Clinical and Translational Radiation Oncology 33 (2022) 153–158

Contents lists available at ScienceDirect

Clinical and Translational Radiation Oncology

journal homepage: www.sciencedirect.com/journal/clinical-and-translational-radiation-oncology

Cleaning radiotherapy contours for radiomics studies, is it worth it? A head and neck cancer study

Pierre Fontaine [a,b], Vincent Andrearczyk [b], Valentin Oreiller [b,c], Daniel Abler [b,c], Joel Castelli [a], Oscar Acosta [a], Renaud De Crevoisier [a], Martin Vallières [d], Mario Jreige [c], John O. Prior [c], Adrien Depeursinge [b,c]

[a] Univ Rennes, CLCC Eugene Marquis, INSERM, LTSI - UMR 1099, F-35000 Rennes, France
[b] Institute of Information Systems, School of Management, HES-SO Valais-Wallis University of Applied Sciences and Arts Western Switzerland, Sierre, Switzerland
[c] Department of Nuclear Medicine and Molecular Imaging, Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne, Switzerland
[d] Department of Computer Science, Université de Sherbrooke, Sherbrooke, QC, Canada

ARTICLE INFO

Keywords:
Radiomics
Head and neck cancer
Survival analysis

ABSTRACT

A vast majority of studies in the radiomics field are based on contours originating from radiotherapy planning. This kind of delineation (*e.g.* Gross Tumor Volume, GTV) is often larger than the true tumoral volume, sometimes including parts of other organs (*e.g.* trachea in Head and Neck, H&N studies) and the impact of such over-segmentation was little investigated so far. In this paper, we propose to evaluate and compare the performance between models using two contour types: those from radiotherapy planning, and those specifically delineated for radiomics studies. For the latter, we modified the radiotherapy contours to fit the true tumoral volume. The two contour types were compared when predicting Progression-Free Survival (PFS) using Cox models based on radiomics features extracted from FluoroDeoxyGlucose-Positron Emission Tomography (FDG-PET) and CT images of 239 patients with oropharyngeal H&N cancer collected from five centers, the data from the 2020 HECKTOR challenge. Using *Dedicated* contours demonstrated better performance for predicting PFS, where Harell's concordance indices of 0.61 and 0.69 were achieved for *Radiotherapy* and *Dedicated* contours, respectively. Using automatically *Resegmented* contours based on a fixed intensity range was associated with a C-index of 0.63. These results illustrate the importance of using clean dedicated contours that are close to the true tumoral volume in radiomics studies, even when tumor contours are already available from radiotherapy treatment planning

Introduction

With the recent advances in computational science, the emergence of *precision medicine* is moving one step further to the clinical world. Radiomics allows quantitative analyses from radiological and nuclear medicine images with high throughput extraction to obtain prognostic patient information[1]. Unlike biopsies, radiomics does not require invasive sampling inside the tumor. It can provide an exhaustive and quantitative evaluation of lesion phenotype based on medical images that were acquired during diagnosis and treatment course. Established links between the radiomics features and outcomes of interest (*e.g.* staging, response to treatment) can be leveraged to assist clinical decisions prospectively. Radiomics features quantify the intensity, texture, and shape properties of provided Volumes of Interest (VOI)[2]. VOIs are necessary to focus the radiomics analysis on relevant biological structures, such as the tumoral volume. This contouring process, among others, is known to have a strong impact on the performance (*e.g.* precision, robustness) of the models[3]. Thus, the VOI must be as close as possible to the true tumoral volume if the latter is considered as the main source of information concerning the targeted outcomes.

Related work

Radiomics studies on Head and Neck cancer (H&N) are based on various kinds of delineations to obtain the VOIs, including the direct reutilization of those used for radiotherapy planning, (semi-) automatically generated (*e.g.* based on metabolic activity thresholding), or dedicated to the study using expert manual contours. Combinations of approaches are also used in some cases, such as manual contouring refined using automatic re-segmentation[2]. Unfortunately, the delineation approach is often not clearly reported in the literature. Table 1 lists the types of delineation methods used in several H&N radiomics studies. The direct reutilization of VOIs created in the context of radiotherapy planning was used in [4–8]. This allows performing

https://doi.org/10.1016/j.ctro.2022.01.003
Received 24 January 2022; Accepted 27 January 2022
Available online 1 February 2022

Table 1
VOI delineation methods used in H&N radiomics studies.

Authors	delineation purpose	delineation method	imaging modalities
(Castelli *et. al* 2019)[5]	radiotherapy	manual	PET/CT
(Leger *et. al* 2019) [9]	radiotherapy	manual + re-segmentation	CT
(Parmar *et. al* 2015)[16]	unknown	manual	CT
(Zhang *et. al* 2008)[17]	unknown	semi-auto	Sonograms
(Bogowicz *et. al* 2017a)[11]	radiotherapy	manual + re-segmentation	CT
(Leijenaar *et. al* 2018)[6]	radiotherapy	manual	CT
(Al Ajmi *et. al* 2018)[18]	unknown	manual	Dual-energy CT
(Wang *et. al* 2018) [19]	radiomics	manual	MRI
(Zhang *et. al* 2017)[20]	radiomics	manual	MRI
(Leijenaar *et. al* 2015)[7]	radiotherapy	manual	CT
(Bogowicz *et. al* 2017b)[12]	radiotherapy	manual (CT) + automatic (PET)	PET/CT
(Vallières *et. al* 2017)[8]	radiotherapy	manual	PET/CT
(Ouyang *et. al* 2017)[21]	radiotherapy	manual	MRI
(Van Dijk *et. al* 2018)[4]	radiotherapy	manual	MRI
(Wenbing *et. al* 2021)[10]	radiotherapy	manual	PET/CT

radiomics studies without the need for re-annotating the images specifically for these tasks. The contours made for radiotherapy are, however, very large as compared to the true tumoral volumes and frequently include non-tumoral tissues and parts of other organs (*e.g.* trachea, see Fig. 1).

A few recent studies used a re-segmentation step of the initial VOI, (*e.g.* Leger *et al.* 2019[9] and Wenbing *et al.* 2021[10]) to remove air and only keep soft tissue. Moreover, several studies including Bogowicz *et al.* 2017a[11] and Bogowicz *et al.* 2017b[12] performed a resegmentation step by manually removing slices that contain artifacts and excluding voxels outside the soft tissue window based on Hounsfield Units (HU). The performance evaluation of using automatically generated segmentation for building deep and traditional prognostic models was studied in [13–15]. Those two studies showed a comparison analysis between the use of manually and automatically generated VOIs. It was reported that fully automatic prognostic models achieved slightly better performance.

Beyond the specific domain of H&N radiomics, several studies investigated the stability of radiomics features with regard to VOI delineation. The tumor segmentation step is a critical stage of the radiomics workflow [22]. Information extracted from those delineations and is crucial to extract relevant biomarkers within the VOI while avoiding the inclusion of peripheral non-informative regions or other information than tumoral site [10]. Even more so, most of the features extracted from the VOI are aggregated into a scalar value via an integrative operation [23], with a risk of decreasing the prognostic power of features via the dilution of relevant localized patterns with other unrelated tissue.

In Depeursinge *et al.* 2015 [24], authors used artificial contour perturbations and observed that their model for predicting lung adenocarcinoma recurrence remained stable as long as VOI perturbations are under 4 mm. Other studies investigated the impact of inter-observer delineation on radiomics features [25,26]. Both studies, based on a single center dataset, demonstrated that most of the radiomics features are unstable under delineation variations. The results show that for different kinds of tumor (*e.g.,* H&N squamous cell carcinoma, non-small cell lung cancers, or malignant pleural mesothelioma) it is possible to find a subset of stable features. However, the prognosis power of this subset was not studied. Huang *et al.* 2017 [27] observed that both the number of stable features with high prognostic value and their predictive value differed across delineations from three radiologist observers. In this study, we evaluate and compare the Progression-Free Survival (PFS) prognosis performance between radiomics models based on two different VOIs types. We use *Radiotherapy* delineations which were used for treatment planning as well as *Dedicated* VOIs. The latter result from the manual re-segmentation of the initial *Radiotherapy* VOIs to fit the primary tumor as perfectly as possible when based on a fusion of FluoroDeoxyGlucose-Positron Emission Tomography (FDG-PET) and Computed Tomography (CT) images.

Material and methods

Patient data

The dataset used in this work includes the training and test sets of the HEad and NeCK TumOR segmentation in PET/CT images (HECKTOR) 2020 challenge [28], organized as a satellite event of the 23rd International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). The dataset was assembled from five centers and includes 239 cases[1]. It contains PET/CT images of patients with H&N cancer located in the oropharynx region. The clinical characteristics of the dataset are detailed in Table 2.

For each patient, a PET/CT image series and two primary Gross Tumor Volume (GTVt) contours are available. We refer to these two types of delineations as *Radiotherapy* and *Dedicated*. The former was made for radiotherapy planning by experts in radiotherapy. Details about these annotations can be found in [8,28]. The *Radiotherapy* contours are potentially not suitable for radiomics studies as they are often larger than the true tumoral volume, considering peripheral tissues and trachea. For this reason, these contours were re-delineated as close as possible to the true tumoral volume in the context of the HECKTOR 2020 challenge [28]. The re-delineation aims at contouring the entire edges of the morphological anomaly, visualized as a mass effect in the non-enhanced CT, for the corresponding hypermetabolic volume in the PET. The contouring excludes the hypermetabolic activity projecting outside the physical limits of the lesion, *e.g.,* lumen of the airway or bony structures with no morphologic evidence of local invasion.

Feature extraction

In this section, we describe the extraction of features from the PET/CT images prior to model building. We preprocessed both PET and CT images with *iso*-resampling of $2 \times 2 \times 2$ mm voxels using linear interpolation. This step is performed before feature extraction.

In order to compare the performance using either *Radiotherapy* or *Dedicated* contours in the context of survival analysis, we used a classical radiomics pipeline. Following the preprocessing step, we extracted features from both PET and CT image series based on either *Radiotherapy* or *Dedicated* VOIs using the PyRadiomics library [29]. In addition, we extracted features with a *Resegmented* VOI initially based on *Radiotherapy* VOI. The re-segmentation step was achieved by thresholding CT images between [−300,200] HU to only keep soft tissue. This re-segmentation step was used to investigate the importance of expert knowledge when contouring the true tumoral volume when compared to *e.g.*, simple air and high-density tissue removal. An example of this new segmentation is illustrated (in purple) in Fig. 1. Table 3 details the

[1] The HECKTOR data contains 254 cases, but for 13 of the test cases, the initial radiotherapy contours were not available. Two other patients were excluded because the follow-up was shorter than 3 months

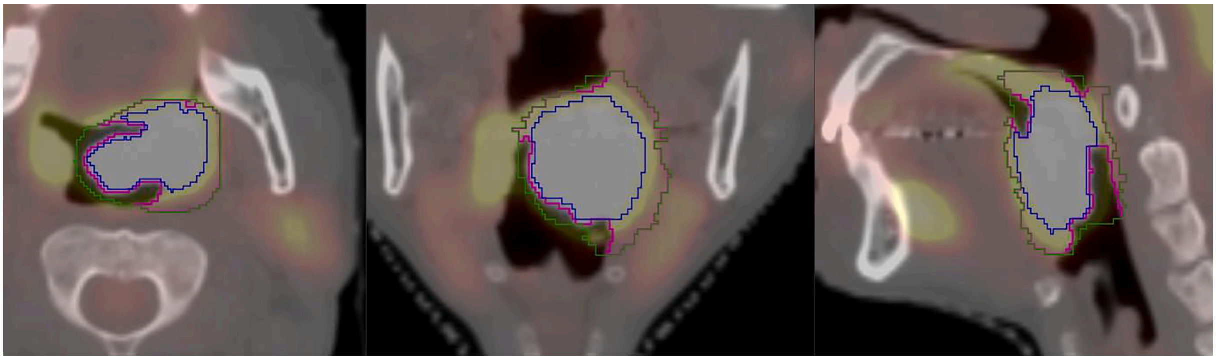


Fig. 1. Example of VOI delineation: *Radiotherapy* (green), *Resegmented* (purple), and *Dedicated* (blue) overlayed on a fused FDG-PET/CT image. The blue contour is closer to the true volume of the primary tumor. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Overview of the dataset. The centers include Hôpital Général Juif (HGJ), Montréal, CA; Centre Hospitalier Universitaire de Sherbooke (CHUS), Sherbrooke, CA; Hôpital Maisonneuve-Rosemont (HMR), Montréal, CA; Centre Hospitalier de l'Université de Montréal (CHUM), Montréal; Centre Hospitalier Universitaire Vaudois (CHUV), CH.

Center	patient	Gender	Age	(avg.)	T classification		N classification	Follow-up	(avg. days)	events
HGJ	55	Male	43	62	T1	12	N0	7	1339	11
			12		T2	18	N1	7		
		Female			T3	16	N2	39		
					T4	9	N3	2		
CHUS	71	Male	50	62	T1	6	N0	19	1246	13
			21		T2	36	N1	4		
		Female			T3	17	N2	45		
					T4	12	N3	3		
HMR	18	Male	14	69	T1	0	N0	1	1274	4
			4		T2	2	N1	0		
		Female			T3		N2	16		
					T4	8	N3	1		
CHUM	55	Male	41	64	T1	8	N0	4	1120	7
			14		T2	25	N1	8		
		Female			T3	17	N2	36		
					T4	5	N3	7		
CHUV	40	Male	35	63	T1	5	N0	10	705	7
			5		T2	14	N1	24		
		Female			T3	17	N2			
					T4	4	N3	3		

Table 3
List of the different combinations of parameters and features.

Image	Preprocessing	Binning	Features
CT	Iso-resampling 2x2x2mm Linear interpolation	FBN = 32 FBS = 50	GLCM (24) GLRLM (16) GLSZM (16) First Order (18) Shape (14)
PET	Iso-resampling 2x2x2mm Linear interpolation	FBN = 8 FBS = 1	GLCM (24) GLRLM (16) GLSZM (16) First Order (18)

features families and extraction parameters used in this study. A total of 130 features were extracted per modality with additional 14 shape features. For each patient and for each contour type, we, therefore, computed a total of 274 features[2]. From those two modalities per patient (CT and PET), we extracted features from the first-order (18 features) and second-order (56 features) families. Regarding the second-order, we extracted the 56 features using two different binning strategies based on Fixed Bin Number (FBN) and Fixed Bin Size (FBS) (as detailed in Table 3). Those 56 features were divided into three subfamilies, namely Grey Level Co-occurrence Matrix(GLCM), Grey Level Run Length Matrix (GLRLM), and Grey Level Size Zone Matrix (GLSZM). Finally, we computed 14 shape features.

Univariate analysis

To compare the two types of delineation, we first performed a univariable analysis to investigate the stability of radiomics features regarding the type of VOI used. This analysis is independent of the radiomics model workflow. We computed the two-way mixed single measure Intraclass Correlation Coefficient (ICC(3,1)) [30] for every single feature and for both modalities to assess their stability when extracted from either *Dedicated* or *Radiotherapy* VOIs. The ICC is a statistical indicator that gives information about the consistency of feature measurements. A value of zero indicates no reliability whereas a value of one means that the measurements are perfectly stable. This univariable analysis allows revealing which kind of feature is more affected by a change of VOI.

We also computed the univariable C-index value of each feature to quantify its association with the PFS outcome. We also further used the results of these univariable C-indexes to select features for the multivariable model.

[2] We can unconventionally detail the number of features as follows: 274 = 2 modalities × (2 binning × 56 s-order + 18 first-order)

Multivariable analysis

The pipeline of the multivariable radiomics analysis used to estimate the influence of using *Radiotherapy* or *Dedicated* contours on the PFS prediction performance is depicted in Fig. 2.

First (1), we pooled the image data from the five centers and randomly divided into a training/validation (80%) cohort and a testing (20%) cohort using a stratified shuffling method where the stratification criterion is the PFS outcome. This first split was repeated 100 times and we used the same splits of each repetition to statistically compare the results between the two contour types. Second (2), we computed the univariable C-index [31] of each feature based on the training dataset and (3) transformed this value (*i.e.* $|C_{index} - 0.5)|$) to keep both concordant and anti-concordant features. (4) We used the resulting C-index to rank the features based on concordance with the outcome and retained the top 20 concordant features. The number 20 was used to respect a ten to one ratio between the number of features and the number of patients. We then used a grid-search (5) method to determine the feature correlation threshold value: $t \in \{0.6, 0.65, 0.70, 0.75, 0.80\}$. We used a stratified 5-folds cross-validation method to divide the sub-dataset into a train (80%) and a validation (20%) dataset. This step avoids basing the models on highly correlated feature sets. Based on this feature set, we trained a Cox proportional hazards model [32] (from scikit-survival [33] V0.14.1 in Python) on the training set to predict the hazard score and further computed the C-index on the validation set, as the performance measure to estimate the performance of this survival analysis. After selecting the best performing model during grid-search, (6) we applied it to the test set, and (7) computed the test C-index value. The code used to compute this pipeline is available on GitHub (https://github.com/Pierre1d6/CleanedContours.git).

Results

Influence of VOI types on feature stability

We first compared the stability of the features across the *Radiotherapy* and *Dedicated* types of VOIs, grouping features based on their family and image modality. The significance of stability comparisons between feature families, imaging modalities, and VOI types is assessed using a Student *t*-test. The associated results are detailed in Fig. 3. We observe that features from PET images are more stable than those from CT images ($p < 0.001$, see Fig. 3a). When further looking at stability differences between feature families, we observe that shape features are the most stable across the five families with a median ICC around 0.7. Fig. 3b confirms the better stability of features regardless of their family when extracted from PET images. GLSZM features achieved the lowest stability (median ICC3 < 0.4) both in PET and CT images. These observations are further interpreted in Section 5.

Multivariable prognostic models

We applied the multivariable radiomics workflow described in Section 3.4 and report the results in Fig. 4.

Discussions and conclusion

In this work, we studied the impact of using *Dedicated* VOIs in the context of H&N radiomics studies in PET/CT that are specifically fitted to the GTVt volume, as compared to reusing VOIs directly from radiotherapy treatment planning.

We first investigated the stability of the features regarding their family type and imaging modality. Fig. 3a and 3c suggest that the features are overall more stable when computed on PET images. This can be explained by the difference in terms of value range between PET (≈[0,25] Standardized Uptake Value, SUV) and CT (≈[-1000, 1000] HU when including air from the trachea). Therefore, including peritumoral regions has a stronger impact on features extracted from the CT images, with air contained in the trachea around GTVt having much lower values in CT (-1000 HU) than in PET (0 SUV) when compared to voxel statistics inside GTVt. In addition, spatial deviations of the contours result in smaller differences in the PET because of the lower resolution when compared to CT.

Fig. 3c reports the stability of features per family and across modalities (PET or CT). In PET and for first-order features, a high median value and high variability are observed. When focusing on specific first-order features, we observed that the *maximum* was the most stable feature (ICC3 = 0.98) because there is no high SUV activation around the tumor and the maximum SUV is almost always in both VOIs. However, the *minimum* was one of the least stable features (ICC3 = 0.2), which can be explained by the fact that the *Radiotherapy* VOI is generally larger than the *Dedicated* VOI and therefore includes lower SUV values. Regarding the second-order families, all GLCM, GLRLM, and GLSZM

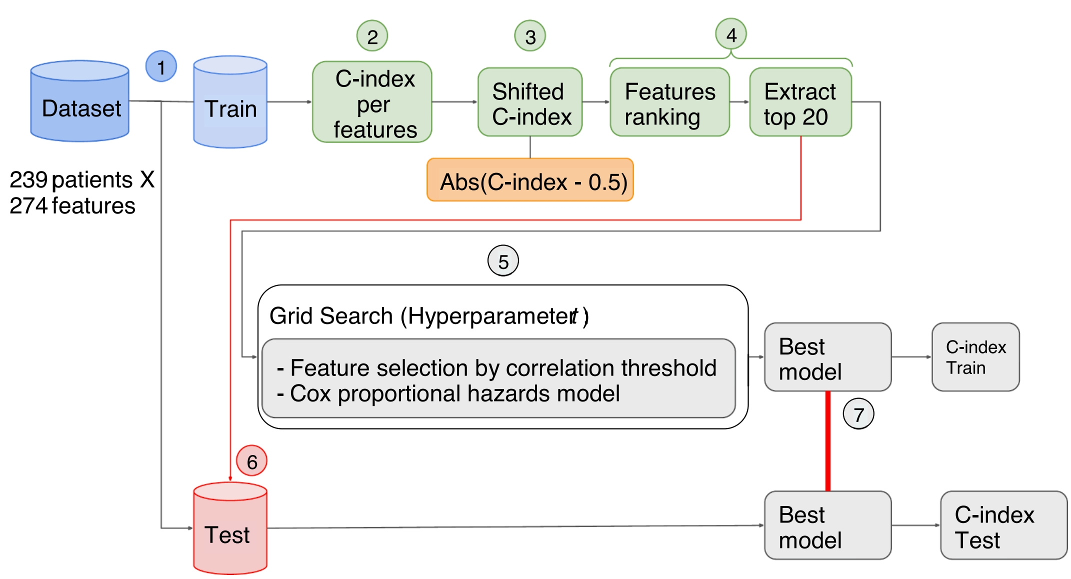
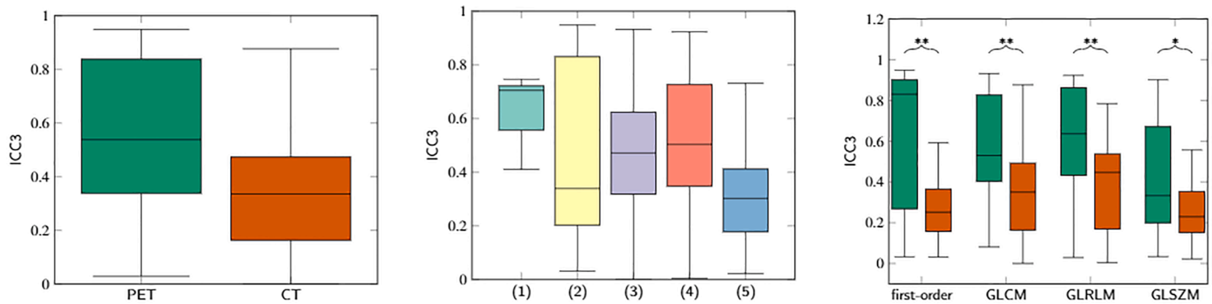


Fig. 2. Flow chart of the proposed radiomics analysis. Univariable steps are shown in green and multivariable analyses in gray. We repeated those steps 100 times with random splits to define training/validation (80%) and test (20%) sets using a stratified shuffle split method. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



(a) Features stability with respect to imaging modality ($p < 0.001$).

(b) Features stability with respect to family. Imaging modalities (PET, CT) are combined. (1) Shape, (2) First-order, (3) GLCM, (4) GLRLM and (5) GLSZM.

(c) Features stability with respect to family extracted from either PET (green) or CT (orange) images. $**$: $p < 0.01$ and $*$: $p < 0.05$.

Fig. 3. Feature stability comparison when extracted from either *Radiotherapy* or *Dedicated* VOIs.

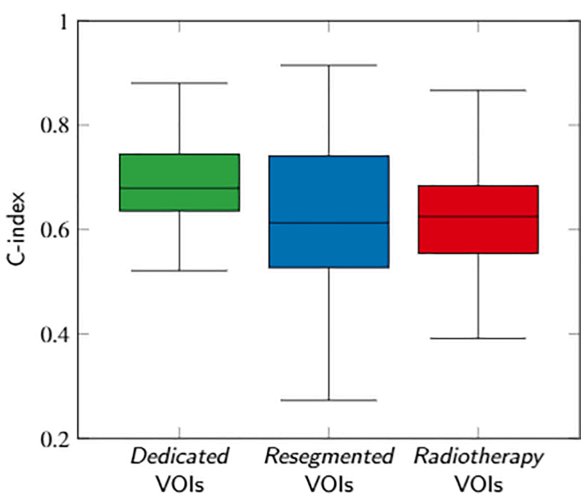


Fig. 4. C-index values for the three VOI types. These results are obtained from 100 repetitions of the radiomics pipeline depicted in Fig. 2.

feature sets were overall unstable (see Fig. 3b). When looking closely at Fig. 3c, however, the stability was larger in PET images, particularly for GLCM and GLRLM features. For GLSZM, the stability was mostly low in both imaging modalities. No specific parameter optimization was performed in the feature extraction step. Therefore, the use of default parameters may explain the poor stability of those texture features.

In this context of H&N cancer, we observed that survival models based on *Dedicated* contours achieved better performance for predicting PFS and led to improved patient risk stratification in comparison to using *Radiotherapy* contours. It is worth noting that using the standard uncorrected student's *t*-test yielded a *p*-value close to 0 ($8.51{\cdot}10^{-8}$). We feel that reporting the latter is important as many studies in the field do not use corrections, breaking the independence assumption of the *t*-test as the repeated random splits are containing overlapping observations. Therefore, according *Benavoli et. al* [34], we performed a Bayesian approach to assess the performance significance between those two model. Thus, we computed the probability density function of the difference between the results of each model (C-index dedicated contours – C-index radiotherapy contours). Then we calculated the integral of the posterior on the interval $(0, +\infty)$ and we obtained a value of 0.893. In other words: the probability of dedicated VOI model being more accurate (C-index) than Radiotherapy VOI model is 89.3%, suggesting that 9 times over 10, a model based on dedicated ROIs will outperform the model based on radiotherapy ROIs. And so, by using this more appropriate approach we can conclude from the statistical analysis that the use of *dedicated* VOIs significantly improved the prediction performance.

It is also worth noting that the cleaning process was based on manual re-segmentation and may not be suitable for large-scale studies. We estimated duration of 20 to 30 min to perform the VOI cleaning stage for one patient. Moreover, adding an automatic re-segmentation step (*Resegmented* VOIs) based on fixed ranges of values did not improve the overall performance. The average C-index was higher than when we use the *Radiotherapy* VOIs but the Inter Quartile Range (IQR) is almost 2 times bigger and the average was lower.

We also recognize some limitations of this work. First, the workflow proposed in this study may not be fully optimized for this task. As an example, we did not explore filter-based radiomics features [35,36]. Liu *et al.* [37] and other studies reported a better predictive performance to model PFS in H&N cancer. However, while the performance can- not be directly compared, the goal of this study was not to find the best model to predict PFS but to focus on the performance comparison between *Dedicated* and *Radiotherapy* contours using the classical radiomics approach.

In future work, we will apply this workflow to combine clinical patient data (*e.g.* age, gender, smoking status, tumor site) and radiomics features in order to further improve the prognosis performance of the model.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Swiss National Science Foundation (SNSF, grant 205320_179069) and the Swiss Personalized Health Network (SPHN via the IMAGINE and QA4IQI projects). Martin Vallières acknowledges funding from the Canada CIFAR AI Chairs Program.

References

[1] Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than pictures, they are data. Radiology 2016;278(2):563–77. https://doi.org/10.1148/radiol.2015151169.

[2] Zwanenburg A, et al. The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology 2020;295(2):328–38. https://doi.org/10.1148/RADIOL.2020191145/ASSET/IMAGES/LARGE/RADIOL.2020191145.FIG5.JPEG.

[3] Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017;14(12):749–62.

[4] van Dijk LV, Thor M, Steenbakkers RJHM, Apte A, Zhai T-T, Borra R, et al. Parotid gland fat related Magnetic Resonance image biomarkers improve prediction of late radiation-induced xerostomia. Radiother. Oncol. 2018;128(3):459–66. https://doi.org/10.1016/j.radonc.2018.06.012.

[5] Castelli J, Depeursinge A, Devillers A, Campillo-Gimenez B, Dicente Y, Prior JO, et al. PET-based prognostic survival model after radiotherapy for head and neck cancer. Eur. J. Nucl. Med. Mol. Imaging 2019;46(3):638–49. https://doi.org/10.1007/s00259-018-4134-9.

[6] Leijenaar RTH, Bogowicz M, Jochems A, Hoebers FJP, Wesseling FWR, Huang SH, et al. Development and validation of a radiomic signature to predict HPV (p16) status from standard CT imaging: A multicenter study. Br. J. Radiol. 2018: 20170498. https://doi.org/10.1259/bjr.20170498.

[7] Leijenaar RTH, Carvalho S, Hoebers FJP, Aerts HJWL, van Elmpt WJC, Huang SH, et al. External validation of a prognostic CT-based radiomic signature in oropharyngeal squamous cell carcinoma. Acta Oncol. (Madr) 2015;54(9):1423–9. https://doi.org/10.3109/0284186X.2015.1061214.

[8] Vallières M, Kay-Rivest E, Perrin LJ, Liem X, Furstoss C, Aerts HJWL, et al. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. Sci. Rep. 2017;7(1). https://doi.org/10.1038/s41598-017-10371-5.

[9] Leger S, Zwanenburg A, Pilz K, Zschaeck S, Zöphel K, Kotzerke J, et al. CT imaging during treatment improves radiomic models for patients with locally advanced head and neck cancer. Radiother. Oncol. 2019;130:10–7. https://doi.org/10.1016/j.radonc.2018.07.020.

[10] W. Lv H. Feng D. Du J. Ma L. Lu "Complementary value of intra-and peri-tumoral PET/CT radiomics for outcome prediction in head and neck cancer", IEEE Access XX 2021 1 1 10.1109/ACCESS.2021.3085601.

[11] Bogowicz M, Riesterer O, Ikenberg K, Stieb S, Moch H, Studer G, et al. Computed Tomography Radiomics Predicts HPV Status and Local Tumor Control After Definitive Radiochemotherapy in Head and Neck Squamous Cell Carcinoma. Int. J. Radiat. Oncol. Biol. Phys. 2017;99(4):921–8. https://doi.org/10.1016/j.ijrobp.2017.06.002.

[12] Bogowicz M, Riesterer O, Stark LS, Studer G, Unkelbach J, Guckenberger M, et al. Comparison of PET and CT radiomics for prediction of local tumor control in head and neck squamous cell carcinoma. Acta Oncol. (Madr) 2017;56(11):1531–6. https://doi.org/10.1080/0284186X.2017.1346382.

[13] V. Andrearczyk *et al.*, "Multi-task Deep Segmentation and Radiomics for Automatic Prognosis in Head and Neck Cancer," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12928 LNCS, pp. 147–156, Oct. 2021, doi: 10.1007/978-3-030-87602-9_14.

[14] P. Fontaine *et al.*, "Fully Automatic Head and Neck Cancer Prognosis Prediction in PET/CT," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 13050 LNCS, pp. 59–68, Oct. 2021, doi: 10.1007/978-3-030-89847-2_6.

[15] [15] G. Zhang *et al.*, "Comparable Performance of Deep Learning-Based to Manual-Based Tumor Segmentation in KRAS/NRAS/ BRAF Mutation Prediction With MR-Based Radiomics in Rectal Cancer," doi: 10.3389/fonc.2021.696706.

[16] Parmar C, et al. Radiomic feature clusters and Prognostic Signatures specific for Lung and Head & neck cancer. Sci. Rep. 2015;5:1–10. https://doi.org/10.1038/srep11044.

[17] Zhang J, Wang Y, Dong Y, Wang Y. Computer-aided diagnosis of cervical lymph nodes on ultrasonography. Comput. Biol. Med. 2008;38(2):234–43. https://doi.org/10.1016/j.compbiomed.2007.10.005.

[18] Al Ajmi E, Forghani B, Reinhold C, Bayat M, Forghani R. Spectral multi-energy CT texture analysis with machine learning for tissue classification: An investigation using classification of benign parotid tumours as a testing paradigm. Eur. Radiol. 2018;28(6):2604–11. https://doi.org/10.1007/s00330-017-5214-0.

[19] [19] G. Wang, L. He, C. Yuan, Y. Huang, Z. Liu, and C. Liang, "Pretreatment MR imaging radiomics signatures for response prediction to induction chemotherapy in patients with nasopharyngeal carcinoma," *Eur. J. Radiol.*, vol. 98, no. October 2017, pp. 100–106, 2018, doi: 10.1016/j.ejrad.2017.11.007.

[20] Zhang B, Tian J, Dong Di, Gu D, Dong Y, Zhang Lu, et al. Radiomics features of multiparametric MRI as novel prognostic factors in advanced nasopharyngeal carcinoma. Clin. Cancer Res. 2017;23(15):4259–69. https://doi.org/10.1158/1078-0432.CCR-16-2910.

[21] Ouyang F-S, Guo B-L, Zhang B, Dong Y-H, Zhang Lu, Mo X-K, et al. Exploration and validation of radiomics signature as an independent prognostic biomarker in stage III-IVb nasopharyngeal carcinoma. Oncotarget 2017;8(43):74869–79. https://doi.org/10.18632/oncotarget.20423.

[22] Rizzo S, Botta F, Raimondi S, Origgi D, Fanciullo C, Morganti AG, et al. Radiomics: the facts and the challenges of image analysis. Eur Radiol Exp 2018;2(1). https://doi.org/10.1186/s41747-018-0068-z.

[23] Fontaine P, Acosta O, Castelli J, De Crevoisier R, Müller H. The importance of feature aggregation in radiomics : a head and neck cancer study. Sci. Rep. 2020: 1–11. https://doi.org/10.1038/s41598-020-76310-z.

[24] Depeursinge A, Yanagawa M, Leung AN, Rubin DL. Predicting adenocarcinoma recurrence using computational texture models of nodule components in lung CT. Med. Phys. 2015;42(4):2054–63. https://doi.org/10.1118/1.4916088.

[25] Yang F, Simpson G, Young L, Ford J, Dogan N, Wang L. Impact of contouring variability on oncological PET radiomics features in the lung. Sci. Rep. 2020;10(1): 1–10. https://doi.org/10.1038/s41598-019-57171-7.

[26] Pavic M, Bogowicz M, Würms X, Glatz S, Finazzi T, Riesterer O, et al. Influence of inter-observer delineation variability on radiomics stability in different tumor sites. Acta Oncol. (Madr) 2018;57(8):1070–4. https://doi.org/10.1080/0284186X.2018.1445283.

[27] Huang Q, Lu L, Dercle L, Lichtenstein P, Li Y, Yin Q, et al. Interobserver variability in tumor contouring affects the use of radiomics to predict mutational status. J. Med. Imaging 2018;5(01):1. https://doi.org/10.1117/1.JMI.5.1.011005.

[28] Andrearczyk V, et al. Overview of the HECKTOR Challenge at MICCAI 2020: Automatic Head and Neck Tumor Segmentation in PET/CT. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 2021; vol. 12603 LNCS:1–21. https://doi.org/10.1007/978-3-030-67194-5_1.

[29] van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Van Griethuysen *et al.*, "Computational Radiomics System to Decode the. Radiographic Phenotype" 2017;77(21):e104–7. https://doi.org/10.1158/0008-5472.CAN-17-0339.

[30] Koo TK, Li MY. "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for. Reliability Research" 2016;15(2):155–63. https://doi.org/10.1016/j.jcm.2016.02.012.

[31] Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat. Med. Feb. 1996;15(4):361–87. https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4.

[32] Cox DR. Regression Models and Life-Tables. "Regression Models and Life-Tables" 1972;34(2):187–202.

[33] S. Pölsterl, "scikit-survival: A Library for Time-to-Event Analysis Built on Top of scikit-learn," *J. Mach. Learn. Res.*, vol. 21, no. 212, pp. 1–6, 2020, Accessed: Dec. 21, 2021. [Online]. Available: http://jmlr.org/papers/v21/20-729.html.

[34] A. Benavoli, G. Corani, J. Demšar, and M. Zaffalon, "Time for a Change: a Tutorial for Comparing Multiple Classifiers Through Bayesian Analysis" *J. Mach. Learn. Res.*, vol. 18, pp. 1–36, 2017, Accessed: Dec. 21, 2021. [Online]. Available: http://jmlr.org/papers/v18/16-305.html.

[35] Soufi M, Arimura H, Nagami N. Identification of optimal mother wavelets in survival prediction of lung cancer patients using wavelet decomposition-based radiomic features. Med. Phys. 2018;45(11):5116–28.

[36] Yang Z, He B, Zhuang X, Gao X, Wang D, Li M, et al. CT-based radiomic signatures for prediction of pathologic complete response in esophageal squamous cell carcinoma after neoadjuvant chemoradiotherapy. J. Radiat. Res. 2019;60(4): 538–45. https://doi.org/10.1093/jrr/rrz027.

[37] Liu Z, Cao Y, Diao W, Cheng Y, Jia Z, Peng X. Radiomics-based prediction of survival in patients with head and neck squamous cell carcinoma based on pre- and post-treatment 18F-PET/CT. Aging 2020;12(14):14593–619.