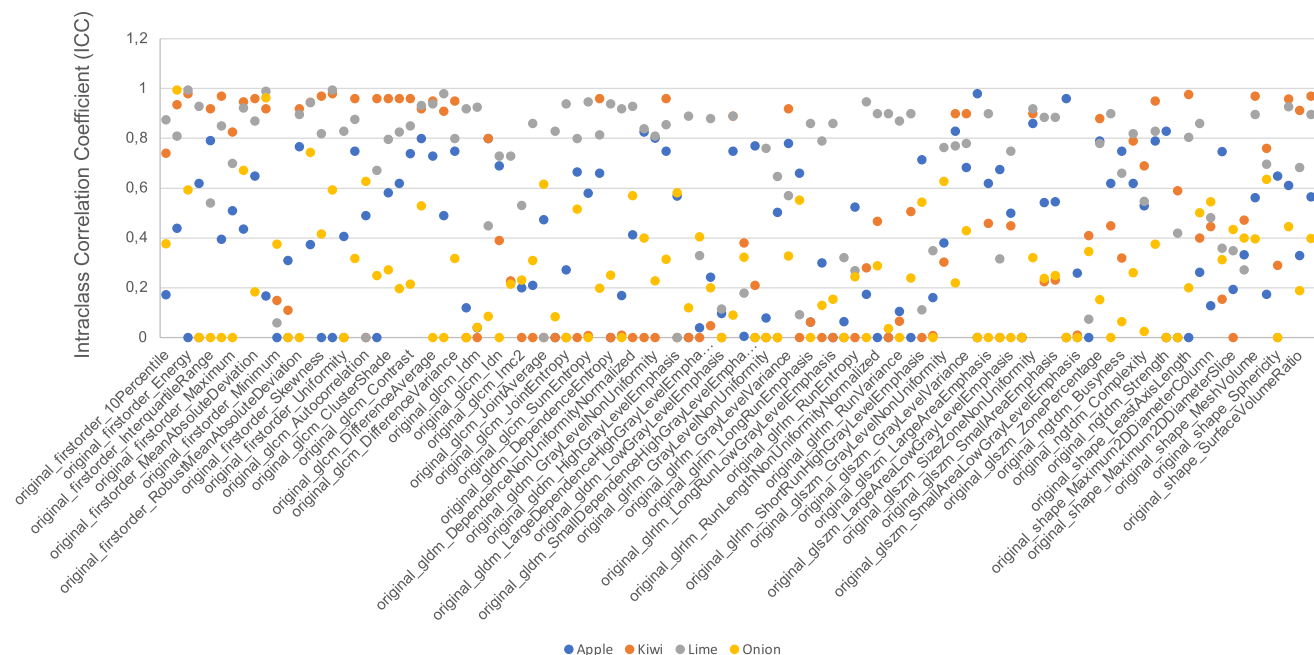
The created phantom selective heatmap with all extracted radiomics features is shown in Fig. 5. The different phantom groups can be visually distinguished as individual clusters. Figure 6 shows a heatmap with the features determined in the ICC analyses (step III) that are stable over at least 3 phantom groups.

## Discussion

In this study, we demonstrated that the vast majority of the extracted radiomics features (70% of the parameters) have excellent test–retest stability on scans in the novel photon-counting CT. Several shape-based features and first-order features show a CCC value of  $> 0.999$ . *DependenceNonUniformity*, as a measurement of the similarity of dependence throughout the image, is the non-shape-based feature with the highest CCC value (0.99929).

**Table 1** Stable features defined as ICC  $> 0.75$  (analysis of the radiomics stability between the individual phantoms of a phantom group) in at least 3 phantom groups

Feature	Apple	Kiwi	Lime	Onion
<i>original_firstorder_90Percentile</i>	0.44	<b>0.936</b>	<b>0.81</b>	<b>0.995</b>
<i>original_firstorder_Median</i>	0.167	<b>0.92</b>	<b>0.99</b>	<b>0.9636</b>
<i>original_firstorder_RobustMeanAbsoluteDeviation</i>	<b>0.767</b>	<b>0.92</b>	<b>0.897</b>	$< 0.001$
<i>original_glcm_Correlation</i>	<b>0.8</b>	<b>0.92</b>	<b>0.932</b>	0.529
<i>original_glszm_GrayLevelNonUniformityNormalized</i>	<b>0.83</b>	<b>0.9</b>	<b>0.77</b>	0.22
<i>original_glszm_SizeZoneNonUniformity</i>	<b>0.86</b>	<b>0.9</b>	<b>0.92</b>	0.322
<i>original_glszm_ZonePercentage</i>	<b>0.79</b>	<b>0.88</b>	<b>0.781</b>	0.1535
<i>original_ngtdm_Contrast</i>	<b>0.79</b>	<b>0.95</b>	<b>0.83</b>	0.376



**Fig. 3** ICC analysis between the individual phantoms of a phantom group

In step III, 8 features could be identified, which were stable in 3 of 4 phantom groups, respectively (see Table 1). It shows that the selected first-order features in apples do not seem to be stable (possibly due to a larger individual deviation of shape and size within the apple group compared to the other phantom groups) as well as instability of all selected second-order features in onions. This may be explained by the multilayer structure of the onions with air between the individual layers, which could lead to larger differences in the calculation of the non-shape features. Given the varying feature stability measured as ICC between the phantom groups, an inherent phantom-related heterogeneity cannot be fully ruled out. Therefore, the overall feature stability may be underestimated.

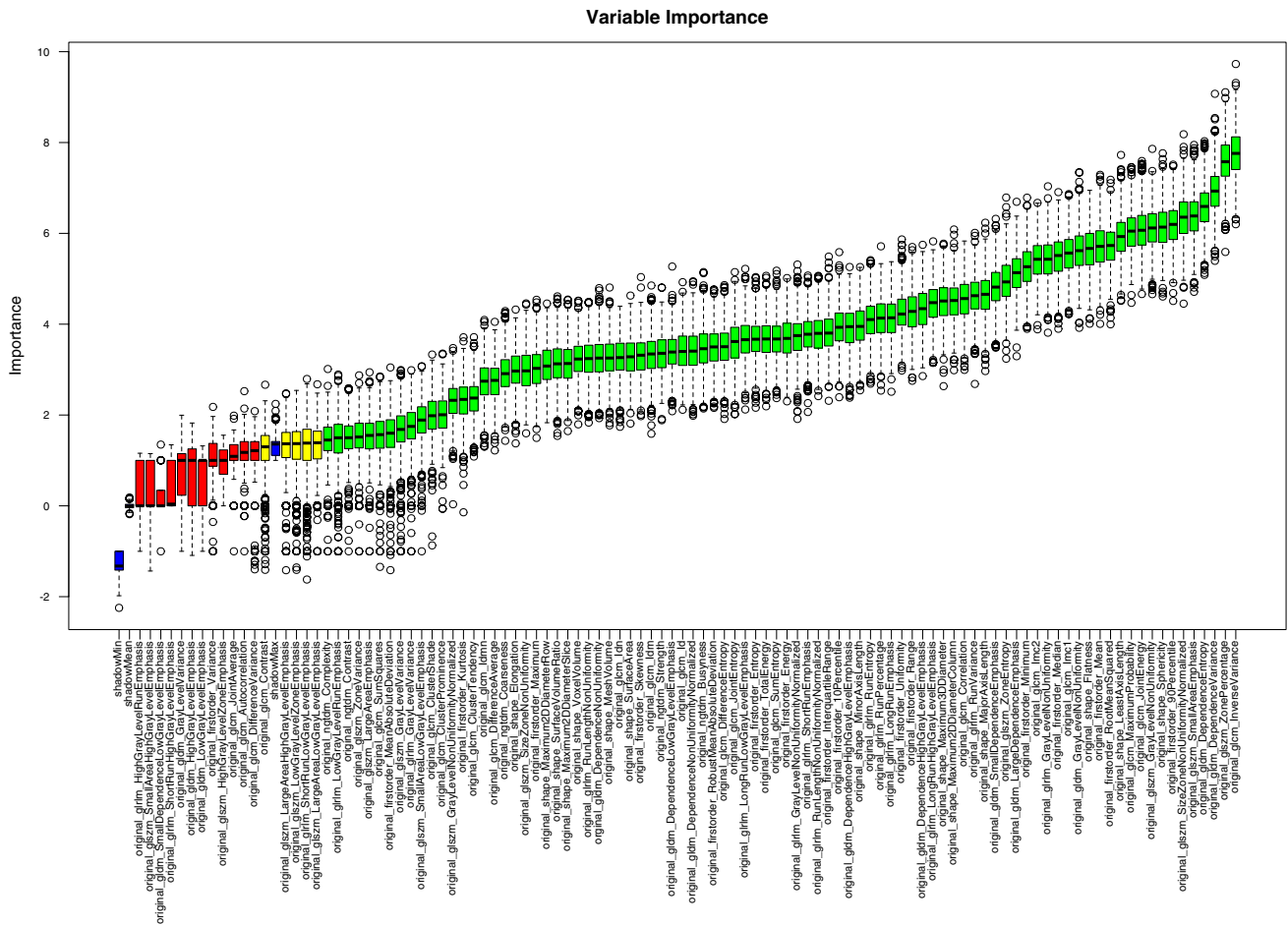
To date, radiomics analysis has not yet been established in daily clinical practice due to the lack of comparability. In the past, multiple studies showed the influence of variable parameters on radiomics texture analysis. Zhao et al investigated the influence of different slice thicknesses on radiomics features using a chest phantom. In comparison to 5-mm slice thickness, 1.25- and 2.5-mm slice thicknesses were

better suitable for homogeneity, volume, density means, and density SD gray level co-occurrence matrix (GLCM) energy. Additionally, the influence of different reconstruction algorithms on radiomics features was analyzed, outlining the standard reconstruction algorithms being best for density SD, whereas the lung reconstruction algorithms for density to mean, respectively [37].

In a phantom study by Jensen et al, radiomics features differed between ROI sizes and volume in both MRI and CT even though a homogenous phantom without any texture differences was used. Comparing both techniques, CT features were more stable than MRI features when tested for significance with the Mann–Whitney *U* test. On the other hand, OCCCs show excellent (> 0.90) agreement for different MR-derived features, but not for CT-derived features [38]. Mackin et al investigated the influence of sixteen different CT scanners, produced by four different manufacturers, and different acquisition parameters on a CT phantom in 2018. They outlined a changeability of radiomics features between the different scanners, suggesting that the comparability of radiomics studies depends on the consistency of

**Table 2** Distribution of stable features by test type (test–retest stability, reposition stability, and mAs stability) and by feature category (shape features, first-order features, and second-order features)

	Stable shape features in %	Stable first-order features in %	Stable second-order features in %
Test–retest stability (CCC)	100	77.8	63
Reposition stability (CCC)	85	66.7	61.6
mAs stability (ICC)	100	94.4	65.8



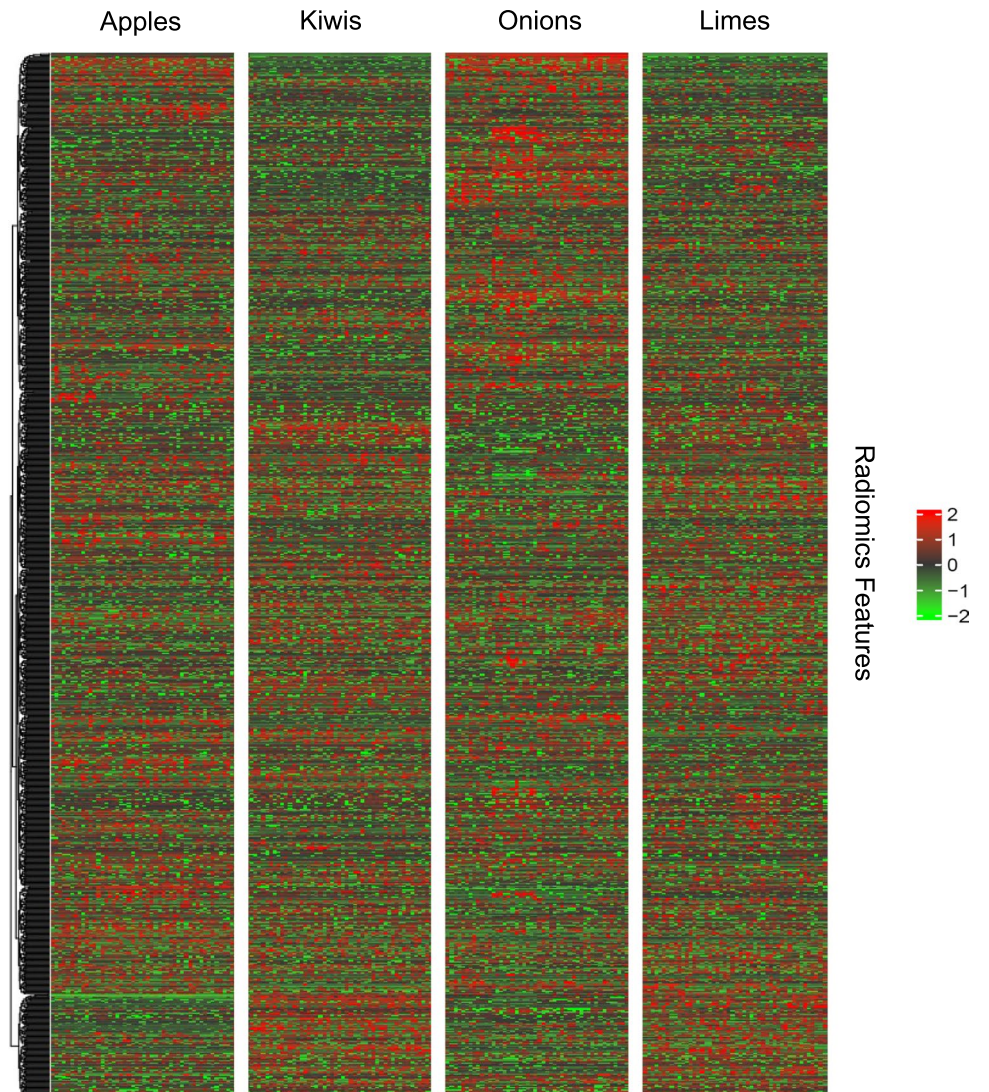
**Fig. 4** Output of the random forest feature selection ranked by importance using the Boruta algorithm

image acquisition and reconstruction. Additionally, Mackin et al compared the radiomics features of a phantom with radiomics features derived from non-small cell lung cancer (NSCLC) tumors and proved comparability of variability in radiomics features between both groups [39].

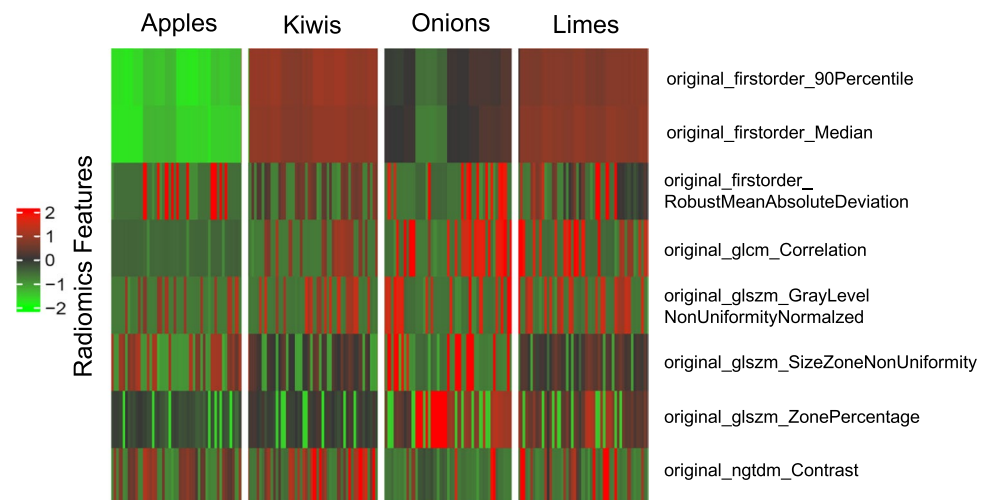
Recently, Dunning et al scanned different organic samples (zucchini, onions, and oranges) on photon-counting detector CT (PCD-CT) and energy-integrating detector CT (EID-CT) at two dose levels. They calculated fourteen relevant radiomics features in each image for each sample and compared those between both scanners. Their results outlined a notable change by at least 50% of 13 out of 14 radiomics features at both dose levels by improved resolution of PCD-CT. Additionally, the higher Dunn Index of radiomics features derived from PCD-CT indicates better cluster separation for classification (Dunn Index 1.99 and 1.80 for PCD-CT and EID-CT respectively at 10 mGy and 2.44 (PCD-CT) and 1.65 (EID-CT) at 60 mGy), indicating an impact of high-resolution PCD-CT on radiomics analysis [30]. Our study goes one step further and outlines that this potential impact of high-resolution PCCT on radiomics

analysis offers excellent test–retest stability on an organic phantom. In comparison, Peng et al detected stability of radiomics features in a test–retest setting between different time points on different conventional EID-CT in only 20.93% (CCC  $0.56 \pm 0.31$ ) using an anthropomorphic thorax phantom consisting of three parts: body model, the internal structure of the lung, and simulated nodules. In further investigation steps, they compared the stability on different CT scanners and with varying image acquisition and reconstruction parameters including pitch, rotation time, tube voltage, tube current, FOV, slice thickness, reconstruction kernel, and iteration level. The tube current (mA/s) showed an ICC of  $0.54 \pm 0.32$  (34.29%) [40]. Another study recently compared feature properties of non-scarred left ventricular myocardium between EICT and PCCT scanners. While mean and standard deviation derived from first-order features were mostly comparable between both detector types, higher-order features showed marked differences in mean and standard deviation, outlining the potential influence of the PCD through higher spatial resolution, better signal-to-noise ratio, and better detection of lower-energy photons

**Fig. 5** Split by phantom group heatmap of all radiomics parameters



**Fig. 6** Inter-phantom stable features heatmap (ICC > 0.75)





on texture analysis [29]. In comparison to our study, this study did not focus on test–retest variability and not on intra-scanner comparability of radiomics features.

Varghese et al evaluated the reproducibility and robustness of computed tomography–based texture analysis metrics in 2021 using CT images of a customized 3D-printed texture phantom consisting of six different texture patterns. The 3D phantom was scanned on four CT scanners of different manufacturers to assess reproducibility. For robustness assessment, varying CT imaging parameters were used (slice thickness, field of view, tube voltage, and tube current). In total, 23.2% of the features showed excellent robustness and reproducibility ( $ICC \geq 0.9$ ) [41]. Even though our study shows better reproducibility, we used only one scanner and one varying image parameter (tube current), so further investigations on PCCT must follow.

The fact that the benefits of the novel technology of PCCT have already found their way into the clinical routine is already shown by a proof-of-concept study demonstrating that the spectral information in PCCT data sets can be used to help to detect and quantify anemia on contrast-enhanced scans [42] as well as preliminary studies outlining potential imaging biomarkers for arteriosclerosis in periaortic adipose tissue [12] and diffuse myocardial damage through coronary artery calcifications [13].

Our study also has several limitations. As a single-center study, no comparison of features between two different PCCT scanners was possible. Additionally, no analysis regarding feature stability between different types of scanners was obtained. Both issues should be addressed in further studies in the future to show a possible more stable feature analysis on PCCT in comparison to conventional EID-CT.

In conclusion, this study outlines the high stability of a large proportion of radiomics features using PCCT in a phantom model. In the past, radiomics analysis was hampered by the lack of comparability, which may be overcome by PCCT technology and therefore foster the implementation of a more stable radiomics analysis in clinical routine.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00330-023-09460-z>.

**Funding** Open Access funding enabled and organized by Projekt DEAL. This project is part of the Photon-Counting Consortium Baden-Württemberg (PC3) funded by Gesundheitsstandort Baden-Württemberg.

## Declarations

**Guarantor** The scientific guarantor of this publication is Dr. Isabelle Ayx.

**Conflict of interest** The Department of Radiology and Nuclear medicine Medical Faculty Mannheim has a research agreement with Siemens Healthcare GmbH.

**Statistics and biometry** One of the authors has significant statistical expertise.

**Informed consent** Written informed consent was acquired as part of this study.

**Ethical approval** Institutional review board approval was obtained (ID 2021–659).

## Methodology

- retrospective
- experimental
- performed at one institution

**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/>.

## References

1. Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: images are more than pictures, they are data. *Radiology* 278:563–577. <https://doi.org/10.1148/radiol.2015151169>
2. van Griethuysen JJM, Fedorov A, Parmar C et al (2017) Computational radiomics system to decode the radiographic phenotype. *Cancer Res* 77:e104–e107. <https://doi.org/10.1158/0008-5472.CAN-17-0339>
3. Huang Y, Liu Z, He L et al (2016) Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (I or II) non-small cell lung cancer. *Radiology* 281:947–957. <https://doi.org/10.1148/radiol.2016152234>
4. Woźnicki P, Westhoff N, Huber T et al (2020) Multiparametric MRI for prostate cancer characterization: combined use of radiomics model with PI-RADS and clinical parameters. *Cancers* 12:1767. <https://doi.org/10.3390/cancers12071767>
5. Xie T, Wang X, Li M et al (2020) Pancreatic ductal adenocarcinoma: a radiomics nomogram outperforms clinical model and TNM staging for survival estimation after curative resection. *Eur Radiol* 30:2513–2524. <https://doi.org/10.1007/s00330-019-06600-2>
6. Nakanishi R, Akiyoshi T, Toda S et al (2020) Radiomics approach outperforms diameter criteria for predicting pathological lateral lymph node metastasis after neoadjuvant (chemo)radiotherapy in advanced low rectal cancer. *Ann Surg Oncol* 27:4273–4283. <https://doi.org/10.1245/s10434-020-08974-w>
7. Mühlberg A, Holch JW, Heinemann V et al (2021) The relevance of CT-based geometric and radiomics analysis of whole liver tumor burden to predict survival of patients with metastatic colorectal cancer. *Eur Radiol* 31:834–846. <https://doi.org/10.1007/s00330-020-07192-y>
8. Jiang Y, Chen C, Xie J et al (2018) Radiomics signature of computed tomography imaging for prediction of survival and chemotherapeutic benefits in gastric cancer. *EBioMedicine* 36:171–182. <https://doi.org/10.1016/j.ebiom.2018.09.007>
9. Liu Z, Wang S, Dong D et al (2019) The applications of radiomics in precision diagnosis and treatment of oncology: opportunities and challenges. *Theranostics* 9:1303–1322. <https://doi.org/10.7150/thno.30309>

10. Enke JS, Moltz JH, D'Anastasi M et al (2022) Radiomics features of the spleen as surrogates for CT-based lymphoma diagnosis and subtype differentiation. *Cancers* 14:713. <https://doi.org/10.3390/cancers14030713>
11. Schoepf UJ, Emrich T (2021) A brave new world: toward precision phenotyping and understanding of coronary artery disease using radiomics plaque analysis. *Radiology* 299:107–108. <https://doi.org/10.1148/radiol.2021204456>
12. Tharmaseelan H, Froelich MF, Nörenberg D et al (2022) Influence of local aortic calcification on periaortic adipose tissue radiomics texture features—a primary analysis on PCCT. *Int J Cardiovasc Imaging*. <https://doi.org/10.1007/s10554-022-02656-2>
13. Ayx I, Tharmaseelan H, Hertel A et al (2022) Myocardial radiomics texture features associated with increased coronary calcium score—first results of a photon-counting CT. *Diagnostics* 12:1663. <https://doi.org/10.3390/diagnostics12071663>
14. Martini K, Baessler B, Bogowicz M et al (2021) Applicability of radiomics in interstitial lung disease associated with systemic sclerosis: proof of concept. *Eur Radiol* 31:1987–1998. <https://doi.org/10.1007/s00330-020-07293-8>
15. Stefano A, Gioè M, Russo G et al (2020) Performance of radiomics features in the quantification of idiopathic pulmonary fibrosis from HRCT. *Diagnostics* 10:306. <https://doi.org/10.3390/diagnostics10050306>
16. Xun Y, Chen M, Liang P, et al (2020) A novel clinical-radiomics model pre-operatively predicted the stone-free rate of flexible ureteroscopy strategy in kidney stone patients. *Front Med* 7:576925. <https://doi.org/10.3389/fmed.2020.576925>
17. Li L, Wang L, Zeng F et al (2021) Development and multicenter validation of a CT-based radiomics signature for predicting severe COVID-19 pneumonia. *Eur Radiol* 31:7901–7912. <https://doi.org/10.1007/s00330-021-07727-x>
18. Liu H, Ren H, Wu Z et al (2021) CT radiomics facilitates more accurate diagnosis of COVID-19 pneumonia: compared with CO-RADS. *J Transl Med* 19:29. <https://doi.org/10.1186/s12967-020-02692-3>
19. Vaidya P, Bera K, Gupta A et al (2020) CT derived radiomic score for predicting the added benefit of adjuvant chemotherapy following surgery in stage I, II resectable non-small cell lung cancer: a retrospective multicohort study for outcome prediction. *The Lancet Digital Health* 2:e116–e128. [https://doi.org/10.1016/S2589-7500\(20\)30002-9](https://doi.org/10.1016/S2589-7500(20)30002-9)
20. Kim H, Park CM, Park SJ et al (2016) Temporal changes of texture features extracted from pulmonary nodules on dynamic contrast-enhanced chest computed tomography: how influential is the scan delay? *Invest Radiol* 51:569–574. <https://doi.org/10.1097/RLL.0000000000000267>
21. Mayerhoefer ME, Materka A, Langs G et al (2020) Introduction to radiomics. *J Nucl Med* 61:488–495. <https://doi.org/10.2967/jnumed.118.222893>
22. Liu R, Elhalawani H, Radwan Mohamed AS et al (2020) Stability analysis of CT radiomic features with respect to segmentation variation in oropharyngeal cancer. *Clin Transl Radiat Oncol* 21:11–18. <https://doi.org/10.1016/j.ctro.2019.11.005>
23. Korte JC, Cardenas C, Hardcastle N et al (2021) Radiomics feature stability of open-source software evaluated on apparent diffusion coefficient maps in head and neck cancer. *Sci Rep* 11:17633. <https://doi.org/10.1038/s41598-021-96600-4>
24. Mayerhoefer ME, Szomolanyi P, Jirak D et al (2009) Effects of MRI acquisition parameter variations and protocol heterogeneity on the results of texture analysis and pattern discrimination: an application-oriented study: effects of MRI acquisition parameters on texture analysis. *Med Phys* 36:1236–1243. <https://doi.org/10.1118/1.3081408>
25. Jirák D, Dezortová M, Hájek M (2004) Phantoms for texture analysis of MR images Long-term and multi-center study. *Med Phys* 31:616–622. <https://doi.org/10.1118/1.1646231>
26. Waugh SA, Lerski RA, Bidaut L, Thompson AM (2011) The influence of field strength and different clinical breast MRI protocols on the outcome of texture analysis using foam phantoms: influence of different MRI protocols on texture analysis. *Med Phys* 38:5058–5066. <https://doi.org/10.1118/1.3622605>
27. Willemink MJ, Persson M, Pourmorteza A et al (2018) Photon-counting CT: technical principles and clinical prospects. *Radiology* 289:293–312. <https://doi.org/10.1148/radiol.2018172656>
28. Rajendran K, Petersilka M, Henning A et al (2022) First clinical photon-counting detector CT system: technical evaluation. *Radiology* 303:130–138. <https://doi.org/10.1148/radiol.212579>
29. Ayx I, Tharmaseelan H, Hertel A et al (2022) Comparison study of myocardial radiomics feature properties on energy-integrating and photon-counting detector CT. *Diagnostics* 12:1294. <https://doi.org/10.3390/diagnostics12051294>
30. Dunning CA, Rajendran K, Fletcher JG et al (2022) Impact of improved spatial resolution on radiomic features using photon-counting-detector CT. In: Išgum I, Colliot O (eds) *Medical imaging 2022: image processing*. SPIE, San Diego, United States, p 70
31. Nolden M, Zelzer S, Seitel A et al (2013) The Medical Imaging Interaction Toolkit: challenges and advances: 10 years of open-source development. *Int J Comput Assist Radiol Surg* 8:607–620. <https://doi.org/10.1007/s11548-013-0840-8>
32. R Studio Team (2020) R Studio, version 1.3.1093, Boston, MA, USA, <http://www.rstudio.com>
33. McBride GB A proposal for strength-of-agreement criteria for Lin's concordance correlation coefficient, NIWA Client Report: HAM2005–062, May 2005
34. Hamre C, Botolfson P, Tangen GG, Helbostad JL (2017) Inter-rater and test-retest reliability and validity of the Norwegian version of the BESTest and mini-BESTest in people with increased risk of falling. *BMC Geriatr* 17:92. <https://doi.org/10.1186/s12877-017-0480-x>
35. Koo TK, Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 15:155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
36. Kursa MB, Rudnicki WR (2010) Feature selection with the Boruta package. *J Stat Soft* 36:. <https://doi.org/10.18637/jss.v036.i11>
37. Zhao B, Tan Y, Tsai WY et al (2014) Exploring variability in CT characterization of tumors: a preliminary phantom study. *Translational Oncology* 7:88–93. <https://doi.org/10.1593/tlo.13865>
38. Jensen LJ, Kim D, Elgeti T et al (2021) Stability of radiomic features across different region of interest sizes—a CT and MR phantom study. *Tomography* 7:238–252. <https://doi.org/10.3390/tomography7020022>
39. Mackin D, Fave X, Zhang L et al (2015) Measuring computed tomography scanner variability of radiomics features. *Invest Radiol* 50:757–765. <https://doi.org/10.1097/RLL.0000000000000180>
40. Peng X, Yang S, Zhou L et al (2022) Repeatability and reproducibility of computed tomography radiomics for pulmonary nodules: a multicenter phantom study. *Invest Radiol* 57:242–253. <https://doi.org/10.1097/RLL.0000000000000834>
41. Varghese BA, Hwang D, Cen SY et al (2021) Identification of robust and reproducible CT-texture metrics using a customized 3D-printed texture phantom. *J Appl Clin Med Phys* 22:98–107. <https://doi.org/10.1002/acm2.13162>
42. Decker JA, Huber A, Senel F, et al (2022) Anemia detection by hemoglobin quantification on contrast-enhanced photon-counting CT data sets. *Radiology* 220063. <https://doi.org/10.1148/radiol.220063>

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