Physica Medica 57 (2019) 41–46

Contents lists available at ScienceDirect

Physica Medica

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

Original paper

Ct radiomic features of pancreatic neuroendocrine neoplasms (panNEN) are robust against delineation uncertainty

Martina Mori[a], Giulia Benedetti[b], Stefano Partelli[c,d], Carla Sini[a], Valentina Andreasi[c], Sara Broggi[a], Maurizio Barbera[b], Giovanni M. Cattaneo[a], Francesca Muffatti[b,d], Marta Panzeri[b], Massimo Falconi[c,d], Claudio Fiorino[a,*], Francesco De Cobelli[b,d]

[a] Medical Physics, San Raffaele Scientific Institute, Milano, Italy
[b] Radiology, San Raffaele Scientific Institute, Milano, Italy
[c] Pancreatic Surgery Unit, Pancreas Translational & Clinical Research Center, San Raffaele Scientific Institute, Milano, Italy
[d] Vita-Salute University, Milano, Italy

ARTICLE INFO

Keywords:
Radiomics
CT imaging
Pancreatic tumors
Neuroendocrine tumors
Inter-observer variability

ABSTRACT

Purpose: The aim of this study was to quantify the impact of CT delineation uncertainty of pancreatic neuroendocrine neoplasms (panNEN) on Radiomic features (RF).
Methods: Thirty-one previously operated patients were considered. Three expert radiologists contoured panNEN lesions on pre-surgical high-resolution contrast-enhanced CT images and contours were transferred onto pre-contrast CT. Volume agreement was quantified by the DICE index. After images resampling and re-binning, 69 RF were extracted and the impact of inter-observer variability was assessed by Intra-Class Correlation (ICC): ICC > 0.80 was considered as a threshold for "very high" inter-observer agreement.
Results: The median volume was 1.3 cc (range: 0.2–110 cc); a satisfactory inter-observer volume agreement was found (mean DICE = 0.78). Only 4 RF showed ICC < 0.80 (0.48–0.73), including asphericity and three RFs (of five) of the neighborhood intensity difference matrix (NID).
Conclusions: The impact of inter-observer variability in delineating panNEN on RF was minimum, with the exception of the NID family and asphericity, showing a moderate agreement. These results support the feasibility of studies aiming to assess CT radiomic biomarkers for panNEN.

1. Introduction

Pancreatic neuroendocrine neoplasm (panNEN) is a rare malignancy with relatively indolent biologic behavior compared to pancreatic adenocarcinoma [1]. Its incidence sharply increased in recent years, possibly due to a higher detection rate with the proliferation of modern imaging and endoscopic technologies [2]. Radiomics is an emerging field of investigation, with great potential in identifying associations between image-based features and biological or clinical endpoints, thus supporting the way for personalized cancer therapies [3]. Predictive models incorporating radiomic features (RF) are expected to characterize neoplasia and correlate with overall survival [4,5,7–11], gene expression patterns [4,7,9], pathologic findings [12,13] and cancer stage [6], with the promise of improving our predictive ability compared to conventional clinical risk factors [14–17].

CT imaging is the most widely used modality in Radiomics due to its ability to assess tissue density, shape, texture and size before, during and after therapy [5–7]. On the other hand, the "Radiomics workflow" (i.e.: acquisition of high quality images, segmentation of the tumor and, finally, feature extraction from the defined tumor region), is subject to uncertainties which may likely influence the RF values. In this sense the rush to construct RF-based models seems to have left many fundamental questions unexplored. The selection of RFs to be used in predictive models requires that RFs should be both reliable when acquired under the same imaging settings and robust in terms of reproducibility.

Sources of loss of variability are imaging scanners, image acquisition/reconstruction parameters, and contour delineation.

Although as yet poorly investigated, the existence of an impact of inter-scanner variability has been reported and few studies have investigated the dependence of RF on different acquisition protocols and different reconstruction parameters settings [18–20]. Standardized acquisition and reconstruction protocols are needed to smooth out any input data variability, especially in the case of multicenter studies, which are expected to create the most generalizable models [21].

* Corresponding author at: Medical Physics, San Raffele Scientific Institute, Via Olgettina 60, 20132 Milano, Italy.
E-mail address: fiorino.claudio@hsr.it (C. Fiorino).

https://doi.org/10.1016/j.ejmp.2018.12.005
Received 13 September 2018; Received in revised form 23 November 2018; Accepted 8 December 2018
Available online 18 December 2018

Among the various sources of uncertainties, volume definition is likely to represent one of the most critical, notwithstanding the fact that manual delineation by an expert radiologist, although subject to inter-observer variability, is still considered to be the "gold standard" [22–24].

The promise of Radiomics in the specific context of PanNEN is highly relevant due to several reasons: first, most tumors have an indolent behavior but no reliable diagnostic tools are available in precisely classifying patients that may benefit from avoiding surgery [1,2] so that Radiomics may be a good candidate to close this gap. Second, most patients show small volumes at diagnosis, which is challenging from the point of view of extracting reliable RF. Finally, PanNEN are generally well visible at contrast-enhanced CT (CECT), which is a positive pre-requisite for Radiomics for the assessment of biomarkers to improve the diagnostic and predictive power of CT images.

To the best of our knowledge, the impact of inter-observer delineation variability on the reliability of CT RF for PanNEN patients, including HU values-, shape-, and texture-based features, has not yet been assessed. Therefore, the aim of this study was to quantify inter-observer variability in delineating PanNEN on CT images and its impact on RF.

2. Methods and materials

2.1. Patient data, acquisition

The group included 31 patients previously operated at our Institute in the period 2009–2016, randomly chosen from an Institutional database. All patients underwent a CT-scan with contrast injection before surgery. A non-ionic contrast agent (Iopromide; Ultravist 370, Bayer Schering Pharma, Berlin, Germany; 120 mL) was injected intravenously followed by a 20 mL saline flush at a rate of 3 mL/s. Phases acquired for all CT-scans were: unenhanced phase; arterial (an ROI is traced and the scan starts when a determined HU level is reached after contrast injection), covering an area from the supraclavicular fossa to the iliac crest; portal venous (70 s), covering an area from the dome of the diaphragm to the pubis; and late phase of the upper abdomen. Two different CT scanners were used for image acquisition: a 64-channels MDCT scanner (Philips Brilliance 64, Philips Medical System, Best, The Netherlands, n = 25/31) and a Toshiba Aquilion 16 Slices CT scanner (n = 6/31). The MDCT scanning parameters were: 64 detector rows, 64 × 0.62 mm beam collimation, 0.983 pitch, 100–120 kVp, 138–534 mAs depending on patient body weight, 2–3 mm slice thickness, 1.0 mm gap. For the Aquilion scanner, the parameters were: 16 detector rows, 4 × 4 mm beam collimation, 0.938 pitch, 100–120 kVp, 138–534 mAs depending on patient body weight, 2–3 mm slice thickness, 1.0–1.5 mm gap. All reconstructed CT images had a matrix size of 512 × 512 pixels with a pixel size ranging between 0.598 and 0.888 mm. The slice thickness was 2 or 3 mm and the voxel size ranged from 0.819 to 2.206 mm^3.

2.2. Delineation

Three blinded radiologists were recruited to contour volumes of all patients. Contours were delineated in arterial or venous phase, depending on where the tumor was most visible. For one patient, calcifications were present inside the lesion and were excluded from the delineated volume. A rigid registration based on a box ROI surrounding the pancreas and including the nearest structures was performed between the pre-contrast and the arterial and/or venous phase CT: a mutual information algorithm was applied followed by fine manual tuning, by visually overlaying images with and without contrast. The contours were copied and pasted on the co-registered pre-contrast CT and were possibly adjusted on the pre-contrast CT by the same radiologists to correct for minor local anatomical discrepancies due to respiration and organ motion between different phases of acquisition (Fig. 1). The contouring and rigid registration of images were performed using the Eclipse Treatment Planning System (Varian Medical Systems, Inc.). The choice of focusing on RFs extracted from the pre-contrast CT, was justified by the expected inter-patient variability of contrast enhancement and intra-tumor density differences that may impact on RF calculation.

2.3. Image processing and standardization

To guarantee feature extraction under the same analysis conditions, CT images were resampled at the same slice spacing of 3 mm, and same pixel spacing of 0.783 mm, chosen by virtue of being the median values of the original CT set. This point is of paramount importance, as many RF definitions are based on the sum over the entire number of voxels in the volume of interest, which depends on the voxel size. Images were then re-binned into 64 bins, which was previously suggested as a good value to limit noise [25]. Given the volumes of the considered lesions, a minimum number of voxels well above 10^2 was available for each lesion, with the very large majority of them ranging between 10^3 and 10^5, suitable for RF calculation.

2.4. Features extraction

The features extraction was performed using the open access CGITA software (Chang-Gung Image Texture Analysis, [26] v.1.4) implemented in the MatLab environment (MathWorks, Boston, USA); CGITA may extract many RFs, as summarized in Table 1.

Among the 47 HU statistics (first order features), we chose to ignore those found to be highly inter-correlated based on previous results [27], considered only: HU maximum, minimum, mean, variance, standard deviation, skewness, kurtosis, asphericity, tumor volume and entropy. Higher order features (N = 58), grouped into 9 "parent" matrices (as shown in Table 1), were calculated. We therefore considered only 69 RFs, including both first and higher order features.

2.5. Analyses

Inter-observer contouring variability was first assessed by the DICE-index. The DICE coefficient quantifies the volume agreement between two observers: for each observers, we considered the mean DICE value. The significance of the differences between pairs of observers was assessed by the Wilcoxon test for both DICE and volume values. Volume differences between pairs of observers were also tested by the Mann-Whitney test.

Finally, RF robustness was investigated against contour variability by calculating Spearman R (R_S) and intra-class correlation (ICC) values [28,29]. All statistical computations were performed by using MedCalc Statistical Software [30]. The R_S of each single feature was calculated by considering its values from each pair of observers, for all contours, resulting in three R_S for each RF. An $R_S = 1$ stands for a perfect agreement of RF between different contours. The mean value of R_S was considered as a good indicator of the robustness of RF values against contouring uncertainty.

ICC was calculated for each RF: ICC is a metric based on the analysis of variance (ANOVA), ranging between 0 and 1 (which indicates, respectively, null and perfect reproducibility between observers). The ICC of a single feature, was calculated from feature values obtained from all contours and all patients, using a one-way random single-measure model. In accordance with other similar studies [27,29], ICC > 0.80 was considered as an indicator of "very good" inter-observer agreement between.

3. Results

The main patients' characteristics are reported in Table 2. The cohort was characterized by a heterogeneous lesion volume distribution,

Fig. 1. The tumor is delineated on contrast-enhanced CT (top) and then projected onto the pre-contrast CT (bottom): if necessary, the contours are manually adjusted to correct small anatomical discrepancies.

with a mean value of 14.7 cc and prevalence of small volumes (< 2 cc) resulting in a median volume of 1.3 cc. Satisfactory agreement (mean DICE = 0.78; SD 0.09) was found for inter-observer variability in delineating contours with no significant differences between pairs of observers.

In Fig. 2 the central axial/sagittal/coronal slices referred to the patients with the lowest and the largest inter-observer variability are shown: in general, the contrast between tumor and surrounding tissues varied among the patients and was likely to be the most relevant factor influencing DICE agreement, together with volume.

Table 1
Radiomic features extracted by CGITA software. First order features are summarized under the parent matrix "Voxel statistics". Parent matrices of second and third order are: Cooc (Cooccurrence), VA (voxel-alignment), NID (neighborhood intensity difference,) ISZ (intensity size zone), nCOC (normalized Cooccurence matrix), TS (texture spectrum), TFC (texture feature coding), TFCC (texture feature coding co-occurrence) and NGLD (neighborhood grey level dependence).

Parent Matrix	Feature Name
Cooccurrence Matrix	Second angular moment, Contrast, Entropy, Homogeneity, Correlation, Dissimilarity, Inverse difference moment
Voxel-alignment Matrix	Short run emphasis, Long run emphasis, Intensity variability, Run-length variability, Run percentage, Low-intensity run emphasis, High-intensity run emphasis, Low-intensity short-run emphasis, High-intensity short-run emphasis, Low-intensity long-run emphasis, High-intensity long-run emphasis
Neighborhood intensity-difference Matrix	Coarseness, Contrast, Busyness, Complexity, Strength
Intensity-size-zone Matrix	Short-zone emphasis, Large-zone emphasis, Intensity variability, Size-zone variability, Zone percentage, Low-intensity zone emphasis, High-intensity zone emphasis, Low-intensity short-zone emphasis, High-intensity short-zone emphasis, Low-intensity large-zone emphasis, High-intensity large-zone emphasis
Normalized Cooccurance Matrix	Second angular moment, Contrast, Entropy, Homogeneity, Dissimilarity, Inverse difference moment
Voxel statistics	Minimum HU, Maximum HU, Mean HU, HU Variance, HU SD, HU Skewness, HU Kurtosis, HU bias-corrected Skewness, HU bias-corrected Kurtosis, Tumor volume, Entropy
Texture Spectrum	Max spectrum, Black-white symmetry
Texture Feature Coding	Coarseness, Homogeneity, Mean convergence
Texture Feature Coding Cooccurance Matrix	Variance, Second angular moment, Contrast, Entropy, Homogeneity, Intensity, Inverse difference moment, Correlation, Variance, Code Similarity
Neighboring Gray Level Dependence	Small number emphasis, Large number emphasis, Number nonuniformity, Second moment, Entropy

Table 2
Main patient's characteristics.

Age (median/range)	Median 62 (30–86)
Sex	M: 21 (68%) F: 10 (32%)
Grading	
G1	21 (68%)
G2	8 (26%)
G3	2 (6%)
T code	
T1	17 (55%)
T2	4 (13%)
T3	9 (29%)
T4	1 (3%)
N	
N0	24 (77%)
N1	7 (23%)
M	
M0	26 (84%)
M1	5 (16%)
Ki67 (median/range) %	6 (0.5–65)
CT max lateral diameter (median/range) mm	16 (6–60)
CT max AP diameter (median/range) mm	16 (5–85)
Necrosis	
No	26 (84%)
Yes	5 (16%)
Calcification	
No	No: 30 (97%)
Yes	Yes: 1 (3%)
Arterial ipervascularization	
No	No: 10 (32%)
Yes	Yes: 21 (68%)

No significant differences were found between the volume values delineated by the different observers: the detailed results are shown in Table 3.

Regarding the RF robustness, the summary of ICC and R_S results are presented in Fig. 3. Specifically, 63/69 RFs were found to be very robust with an ICC $>$ 0.9, while only 4 RFs showed an ICC $<$ 0.80, including Asphericity (ICC = 0.65) and three RFs of the neighborhood intensity difference (NID) matrix. The Spearman test confirmed the good agreement with an average R_S values of 0.88; the feature with the lowest R_S value was Asphericity (0.55).

4. Discussion

Quantitative imaging is likely to become one of the fundamental building blocks of clinical oncology, thanks to recent developments in computational imaging, data mining and predictive analysis, which allow the prediction of patient outcome and tumor characteristics from image-derived data. Radiomics represents a recent computational approach for image quantification that could have a decisive role in this setting.

In the particular context of panNEN, Radiomics could help in the assessment of staging and prognosis, reflecting histopathological characteristics, and thus indicating the appropriate choice of treatment.

According to the World Health Organization (WHO), PanNENs are categorized into four grades based on the Ki-67 index, a proliferation index obtained from histology and related to different prognoses [31].

While panNEN detection largely improved owing to advances in imaging techniques [32], none of the available imaging-based diagnostic procedures was demonstrated to be highly specific in the assessment of the grade, Ki-67 expression and other histopathological characteristics, resulting in a limited value in suggesting the most appropriate therapeutic approach.

Surgical resection, when feasible, is usually the only option recommended for these neoplasms [33]. Nevertheless, watchful waiting may be considered for low-grade tumors, given the high rates of

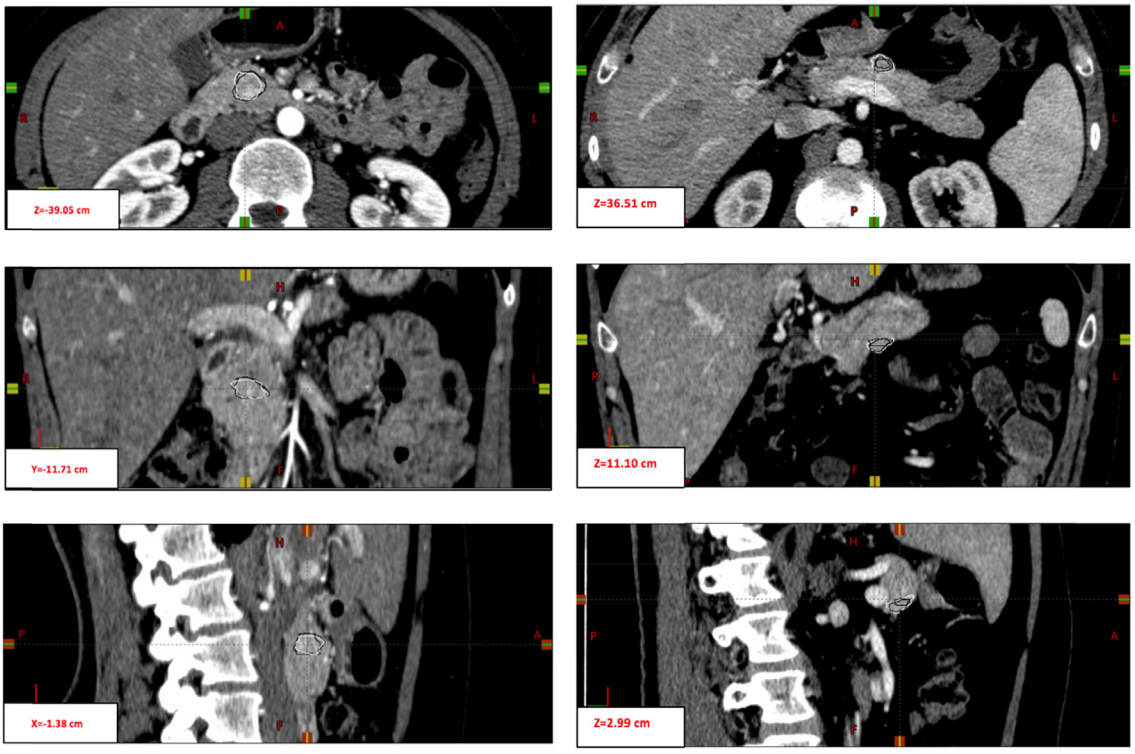

Fig. 2. The "best" (a) and "worst" (b) patients in terms of inter-observer agreement are shown depicting how the consistency is depending on the inter-patient variability of the contrast between tumor borders and normal tissue.

Table 3
The values of the tumor volumes of all considered patients are shown for each observer, including mean, median and SD. The differences between observers are relatively small and not statistically significant (Mann-Whitney test).

patient	Obs 1*	Obs 2*	Obs 3*	Mean
1	1,8	1,3	1,3	1,5
2	1,8	1,6	1,9	1,7
3	1,6	1,2	1,6	1,5
4	26,0	25,9	24,2	25,3
5	2,4	1,5	1,5	1,8
6	0,2	0,2	0,3	0,3
7	1,4	1,4	1,2	1,3
8	0,9	0,4	1,0	0,8
9	45,7	41,8	36,1	41,2
10	1,1	0,9	1,1	1,0
11	0,7	0,7	1,2	0,9
12	0,3	0,5	0,9	0,6
13	0,5	0,4	0,4	0,4
14	1,6	1,3	1,8	1,5
15	1,3	1,0	1,3	1,2
16	52,2	69,1	63,0	61,4
17	1,8	3,5	3,0	2,8
18	16,6	13,0	11,7	13,8
19	66,0	76,7	78,9	73,8
20	24,2	32,0	30,2	28,8
21	1,2	1,3	1,5	1,3
22	0,3	0,1	0,1	0,1
23	1,0	0,7	0,6	0,8
24	0,5	0,5	0,3	0,4
25	113,9	100,7	95,2	103,3
26	0,7	0,8	0,6	0,7
27	0,3	0,3	0,4	0,3
28	8,9	6,0	6,1	7,0
29	57,6	70,3	65,0	64,3
30	0,5	0,6	0,7	0,6
31	0,2	0,1	0,2	0,2
Mean	14,0	14,7	14,0	14,2
Median	1,4	1,3	1,3	1,3
SD	25,3	26,3	25,0	25,4

* Obs 1 vs Obs 2: p = 0.69; Obs 1 vs Obs 3: p = 0.90; Obs 2 vs Obs 3: p = 0.75.

morbidity and mortality following pancreatectomy [34–36] and the availability of diagnostic tools able to identify low risk patients would greatly help in better selecting patients to be submitted to watchful waiting, avoiding unnecessary surgical interventions.

Recent studies have attempted to explore CT texture analysis to predict panNEN characteristics and their outcome after surgery [15,16]: in both studies, post-contrast CT phases were used to assess predictors, possibly introducing additional uncertainty to the picture.

Accurate and efficient tumor segmentation is one of the main challenges for the extraction of "robust" RFs and manual segmentation suffers from inter-observer variability, negatively impacting on RF reliability. Furthermore, the choice to extract RFs from CECT images is debatable, due to the variable parameters involved in the administration of the contrast agent and in the physiological process of uptake. The relatively poor results reported by Canellas et al (AUC around 0.65 for the best performing RF [15]) could also reflect this uncertainty. Conversely, we used CECT images for image delineation and then performed texture analysis on pre-contrast CTs after an accurate rigid registration. Starting from a set of pre-contrast CTs processed in the same way, namely with the same pixel size, slice thickness and grey level discretization, we investigated the robustness of RFs among three observers. Results show a quite satisfactory inter-observer variability in delineating panNEN with little impact on almost all RFs considered. Intriguingly, 'Asphericity', which was found to be a predictor of tumor grade in one of the abovementioned studies [16], turned out to be the worst factor in term of robustness, although with a moderate inter-observer agreement. It may be worth mentioning that, given the quite irregular shape of tumors and the prevalence of small volumes, this results is not unexpected reflecting the higher robustness of features that are not related to shape. Interestingly, this point could benefit from the use of automatic segmentation which, however, would need to be specifically optimized and validated in current situation.

5. Conclusion

The results of current exploratory investigation are highly promising, despite the prevalence of small volumes (< 2 cc) in our cohort of patients; nevertheless, we verified that small volumes can be accurately delineated thanks to the high contrast-uptake in the CECT, with more than acceptable inter-observer contouring variability.

In conclusion, our findings suggest the possibility to robustly extract RF in large retrospective panNEN population, with the objective of assessing CT-based radiomic biomarkers: pooling multi-center data will be of paramount importance in collecting sufficiently large groups of patients to robustly assess RFs to be used as biomarkers before surgery, with the aim of assisting the surgeons in the reliable individualization of treatment choice. More investigation on the impact of inter-scanner and inter-institution variability is warranted.

Acknowledgment

The study was supported by an AIRC grant (IG18965).

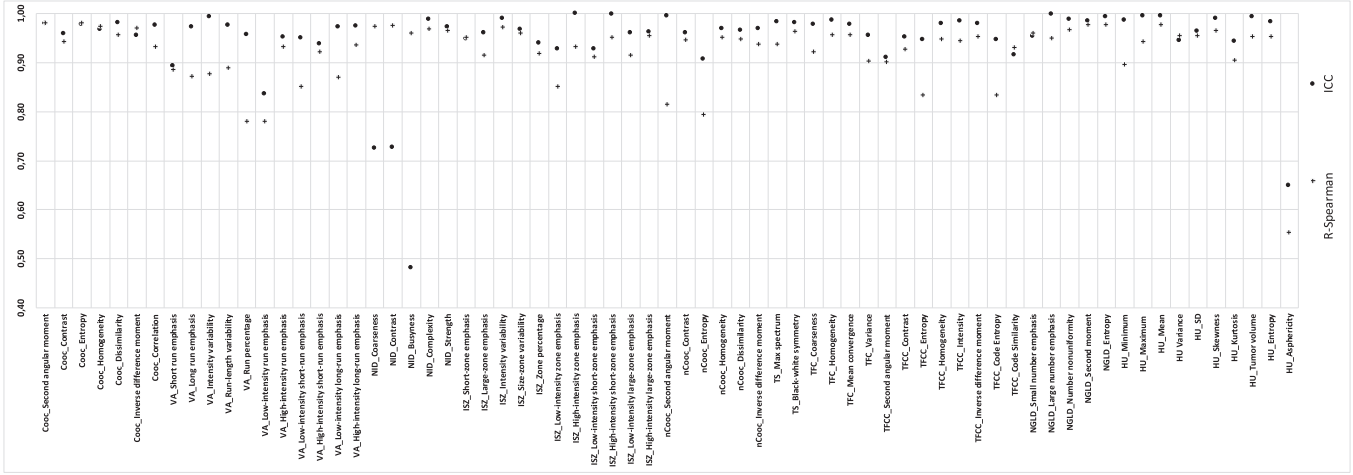

Fig. 3. The average values of the Spearman correlation coefficient (R_s) and of the intra-class-correlation (ICC) coefficients of inter-observer variability of all considered features are plotted. Results show the very high inter-observer consistency for almost all features.

References

[1] Panzuto F, Boninsegna L, Fazio N, et al. Metastatic and locally advanced pancreatic endocrine carcinomas: analysis of factors associated with disease progression. J Clin Oncol 2011;29:2372–7.

[2] Yao JC, Hassan M, Phan A, et al. One hundred years after "carcinoid": Epidemiology of and prognostic factors for neuroendocrine tumors in 35,825 cases in the United States. J Clin Oncol 2008;26:3063–72.

[3] Lambin P, Roelofs E, Reymen B, et al. Rapid learning health care in oncology—an approach towards decision support systems enabling customised radiotherapy. Radiother Oncol 2013;109:159–64.

[4] Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by non-invasive imaging using a quantitative radiomics approach. Nat Commun 2014;5:4006.

[5] Balagurunathan Y, Gu Y, Wang H, et al. Reproducibility and prognosis of quantitative features extracted from CT images. Transl Oncol 2014;7:72–87.

[6] Ganeshan B, Panayiotou E, Burnand K, et al. Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: a potential marker of survival. Eur Radiol 2012;22:796–802.

[7] Weiss GJ, Ganeshan B, Miles KA, et al. Noninvasive image texture analysis differentiates K-ras mutation from pan-wildtype NSCLC and is prognostic. PLoS One 2014;9:e100244.

[8] Win T, Miles KA, Janes SM, et al. Tumor heterogeneity and permeability as measured on the CT component of PET/CT predict survival in patients with non-small cell lung cancer. Clin Cancer Res 2013;19:3591–9.

[9] Gevaert O, Xu J, Hoang CD, et al. Non-small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data–methods and preliminary results. Radiology 2012;264:387–96.

[10] McNitt-Gray MF, Hart EM, Wyckoff N, Sayre JW, Goldin JG, Aberle DR. A pattern classification approach to characterizing solitary pulmonary nodules imaged on high resolution CT: preliminary results. Med Phys 1999;26:880–8.

[11] Huang Y, Liu Z, He L, Chen X, Pan D, Ma Z, et al. Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (I or II) non-small cell lung cancer. Radiology 2016;152234.

[12] Basu S, Hall LO, Goldgof DB, et al. editors. Developing a classifier model for lung tumors in CT-scan images. Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference On. Anchorage:IEEE;2011:1306–12.

[13] 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.

[14] Fried DV, Tucker SL, Zhou S, et al. Prognostic value and reproducibility of pretreatment CT texture features in stage III non-small cell lung cancer. Int J Radiat Oncol Biol Phys 2014;90:834–42.

[15] Canellas R, Burk KS, Parakh A, Sahani DV. Prediction of pancreatic neuroendocrine tumor grade based on CT features and texture analysis. AJR 2018;210:341–6.

[16] Choi TW, Kim JH, Yu MH, Park SJ, Han JK. Pancreatic neuroendocrine tumor: prediction of the tumor grade using CT findings and computerized texture analysis. Acta Radiol 2018;59(4):383–92.

[17] Altazi BA, Fernandez DC, Zhang GG, Hawkins S, Naqvi SM, Kim Y, et al. Investigating multi-radiomic models for enhancing prediction power of cervical cancer treatment outcomes. Phys Med 2018;46:180–8.

[18] 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.

[19] Shafiq-ul-Hassan M, Zhang GG, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. Med Phys 2017;44(3):1050–62.

[20] Fave X, Zhang L, Yang J, Mackin D, Balter P, Gomez D, et al. Impact of image preprocessing on the volume dependence and prognostic potential of radiomics features in non-small cell lung cancer. Transl Cancer Res 2016;5(4):349–63.

[21] Avanzo M, Stancanello J, El-Naqa I. Beyond imaging: the promise of radiomics. Phys Med 2017;38:122–39.

[22] 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.

[23] Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: the process and the challenges. Magn Reson Imaging 2012;30:1234–48.

[24] Doumou G, Siddique M, Tsoumpas C, et al. The precision of textural analysis in 18 F-FDG-PET scans of oesophageal cancer. Eur Radiol 2015;25:2805–12.

[25] Desseroit MC, Tixier F, Weber WA, Siegel BA, Le Rest CC, Visvikis D, et al. 18F-Reliability of PET/CT shape and heterogeneity features in functional and morphological components of Non-Small Cell Lung Cancer tumors: a repeatability analysis in a prospective multi-center cohort. J Nucl Med 2017;58(3):406–11.

[26] Fang YHD, Lin CY, Shih MJ, et al. Development and evaluation of an open-source software package "CGITA" for quantifying tumor heterogeneity with molecular images. Biomed Res Int 2014;5:248505–9.

[27] Belli ML, Mori M, Broggi S, Cattaneo GM, Bettinardi V, Dell'Oca I, et al. Quantifying the robustness of [18F]FDG-PET/CT radiomic features with respect to tumor delineation in head and neck and pancreatic cancer patients. Phys Med 2018;49:105–11.

[28] Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. Psychol Bull 1979;86:420–8.

[29] Leijenaar RT, Carvalho S, Velazquez ER, Van Elmpt WJ, Parmar C, Hoekstra OS, et al. Stability of FDG-PET Radiomics features: An integrated analysis of test- retest and inter-observer variability. Acta Oncol 2013;52:1391–7.

[30] https://www.medcalc.org/.

[31] Kim JY, Sm Hong, Ro JY. Recent updates on grading and classification of neuroendocrine tumors. Ann Diagn Pathol 2017;29:11–6.

[32] Yang G, Ji M, Chen J, et al. Surgery management for sporadic small ($\leq$2 cm), nonfunctioning pancreatic neuroendocrine tumors: a consensus statement by the Chinese Study Group for Neuroendocrine Tumors (CSNET). Int J Oncol 2017;50:567–74.

[33] Kazanjian KK, Reber HA, Hines OJ. Resection of pancreatic neuroendocrine tumors: results of 70 cases. Arch Surg 2006;141:765–9.

[34] Smith JK, Ng SC, Hill JS, et al. Complications after pancreatectomy for neuroendocrine tumors: a national study. J Surg Res 2010;163:63–8.

[35] Falconi M, Eriksson B, Kaltsas G, Bartsch DK, Capdevila J, Caplin M, et al. Vienna consensus conference participants. ENETS consensus guidelines update for the management of patients with functional pancreatic neuroendocrine tumors and non-functional pancreatic neuroendocrine tumors. Neuroendocrinology 2016;103(2):153–71.

[36] Falconi M, Bartsch DK, Eriksson B, Klöppel G, Lopes JM, O'Connor JM, et al. Vullierme MP, O'Toole D; Barcelona Consensus Conference participants. ENETS Consensus Guidelines for the management of patients with digestive neuroendocrine neoplasms of the digestive system: well-differentiated pancreatic non-functioning tumors. Neuroendocrinology 2012;95(2):120–34.