

Predicting the Need for Adaptive Radiotherapy in Head and Neck Patients from CT-Based Radiomics and Pre-treatment Data

Natália Alves[1](✉), Joana Dias[2,3], Tiago Ventura[4], Josefina Mateus[4], Miguel Capela[4], Leila Khouri[5], and Maria do Carmo Lopes[2,4]

[1] Faculty of Science and Technology, Physics Department, University of Coimbra, Coimbra, Portugal
[2] Institute of Computer and Systems Engineering, Coimbra, Portugal
[3] Faculty of Economics, University of Coimbra, Coimbra, Portugal
[4] Medical Physics Department, Instituto Português de Oncologia de Coimbra Francisco Gentil, EPE, Coimbra, Portugal
[5] EPE, Radiotherapy Department, Instituto Português de Oncologia de Coimbra Francisco Gentil, Coimbra, Portugal

Abstract. Although adaptive radiotherapy can reduce the negative dosimetric and clinical impacts of anatomical changes during head and neck treatments, evidence shows that it is not equally beneficial for all patients. This makes it important to electively schedule adaptation ahead of time to optimize clinical resources and patient benefit. The purpose of this study is to assess the feasibility of using both pre-treatment patient features and radiomic features extracted from a pre-treatment contrast enhanced computed tomography scan to predict the need for adaptive radiotherapy. Seventy-two patients were included in the analysis, of which 36 required adaptation. 36 pre-treatment semantic features as well as 351 radiomic features extracted from the gross target volume were considered. Three support vector machine models were developed: 1) considering only semantic features; 2) considering only radiomic features; 3) using a combination of features from 1 and 2. A robustness analysis of the selected radiomic features was also conducted. The best classification results were obtained considering 6 features (4 semantic and 2 radiomic) with median accuracy and area under the receiver operating characteristic curve of 0.821 and 0.843, respectively.

Keywords: Adaptive radiotherapy · Radiomics · Head and neck cancer

1 Introduction

Head and neck (H&N) cancer is one of the most common type of cancer worldwide, with approximately 1.5 million new cases and 0.9 million deaths in 2018 alone [1]. In current clinical practice, the vast majority of patients with locally advanced H&N cancer require radiotherapy (RT), with or without concurrent chemotherapy. Intensity modulated radiotherapy (IMRT) is the gold standard treatment technique [2]. The highly

© Springer Nature Switzerland AG 2021
O. Gervasi et al. (Eds.): ICCSA 2021, LNCS 12953, pp. 429–444, 2021.
https://doi.org/10.1007/978-3-030-86976-2_29

conformal dose distributions produced by IMRT lead to steep dose gradients in the borders of the target volumes, which are extremely sensitive to positional errors and anatomic changes, especially in H&N cases, since there are several critical structures very close to the tumor.

Several studies have shown that significant anatomical alterations in both the target volumes and the organs at risk (OAR) may occur during RT treatments, leading to deviations between the planned and the delivered doses [3–6]. These might cause underdose the target volume, with loss of tumor control, and/or overdose of the normal structures, originating unexpected side effects [6, 7].

In order to correct for day-to-day positioning errors relative to the planning CT (p-CT), most centers perform Image Guided Radiotherapy (IGRT). This consists of daily position verification through imaging, and the performance of couch shifts to correct for deviations between the daily image and the p-CT. However, one of the main limitations of IGRT is that internal changes in size, shape or relative position of the target volumes or OARs compared with the initial p-CT cannot be corrected by rigid translational and/or rotational couch shifts [8].

A possible solution is Adaptive Radiotherapy (ART), which aims to correct for the anatomical modifications by adapting the initial target volumes and planning to the current patient status [4–9]. This process requires repeated imaging with sufficient quality for treatment planning, re-contouring and re-planning, and can be done either offline, between treatment fractions, or online immediately prior to a fraction, while the patient is laying on the treatment couch [9, 10]. The most frequently implemented type of ART is offline ART since online ART requires advanced tools such as automatic segmentation and planning which are not yet commonly available in most centers. In this work, the term ART refers to offline ART.

Although several studies have successfully established both the clinical and dosimetric benefits of ART application [11–17], its implementation into the clinical practice is still a rather labor-intensive and time-consuming process, making it impossible for most clinics to implement it for every patient [9]. Furthermore, evidence shows that ART is not equally important for all patients, potentially being an unnecessary effort for some [18], with the percentage of patients that benefit from ART reported to range from 21% to 66% [19, 20]. This means that electively scheduling adaptive replanning when initiating treatment would be extremely important to optimize clinical resources and the benefit for the patient.

With this goal, many studies have investigated the statistical relationship between different pre-treatment factors and the need for ART [4], but to this date there is relatively few work on developing multi-variable models with the objective of predicting the need for ART [21–23]. The main issue with performing the selection of predictive factors based on correlation tests alone is that these only take into account the linear relationship between individual variables and the endpoint, not-considering multi-variable or non-linear interactions that could potentially improve prediction.

Radiomics is a field with rapidly increasing popularity in radiation oncology, consisting of the high-throughput extraction of quantitative biomarkers (features) from medical images, which can then be applied in mathematical models for several purposes, such as aid in diagnosis or prediction of outcome [24–29].

The purpose of this study is to assess the feasibility of using both pre-treatment patient features and radiomic features from a pre-treatment contrast enhanced CT scan to predict the need for ART in H&N patients. Three models were developed: 1) considering only pre-treatment clinical data from the patient; 2) considering only radiomic features extracted from the gross target volume (GTV); 3) using a combination of features from 1 and 2.

2 Materials and Methods

2.1 Sample Description

The study sample consisted of 72 (66 male and 6 female) H&N patients from the Portuguese Institute of Oncology Clinic Francisco Gentil in Coimbra, Portugal, which were treated with helical IMRT delivered by a Tomotherapy HD (Accuray) unit, with (58 patients) or without (14 patients) concomitant chemotherapy.

The vast majority (63 out of 72) of patients were prescribed 69.96 Gy to the tumor planning target volume (PTV-T), with 2 patients being prescribed 59.4 Gy. The prescription to the lymphatic nodes' PTVs (PTV-N) was either 50.4 Gy, 54 Gy or 59.4 Gy. Regarding the adenopathies, 48 patients were prescribed 69.96 Gy, 4 patients were prescribed 66 Gy, one patient was prescribed 50 Gy and one patient 59,4 Gy. The remaining patients did not present any adenopathies. There were 3 patients with a prescription of 50.4 Gy for both the PTV-T and the PTV-N, and 4 patients with dose prescribed to the PTV-T only (50 Gy for 3 patients and 70Gy for 1 patient). Depending on the prescription scheme the treatment was delivered in either 33 (65 patients), 28 (4 patients) or 20 (3 patients) fractions.

2.2 Patient Imaging

For diagnostic purposes, patients underwent a contrast enhanced CT (d-CT) prior to the first appointment with the radiation oncologist (at a median of 27 days before the planning CT). The d-CT scan is often acquired at different external centers, leading to varying acquisition parameters, namely slice thickness and reconstruction kernels. The d-CT was available for 67 patients.

All patients had a planning CT (p-CT) scan in treatment position acquired at a median of 14.5 days before the start of the treatment, on which the target volumes and OAR were delineated either manually (targets) or through an auto-segmentation software (OAR). This CT scan was acquired using the scanner Somatom Sensation Open from Siemens, with a 3 mm slice thickness.

2.3 Adaptive Radiotherapy Scheme

All patients underwent a daily Megavoltage CT (MV-CT) scan, before each treatment fraction, to assess for differences between the current position and the one in the p-CT. A rigid registration was then performed between the MV-CT and the p-CT, and a transformation vector calculated, which translated into rigid couch and roll angle shifts to correct for positional errors.

If significant discrepancies could be observed after the rigid transformation, the medical physicist was alerted, and a dose calculation was performed on the daily scan using the Planned Adaptive software module from Accuray. This dose distribution was then revised by the responsible radiation oncologist, which determined the need for a new treatment plan. If a replan was required, a second CT (re-CT) was acquired, and both the target volumes and OAR were propagated from the p-CT to the re-CT by rigid and deformable image registration. These volumes were then manually corrected by the radiation oncologist and a new plan was designed by the responsible medical physicist.

From the 72 patients included in the study, 36 had at least one replan during treatment (ART group), with 3 requiring 2 replans. The remaining 36 patients were used as the control group for comparison and model development.

2.4 Semantic Features

Thirty-six pre-treatment factors (features), which could potentially be predictive for the need of ART were acquired for each patient.

Initially, statistical tests were performed to determine whether there were significant differences between the replan and control groups for each factor. For continuous features (as initial weight, GTV volume/surface area, parotid glands volume/surface area, or prescription doses) first the normality and homogeneity of the variance were assessed using the Shapiro-Wilk normality test and Levene's test respectively, in order to determine whether parametric or non-parametric tests were suitable. Then, the t-test and Wilcoxon sign ranked test were used to assess statistical significance in the parametric and non-parametric continuous features, respectively.

For categorical features (as drinking habits, smoking habits, TNM staging, or chemotherapy) the chi-squared test was used. As a real clinical dataset was considered in this study, there was some missing data and not all parameters had the information for all patients.

2.5 Radiomics

The radiomic analysis was performed in MATLAB R2019b. The adopted workflow was based on the guidelines provided by the Image Biomarker Standardization Initiative (IBSI) [30], regarding both image pre-processing and feature extraction, as implemented by the Standardized Environment for Radiomics Analysis (SERA) [31]. SERA is a MATLAB-based framework developed at Johns Hopkins University in 2019, that can handle CT, PET, SPECT or MRI images, and extracts standardized radiomic features in compliance with the IBSI's guidelines. A simplified global representation of the adopted radiomics workflow can be seen in Fig. 1.

Image segmentation was done from the manual contour of the GTV by the radiation oncologist in the p-CTs. Rigid and deformable image registration of the GTV from the p-CT was performed using the Velocity AI v. 3.2. software from Varian to define the region of interest (ROI) in the d-CT.

To correct for differences in voxel spacing and slice thickness on the d-CTs, all images were interpolated to an isotropic voxel spacing of 1 mm x 1 mm x 1 mm, using

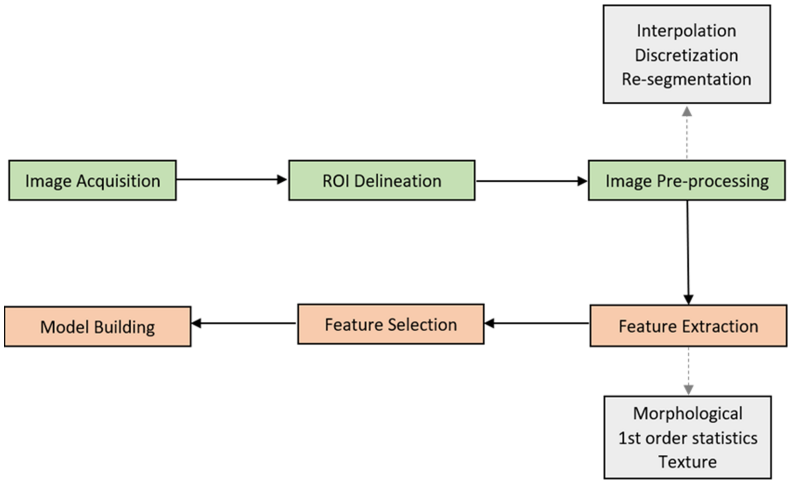

Fig. 1. General Radiomics workflow. The boxes in green represent steps where the actual images and/or ROI mask are considered. The orange boxes represent steps where vectors of the quantitative features extracted from the images are considered.

the tri-linear interpolation algorithm. Furthermore, the intensity range of the ROI was re-segmented to include only voxels between -150 and 180 Hounsfield units, eliminating all bone and air voxels [32], and voxels with outlier intensities (outside the range of ± 3 standard deviations) were removed. Lastly, intensities inside the ROI were discretized in a fixed bin number approach considering 32 bins.

Having completed these pre-processing steps, 351 radiomic features (as defined by the IBSI [30]) were extracted from each ROI, which are summarized in Table 1.

2.6 Feature Selection and Model Building

Feature selection was made by a greedy search algorithm, which at each iteration eliminates one feature based on the performance of the classification. This was done considering four different classifiers: support vector machines (SVM), naïve-Bayes, decision trees (DT) and multi adaptive regression splines (MARS). At each iteration, the impact of removing each available feature was tested using leave one out cross-validation (LOO) and assessing the accuracy, true positive rate (TPR) and true negative rate (TNR) of the classification. The feature set with better performance moves on to the next iteration of the algorithm and the whole process is repeated until there is only one feature left. Applying this method with different classifiers will produce possibly different feature selected sets, since the classification performance will be the best one obtained for the particular classifier used.

This process was done for the semantic features alone, for the radiomic features alone, and then for a combination of both, using the 20 last-standing features from each. For the semantic features, all the missing values were replaced by the mean value of the existing data for that factor.

Regarding the radiomic features, all features were normalized (z-normalization) after extraction and hierarchical clustering was performed to prevent redundancy. The clustering was based on the Spearman correlation coefficient, meaning that highly correlated features were assigned to the same cluster, and all feature clusters with intra-cluster correlation > 0.9 were replaced by a meta-feature which corresponded to the average of all features in the cluster. This reduced set of non-redundant features was finally fed to the feature selection algorithm previously described.

After determining a suitable set of variables, hyper-parameter optimization was conducted for the SVM using only the selected variables. The parameters considered for optimization were the box-constraints and the kernel scale, and a Bayesian optimization algorithm implemented in the built-in SVM function of MATLAB was used, considering a 5-fold cross-validation.

Table 1. Extracted radiomic features.

Feature family	Number
Morphological	29
Local Intensity	2
Intensity-based statistics	18
Intensity Histogram	23
Intensity-Volume Histogram	7
Gray Level Co-occurrence Matrix (GLCM) 2D/2.5D/3D	100
Gray Level Run Length Matrix (GLRLM) 2D/2.5D/3D	64
Gray Level Size Zone Matrix (GLSZM) 2D/3D	32
Gray Level Distance Zone Matrix (GLDZM) 2D/3D	32
Neighbourhood Grey Tone Difference Matrix (NGTDM) 2D/3D	10
Neighbouring Grey Level Dependence Matrix (NGLDM) 2D/3D	34
TOTAL	351

The classification was then performed with 30 repetitions of leave-one-out cross-validation. Mean accuracy, sensitivity, specificity, and area under the ROC curve were evaluated. The presented workflow is schematically represented in Fig. 2.

2.7 Robustness Analysis

Radiomic feature's robustness to variations in patient positioning, different degrees of noise in the image, and variation in the ROI delineation was assessed using the framework proposed by Zwanenburg, et al. [33]. Based on the results obtained by the authors for a H&N squamous cell carcinoma cohort, four types of image perturbations were selected: rotation, gaussian noise addition, volume adaption an contour randomization [33]. These

perturbations were chained using varying degrees for the angle θ of rotation around the z-axis, namely $\theta \in [-10°, -6°, -2°, 2°, 6°, 10°]$, as well as varying fractions β for growth/shrinkage of the ROI mask in volume adaption, namely $\beta \in [-0.2, -0.1, 0, 0.1, 0.2,]$. The combination of these parameters resulted in 30 perturbed images per patient, on which features selected in the previous step were calculated. The intra-class correlation coefficient ICC(1,1) [34] was then computed to assess feature robustness.

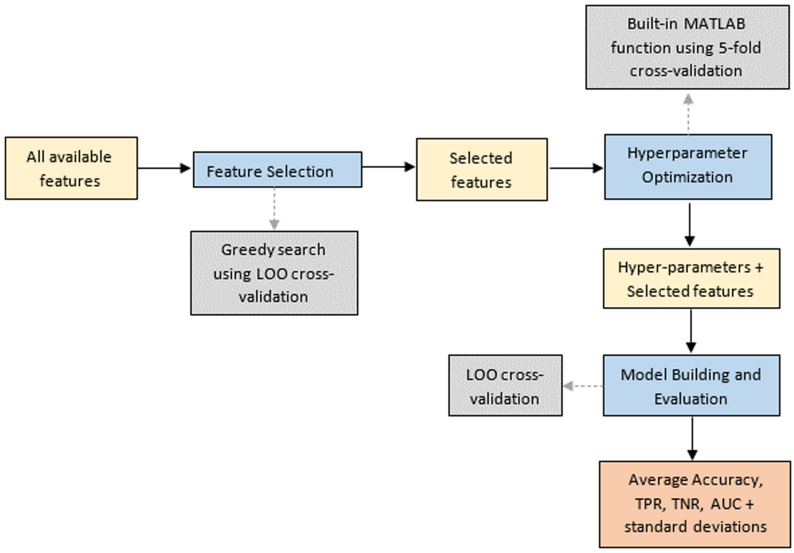

Fig. 2. Adopted workflow from feature selection to model building. The yellow boxes represent features and hyper-parameters, which are inputs/outputs of the different processes; the blue boxes represent the algorithmic processes, with a brief description in the grey box, and the orange box represents the final result. (Color figure online)

3 Results

3.1 Statistical Analysis

From all the analyzed variables the only ones that showed a statistically significant difference between the control and the ART groups were the planned minimum dose to the GTV (Dmin GTV) and both the volume and the surface area of the initial GTV. These results as well as some other pre-treatment features that were considered relevant are shown in Table 2.

Table 2. Statistical analysis for pre-treatment factors. The mean ± standard deviation for each group is shown for the continuous variables age, initial weight, Dmin GTV, prescribed dose to the adenopathies (Dpresc ADNs) and GTV volume (vol)/surface area (sa). The drinking habits correspond to 1- marked drinking habits, 2- moderate drinking habits, 3- no drinking habits and 4- ex-drinking habits.

Feature	Control	ART	p-value
Age	57.28 ± 8.77	61.19 ± 10.24	0.102
Gender (M/F)	34/2	32/4	0.157
T stage (1/2/3/4)	2/4/12/18	1/1/10/24	0.243
N stage (0/1/2/3/X)	4/5/25/1/1	4/1/26/5/0	0.242
Drinking Habits (1/2/3/4)	6/10/3/17	14/3/8/11	0.213
Initial Weight (Kg)	67.05 ± 14.14	62.39 ± 14.59	0.140
Dpresc ADNs (Gy)	52.14 ± 30,55	52.8 ± 29,50	0,889
Dmin GTV (Gy)	67.47 ± 4.44	64.78 ± 6.67	0.031*
GTV vol (cm3)	47.10 ± 46.02	85.56 ± 69.95	0.006*
GTV sa (mm2)	4284.08 ± 3157,49	6367.05 ± 3778.02	0.007*

3.2 Feature Selection and Model Building

For the semantic features, after performing several experiments using SVM, DT, MARS, and Bayes classification methods (referred in Sect. 2.6) the SVM presented the best results so it was the selected classifier.

The best balance between accuracy, TPR and TNR was found with six selected variables using the greedy search method with LOO cross-validation. These features were: age, initial weight, prescription dose to adenopathies (Dpresc ADNs), Dmin GTV, and the binary variables T stage = 1 and marked drinking habits.

After optimizing the box constraints and kernel scale parameters for the SVM classifier with the selected variables, the results achieved for 30 repetitions of LOO cross validation for the accuracy, TPR, TNR, and area under the receiver operating characteristic (ROC) curve (AUC) can be seen in Table 3.

Table 3. Results of the classification using 6 semantic features.

Accuracy			TPR		
Mean	Med	Std	Mean	Med	Std
0.799	0.806	0.014	0.778	0.778	0.023
TNR			**AUC**		
Mean	Med	Std	Mean	Med	Std
0.820	0.819	0.016	0.803	0.803	0.004

For the radiomic features, after the hierarchical clustering, the original set of 351 features extracted from the GTV delineated in the d-CT was reduced to a non-redundant initial set of 88 features. Several tests for feature selection were done with three different classifiers: DT, Bayes and SVM, with the best results arising for SVM, which was again the selected method.

The best balance between accuracy, TPR and TNR was found for six selected features (R1-R6 Table 4). From the six selected features, features R4, R5 and R6 corresponded to original extracted features and R1, R2 and R3 to clustered features. Feature R2 consists of a cluster containing eighteen highly correlated features Table 5 presents these results, as well as the ones obtained considering only 3 features.

Table 4. 6 selected features from the GTV in the dCT.

Feature(s)		Feature Family
R1	Grey level non uniformity normalised	GLSZM 2D GLDZM 2D
R2	Cluster with 18 features (described in the text)	GLRLM 2D/2.5D/3D (8 features) GLSZM 2D/3D (4 features) GLDZM 2D/3D (2 features) NGLDM 2D/3D (4 features)
R3	10th percentile Intensity at 90	Intensity-based statistics Intensity-volume histogram
R4	Centre of mass shift	Morphology
R5	Flatness	Morphology
R6	Small distance low grey level emphasis	GLDZM 3D

Figure 3 shows the ROC curves obtained for each set of variables considering LOO cross-validation from 3 to 6 features in total. The 3-feature model corresponds to.

features R1 to R3 from Table 4, the 4-feature model to features R1 to R4, and so on for the models with five and six features.

As for the combination of semantic and radiomic features, the best results for the feature selection were again achieved with six selected features, from which four were the semantic features: age, prescription dose to ADNs, Dmin to the GTV and marked drinking habits, and two were the radiomic features: a clustered feature comprising the zone-size non-uniformity normalized and the small-zone emphasis from the GLSZM 3D (feature R7), and feature R1 from Table 4. The results obtained for 30 repetitions LOO cross-validation for the models with 6 and 3 features are shown in Table 6. The last three features standing after the greedy search were the Dmin to the GTV, R1 and R7.

3.3 Robustness Analysis

Regarding the robustness analysis, the ICC (1,1) was above 0.77 for all radiomic features except R1, which presented an ICC (1,1) of 0.602 (95% CI of 0.523–0.690). Features R2 and R5 were the most reproducible across the different image perturbations with ICCs (1,1) of 0.985 (95% CI of 0.980–0.990]) and 0.954 (95% CI of 0.937–0.968) respectively. Features R7 and R4 also obtained good reproducibility with ICCs of 0.887 (95% CI of 0.850–0.920) and 0.858 (95% CI of 0.813–0.898) respectively. Finally, the ICC values for features R3 and R6 were 0.785 (95% CI of 0.725–0.842) and 0.775 (95% CI of 0.713–0.834).

Table 5. Results obtained for the 6-feature and 3-feature radiomic models

Accuracy				TPR		
	Mean	Med	Std	Med	Mean	Std
6 features	0.783	0.776	0.008	0.755	0.758	0.009
3 features	0.761	0.761	0.000	0.727	0.727	0.000
TNR				**AUC**		
	Mean	Med	Std	Mean	Med	Std
6 features	0.810	0.824	0.015	0.798	0.799	0.004
3 features	0.794	0.794	0.000	0.765	0.765	0.003

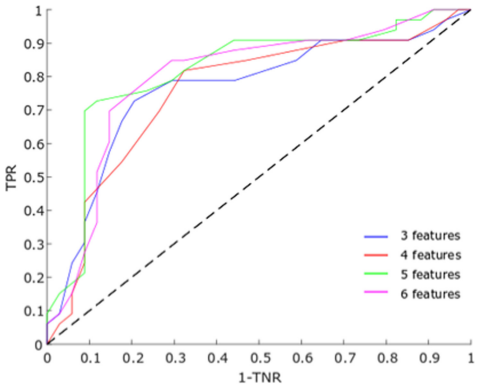

Fig. 3. ROC curves obtained for the models with 3 to 6 radiomic features using LOO cross-validation.

Table 6. Results obtained for the 6-feature and 3-feature mixed models

Accuracy				TPR		
	Mean	Med	Std	Mean	Med	Std
6 features	0.817	0.821	0.016	0.778	0.788	0.029
3 features	0.726	0.731	0.009	0.672	0.667	0.011
TNR				**AUC**		
	Mean	Med	Std	Mean	Med	Std
6 features	0.856	0.853	0.014	0.844	0.843	0.005
3 features	0.779	0.779	0.015	0.777	0.776	0.005

4 Discussion

In this study, three separate models based on SVM were built with the purpose of predicting the need for ART in H&N patients prior to the beginning of treatment: 1) considering only pre-treatment clinical data from the patient; 2) considering only radiomic features extracted from pre-treatment CT images; 3) using a combination of features from 1 and 2.

The idea of applying radiomics to predict the need for ART is very recent, with only two published papers on the topic so far [35, 36]. The first study to explore radiomics for ART patient eligibility was published by Ramella et al. [35] in 2018 and investigated the ability to predict tumor reduction during chemoradiotherapy for a cohort of 91 non-small cell lung cancer patients. The authors considered both radiomic and semantic features, and were able to obtain an AUC of 82% (95% CI of 73%-91%), an accuracy of 78% (95% CI of 69.5%-86.5%), a precision of 77.8% (95% CI 69.95%-86%), a sensitivity of 84% (95% CI of 75.7%-92.2%), and a PPV and NPV of 65.7% (95% CI of 60.7%-70.7%) and 86.9% (95% CI 83.4%-90.4%) respectively, considering 12 (5 semantic and 7 radiomic) features. It was also reported that the scores obtained using the semantic features alone were considerably lower than with the radiomic signature, showing a decrease of 4% in AUC and 6% in accuracy [35]. Another more recent study, by Yu et al. [36], was published in October 2019 and had the objective of predicting ART eligibility for advanced nasopharyngeal carcinoma patients (NPC), using radiomic features extracted from multi-parametric MR images. The study cohort included 70 NPC patients treated with radical radiotherapy, from which 13 had a replan sometime during treatment. Using Least Absolute Shrinkage and Selection Operator algorithm for both feature selection and for classification, the authors obtained an average AUC value in the testing sets of 93% (95%CI of 92.8%-93.3%) with 6 selected features. No semantic features were included in the predictive models for this study. These studies show that there is potential in exploring radiomic and semantic features to build models capable of predicting the need for ART.

As can be seen from the statistical analysis (Table 2), the planned minimum dose to the GTV is significantly lower in the ART group. However, when comparing the ratio between the Dmin to the GTV and the prescribed dose to the PTV-T no differences

were found between the two groups, which indicates that the PTV-T underdosage is not significantly larger in the ART group than it is in the control group.

Furthermore, statistically significant differences were found for the initial GTV volume and surface area between the two groups, with the ART group presenting on average an 82% higher initial volume when compared to the control group. These results are in line with several published studies that identify higher initial volumes to be associated with greater anatomical changes in the course of treatment [4–7, 22, 23]. Nevertheless, the initial GTV volume was not identified as a predictive factor in any of the models, which suggests that it would not add additional predictive information to the interactions between the selected variables.

Regarding the first predictive model, the features selected by the greedy-search algorithm using only pre-treatment semantic features were age, initial patient weight, the prescribed dose to the ADNs, T stage, the planned minimum dose to the GTV and whether the patient had marked drinking habits or not. These features are clinically relevant and in conformance with previously published results. Different studies have shown patient age to be predictive of both tumor [22] and parotid [37] shrinkage, as well as reported tumor staging to be significantly correlated with anatomical/dosimetric changes [4, 23].

As is shown in Table 3, using the six selected variables in an SVM classifier and performing 30 repetitions of LOO cross-validation we were able to obtain a median accuracy of 80.6%, TPR of 77.8%, TNR of 81.9% and AUC of 80.3%. These results demonstrate that it is possible to obtain a reliable prediction for the need of ART using only pre-treatment semantic variables. LOO cross-validation was chosen since the study cohort is relatively small and it is important to use as much as possible the available samples for training. Furthermore, using, for instance, an 80/20 split could make the results too biased towards the training set. As the final goal is to classify new samples using the whole of the cohort for training, the LOO cross-validation results are a good representation of the classifier performance.

Moving to the radiomics model, the best predictive performance was obtained for six selected features from the GTV volume in the d-CT. As the selected radiomic variables are highly dependent on the imaging modality, the cancer site and classification endpoint, and this is the first study to assess CT radiomics for prediction of ART, more studies of this nature with different cohorts are needed to assess the consistency of the selected features.

The results obtained with LOO cross-validation for these six features presented a median accuracy of 77.6%, a median TPR of 75.8%, a median TNR of 82.4% and finally a median AUC of 79.9%, all with standard deviations ranging from 0.4% to 1.5%.

It is very interesting to note that only three radiomic features are needed, namely the three previously mentioned meta-features, to reach very similar results: a median accuracy of 76.1%, TPR of 72.7%, TNR of 79.4% and AUC of 76.5% with standard deviations lower than 0.3%. This observation is not valid considering semantic features, where it is clear that much more information is lost when the total number of features is reduced. It seems reasonable to conclude that the radiomic model could have a higher ability of generalization considering new data, since the use of fewer features decreases the possibility of overfitting. Nevertheless, the results obtained using just radiomics are in general weaker than the ones considering just the semantic features.

A mixed model was also built using the information obtained from the feature selection procedure based on both the semantic features and the radiomic features. The feature selection algorithm showed the best performance for a set of six features, including four semantic and two radiomic features. The selected semantic features were also selected for the first model, confirming their predictive power. As for the radiomic features, one of them was the meta-feature including the grey level non uniformity normalized obtained from both the 2D grey level size and distance zone matrixes, which had also been selected in the radiomic-only model and which was the last standing feature in the radiomic greedy search, indicating its high predictive capacity. The other selected radiomic feature was also a clustered feature comprising the zone-size non-uniformity normalized and the small-zone emphasis from the GLSZM 3D, which on the contrary was not selected for the radiomic-model alone. This mixed model presented the best results out of the three models, indicating that a mixture of both semantic and radiomic variables is the most likely to be able to predict whether a given patient will need ART during treatment. Nonetheless, the model built using only the three last standing variables, and the two radiomic features could not match the performance of the 3-feature model using only radiomic features.

One of the potential drawback of using a radiomics approach is that some features' values have shown to be affected by factors such as patient positioning, image acquisition parameters and segmentation [26, 38]. In this way, it could be difficult to achieve a set of reproducible features, which in turn are able to generate generalizable models, if a robustness analysis is not included in the radiomic studies.

In this study, a robustness analysis was performed by applying different image perturbations, namely noise addition, rotations, volume adaptations and contour randomizations, to mimic the effects of the above mentioned factors on the calculated features, as proposed by Zwanenburg et al. [33]. From the results of this analysis, 6 out of the 7 radiomic features included in the models presented an ICC higher than 0.75, indicating a good reproducibility across the perturbed images for each patient.

The main limitation of this study is the fact that the sample cohort has a small number of patients and they were all treated at the same health institution. It would be interesting to assess the predictive capability of these models on unseen patients coming from different clinics. Despite this fact, the number of selected variables was six, which is a relatively low number for the size of the dataset, respecting the rule of thumb of having 10–15 samples per feature [39]. Furthermore, because the diagnostic CTs were used, which are acquired at different centers with different equipment and, consequently, varying acquisition parameters, it is reasonable to infer that our models are likely not overfitting to a given set of imaging parameters, which is a major concern in radiomics studies as the extracted features have shown to be affected by these factors [40]. Additionally, by performing the robustness analysis it is also guaranteed that the selected radiomic features are not dependent on specific image orientations (rotation perturbation), contours of the region of interest (volume adaption/ contour randomization perturbation), or overfitting to noise in the images, since they show good robustness to varying degrees of added random gaussian noise.

To the best of our knowledge this is the first study to use radiomic features from CT images to build a model to predict the need for ART in H&N patients. Furthermore, we

compared the radiomic model with both a non-radiomic model built only with semantic features and a mixed model using a combination of radiomic and non-radiomic features. This is a hypothesis-generating study as although several steps were taken to improve the generalizability of the developed models, these do not eliminate the need for external validation.

These results are very promising and in conformity with the ones reported by Ramella et al. [35] and Yu et al. [36], suggesting the ability of using radiomics and pre-treatment factors to predict the need for ART in a clinical H&N cancer dataset.

5 Conclusion

The presented study successfully demonstrated the ability to predict the need for ART in H&N patients from both pre-treatment semantic features and radiomic features extracted from the GTV of a contrast-enhanced CT scan acquired prior to the beginning of treatment. The best classification results were obtained considering 6 features, of which 4 were semantic and 2 radiomic features. A comparable result was achieved considering only 3 radiomic features. A robustness analysis showed that 6 out of the 7 radiomic features that were used in the models presented good robustness across perturbed images. These models could be valuable to optimize the ART workflow and the clinic's resources. Future work is needed to validate the proposed models using an independent patient cohort preferably from a different institution, in order to incorporate them into clinical practice.

References

1. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R., Torre, L., Jemal, A.: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer J. Clini. **68**, 394–424 (2018)
2. Chin, D., Boyle, G., Porceddu, S., Theile, D., Parsons, P., Coman, W.: Head and neck cancer: past, present and future. Expert Rev. Anticancer Ther. **6**, 1111–1118 (2006)
3. Castadot, P., Lee, J., Geets, X., Grégoire, V.: Adaptive radiotherapy of head and neck cancer. Semin. Radiat. Oncol. **20**, 84–93 (2010)
4. Brouwer, C., Steenbakkers, R., Langendijk, J., Sijtsema, N.: Identifying patients who may benefit from adaptive radiotherapy: does the literature on anatomic and dosimetric changes in head and neck organs at risk during radiotherapy provide information to help? Radiother. Oncol. **115**, 285–294 (2015)
5. Mnejja, W., et al.: Dosimetric impact on changes in target volumes during intensity-modulated radiotherapy for nasopharyngeal carcinoma. Rep. Pract. Oncol. Radiother. **25**, 41–45 (2020)
6. Yousuf, A., Qureshi, B., Hussain, A., Qadir, A., Abbasi, N.: SU-E-J-81: Adaptive Radiotherapy for IMRT Head & Neck Patient in AKUH. Med. Phys. **42**, 3282 (2015)
7. Barker, J., et al.: Quantification of volumetric and geometric changes occurring during fractionated radiotherapy for head-and-neck cancer using an integrated CT/linear accelerator system. Int. J. Radiat. Oncol.* Biol.* Phys. **59**, 960–970 (2004)
8. Yan, D., et al.: Computed tomography guided management of interfractional patient variation. Semin. Radiat. Oncol. **15**, 168–179 (2005)
9. Böck, M.: On adaptation cost and tractability in robust adaptive radiation therapy optimization. Med. Phys. **47**, 2791–2804 (2020)

10. Green, O., Henke, L., Hugo, G.: Practical clinical workflows for online and offline adaptive radiation therapy. Semin. Radiat. Oncol. **29**, 219–227 (2019)
11. Schwartz, D., et al.: Adaptive radiotherapy for head and neck cancer—dosimetric results from a prospective clinical trial. Radiother. Oncol. **106**, 80–84 (2013)
12. Zhao, L., Wan, Q., Zhou, Y., Deng, X., Xie, C., Wu, S.: The role of replanning in fractionated intensity modulated radiotherapy for nasopharyngeal carcinoma. Radiother. Oncol. **98**, 23–27 (2011)
13. Zhang, P., et al.: Optimal adaptive IMRT strategy to spare the parotid glands in oropharyngeal cancer. Radiother. Oncol. **120**, 41–47 (2016)
14. Yang, H., Hu, W., Wang, W., Chen, P., Ding, W., Luo, W.: Replanning During Intensity Modulated Radiation Therapy Improved Quality of Life in Patients with Nasopharyngeal Carcinoma. Int. J. Radiat. Oncol. *Biol.*Phys. **85**, e47-e54 (2013)
15. Lindsay, P., et al.: SU-GG-T-50: dosimetric impact of anatomy variations and benefits of mid-course replanning for head and neck IMRT. Med. Phys. **35**, 2737 (2008)
16. Poon, E., Shenouda, G., Parker, W.: SU-E-J-209: dosimetric benefits of replanning for IMRT treatment of head and neck cancer. Med. Phys. **40**, 199 (2013)
17. Shang, Q., et al.: SU-E-J-74: Dosimetric advantages of adaptive radiotherapy for head and neck cancer are confirmed with weekly CBCT images. Med. Phys. **42**, 3281 (2015)
18. Capelle, L., Mackenzie, M., Field, C., Parliament, M., Ghosh, S., Scrimger, R.: Adaptive radiotherapy using helical tomotherapy for head and neck cancer in definitive and postoperative settings: initial results. Clin. Oncol. **24**, 208–215 (2012)
19. Ahn, P., et al.: Adaptive planning in intensity-modulated radiation therapy for head and neck cancers: single-institution experience and clinical implications. Int. J. Radiat. Oncol.*Biol.*Phys. **80**, 677–685 (2011)
20. Hansen, E., Bucci, M., Quivey, J., Weinberg, V., Xia, P.: Repeat CT imaging and replanning during the course of IMRT for head-and-neck cancer. Int. J. Radiat. Oncol.*Biol.*Phys. **64**, 355–362 (2006)
21. Brown, E., et al.: Predicting the need for adaptive radiotherapy in head and neck cancer. Radiother. Oncol. **116**, 57–63 (2015)
22. Surucu, M., et al.: Decision trees predicting tumor shrinkage for head and neck cancer. Technol. Cancer Res. Treat. **15**, 139–145 (2015)
23. Brouwer, C., et al.: Selection of head and neck cancer patients for adaptive radiotherapy to decrease xerostomia. Radiother. Oncol. **120**, 36–40 (2016)
24. Lambin, P., et al.: Radiomics: Extracting more information from medical images using advanced feature analysis. Eur. J. Cancer **48**, 441–446 (2012)
25. Gillies, R., Kinahan, P., Hricak, H.: Radiomics: images are more than pictures they are data. Radiology **278**, 563–577 (2016)
26. Kumar, V., et al.: Radiomics: the process and the challenges. Magn. Reson. Imaging **30**, 1234–1248 (2012)
27. Song, J., et al.: TU-AB-BRA-10: prognostic value of intra-radiation treatment FDG-PET and CT imaging features in locally advanced head and neck cancer. Med. Phys. **42**, 3588–3589 (2015)
28. Bagher-Ebadian, H., et al.: Application of radiomics for the prediction of HPV status for patients with head and neck cancers. Med. Phys. **47**, 563–575 (2020)
29. Oh, J., et al.: WE-E-17A-03: FDG-PET-based radiomics to predict local control and survival following radiotherapy. Med. Phys. **41**, 507–508 (2014)
30. Zwanenburg, A., et al.: The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology **295**, 328–338 (2020)
31. Ashrafinia, S.: Quantitative Nuclear Medicine Imaging using Advanced Image Reconstruction and Radiomics, Ph.D. Dissertation, Johns Hopkins University (2019)

32. Leger, S., et al.: CT imaging during treatment improves radiomic models for patients with locally advanced head and neck cancer. Radiother. Oncol. **130**, 10–17 (2019)
33. Zwanenburg, A., et al.: Assessing robustness of radiomic features by image perturbation. Sci. Rep. **9**(1) (2019)
34. Shrout, P.E., Fleiss, J.L.: Intraclass correlations: uses in assessing rater reliability. Psychol Bull. **86**(2), 420–428 (1979). https://doi.org/10.1037//0033-2909.86.2.420
35. Ramella, S., et al.: A radiomic approach for adaptive radiotherapy in non-small cell lung cancer patients. PLOS ONE **13**(11), p. e0207455 (2018)
36. Yu, T., et al.: Pretreatment prediction of adaptive radiation therapy eligibility using MRI-based radiomics for advanced nasopharyngeal carcinoma patients. Front. Oncol. **9** (2019)
37. Sanguineti, G., Ricchetti, F., Thomas, O., Wu, B., McNutt, T.: Pattern and predictors of volumetric change of parotid glands during intensity modulated radiotherapy. Br. J. Radiol. **86**(1031), 20130363 (2013)
38. Traverso, A., Wee, L., Dekker, A. and Gillies, R.: Repeatability and reproducibility of radiomic features: a systematic review. Int. J. Radiat. Oncol.*Biol.*Phys. **102**(4), 1143–1158 (2018)
39. Yip, S., Aerts, H.: Applications and limitations of radiomics. Phys. Med. Biol. **61**(13), R150–R166 (2016)
40. Rai, R., et al.: Multicenter evaluation of MRI-based radiomic features: a phantom study. Med. Phys. **47**(7), 3054–3063 (2020)