LIMITED IMPACT OF DISCRETIZATION/INTERPOLATION PARAMETERS IN PREDICTING OUTCOME OF PANCREATIC ADENOCARCINOMA BASED ON CT RADIOMICS

Sara Loi[1], Martina Mori[1], Diego Palumbo[2], Stefano Crippa[3,4], Gabriele Palazzo[1], Emiliano Spezi[5], Antonella Del Vecchio[1], Massimo Falconi[3,4], Francesco De Cobelli[2,4], Claudio Fiorino[1]

[1] Medical Physics, San Raffaele Scientific institute, Milano, Italy

[2] Radiology, San Raffaele Scientific institute, Milano, Italy

[3] Pancreatic Surgery Unit, Pancreas Translational and Clinical Research Center, San Raffaele Scientific institute, Milano, Italy

[1] Università Vita-Salute, Milano, Italy

[1] School of Engineering, Cardiff University, Cardiff, Wales, United Kingdom

Corresponding author:

C. Fiorino, Medical Physics, San Raffaele institute,

Via Olgettina 690, 20132 Milano; e-mail: fiorino.claudio@hsr.it

---

[1] Medical Physics, San Raffaele Scientific institute, Milano, Italy

[2] Radiology, San Raffaele Scientific institute, Milano, Italy

[3] Pancreatic Surgery Unit, Pancreas Translational and Clinical Research Center, San Raffaele Scientific institute, Milano, Italy

[4] Università Vita-Salute, Milano, Italy

[5] School of Engineering, Cardiff University, Cardiff, Wales, United Kingdom

Abstract

Purpose: To explore the variation of the discriminative power of CT (Computed Tomography) radiomic features (RF) against image discretization/interpolation in predicting early distant relapses (EDR) after upfront surgery.

Materials and methods: Data of 144 patients with pre-surgical high contrast CT were processed consistently with IBSI (Image Biomarker Standardization Initiative) guidelines. Image interpolation/discretization parameters were intentionally changed, including cubic voxel size (0.21-27 mm^3) and binning (32-128 grey levels) in a 15 parameter's sets. After excluding RF with poor inter-observer delineation agreement (ICC<0.80), the variation of 88 RF against discretization/interpolation was first quantified. Then, their ability in classifying patients with early (<10 months) distant relapses (EDR) was investigated in terms of AUC (Area Under Curve) variation for those RF significantly associated to EDR.

Results: Despite RF variability against discretization/interpolation parameters was large and only 36/88 RF showed %COV<20%, AUC changes were limited and the analysis on 32 RF significantly associated with EDR (AUC values around 0.6-0.7), the mean values of SD of AUC variability and AUC range were 0.02 and 0.05 respectively. AUC ranges were between 0.00 and 0.11, with values ≤0.05 in 16/32 RF. These variations were further reduced when excluding the extreme values of 32 and 128 for grey levels (Average AUC range 0.04, with values between 0.00 and 0.09).

Conclusions: The discriminative power of CT RF in the prediction of early distant relapses after upfront surgery for pancreatic cancer is relatively invariant against image interpolation/discretization within a large range of voxel sizes and binning.

Keywords

Radiomics, adenocarcinoma, pancreatic tumors, CT imaging

1. Introduction

The therapeutic approach and the prognosis of pancreatic adenocarcinoma are mainly based on the resectability status, that should be determined by a multidisciplinary team and that includes the discussion of findings of contrast enhanced CT scan, according to the 2019 NCCN (National Comprehensive Cancer Network) guidelines [1].

Nevertheless, approximately 40% of patients undergoing upfront surgery experience distant disease recurrence within 12 months [2], resulting in poor outcome [2-3]. This suggests that upfront surgery is not the best treatment approach for many of those patients with tumor classified as resectable, which could instead benefit from neoadjuvant chemotherapy [4-5]. It is indeed important to find clinical, pathological and biological features helping in identifying patients who would not benefit from upfront surgery, even when a resectable tumour is identified. Poor differentiation, high CA19.9 values and long standing symptoms are reported to identify patients expected to have a poor outcome after primary surgery [2,5-9]. However, around 10% [10] of patients does not express CA 19.9, and its value may be affected by concurrent common occurrences in this population (e.g. jaundice and/or cholangitis). Moreover, other pathological findings, such as perineural and/or micro vascular infiltration, presence of lymphadenopathies and resection margins status, can also predict disease-free and overall survival [11], but they can be assessed only after resection.

The great potential of radiomics in providing a non-invasive, deep insight into tumor microenvironment, currently under investigation in many different settings, could help in defining imaging biomarkers for the currently considered clinical scenario. Radiomics was seldomly investigated in pancreatic cancer to predict tumor grade [12] and for selecting appropriate treatment strategies [13], such as to identify those patients upfront resectable but at high risk for early relapse after surgery [14], who could instead benefit from neoadjuvant chemotherapy, and for predicting individual patients radiotherapy response [15-16].

Any predictive radiomic-based model requires extensive validation to be robustly translated into the clinic; clinical application is still very limited also due to lack of repeatability and reproducibility of radiomic features as well as to the difficult clinical interpretation of predictive radiomic signatures [16-25].

Several papers investigated the robustness of features calculated from patient CT images against delineation uncertainty [17,19-20], reconstruction [21] and interpolation [22-23] algorithms, voxel sizes and bin widths [22-23], filtering and slice thickness [24] and organ motion [25]. Moreover, few works recently explored harmonization methods to counteract some of these problems, such as inter-scanner variability [26-27]. Surprisingly, very few studies focused on the impact of RF

variability on the predictive/discriminative power in clinical models [28-29]: in fact, the informative content could in principle be robustly independent on the image processing phase, and this could be beneficial for the generalizability of radiomic-based diagnostic/prognostic/prediction models. Our group previously reported promising results [30] showing a limited impact of different image discretization/interpolation procedures on the discriminative power of CT RF in assessing clinical and histology characteristics of a relatively small cohort of pancreatic neuroendocrine neoplasms (panNEN). Aim of current study was to quantify the impact of image discretization/interpolation on the ability of RF extracted from CT images of pancreatic adenocarcinoma patients treated by upfront surgery in classifying patients experiencing early distant metastases; the available large group of patients was previously analyzed with a fixed set of image processing parameters [14], according to IBSI guidelines.

2. Materials and Methods

2.1 Patient data, acquisition

The study was performed on a set of CT images of patients from a single-center retrospective study conducted at San Raffaele Scientific Institute (Milan, Italy). Within a prospectively acquired database, all consecutive patients with pancreatic adenocarcinoma who underwent upfront pancreaticoduodenectomy (PD) between January 2015 and December 2019 were identified; patients who were evaluated with at least one multiphase, contrast-enhanced CT scan within 30 days before index surgery were enrolled into our study. Patients who died within 90 days after index surgery were excluded from further analysis.

The study was performed on the preoperative CT images of 144 patients, without contrast medium injection. CTs were acquired on 64-row multidetector CT scanners (scanner 1: SOMATOM Definition Flash Dual Source CT, Siemens Healthcare; scanner 2: BRILLIANCE, Philips medical system). Scanning parameters were as follows: detector collimation: 64 × 0.62 mm or 128 × 0.6 mm, rotation time: 0.5–0.6, tube voltage: 120 kVp with automated tube current modulation, tube current: 230–300 mAs, slice thickness: 2–3 mm, gap: 1 mm.

2.2 Delineation

The Gross Target Volumes (GTVs) were delineated by the same radiologist; the robustness of CT radiomic features against interobserver contouring variability was preliminarily assessed on a subgroup of 29 patients by the same three readers and found to be very limited, suggesting that the results should be generalizable to other observers

[14]. Contouring was performed using the MIM Software (v. 6.8.2).

2.3 Image processing and radiomic features extraction

SPAARC Pipeline for Automated Analysis and Radiomics Computing complying with IBSI [31] was used to process images for RF extraction.

All images were resampled at cubic voxels of the same size with a bilinear interpolation. This procedure was implemented to reduce directional bias when voxel sizes were not already isotropic, according to the specific recommendation of IBSI. 5 voxel size were considered (0.6, 0.8, 1.0, 2.0 and 3.0 mm), covering and reasonably extending the range of voxel sizes encountered in current population (being 0.6 mm the minimum and 1.0 mm the maximum pixel size and being 1.0 mm the minimum and 5 mm the maximum slice thickness of the population).

Regarding grey level discretization, binning values equal to 32, 64 and 128 were considered, being 64 bins the value commonly suggested as an optimal compromise to preserve the quantitative information while limiting the impact of noise [17,32]. In total, the combination of different voxel size and binning resulted into 15 different conditions. As an example, a comparison between a patient image resampled at 0.6 and at 3 mm is shown in Fig. 1. Resampled and rebinned DICOM files were imported to MATLAB using the Computational Environment for Radiological Research (CERR) for RF calculation. Only RF showing good robustness (ICC > 0.8) against inter-observer contouring variability were considered resulting in 88 out of 182 RF of first and higher order, belonging to the following families: Morphology, Statistical, Intensity Histogram, Grey Level Co-occurrence Matrix 3D_average (GLCM3D_avg), Grey Level Co-occurrence Matrix 3D_combined (GLCM3D_comb), Grey Level Run Length 3D_average (GLRL3D_avg), Grey Level Run Length 3D_combined (GLRL3D_comb), Grey Level Size Zone Matrix 3D (GLSZM3D), Neighbour Grey Tone Difference Matrix 3D (NGTDM3D), Grey Level Distance Zone Matrix 3D (GLDZM3D). RF were calculated for the 144 patients, for each of the 15 parameter's set.

2.4 Features variability against interpolation/discretization

First, the impact of voxel size and binning variability on the RF was evaluated in order to assess the RF robustness. The impact of image discretization was estimated in terms of the parameter %COV=100*STDEV/MEAN, where the standard deviation (STDEV) and the average (MEAN) were calculated for each of the 144

patients over the 15 resampling and re-binning conditions.

2.5 Discrimination against interpolation/discretization

For the whole set of 144 patients, the robustness of the discriminative power of CT features was evaluated through the AUC derived from the Receiver Operator Characteristic (ROC) curves. Following the study [14], the early recurrence of distant metastases was considered as end point in the ROC analysis, considering an early recurrence when it appeared before the quartile (10 months) of the recurrence time set for the 144 patients. First, the Mann-Whitney test of the null hypothesis was applied to each RF, testing if RF values were different between patients with the considered event or without the event (considering $p<0.05$ as a threshold for significance): the same methodology was repeated in each of the 15 sets. RF with p-value <0.05 in all 15 conditions were considered and ROC curves and AUC values were obtained; the variation of the discriminative value within the 15 conditions was evaluated and RF were ranked in terms of AUC variability (%COV). The Bonferroni condition for the test of the null hypothesis was intentionally not applied in this study, since it is applicable in the clinical model after the final selection of radiomic features for each end point. For repetitive tests, an automatic procedure was set up through a Matlab (MathWorks, Boston, USA) script, starting from the Matlab ROC function. Similarly, a Matlab script was set up for repetitive Mann-Whitney testing.

3. Results

3.1 RF variability

The ranking of the RF robustness against the variability due to the resampling and re-binning is shown in Fig. 2, showing, as expected, a relevant impact. Only 36 RF showed in fact an absolute value of the mean (over the patients) $\%COV < 20$, while 44 RF showed poor reproducibility ($\%COV > 50$), and 8 intermediate reproducibility ($20 < \%COV < 50$).

3.2 AUC variability

The events for the end point "early recurrence" (before 10 months) of distant metastases were 36. The impact of RF variability in the 15 cases is shown in Table 1 and in Fig. 3. The analysis on 32 RF significantly associated with EDR shows typical AUC values for single RF around 0.6-0.7. In general, the variation of AUC for the large majority of RF was relatively low: the average AUC standard deviation and range over all RF were 0.02 and 0.05 respectively (SD: 0.00-0.04, range: 0.00-0.11) with mean and maximum AUC %COV equal to 0.1% and 5.9% respectively.

Smaller variability of AUC range was found for first order RF (0.00-0.07) against RF of higher order (0.03-0.11). These variations were further reduced when excluding the extreme values of 32 and 128

for grey levels, with the maximum AUC range improved from 0.11 to 0.09 and the maximum standard deviation reduced from 0.04 to 0.03), as shown in Fig. 4.

4. Discussion

In the first part of this study, the results on the RF variations against image interpolation and discretization (voxel size and grey levels) showed a relatively poor robustness for most RF, confirming the expected results, due to their mathematical definition. Similar results have been reported elsewhere [17-24,33], although frequently on phantoms and, to our knowledge, never for patients affected by pancreatic adenocarcinoma.

The second part of the work showed that, despite the large variations of RF values, the discriminative power of radiomic parameters for early recurrence of distant metastases prediction depends on image interpolation/discretization in a quite limited way. And this was demonstrated while extending the range of cubic sizes to much larger values (up to 27 mm^3) than the ones referred to the original population (within 5 mm^3).

To our knowledge, the impact of image interpolation/discretization on the discriminative power of RF has rarely been investigated in clinical series and never for pancreatic cancer patients. Lan He et al [28] studied the diagnostic performance of radiomics signature for discriminating benign and malignant solitary pulmonary nodule (SPN). They investigated the impact of the reconstruction slice thickness, the convolution kernel and the contrast enhancement, in order to maximize the diagnostic performance. Larue et al [29] studied the robustness of radiomic features in non-small cell lung cancer (NSCLC) and found no significant correlation between feature stability and prognostic value. Qiu et al [34] studied the discrimination of hepatocellular carcinoma (HCC) based on the RF extracted from CT scans. They reported different AUC values, reproducibility and redundancy of the RF for different tumor segmentation methods.

In our study, the reproducibility of radiomic features against image discretization and its impact on the discriminative power, in terms of AUC stability, were investigated and a limited AUC variability was found. These variations were further reduced when considering a more realistic range of variability (64 bins for the gray levels), with the maximum AUC range improved from 0.11 to 0.09 and the maximum standard deviation reduced from 0.04 to 0.03.

Despite the expected variability of RF values when changing the interpolation/discretization conditions, the discriminative power seems to be relatively invariant within a reasonably large range of parameters. A quite consistent result was found by Lv et al [35], investigating the robustness of radiomic features in nasopharyngeal PET/CT images against the feature definition and the impact

on disease differentiation in terms of AUC. In agreement with our results, they concluded that robustness of radiomics features should not be overemphasized for removal of features towards assessment of clinical tasks, since the reproducibility of the features is not always correlated to AUC stability. In a study on the renal clear cell carcinomas, Kocak et al [36] showed that, despite yielding fewer features with excellent reproducibility, use of the contour-focused segmentation method provided better classification performance (AUC) for distinguishing nuclear grade.

In a previous investigation from our group, Loi et al [30] reported promising results showing the limited impact of image interpolation/discretization on the discriminative power of CT RF in predicting clinical parameters, for a relatively small set of 39 patients affected by panNEN. Current results are consistent with those preliminary results, showing limited impact of image interpolation and discretization on the RF discriminative power of early recurrence of distant metastases after pancreatic duodenectomy in patients affected by pancreatic adenocarcinoma. Results are even more important since they confirm, for a different disease, for IBSI compliant RF definition, and for a statistically more robustly large set of patients, the important result that not only the RF robustness should be investigated, but also its impact on the resulting predictive models.

5. Conclusions

Similar results in assessing the discriminative power of CT (IBSI compliant) radiomics predictors for pancreatic adenocarcinoma may be expected when applying different image discretization procedures: keeping the voxel size equal of less than 3x3x3 mm^3 and fixing binning to 64, seems to guarantee limited AUC variations when changing voxel size in a quite large range. Moreover, results are not dramatically different even when extending the range of binning to include 32 and 128 grey levels. Of course, this result cannot be extended to all clinical scenarios. However, it suggests that the informative content of CT RF seems to be relatively invariant with respect to interpolation/binning.

Acknowledgments

Martina Mori was supported by AIRC (Italian Association for Cancer Research, IG23150).

References

[1] NCCN (2019) NCCN Clinical Practice Guidelines in Oncology. Pancreatic adenocarcinoma; Version 1. Journal of the National Comprehensive Cancer Network: JNCCN 5(10):998-1033. https://doi.org/10.6004/jnccn.2007.0085

[2] Petrelli F, Inno A, Barni S et al (2017) Borderline resectable pancreatic cancer: More than an anatomical concept. Dig Liver Dis 49(2):223–226. https://doi.org/10.1016/j.dld.2016.11.010

[3] Matsumoto I, Murakami Y, Shinzeki M et al (2015) Proposed preoperative risk factors for early recurrence in patients with resectable pancreatic ductal adenocarcinoma after surgical resection: A multi-center retrospective study. Pancreatology 15(6):674–680. https://doi.org/10.1016/j.pan.2015.09.008

[4] Lee Y, Lee J, Yang S et al (2019) Neoadjuvant therapy versus upfront surgery in resectable pancreatic cancer according to intention-to-treat and per-protocol analysis: A systematic review and meta-analysis. Sci Rep 9:15662. https://doi.org/10.1038/s41598-019-52167-9

[5] Barugola G, Partelli S, Marcucci S et al (2009) Resectable pancreatic cancer: Who really benefits from resection? Ann Surg Oncol 16:3316–3322. https://doi.org/10.1245/s10434-009-0670-7

[6] Bergquist JR, Puig CA, Shubert CR et al (2016) Carbohydrate Antigen 19-9 Elevation in Anatomically Resectable, Early Stage Pancreatic Cancer Is Independently Associated with Decreased Overall Survival and an Indication for Neoadjuvant Therapy: A National Cancer Database Study. J Am Coll Surg 223:52–65. https://doi.org/10.1016/j.jamcollsurg.2016.02.009

[7] Nakamura T, Asano T, Okamura K et al (2019) A Preoperative Prognostic Scoring System to Predict Prognosis for Resectable Pancreatic Cancer: Who Will Benefit from Upfront Surgery? J Gastrointest Surg 23(5):990–996. https://doi.org/10.1007/s11605-018-3972-x

[8] Groot V, Rezaee N, Wu W et al (2018) Patterns, Timing, and Predictors of Recurrence Following Pancreatectomy for Pancreatic Ductal Adenocarcinoma. Ann Surg 267(5):936-945. https://doi.org/10.1097/SLA.0000000000002234

[9] Barugola G, Frulloni L, Salvia R et al (2009) Is CA 19-9 a screening marker? Dig Liver Dis 41:325–327. https://doi.org/10.1016/j.dld.2009.02.009

[10] Ballehaninna UK, Chamberlain RS (2012) The clinical utility of serum CA 19-9 in the diagnosis, prognosis and management of pancreatic adenocarcinoma: An evidence based appraisal. J Gastrointest Oncol 3:105–119. https://doi.org/10.3978/j.issn.2078-6891.2011.021

[11] Crippa S, Pergolini I, Javed A et al (2020) Implications of Perineural Invasion on Disease Recurrence and Survival After Pancreatectomy for Pancreatic Head Ductal Adenocarcinoma. Ann Surg 276(2):378-385. https://doi.org/10.1097/SLA.0000000000004464

[12] Karmazanovsky G, Gruzdev I, Tikhonova V et al (2021) Computed tomography-based radiomics approach in pancreatic tumors characterization. Radiologia Medica 126 (11):1388-1395. https://doi.org/10.1007/s11547-021-01405-0

[13] Marti-Bonmati L, Cerdá-Alberich L, Pérez-Girbés A et al (2022) Pancreatic cancer, radiomics and artificial intelligence. British Journal of Radiology 95(1137):20220072. https://doi.org/10.1259/bjr.20220072

[14] Palumbo D, Mori M, Prato F et al (2021) Prediction of Early Distant Recurrence in Upfront Resectable Pancreatic Adenocarcinoma: A Multidisciplinary, Machine Learning-Based Approach. Cancers 13(19):4938. https://doi.org/10.3390/cancers13194938

[15] Mori M, Passoni P, Incerti E et al (2020) Training and validation of a robust PET radiomic-based index to predict distant-relapse-free-survival after radio

chemotherapy for locally advanced pancreatic cancer. Radiother Oncol 153:258-264. https://doi.org/10.1016/j.radonc.2020.07.003

[16] Chen X, Oshima K, Schott D et al (2017) Assessment of treatment response during chemoradiation therapy for pancreatic cancer based on quantitative radiomic analysis of daily CTs: An exploratory study. PLoS ONE. https://doi.org/10.1371/journal.pone.0178961

[17] Desseroit M, Tixier F, Weber WA et al (2017) 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 58:406-411. https://doi.org/10.2967/jnumed.116.180919

[18] Zhao B (2021) Understanding Sources of Variation to Improve the Reproducibility of Radiomics. Frontiers in Oncology 11:633176. https://doi.org/10.3389/fonc.2021.633176

[19] Mori M, Cattaneo GM, Muffatti F et al (2019) CT radiomic features of pancreatic neuroendocrine neoplasms (panNEN) are robust against delineation uncertainty. Physica Medica 57:41-46. https://doi.org/10.1016/j.ejmp.2018.12.005

[20] Van Velden FH, Kramer GM, Frings V et al (2016) Repeatability of Radiomic Features in Non-Small-Cell Lung Cancer [18F]FDG-PET/ CT Studies: Impact of Reconstruction and Delineation. Mol Imaging Biol 18(5):788-95. https://doi.org/10.1007/s11307-016-0940-2

[21] Lu L, Ehmke RC, Schwartz LH et al (2016) Assessing Agreement between Radiomic Features Computed for Multiple CT Imaging Settings. PLOS ONE 11(12):e0166550. https://doi.org/10.1371/journal.pone.0166550

[22] Linsalata S, Borgheresi R, Marfisi D et al (2022) Radiomics of Patients with Locally Advanced Rectal Cancer: Effect of Preprocessing on Features Estimation from Computed Tomography Imaging Biomed Res Int 2022:2003286. https://doi.org/10.1155/2022/2003286

[23] Hosseini S, Shiri I, Hajianfar G et al (2021) The Impact of Preprocessing on the PET-CT Radiomics Features in Non-Small Cell Lung Cancer. Frontiers in Biomedical Technologies 8(4):261-272. https://doi.org/10.18502/fbt.v8i4.7754

[24] Hu P, Wang J, Zhong H et al (2016) Reproducibility with repeat CT in radiomics study for rectal cancer. Oncotarget 7(44):71440–71446. https://doi.org/10.18632/oncotarget.12199

[25] Oliver JA, Budzevich M, Zhang GG et al (2015) Variability of Image Features Computed from Conventional and Respiratory-Gated PET/CT Images of Lung Cancer. Translational Oncology 8(6):524-34. https://doi.org/10.1016/j.tranon.2015.11.013

[26] Sedlaczek O, Kleesiek J, Gallagher F et al (2022) Quantification and reduction of cross-vendor variation in multicenter DWI MR imaging: results of the Cancer Core Europe imaging task force European Radiology 32:8617–8628. https://doi.org/10.1007/s00330-022-08880-7

[27] Ibrahim A, Refaee T, Primakov S et al (2021) The effects of in-plane spatial resolution on ct-based radiomic features' stability with and without combat harmonization. Cancers 13(8):1848. https://doi.org/10.3390/cancers13081848

[28] He L, Huang Y, Ma Z et al (2016) Effects of contrast-enhancement, reconstruction slice thickness and

convolution kernel on the diagnostic performance of radiomics signature in solitary pulmonary nodule. Nature Scientific Reports 6:34921. https://doi.org/10.1038/srep34921

[29] Larue RTHM, Van De Voorde L, Van Timmeren JE et al (2017) 4DCT imaging to assess radiomics feature stability: An investigation for thoracic cancers. Radiotherapy and Oncology 125(1):147-153. https://doi.org/10.1016/j.radonc.2017.07.023

[30] Loi S, Mori M, Benedetti G et al (2020) Robustness of CT radiomic features against image discretization and interpolation in characterizing pancreatic neuroendocrine neoplasms. Physica Medica 76:125–133. https://doi.org/10.1016/j.ejmp.2020.06.025

[31] Zwanenburg A, Vallieres M, Abdalah M et al (2020) The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput imagebased phenotyping. Radiology 95(2):328–338. https://doi.org/10.1148/radiol.2020191145

[32] Tixier F, Le Rest CC, Hatt M et al (2011) Intratumor Heterogeneity Characterized by Textural Features on Baseline 18F-FDG PET Images Predicts Response to Concomitant Radiochemotherapy in Esophageal Cancer J Nucl Med 52(3):369-78. https://doi.org/10.2967/jnumed.110.082404

[33] Traverso A, Wee L, Dekker A et al (2018) Repeatability and reproducibility of radiomic features: a systematic review. Int J Radiat Oncol Biol Phys 102(4):1143–1158. https://doi.org/10.1016/j.ijrobp.2018.05.053

[34] Qiu Q, Duan J, Duan Z et al (2019) Reproducibility and non-redundancy of radiomic features extracted from arterial phase CT scans in hepatocellular carcinoma patients: impact of tumor segmentation variability Quantitative Imaging in Medicine and Surgery 9(3):453-464. https://doi.org/10.21037/qims.2019.03.02

[35] Lv W, Yuan Q, Wang Q et al (2018) Robustness versus disease differentiation when varying parameter settings in radiomics features: application to nasopharyngeal PET/CT European Radiology 28(8):3245-3254. https://doi.org/10.1007/s00330-018-5343-0

[36] Kocak B, Ates E, Durmaz E et al (2019) Influence of segmentation margin on machine learning–based high-dimensional quantitative CT texture analysis: a reproducibility study on renal clear cell carcinomas. European Radiology 29:4765–4775. https://doi.org/10.1007/s00330-019-6003-8

Figures

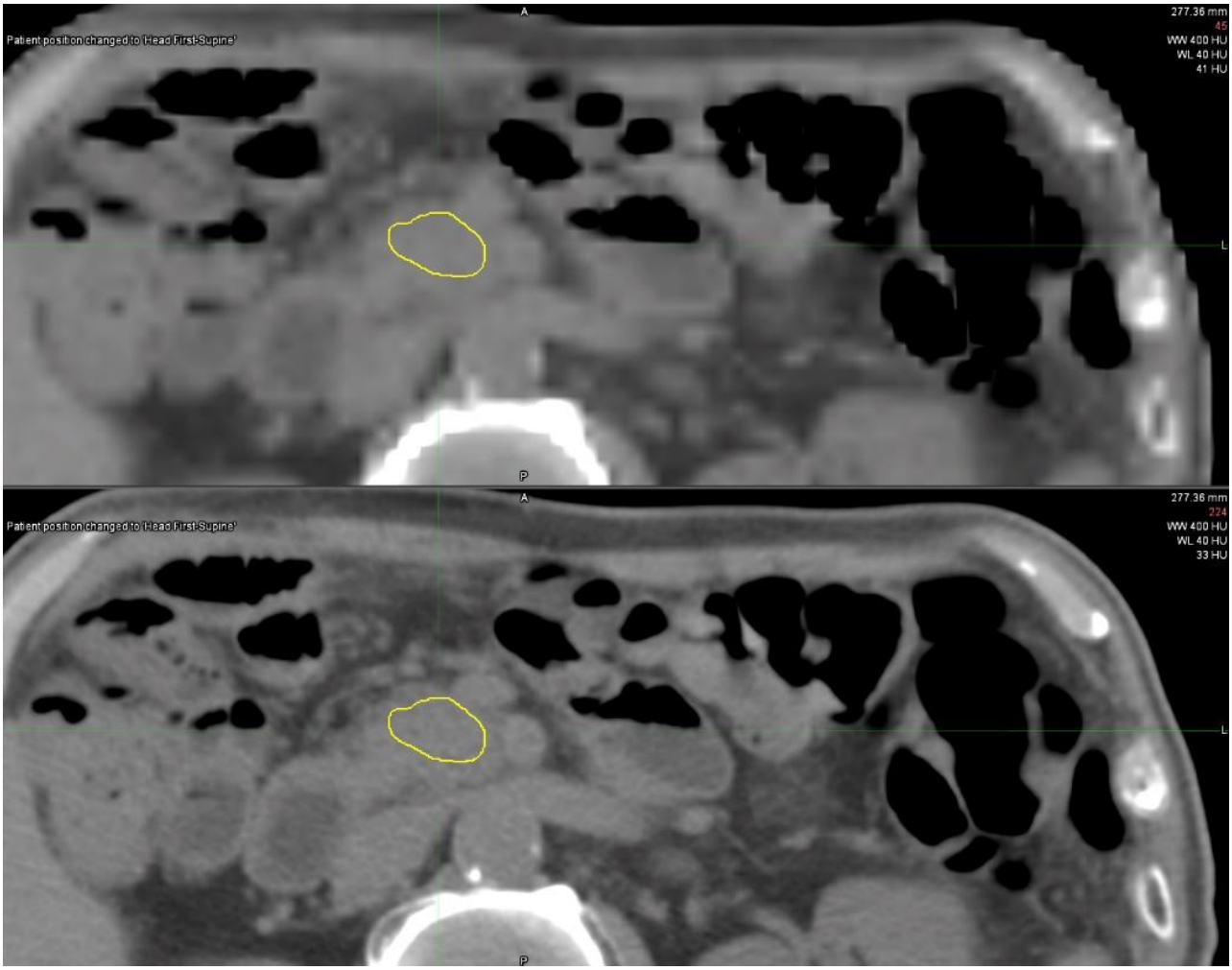

Fig. 1 Example of comparison of a patient image resampled at 0.6 mm (bottom) and at 3 mm (top) cubic voxel, with contoured GTV

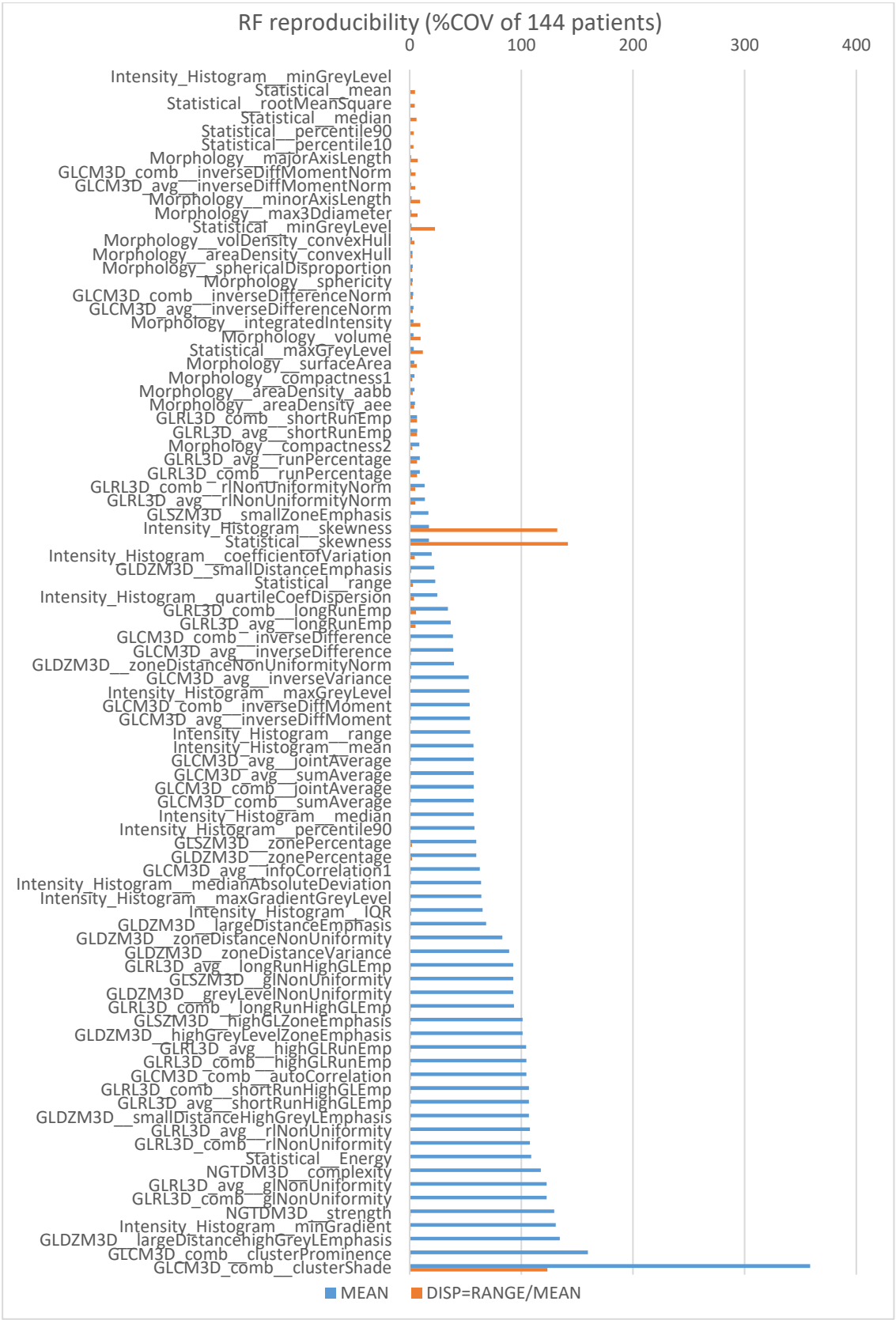

Fig. 2 Ranking of the considered radiomic features, according to their robustness to different resampling or re-binning conditions (%COV parameter)

		AUC robustness					
RF family	**RF**	**Min**	**Max**	**%COV**	**Mean**	**σ**	**Range**
Morphology	Volume	0.69	0.70	0.3%	0.69	0.00	0.01
	SurfaceArea	0.68	0.69	0.2%	0.68	0.00	0.00
	Max3Ddiameter	0.68	0.69	0.2%	0.69	0.00	0.00
	MajorAxisLength	0.68	0.68	0.1%	0.68	0.00	0.00
	MinorAxisLength	0.66	0.66	0.5%	0.66	0.00	0.01
	AreaDensity_aee	0.63	0.68	2.5%	0.65	0.02	0.05
	AreaDensity_convexHull	0.61	0.68	3.9%	0.66	0.03	0.06
	IntegratedIntensity	0.69	0.69	0.2%	0.69	0.00	0.00
Statistical	Energy	0.69	0.69	0.2%	0.69	0.00	0.00
Intensity_Histogram	MedianAbsoluteDeviation	0.63	0.69	3.5%	0.65	0.02	0.07
	CoefficientofVariation	0.62	0.69	3.4%	0.65	0.02	0.07
	MinGradient	0.66	0.73	3.2%	0.69	0.02	0.07
GLCM3D_avg	InverseDifference	0.61	0.69	4.4%	0.64	0.03	0.08
	InverseDifferenceNorm	0.62	0.68	4.0%	0.64	0.03	0.07
	InverseDiffMomentNorm	0.62	0.68	3.7%	0.64	0.02	0.06
GLCM3D_comb	InverseDifference	0.61	0.69	4.4%	0.64	0.03	0.08
	InverseDifferenceNorm	0.61	0.68	4.0%	0.64	0.03	0.07
	InverseDiffMomentNorm	0.62	0.68	3.7%	0.64	0.02	0.06
GLRL3D_avg	GlNonUniformity	0.68	0.73	1.8%	0.70	0.01	0.04
	RlNonUniformity	0.67	0.70	1.4%	0.69	0.01	0.03
GLRL3D_comb	LongRunEmp	0.61	0.72	5.9%	0.65	0.04	0.11
	GlNonUniformity	0.68	0.73	1.8%	0.70	0.01	0.04
	RlNonUniformity	0.67	0.70	1.4%	0.69	0.01	0.04
GLSZM3D	GlNonUniformity	0.62	0.70	3.5%	0.67	0.02	0.08
	ZonePercentage	0.61	0.72	5.8%	0.65	0.04	0.11
GLDZM3D	SmallDistanceEmphasis	0.63	0.68	2.8%	0.65	0.02	0.06
	LargeDistanceEmphasis	0.65	0.69	2.0%	0.68	0.01	0.04
	GreyLevelNonUniformity	0.62	0.70	3.5%	0.67	0.02	0.08
	ZoneDistanceNonUniformityNorm	0.63	0.68	2.8%	0.66	0.02	0.05
	ZonePercentage	0.61	0.72	5.8%	0.65	0.04	0.11
	ZoneDistanceVariance	0.66	0.69	1.5%	0.68	0.01	0.04
NGTDM3D	Strength	0.66	0.69	1.1%	0.68	0.01	0.03

Average	0.64	0.70	2.6%	0.67	0.02	0.05
Maximum	0.69	0.73	**5.9%**	0.70	**0.04**	**0.11**
Minimum	**0.61**	0.66	0.1%	**0.64**	0.00	0.00

Table 1: Impact of the RF variability on the discriminative power (AUC in 15 discretization conditions; end point EDR).

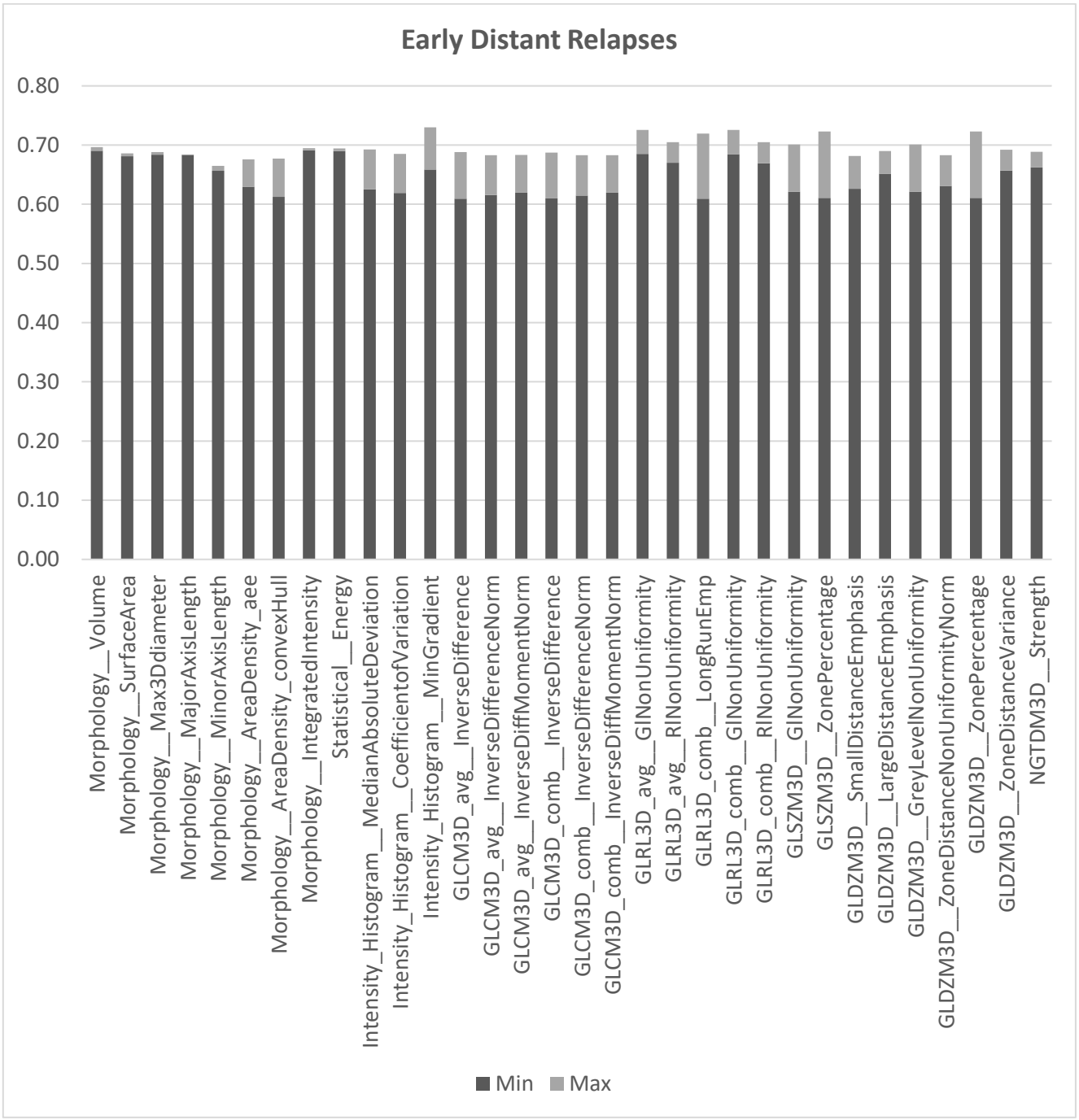

Fig. 3 Range of AUC in the 15 investigated conditions for the features significantly associated to EDR

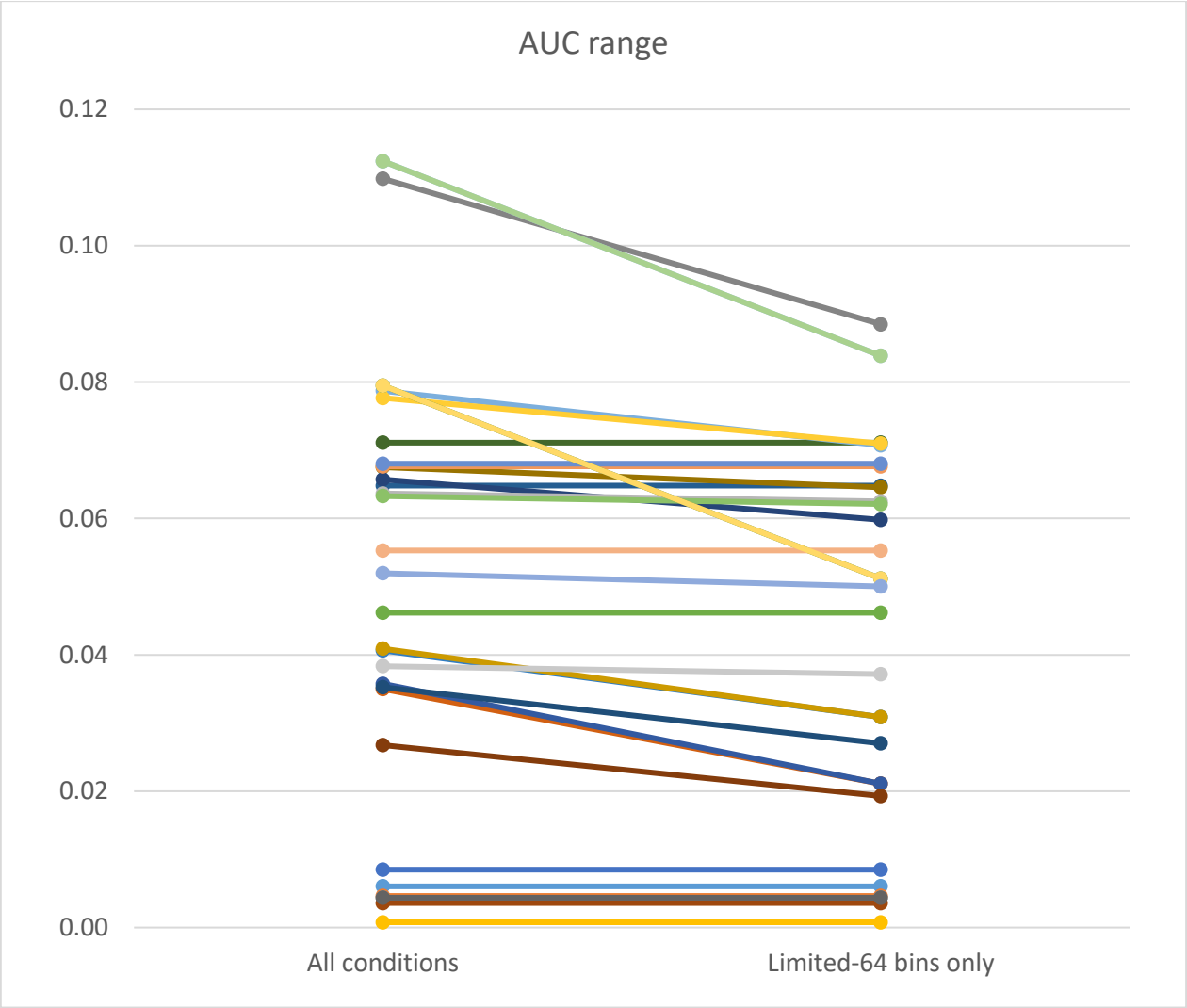

Fig. 4 AUC ranges over the whole range of variability for the RF significantly associated to EDR, compared to the limited range of variability restricting the gray levels to 64 bins