

OPEN

# Delta-radiomics features during radiotherapy improve the prediction of late xerostomia

Lisanne V. van Dijk<sup>1</sup>, Johannes A. Langendijk<sup>1</sup>, Tian-Tian Zhai<sup>1,2</sup>, Thea A. Vedelaar<sup>1</sup>, Walter Noordzij<sup>3</sup>, Roel J. H. M. Steenbakkers<sup>1</sup> & Nanna M. Sijtsema<sup>1</sup>

The response of the major salivary glands, the parotid glands, to radiation dose is patient-specific. This study was designed to investigate whether parotid gland changes seen in weekly CT during treatment, quantified by delta-radiomics features ( $\Delta$ features), could improve the prediction of moderate-to-severe xerostomia at 12 months after radiotherapy ( $Xer_{12m}$ ). Parotid gland  $\Delta$ features were extracted from in total 68 planning and 340 weekly CTs, representing geometric, intensity and texture characteristics. Bootstrapped forward variable selection was performed to identify the best predictors of  $Xer_{12m}$ . The predictive contribution of the resulting  $\Delta$ features to a pre-treatment reference model, based on contralateral parotid gland mean dose and baseline xerostomia scores ( $Xer_{baseline}$ ) only, was evaluated.  $Xer_{12m}$  was reported by 26 (38%) of the 68 patients included. The most predictive  $\Delta$ feature was the contralateral parotid gland surface change, which was significantly associated with  $Xer_{12m}$  for all weeks ( $p < 0.04$ ), but performed best for week 3 ( $\Delta PG-surface_{w3}$ ;  $p < 0.001$ ). Moreover,  $\Delta PG-surface_{w3}$  showed a significant predictive contribution in addition to the pre-treatment reference model (likelihood-ratio test;  $p = 0.003$ ), resulting in a significantly better model performance ( $AUC_{train} = 0.92$ ;  $AUC_{test} = 0.93$ ) compared to that of the pre-treatment model ( $AUC_{train} = 0.82$ ;  $AUC_{test} = 0.82$ ). These results suggest that mid-treatment parotid gland changes substantially improve the prediction of late radiation-induced xerostomia.

Xerostomia is one of the most frequently reported side effects following radiotherapy of head and neck cancer (HNC) patients and affects patient-reported quality of life<sup>1</sup>. For the prediction of late xerostomia, Normal Tissue Complication Probability (NTCP) models have been developed with pre-treatment dose-volume parameters and baseline complaints as most important predictors<sup>2,3</sup>. However, xerostomia NTCP models based on information during treatment are less explored. Since in-treatment parameters contain information on the patient-specific response to treatment, they may resolve some of the unexplained variability that remains for NTCP models that are based on pre-treatment variables only. These in-treatment parameters could therefore be used to improve the prediction of late xerostomia. Adequate prediction supported by in-treatment data may offer new opportunities to guide treatment adaptation aiming at a further reduction of late radiation-induced side effects.

Several studies have investigated changes of the parotid glands during and after treatment in CT images<sup>4-7</sup> and have shown a weak to moderate relationship between parotid gland dose and volume change<sup>4,6</sup>. However, knowledge of the relationship between parotid gland changes and patient-reported xerostomia is still limited. Therefore, in our previous study, we investigated the association between late patient-reported xerostomia and parotid gland changes quantified in radiomics features, or also called image biomarkers, extracted from CT images before and 6 weeks after treatment. That study showed that the parotid gland surface reduction ( $\Delta PG-surface_{6w-postRT}$ ) was strongly associated with the development of xerostomia at 6 and 12 months after radiotherapy<sup>8</sup>.

However, this post-treatment model does not allow for treatment adaptation, as the total prescribed radiation dose has already been administered. Hence, the next step is to investigate parotid gland changes during treatment.

The aim of the current study was to identify quantitative parotid gland changes during treatment that predict the development of late xerostomia. These parotid gland changes were extracted from pre-treatment and weekly

<sup>1</sup>Department of Radiation Oncology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands. <sup>2</sup>Department of Radiation Oncology, Cancer Hospital of Shantou University Medical College, Shantou, China. <sup>3</sup>Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands. Correspondence and requests for materials should be addressed to L.V.D. (email: [l.van.dijk@umcg.nl](mailto:l.van.dijk@umcg.nl))

			Pre-treatment reference model	$\Delta$ feature model 1	$\Delta$ feature model 2
			Xer <sub>baseline</sub>	Xer <sub>baseline</sub>	Xer <sub>baseline</sub>
			PG dose		PG dose
				$\Delta$ PG-Surface <sub>w3</sub>	$\Delta$ PG-Surface <sub>w3</sub>
Training set 56 patients	Apparent	Nagelkerke R2	0.46	0.53	0.59
		Area Under the Curve (AUC)	0.83 (0.70–0.96)	0.88 (0.79–0.97)	0.92 (0.85–0.99)
		Discrimination slope	0.39	0.45	0.48
	Internal validation	Nagelkerke R2	0.41*	0.45	0.51
		AUC	0.82*	0.84	0.89
Test set 14 patients	Validation	Nagelkerke R2	0.36	0.39	0.49
		Area Under the Curve (AUC)	0.80	0.85	0.93

**Table 1.** Performance of NTCP models predicting Xer<sub>12m</sub> with and without  $\Delta$ image biomarkers in the training and the test set. \*No variable selection was performed for internal validation of the reference model.

CT-images during radiotherapy, from which delta radiomics features ( $\Delta$ features) were quantified, representing differences in intensity, texture and geometric characteristics of the parotid glands.

## Results

**Patients.** Moderate-to-severe xerostomia was reported by 26 (38%) of all 68 patients included at 12 months after radiotherapy (22 in the training set and 4 in the test set). At 6 months after radiotherapy, the moderate-to-severe xerostomia reporting rate was 46 (52%) out of a total of 88 patients.

**$\Delta$ features selection.** For the geometric features, a change of contralateral parotid gland surface ( $\Delta$ PG-surface) was the most frequently selected  $\Delta$ feature for all weeks predicting Xer<sub>12m</sub> (see Supplementary Data 4 for frequency plot), except for week 2 where the  $\Delta$ PG-‘bounding box volume’ frequency was slightly higher. The  $\Delta$ PG-surface frequency was especially high in week 3 (725 times selected in the 1000 bootstrap samples). In other weeks, the subsequently selected  $\Delta$ features,  $\Delta$ ‘volume’,  $\Delta$ ‘bounding box volume’ and  $\Delta$ ‘volume times mean intensity’, were highly correlated with  $\Delta$ PG-surface for all weeks with  $\rho = 0.79$ – $0.93$ ,  $\rho = 0.66$ – $0.79$  and  $\rho = 0.82$ – $0.91$ , respectively.

For the intensity and texture  $\Delta$ features, no clear selection of  $\Delta$ features that were most frequently selected for all weeks could be made (see Supplementary Data 4 for frequency plot). Overall, the most frequently selected  $\Delta$ features on average were: coarseness from the neighbourhood grey tone difference matrix (coarseness), kurtosis and the median intensity (median). The  $\Delta$ features kurtosis and median were weakly correlated to coarseness ( $\rho = 0.13$ ,  $0.18$ ), but were stronger correlated to each other ( $\rho = 0.79$ ).

A heatmap of all  $\Delta$ IBMs and histograms of the most frequently selected  $\Delta$ IBMs are depicted in Supplementary Data 5 to illustrate the distribution and relation between the  $\Delta$ IBMs.

**$\Delta$ feature: univariable analysis.** The univariable analysis also showed that  $\Delta$ PG-surface was the most significant  $\Delta$ feature. This geometric  $\Delta$ feature was significantly associated with Xer<sub>12m</sub> at all weeks ( $p < 0.04$ ) but was most significant in week 3 ( $p < 0.001$ ). This week showed the largest regression coefficient of all weeks (Supplementary Data 6).

For the intensity and texture  $\Delta$ features, none of the most frequently selected  $\Delta$ features were significantly associated with Xer<sub>12m</sub> in any of the weeks (coarseness:  $p \geq 0.06$ ; kurtosis  $\geq 0.08$ ; median:  $p \geq 0.07$ ) (Supplementary Data 6).

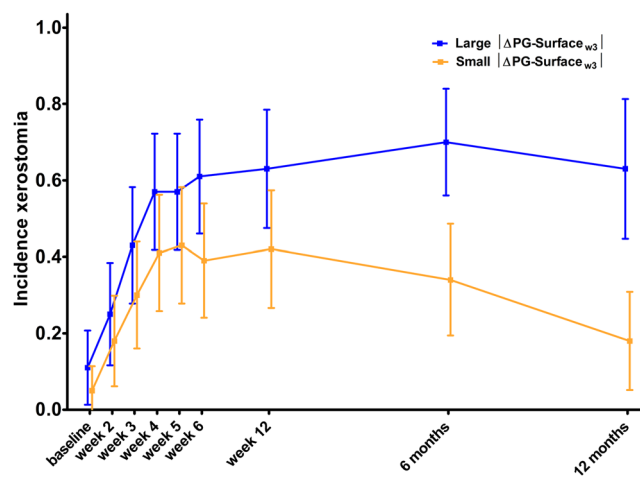
**$\Delta$ features, dose and toxicity: multivariable analysis.** Since the  $\Delta$ features showed the best performance in week 3, the multivariable analysis was performed with the selected week 3  $\Delta$ features only.

In the training set (56 patients), discrimination of the reference ‘pre-treatment’ model (Xer<sub>baseline</sub> and PGdose) was good (AUC<sub>train</sub> = 0.83 (AUC<sub>internal.val.</sub> = 0.82)), yet the geometric  $\Delta$ feature model with  $\Delta$ PG-surface<sub>w3</sub> and Xer<sub>baseline</sub> performed better ( $\Delta$ feature model 1: AUC<sub>train</sub> = 0.88 (AUC<sub>internal.val.</sub> = 0.84)) in predicting Xer<sub>12m</sub> (Table 1). Moreover, the addition of  $\Delta$ PG-surface<sub>w3</sub> to the pre-treatment model (likelihood-ratio test,  $p = 0.003$ ), significantly improved different aspects of performance ( $\Delta$ feature model 2: AUC<sub>train</sub> = 0.92 (AUC<sub>internal.val.</sub> = 0.89); Table 1). Validated in the test set (14 patients), the models showed stable performance (Table 1). Furthermore, also in the test set  $\Delta$ feature model 2 showed the highest performance (AUC<sub>test</sub> = 0.93 and  $R^2 = 0.49$ ). Ultimately, Table 2 gives the model coefficients that were optimized for the entire cohort, with the coefficients corrected for optimism using bootstrapping. The performance measures of the final model are shown in Supplementary Data 7.

Acute xerostomia scores at week 3 (Xer<sub>w3</sub>) significantly improved  $\Delta$ feature model 2 (Xer<sub>baseline</sub>, PGdose and  $\Delta$ PG-surface<sub>w3</sub>) (likelihood-ratio test,  $p = 0.01$ ), but the improvement in performance was relatively small (AUC<sub>train</sub> = 0.93 (AUC<sub>internal.val.</sub> = 0.89)) and decreased when validated in the test set (AUC<sub>test</sub> = 0.88). The relationship between  $\Delta$ PG-surface<sub>w3</sub> and Xer<sub>w3</sub> was not significant ( $p = 0.14$ ). This is also demonstrated in Fig. 1, where patients with a large and small surface reduction at week 3 (median  $\Delta$ PG-surface<sub>w3</sub> =  $-2.73$ ) showed a clear differentiation of actual moderate-to-severe xerostomia incidences at 6 or 12 months after treatment, but not for acute time points.

Model		$\beta$		odds ratio (95% CI)	p-value
		Uncorrected	Corrected		
Pre-treatment reference model	intercept	-3.794	-3.385*		
	Xer <sub>baseline</sub>	2.531	2.280*	12.56 (3.39–46.54)	<0.001
	Parotid gland dose	0.099	0.089*	1.1 (1.03–1.18)	0.005
$\Delta$ feature model 1	intercept	-3.139	-2.515		
	Xer <sub>baseline</sub>	2.533	2.074	12.59 (3.13–50.73)	<0.001
	$\Delta$ PG-surface <sub>w3</sub> (cm <sup>2</sup> )	-0.568	-0.465	0.57 (0.41–0.79)	0.001
$\Delta$ feature model 2	intercept	-4.515	-3.305		
	Xer <sub>baseline</sub>	2.591	1.936	13.35 (3.13–56.95)	<0.001
	Parotid gland dose	0.072	0.054	1.07 (0.77–1.51)	0.074
	$\Delta$ PG-surface <sub>w3</sub> (cm <sup>2</sup> )	-0.481	-0.360	0.62 (0.57–0.67)	0.005

**Table 2.** Estimated coefficients (uncorrected and corrected for optimism) of pre-treatment and  $\Delta$ image biomarkers models fitted to the entire dataset. \*No variable selection was performed for internal validation of the reference model.



**Figure 1.** Actual moderate-to-severe xerostomia incidence and 95% confident intervals at baseline, weekly during, and 6 weeks (week 12), 6 months, and 12 months after treatment for patients, with parotid gland surface reduction at week 3 ( $\Delta$ PG-Surface<sub>w3</sub>) larger (blue) or smaller (yellow) than the median reduction ( $|\text{median}| = 2.73$ ).

No significant relationship was found between  $\Delta$ PG-surface and Xer<sub>baseline</sub> ( $p = 0.17$ ). Xer<sub>w3</sub> was significantly associated with both PGdose ( $p = 0.04$ ) and Xer<sub>baseline</sub> ( $p = 0.03$ ), probably explaining the marginal prediction improvement of Xer<sub>3w</sub> to the  $\Delta$ feature model with PGdose and Xer<sub>baseline</sub>.

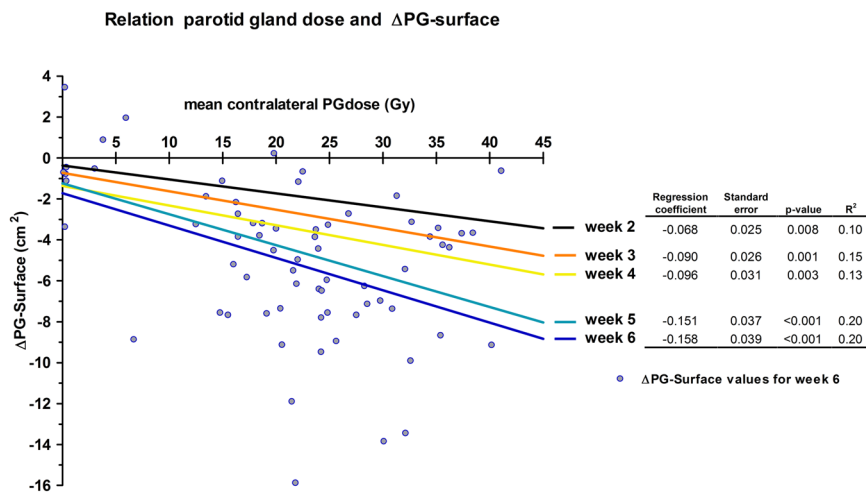
For the secondary endpoint Xer<sub>6m</sub>,  $\Delta$ PG-surface<sub>w3</sub> also added significantly to the pre-treatment model in predicting Xer<sub>6m</sub> (likelihood-ratio test,  $p = 0.02$ ). See Supplementary Data 8 for more details. None of the frequently selected intensity or texture  $\Delta$ features showed any significant improvement either compared to or in addition to the pre-treatment model (likelihood-ratio test,  $p > 0.27$ ) in predicting Xer<sub>12m</sub> or Xer<sub>6m</sub>.

**Parotid gland dose and  $\Delta$ features.** The linear relationship of contralateral parotid gland mean dose and  $\Delta$ PG-surface was significant for all weeks ( $p < 0.008$ ; Fig. 2). Depicted in Fig. 2, the regression coefficients of this linear relationship effectively increased over time, as did the coefficient of determination, but remained weak.

The selected intensity and the texture  $\Delta$ features,  $\Delta$ median and  $\Delta$ coarseness were significantly correlated ( $p < 0.05$ ) to parotid gland dose for week 2, 3, 5, 6 and 5, 6, respectively (Supplementary Data 9). However, the coefficient of determination was relatively low ( $R^2 = 0.00$ – $0.21$ ).  $\Delta$ LZLGE was not significant for any week.

## Discussion

The current study shows that surface change of the contralateral parotid gland ( $\Delta$ PG-surface) assessed during the course of radiotherapy was strongly associated with the development of late xerostomia (Xer<sub>12m</sub> and Xer<sub>6m</sub>). The association of this geometric  $\Delta$ feature was statistically significant during the entire course of treatment but performed best for changes obtained between treatment planning and week 3. This time point is still clinically relevant, as any treatment adaptations could still influence the patient's toxicity outcome.  $\Delta$ PG-surface<sub>w3</sub> did not only show improved predictive performance over PGdose, but it also improved the pre-treatment model performance significantly. The resulting model that was based on Xer<sub>baseline</sub>, PGdose and  $\Delta$ PG-surface<sub>w3</sub> showed



**Figure 2.** Univariable linear regression of contralateral parotid gland mean dose (PGdose) predicting parotid gland surface reduction ( $\Delta$ PG-surface) for different weeks (lines) and regression characteristics (Table). Correlation increases over time, but remained weak. Data point represent  $\Delta$ PG-surface values for week 6.

excellent performance when predicting  $Xer_{12m}$  in both the train (AUC = 0.92), test (AUC = 0.93) and when fitted to the entire cohort (AUC = 0.91). However, these results should be confirmed by external validation in larger patient cohorts.

Castelli *et al.* showed that parotid gland dose could significantly be reduced with an adaptive radiotherapy approach (ART)<sup>9</sup>. By re-planning the dose distribution on weekly CTs, an average NTCP reduction of 11% (maximum 30%) was observed. However, weekly re-planning is time consuming. This highlights the potential of the  $\Delta$ feature NTCP model with  $\Delta$ PG-surface<sub>w3</sub>, since it could select patients during treatment that have a high risk of developing xerostomia. If these high-risk patients could, subsequently, receive less PGdose by re-planning, their risk of xerostomia could be further reduced. Alternatively, the model-based approach has been introduced to select patients for proton therapy. Patients can be selected that have a clinically relevant NTCP-reduction with a proton plan compared to their photon based treatment plan<sup>10</sup>. Proton therapy has the potential to better conform the dose to the tumour while sparing the surrounding normal tissue, due to the intrinsic properties of protons<sup>11</sup>. By incorporating patient-specific  $\Delta$ feature response information in the pre-treatment reference model, patients that do not initially qualify could be reclassified for proton therapy. Accordingly, treatment can be changed from photon to proton therapy, when relevant differences are seen in the new  $\Delta$ NTCP values.

In a previous study, the geometric radiomics feature differences were calculated between 6 weeks post-treatment and prior to treatment ( $\Delta$ feature<sub>6week-postRT</sub>)<sup>8</sup>. The association of  $\Delta$ feature<sub>6week-postRT</sub> with  $Xer_{12m}$  was investigated in a patient cohort (n = 107) independent of the current cohort. Interestingly, the most stable and predictive post-treatment  $\Delta$ feature was also the contralateral  $\Delta$ PG-surface. Similar to the results of the current study, inclusion of  $\Delta$ PG-surface substantially improved the pre-treatment model. Using the same coefficients of the post-treatment model with  $\Delta$ PG-surface<sub>6w-postRT</sub>,  $Xer_{baseline}$  and PGdose in the current cohort, also showed a comparable performance (AUC = 0.89) to that of the model trained in the current cohort (AUC = 0.91). The other way around, using the coefficients of the current model in the previous cohort also resulted in a comparable improvement in performance for the model in the post-treatment cohort (Supplementary Data 10). This suggests that the  $\Delta$ PG-surface<sub>w3</sub> model would also perform well when externally validated in a cohort where  $\Delta$ PG-surface is acquired at week 3. In both studies,  $\Delta$ PG-volume was highly correlated to  $\Delta$ PG-surface, and also performed well in predicting late xerostomia.

In line with other studies that observed a relationship between PGdose and PG shrinkage<sup>4,6,12,13</sup>, linear regression in the current study also showed that there was a weak to moderate correlation between  $\Delta$ PG-surface and PGdose. Interestingly, the correlation between PGdose and  $\Delta$ PG-surface effectively increased over the time of treatment, illustrated by the increasing values of the regression coefficients and R<sup>2</sup> every consecutive week. This suggests that the effect of planning PGdose on  $\Delta$ PG-surface becomes clearer as more dose is administered. However, such an effect was not seen for the association between  $\Delta$ PG-surface and  $Xer_{12m}$ , since the univariable logistic regression coefficient and the performance of  $\Delta$ PG-surface increased from week 2 to 3, but decreased for the subsequent weeks. Hence, we concluded that the best moment for predicting  $Xer_{12m}$  was during week 3. The explanation may be that most parotid glands shrink when irradiated, as reported in previous studies<sup>4-7</sup>, but patients that have a parotid gland that shrinks early in treatment have a higher risk of developing late xerostomia. Therefore,  $\Delta$ PG-surface<sub>week3</sub> could be a marker to differentiate between patients that develop permanent damage of the parotid gland versus those that can recover.

In addition to these observations,  $\Delta$ PG-surface was not associated with acute xerostomia, although it was strongly associated with the development of late xerostomia. Figure 1 also demonstrated this, as  $\Delta$ PG-surface<sub>w3</sub> did not show a clear differentiation between the actual incidences of moderate-to-severe xerostomia at week 3 or any of the other acute time points. In contrast, this differentiation can be clearly seen for 6 and 12 months after radiotherapy. Furthermore, acute xerostomia scores at 3 weeks ( $Xer_{w3}$ ) did significantly add to the model with

$Xer_{baseline}$ ,  $\Delta PG$ -surface and PGdose, although the improvement in performance measures was small. This is probably due to the correlation between  $Xer_{w3}$  and both PGdose and  $Xer_{baseline}$ . Further research needs to be performed on larger datasets in order to investigate whether acute toxicities can contribute to  $\Delta$ feature models.

Changes in intensity or texture features were not related to the development of xerostomia. In contrast, many of these  $\Delta$ features were significantly related to PGdose, even though no relationship was seen with the development of xerostomia. Furthermore, detailed investigation of the most frequently selected intensity or texture  $\Delta$ features showed that these  $\Delta$ features contained one or two outliers that determined the effect. The influence of outliers indicates the importance of evaluating the selected features before presenting them in a final model. In this study, the analysis of  $\Delta$ features was used rather than the features directly extracted per week. The results of these absolute weekly features were not significant. In contrast, in a previous study, a pre-treatment CT feature that indicates tissue heterogeneity, was significantly associated with the development of late xerostomia<sup>14</sup>. It might be that the effect of pre-treatment is too weak to be observed in this relatively small dataset. In addition, using proportional  $\Delta$ features instead of absolute difference  $\Delta$ features did not improve the results of this study either.

The limitations of this study are the low numbers of patients included in this analysis and no direct external validation was performed. Nevertheless, we have validated our results by splitting our dataset in a training set, for which variables were selected, and tested them in a small unseen cohort. The models performed well in both the training as the test cohort. Finally the models were fitted to the entire dataset, to make full use of the total number of patients in estimating the most optimal coefficients given all data. Furthermore, only pre-treatment PGdose rather than the accumulated dose over all weekly CT scans was evaluated, since this was outside the scope of the paper. Brouwer *et al.* showed that accumulated dose calculated on weekly CTs was almost equal to the pre-treatment PGdose<sup>15</sup>. Using accumulated PGdose could improve the predictive performance of PGdose. Additionally, other modalities, such as positron emission tomography and magnetic resonance imaging could potentially provide better information during treatment on function loss of the PG gland. Future studies using these modalities could improve the quantification and understanding of the development of late xerostomia.

In conclusion, contralateral parotid gland surface area reduction during the course of radiotherapy ( $\Delta PG$ -surface) was associated with the development of late xerostomia both at 6 and 12 months after radiotherapy. The model consisting of  $Xer_{baseline}$ , parotid gland dose and  $\Delta PG$ -surface, as assessed at week 3 during treatment ( $\Delta PG$ -surface<sub>w3</sub>), showed the best performance, and substantially improved the pre-treatment model based on parotid gland dose and  $Xer_{baseline}$  only (from AUC of 0.83 to 0.91). This mid-treatment model may be a good candidate to identify patients most at risk of developing late xerostomia and who may benefit from treatment adaptations, but external validation is warranted.

## Method

**Patients and image acquisition.** The study cohort included consecutive HNC patients that were treated with definitive radiotherapy and received weekly CTs between January 2014 and December 2016. The cohort, sorted by treatment start date, was split in a training set (80%; 54 patients) and test set (20%; 14 patients). Radiation plans were adapted where necessary due to anatomical changes causing reduced target coverage. Patients were treated with IMRT or VMAT using a simultaneous integrated boost (SIB) technique, either as a single modality or in combination with concurrent chemotherapy or cetuximab. Plans were optimised to spare the parotid glands and swallowing organs at risk (superior pharyngeal constrictor muscle and supraglottic area) as much as possible without compromising the dose to the target volumes<sup>16</sup>. The primary tumour and pathologic lymph nodes were generally prescribed 70 Gy (2 Gy per fraction) and the cervical lymph node levels were prescribed an elective radiation dose of 54.25 Gy (1.55 Gy per fraction)<sup>17</sup>. More detailed descriptions of the radiation protocols used was reported in previous work<sup>16</sup>. Patient characteristics are listed in Table 3.

Patients were excluded if they had salivary gland tumours, underwent prior surgery and/or underwent re-irradiation. An additional requirement was that patient-rated follow-up information at 6 and/or 12 months after radiotherapy was available.

CT scans (Somatom Sensation Open, Siemens, Forchheim, Germany; voxel size:  $0.94 \times 0.94 \times 2.0$  mm<sup>3</sup>; 100–140 kV) were acquired within 2 weeks prior to treatment ( $CT_0$ ) and weekly during treatment ( $CT_{w1-6}$ ), where  $CT_{w1}$  was generally acquired on the day of the first or second fraction. Patients only received intravenous contrast for  $CT_0$ . All scans were acquired with a thermoplastic mask in their radiotherapy treatment position.

All patients provided written informed consent before starting therapy that their data could be used within the department's research program. The Dutch Medical Research Involving Human Subjects Act is not applicable to data collection as part of routine clinical practice and therefore, the hospital ethics committee granted us a waiver from needing ethical approval for the conduct of studies based on these data. All patients received standard clinical care of adaptive radiotherapy.

**Endpoints.** Patient-rated xerostomia scores were collected prospectively on a routine basis; before, weekly during, and subsequently 6 and 12 months after radiotherapy using the EORTC QLQ-H&N35 questionnaire, as part of the standard follow-up programme (SFP) (NCT02435576)<sup>2,18</sup>. The primary endpoint of this study was moderate-to-severe patient-rated xerostomia at 12 months after radiotherapy ( $Xer_{12m}$ ) and the secondary endpoint was moderate-to-severe patient-rated xerostomia at 6 months ( $Xer_{6m}$ ). This corresponds to the 2 highest scores of the 4-point Likert scale (not, a bit, quite a bit, a lot).

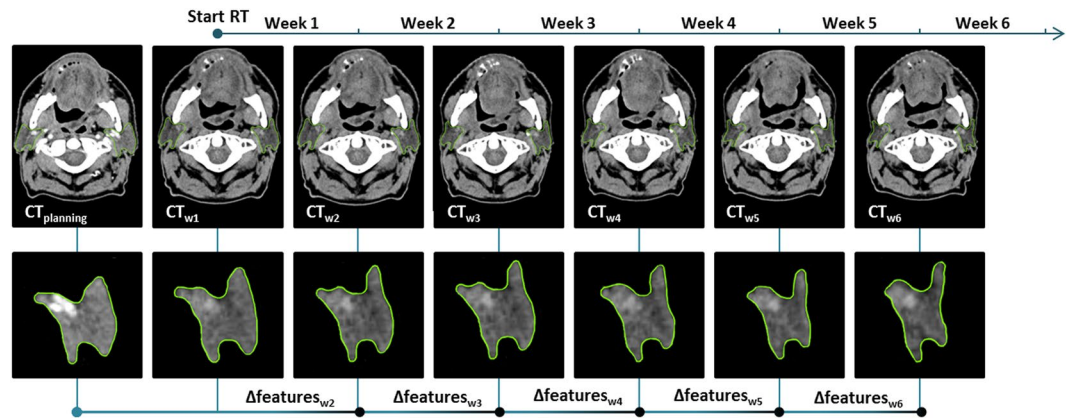
**$\Delta$ features definitions.** Parotid glands were delineated on the  $CT_0$  and  $CT_1$  according to the consensus guidelines of Brouwer *et al.*<sup>19</sup>. Delineations were warped from  $CT_1$  to the weekly CTs using the deformable image registration tool in the treatment planning system RayStation v5.99 (RaySearch Laboratories, Stockholm, Sweden), the warped structures were carefully checked and manually corrected where necessary.

Characteristics	Follow-up info at 12 months		Follow-up info at 6 months	
	N = 68	%	N = 88	%
<b>Sex</b>				
Female	20	29	26	30
Male	48	71	62	70
<b>Age</b>				
18–65	48	71	62	70
>65	20	29	26	30
<b>Tumour site</b>				
Oropharynx	22	32	27	31
Hypopharynx	0	0	1	1
Nasopharynx	5	7	5	6
Larynx	22	32	27	31
Oral cavity	15	22	23	26
Unknown primary	1	1	1	1
Other	3	4	4	5
<b>Tumour classification</b>				
T0	1	1	1	1
T1	11	16	14	16
T2	14	21	19	22
T3	17	25	19	22
T4	23	34	33	38
Unknown	2	3	2	2
<b>Node classification</b>				
N0	23	34	28	33
N1	9	13	13	15
N2abc	31	46	41	47
N3	3	4	4	5
<b>Systemic treatment</b>				
Yes	34	50	47	53
No	34	50	41	47
<b>Treatment technique</b>				
IMRT	27	40	30	34
VMAT	41	60	58	66
<b>Bilateral</b>				
Yes	57	84	72	82
no	11	16	16	18
<b>Baseline xerostomia</b>				
Any	26	38	36	41
None	42	62	52	59

**Table 3.** Patient characteristics of patients that had follow-up information available at 12 and 6 months after treatment.

The radiomics features were extracted from the planning and the weekly CTs with Matlab-based (Mathworks, Natick, MA, USA; version R2014a) in-house developed software. The definitions and formulas were in line with the ‘Image biomarker standardisation initiative’<sup>20</sup>. The geometric changes (geometric  $\Delta$ features) were calculated by subtracting the radiomics features of  $CT_{w2-6}$  from those of  $CT_0$ . This resulted in 15 geometric  $\Delta$ features per weekly CT, that for example represent volume, surface or compactness changes. Intensity features describe first order and histogram characteristics of CT intensities of a parotid gland (e.g. mean or variance). Textural features describe the intensity heterogeneity and were extracted from the grey level co-occurrence matrix (GLCM)<sup>21</sup>, grey level run-length matrix (GLRLM)<sup>22,23</sup>, grey level size-zone matrix (GLSZM)<sup>24</sup> and neighbourhood grey tone difference matrix (NGTDM)<sup>25</sup>. Contrast enhancement was only used for  $CT_0$ , and not for the weekly CT-scans. Since this can affect the intensity and texture feature values,  $CT_{w1}$  (generally acquired before the 2<sup>nd</sup> radiation fraction) was considered to be the baseline CT. Hence, intensity and texture changes were quantified by calculating the difference between  $CT_{w1}$  and the  $CT_{w2-6}$ . As the intensity and texture features can be influenced by metal artefacts, slices with metal artefacts were deleted and features were calculated on the remaining slices only.

Figure 3 depicts the calculation of the  $\Delta$ features and CT time points. For a complete list of the 15 geometric, 17 intensity and 66 texture features refer to the Supplementary Datas 1–3, respectively. Only  $\Delta$ features of the contralateral parotid gland were reported, as they performed better than those of the ipsilateral parotid gland. The geometric  $\Delta$ features were analysed separately from intensity and texture  $\Delta$ features.



**Figure 3.** The difference between features extracted from the weekly CTs ( $CT_w$ ) and either the planning (geometric features) or week 1 CT (intensity and texture features) resulted in  $\Delta$ features per week. Generally, scans were taken at the start of consecutive radiation week.

Clinical variables age and gender were investigated, but not reported as they did not show, to have a significant relation with the development of late xerostomia, which is in line with previous analyses<sup>2,26,27</sup>.

**$\Delta$ feature selection.** To identify the most predictive  $\Delta$ features,  $\Delta$ feature variable selection was performed each week. Firstly,  $\Delta$ features values were normalised by taking the difference between each value and the average across patients, and then dividing by the standard deviation. Secondly, a pre-selection that was based on inter-variable correlation was performed to reduce the number of variables. If the (Pearson) correlation of two variables was larger than 0.80, only the  $\Delta$ feature with the highest univariable association with the endpoint was selected. Thirdly, stepwise forward selection was used to select the most important predictors (likelihood-ratio test:  $p$ -value  $< 0.01$ )<sup>28</sup>. The entire variable selection procedure (normalisation, pre-selection and forward selection) was repeated on 1000 bootstrapped samples (i.e. with replacement), according to the TRIPOD guidelines<sup>29</sup>. The variable selection frequencies were evaluated to identify the most stable predictive variables per week. The Pearson correlation between the selected  $\Delta$ features was also investigated.

**$\Delta$ feature: univariable analysis.** In order to identify the optimum week for predicting  $Xer_{12m}$  with  $\Delta$ features, the univariable associations were investigated for the selected  $\Delta$ features per week in the entire cohort.

**$\Delta$ feature, dose and toxicity: multivariable analysis.** A reference ‘pre-treatment model’ that was based on baseline xerostomia scores ( $Xer_{baseline}$ ; none vs. any) and the contralateral PGdose, was fitted to the current train set<sup>2</sup>. The prediction performance of the ‘pre-treatment model’ was first compared with that of models based on  $Xer_{baseline}$  and the selected  $\Delta$ features. Subsequently, the addition of the selected  $\Delta$ features to the ‘pre-treatment model’ was investigated in terms of significance (likelihood-ratio test) and performance.

The resulting multivariable logistic regression models were tested in the unseen data of the test set. Model discrimination was measured with the area under the receiver operating characteristic curve (AUC) and the discrimination slope. Nagelkerke  $R^2$  was used as a measure for explained variance. Ultimately, the final model coefficient estimations were performed on the entire cohort. Model calibration was tested for these final models with the Hosmer–Lemeshow test and by repeating the entire variable selection on 1000 bootstrap samples, and by calculating the average of all resulting linear predictor slopes and intercepts. The coefficients were corrected for optimism according to this internal validation procedure.

Since our previous study showed that acute xerostomia scores significantly improved the  $\Delta$ feature model 6 weeks after treatment<sup>8</sup>, we also investigated if the addition of acute toxicity as assessed during treatment to the  $\Delta$ feature-models improved model performance.

The relationships between the resulting  $\Delta$ feature predictors,  $Xer_{baseline}$ , PGdose and acute xerostomia scores, were additionally explored with univariable logistic regression.

**Parotid gland dose and  $\Delta$ features.** The relationship of the mean contralateral PGdose and the  $\Delta$ features was investigated with linear regression. Model performance was measured with the coefficient of determination ( $R^2$ ), and normality of the residuals of the regression models was checked.

## References

- Langendijk, J. A. *et al.* Impact of late treatment-related toxicity on quality of life among patients with head and neck cancer treated with radiotherapy. *J. Clin. Oncol.* **26**, 3770–6 (2008).
- Beetz, I. *et al.* NTCP models for patient-rated xerostomia and sticky saliva after treatment with intensity modulated radiotherapy for head and neck cancer: The role of dosimetric and clinical factors. *Radiother. Oncol.* **105**, 101–106 (2012).
- Houweling, A. C. *et al.* A comparison of dose-response models for the parotid gland in a large group of head-and-neck cancer patients. *Int. J. Radiat. Oncol. Biol. Phys.* **76**, 1259–65 (2010).

4. Wang, Z.-H. *et al.* Radiation-induced volume changes in parotid and submandibular glands in patients with head and neck cancer receiving postoperative radiotherapy: a longitudinal study. *Laryngoscope* **119**, 1966–1974 (2009).
5. Vásquez Osorio, E. M. *et al.* Local Anatomic Changes in Parotid and Submandibular Glands During Radiotherapy for Oropharynx Cancer and Correlation With Dose, Studied in Detail With Nonrigid Registration. *Int. J. Radiat. Oncol.* **70**, 875–882 (2008).
6. Reali, A. *et al.* Volumetric and positional changes of planning target volumes and organs at risk using computed tomography imaging during intensity-modulated radiation therapy for head–neck cancer: an “old” adaptive radiation therapy approach. *Radiol. Med.* **119**, 714–720 (2014).
7. Brouwer, C. L., Steenbakkers, R. J. H. M., Langendijk, J. A. & Sijtsema, N. M. 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).
8. van Dijk, L. V. *et al.* Geometric Image Biomarker Changes of the Parotid Gland Are Associated With Late Xerostomia. *Int. J. Radiat. Oncol. Biol. Phys.* <https://doi.org/10.1016/j.ijrobp.2017.08.003> (2017).
9. Castelli, J. *et al.* Impact of head and neck cancer adaptive radiotherapy to spare the parotid glands and decrease the risk of xerostomia. *Radiat. Oncol.* **10** (2015).
10. Langendijk, J. A. *et al.* Selection of patients for radiotherapy with protons aiming at reduction of side effects: the model-based approach. *Radiother. Oncol.* **107**, 267–73 (2013).
11. Lomax, A. Intensity modulation methods for proton radiotherapy. *Phys. Med. Biol.* **44**, 185–205 (1999).
12. Belli, M. L. *et al.* Early changes of parotid density and volume predict modifications at the end of therapy and intensity of acute xerostomia. *Strahlentherapie und Onkol.* **190**, 1001–1007 (2014).
13. Broggi, S. *et al.* A two-variable linear model of parotid shrinkage during IMRT for head and neck cancer. *Radiother. Oncol.* **94**, 206–212 (2010).
14. van Dijk, L. V. *et al.* CT image biomarkers to improve patient-specific prediction of radiation-induced xerostomia and sticky saliva. *Radiother. Oncol.* **122**, 185–191 (2017).
15. Brouwer, C. L. *et al.* Selection of head and neck cancer patients for adaptive radiotherapy to decrease xerostomia. *Radiother. Oncol.* 924–932, <https://doi.org/10.1016/j.radonc.2016.05.025> (2016).
16. Van Der Laan, H. P. *et al.* Swallowing-sparing intensity-modulated radiotherapy for head and neck cancer patients: Treatment planning optimization and clinical introduction. *Radiother. Oncol.* **107**, 282–287 (2013).
17. Grégoire, V. *et al.* CT-based delineation of lymph node levels and related CTVs in the node-negative neck: DAHANCA, EORTC, GORTEC, NCIC, RTOG consensus guidelines. *Radiother. Oncol.* **69**, 227–236 (2003).
18. Vergeer, M. R. *et al.* Intensity-modulated radiotherapy reduces radiation-induced morbidity and improves health-related quality of life: results of a nonrandomized prospective study using a standardized follow-up program. *Int. J. Radiat. Oncol. Biol. Phys.* **74**, 1–8 (2009).
19. Brouwer, C. L. *et al.* CT-based delineation of organs at risk in the head and neck region: DAHANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG Oncology and TROG consensus guidelines. *Radiother. Oncol.* **117**, 83–90 (2015).
20. Zwanenburg, A., Leger, S., Vallières, M. & Löck, S. *Image biomarker standardisation initiative - feature definitions*. *arXiv* **1612**, 07003 (2016).
21. Haralick, R., Shanmugan, K. & Dinstein, I. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics* **3**, 610–621 (1973).
22. Tang, X. Texture information in run-length matrices. *IEEE Trans. Image Process.* **7**, 1602–1609 (1998).
23. Galloway, M. M. Texture analysis using gray level run lengths. *Comput. Graph. Image Process.* **4**, 172–179 (1975).
24. Thibault, G. *et al.* Texture Indexes and Gray Level Size Zone Matrix Application to Cell Nuclei Classification. *Pattern Recognit. Inf. Process.* 140–145, <https://doi.org/10.1142/S0218001413570024> (2009).
25. Amadasun, M. & King, R. Textural features corresponding to textural properties. *IEEE Trans. Syst. Man Cybern.* **19**, 1264–1273 (1989).
26. Deasy, J. O. *et al.* Radiotherapy dose-volume effects on salivary gland function. *Int. J. Radiat. Oncol. Biol. Phys.* **76**, S58–63 (2010).
27. Eisbruch, A. *et al.* Xerostomia and its predictors following parotid-sparing irradiation of head-and-neck cancer. *Int. J. Radiat. Oncol. Biol. Phys.* **50**, 695–704 (2001).
28. Dehing-Oberije, C. *et al.* Development, external validation and clinical usefulness of a practical prediction model for radiation-induced dysphagia in lung cancer patients. *Radiother. Oncol.* **97**, 455–461 (2010).
29. Moons, K. G. M. *et al.* Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. *Ann. Intern. Med.* **162**, W1 (2015).

## Author Contributions

L.v.D., T.Z., J.L., N.S., R.S., T.V. were actively involved with the design of the project, analyses and writing the manuscript. T.V., T.Z. and L.v.D. were responsible for collecting, parotid gland warping, delineation and curation of the weekly CTs., L.v.D. developed radiomics feature extraction and modelling software. W.N. gave provided expert knowledge and all authors reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-48184-3>.

**Competing Interests:** The authors declare no competing interests.

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019