Original Article

Early prediction of acute xerostomia during radiation therapy for nasopharyngeal cancer based on delta radiomics from CT images

Yanxia Liu[1], Hongyu Shi[1], Sijuan Huang[2], Xiaochuan Chen[1], Huimin Zhou[2,3], Hui Chang[2], Yunfei Xia[2], Guohua Wang[1], Xin Yang[2]

[1]School of Software Engineering, South China University of Technology, Guangzhou 510006, China; [2]Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Guangzhou 510060, China; [3]Department of Oncology, the Seventy-fourth Group Army Hospital of the Chinese People's Liberation Army, Guangzhou 510318, China

Correspondence to: Guohua Wang. School of Software Engineering, South China University of Technology, Guangzhou 510006, China. Email: ghwang@scut.edu.cn; Xin Yang. Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Guangzhou 510060, China. Email: yangxin@sysucc.org.cn.

Background: Acute xerostomia is the most common side effect of radiation therapy (RT) for head and neck (H&N) malignancies. Investigating radiation-induced changes of computed tomography (CT) radiomics in parotid glands (PGs) and saliva amount (SA) can predict acute xerostomia during the RT for nasopharyngeal cancer (NPC).

Methods: CT and SA data from 35 patients with stages I–IVB were randomly collected from an NPC clinical trial registered on the clinicaltrials.gov (ID: NCT01762514). All patients received radical treatment based on intensity-modulated RT (IMRT) with a prescription dose of 68.1 Gy in 30 fractions. The patients' ages ranged 24–72 years, and each patient had five CT sets acquired at treatment position: at the 0^{th}, 10^{th}, 20^{th}, 30^{th} fractions during the RT, and at 3-month later after the RT. The PGs for each CT set were delineated by a radiation oncologist and verified independently by another. Patients' saliva was collected every other 10 days during the RT. Acute xerostomia was evaluated based on the RTOG acute toxicity scoring and the SA. In total, 1,703 radiomics features were calculated for PGs from each CT set, including feature value at 0^{th} fraction (FV_{0F}), FV_{10F}, and delta FV (ΔFV_{10F-0F}), respectively. Extensive experiments were conducted to achieve the optimal results. RidgeCV and Recursive Feature Elimination (RFE) were used for feature selection, while linear regression was used for predicting SA_{30F}. Four more patients were added for independent testing.

Results: Substantial changes in various radiomics metrics of PGs were observed during the RT. Eight normalized feature value (NFV), selected from NFV_{0F}, predicted SA_{10F} with a mean square error (MSE) of 0.9042 and a R^2 score of 0.7406. Fourteen NFV, selected from ΔNFV_{10F-0F}, NFV_{0F}, and NFV_{10F} to predict SA_{30F}, showed the best predictive ability with an MSE of 0.0569. The model predicted the level of acute xerostomia with a precision of 0.9220 and a sensitivity of 100%, compared to the clinical observed SA. For the independent test, the MSE of PSA_{30F} was 0.0233.

Conclusions: This study demonstrated that radiation-induced acute xerostomia level could be early predicted based on the SA and radiomics changes of the PGs during IMRT delivery. SA, NFV_{0F}, NFV_{10F}, and especially ΔNFV_{10F-0F} provided the best performance on acute xerostomia prediction for individual patient based on RidgeCV_RFE_LinearRegression method of delta radiomics.

Keywords: Acute xerostomia; delta radiomics; saliva amount prediction

Submitted Apr 30, 2019. Accepted for publication Jul 10, 2019.
doi: 10.21037/qims.2019.07.08
View this article at: http://dx.doi.org/10.21037/qims.2019.07.08

Introduction

Permanent xerostomia is one of the most common side effect of radiation therapy (RT) toxicity for head and neck (H&N) malignancies (1). It is considered as a significant cause of decreased quality of life (QOL) in patients (2,3), by introducing swallowing and chewing hypofunction (4), disturbing speech, taste, and even sleep patterns (5). Xerostomia grade (XG) is commonly scoring via five level scales from Grade 0 to 4 (G0-G4) according to the Radiation Therapy Oncology Group (RTOG) radiation morbidity scoring criteria (3). Recent literatures manifested employing computed tomography (CT) images to predict the dry mouth symptoms of patients with H&N cancer after receiving radiotherapy (6,7). There are two major concerns in such scoring system: (I) inadequate objectiveness and (II) insufficient quantitative description. Most studies measured and graded using subjective and qualitative observer-rated toxicity grading (3), rather than the measurements of major salivary gland output or saliva amount (SA) for description. The separation between the different grades is ambiguous to some extent in the current toxicity grading systems, which also induced the difficulty of non-subjective description.

Quantitatively and accurately predicting SA will facilitate clinical decision-making to prevent xerostomia, while radiomics can aid diagnosis by analyzing medical images (8). Radiomics uses automatic algorithm to extract a large number of feature information from the region of interest (ROI) of medical images as the research object, and further used diversified statistical analysis and data mining methods to extract and reveal the key information that really plays a role in the immense amount of information. Finally, these features were used in the auxiliary diagnosis, classification or grading of diseases (9,10). The specific steps can be divided into image collection, ROI segmentation, feature extraction, statistical analysis and classification prediction (11).

Accurate toxicity prediction could assist clinical decision-making for planning personalized treatment. Recently, some studies had reported the use of radiomics features extracted from medical images [CT (7,12,13), MR (14,15), PET (16,17) and CBCT (18,19)] to improve the prediction models of side effect treatment. The rate of CBCT-measured parotid glands (PGs) image feature changes improved the prediction concerning the dose alone for chronic xerostomia prediction (18). Analysis of CBCT images acquired for treatment positioning might provide an inexpensive monitoring system to support toxicity reducing adaptive RT. Xu *et al.* (20) recognized that the CT number changes in parotids were measurable during the delivery of fractionated radiotherapy for NPC. Feng *et al.* (21) revealed that the CT number could be reduced in tumor and PGs during the course of RT. There was a fair correlation between CT number reduction and radiation doses for a part of patients, whereas the correlation between CT number reductions and volume reductions in GTV and PGs was weak. Recently, Wu *et al.* (6) pointed out that the significant changes in the CT histogram features of the PGs were observed during RT. And a practical method that used the changes of mean CT number and volume of PGs was proposed to predict radiation-induced acute xerostomia in multiple institutions. This method might be helpful for designing adaptive treatment or personalized treatment, such as the submandibular glands transfer prior to RT (22) or the concomitant administration of pilocarpine during radiation increases unstimulated saliva flow rate and reduces clinician-rated XG (23).

We noticed that all the previous work graded the xerostomia either by patients themselves or by physicians, rather than an objective reported as an main end points (24). In this study, a new SA prediction system was proposed based primarily on the changes of radiomics from CT images for patients with nasopharyngeal cancer (NPC) during radiotherapy, combined with SA and associated with xerostomia. As far as to our knowledge, this is the very first paper to quantitatively predict patients' SA.

Methods

This article predicted the probability and extent of the patient's future acute xerostomia by predicting the prospective SA of patient. The SA prediction model consists of five parts as shown in *Figure 1*: (I) data generation; (II) data preprocessing; (III) feature score calculation; (IV) feature selection; and (V) SA prediction.

The data generation (*Figure 1A*) was described in *Patient data*. *Data preprocessing* elucidated the content of *Figure 1B*. *Xerostomia prediction model* explained in detail in three sub-sections: (I) the feature score calculation (*Figure 1C*); (II) feature selection (*Figure 1D*); and (III) SA prediction (*Figure 1E*).

CT simulator protocols

The CT simulator in our department, a Brilliance Big Bore CT scanner (Philips Medical Systems, Cleveland, OH, USA), had monthly/quarterly/yearly QA/QC based on the

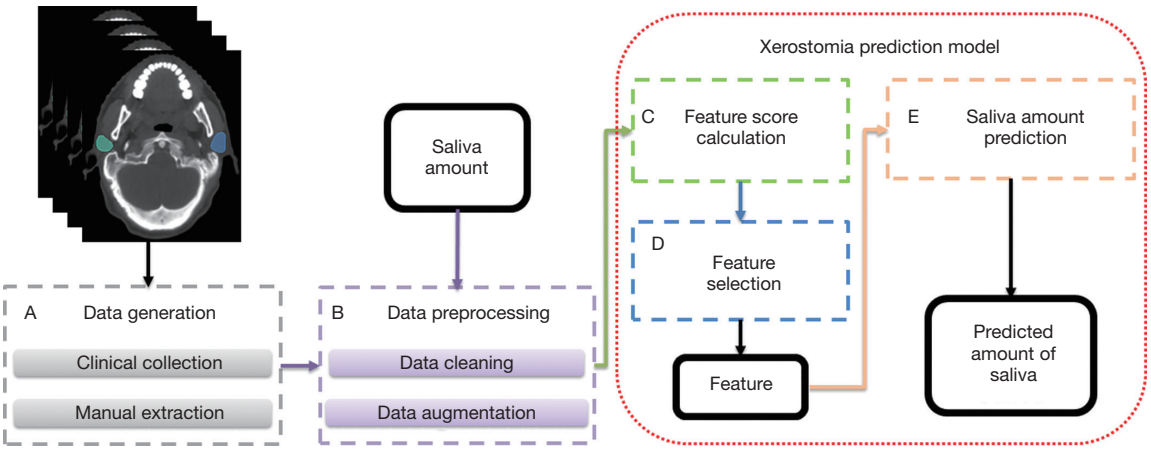

Figure 1 Saliva amount prediction model.

American College of Radiology (ACR) CT accreditation phantom before scanning the patients. And this CT simulator images showed that image quality was within ACR guidelines for all tested scanning protocols.

A pre-defined set of scan protocol was performed on NPC patients using the CT simulator and the image data were reconstructed with a 3 mm thickness. These images were generated by using the more common clinical protocol of 3 mm slices at 3 mm increments with the scanning protocols most commonly used in clinical practice. The detailed parameters for these protocols were given as following: voltage 120 kVp, exposure 300 mAs, slice thickness 3 mm, increment 3 mm, collimation 16 mm × 0.75 mm, display FOV 600 mm, scan FOV 600 mm, reconstruction filter type UB/B, and pitch 0.567.

Patient data

CT images were retrieved to predict the dry mouth symptoms of patients with NPC after receiving radiotherapy. CT and true SA data were collected from 35 NPC patients at the Sun Yat-sen University Cancer Center (SYSUCC), Guangzhou, China. All the data were retrospectively analyzed. The patients with stages I–IVB, were randomly chosen from the control group in the NPC clinical trial, which was registered on the clinicaltrials.gov (ID: NCT01762514). The study was approved by the institutional review board and the ethical review office from the institution, and the data had been submitted to a public Research Data Deposit platform (www.researchdata.org.cn), with an approval RDD number as RDDB2018000256. The 35 patients were used as training and validation sets, and 4 more patients were used as the independent test set.

All patients received radical treatment based on intensity-modulated RT (IMRT) with a prescription dose of 68.1 Gy in 30 fractions. The patients' ages ranged 24–72 years, and each patient had five CT sets acquired using a CT simulator (CT Big Bore; Philips) at treatment position at 0^{th}, 10^{th}, 20^{th}, 30^{th} fractions during RT and 3-month later after the RT. The PGs were delineated on each CT set by a radiation oncologist and verified independently by another. Acute xerostomia was evaluated based on RTOG acute toxicity scoring (3) and all the patients were evaluated in those XG every other ten days by the attending physician based on patients reporting as follows: (I) G0: no change over baseline; (II) G1: mild mouth dryness/slightly thickened saliva/may have slightly altered taste such as metallic taste; (III) G2: moderate to complete dryness/thick, sticky saliva/markedly altered taste (i.e., copious water or other lubricants); (IV) G3: severe dry mouth, no stimulation, often need to wake up at night to drink water; and (V) G4: acute salivary gland necrosis.

Patients saliva were collected every other 10 days during RT. Measurement of saliva output is the most commonly applied objective evaluation of xerostomia. Previous literature (25) comprehensively explored on how to collect saliva. The un-stimulated saliva was collected, which represents the saliva output of the whole-mouth saliva. Saliva collection was performed over a period of 5 minutes at the 0^{th}, 10^{th}, 20^{th}, 30^{th} fractions during RT and at 3-month later after RT. Collection of the whole saliva was performed by drainage and measured via a medical cylinder. All the SA and its changes were documented.

The 0^{th} fraction (0F) CT images and SA_{0F} were used as

the baseline to predict the SA_{10F} in this paper. The result was recorded as PSA_{10F} (predicted SA at 10^{th} fraction). Furthermore, 0F CT images, 10F CT images, SA_{0F}, SA_{10F}, and the corresponding changes were used together to predict PSA_{30F}. The probability and extent of the patient's future acute xerostomia were evaluated according to the PSA. Patient characteristics along with other data were summarized in *Table 1*.

Data preprocessing

The PGs of 35 patients were delineated on each CT set by a radiation oncologist and verified independently by another one. More than 5,000 radiomics features were obtained from each annotated CT (*Figure 1A*) using 3D Slicer (26-28). 3D Slicer has an extension module called Radiomics which provides an interface to the PyRadiomics library (https://github.com/Radiomics/pyradiomics). PyRadiomics is an open-source python package for extracting Radiomics features from medical imaging. Current possible image types for PyRadiomics are Original, Wavelet, LoG, Square, SquareRoot, Logarithm, Exponential, Gradient, LBP2D and LBP3D. In this article, Original (no filter applied) and Wavelet (wavelet filtering, yields 8 decompositions per level, and all possible combinations of applying either a High or a Low pass filter in each of the three dimensions), two types of image were utilized. Various features can be extracted by PyRadiomics. They can be divided into seven categories: First Order, Shape, Gray Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), Gray Level Size Zone Matrix (GLSZM), Neighbouring Gray Tone Difference Matrix (NGTDM) and Gray Level Dependence Matrix (GLDM). All the mentioned features categories were employed in our method.

In all, 1,703 feature values (FVs) were extracted from radiomics features for each CT set by data cleaning, which includes checking data consistency, handling invalid values and missing values. In addition, we performed data augmentation through the normalization process (29) and the differences between those FVs (*Figure 1B*).

In this paper, the FVs at 0^{th} fraction (include SA_{0F}), named FV_{0F}, were employed to predict the SA at 10^{th} fraction (PSA_{10F}). FV_{0F}, FV_{10F} and delta radiomics FVs named as delta FV (ΔFV_{10F-0F}) were also used to predict PSA_{30F}. Besides the original data set, L^2-norm normalization was performed on FV_{0F} and FV_{10F}, to ensure all the features were on the same order of magnitude. Formally, given one feature $x=\{x_1, x_2,..., x_n\}$ where n represents the number of patients, the L^2-norm $\|x\|$ is formulated as:

$$\|x\| = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{\frac{1}{2}} \quad [1]$$

And the normalization formula is deduced as:

$$x_i^{'} = \frac{x_i}{\|x\|} \quad [2]$$

where $x_i^{'}$ is the normalized feature value (NFV) or normalized SA (NSA).

Besides normalization, the feature difference was added between FV_{10F} and FV_{0F} (ΔFV_{10F-0F}) to the data set, while the ΔFV_{10F-0F} was normalized to obtain the ΔNFV_{10F-0F}. The difference feature was also considered as a kind of feature. ΔNFV_{10F-0F} was added to the data set to verify the effect of feature changes on predicting patients' future SA, which can further determine whether the prediction ability of the difference feature was stronger than the ordinary ones or not.

Xerostomia prediction model

After the data preprocessing, the FVs and SA labels were used to calculate the feature scores (*Figure 1C*) by specified functions. Then K features were filtered out based on the scores ranking (*Figure 1D*). Finally, the selected features were applied to predict the SA (*Figure 1E*) by prediction functions. The functions were shown in *Table 2*. The first column was nine score calculation functions, which match with 2 feature selection functions placed in the second column. The third column contained eight SA prediction functions. We tried the various combination of machine learning techniques to determine the best method to predict SA efficiently and accurately.

Feature score calculation

After the data were preprocessed, nine score calculation functions were used to generate the feature scores, as following:

(I) f_regression: this is a univariate linear regression test (30), which is used in feature selection procedure, the formula for calculating the f value is shown as follows:

$$r_j = \frac{\left(X_j - \overline{X_j}\right) * \left(Y - \overline{Y}\right)}{std\left(X_j\right) * std\left(Y\right)} \quad [3]$$

Table 1 Patient demographic and treatment information and xerostomia grade (XG) at 10th and 30th fraction (F: female; M: male)

Patient No.	Age (years)	Sex	Stage	Chemotherapy	XG at 10F	XG at 30F
1	24	M	T2N2M0 III	Yes	0	1
2	49	M	T2N2M0 III	Yes	2	3
3	35	M	T1N2M0 III	Yes	1	2
4	25	M	T2N0M0 II	No	0	1
5	40	M	T2N1M0 II	No	1	2
6	40	F	T2N3M0 IV	Yes	1	2
7	65	M	T3N1M0 III	Yes	1	2
8	32	M	T4N2M0 IV	Yes	1	2
9	33	M	T2N2M0 III	Yes	1	2
10	52	M	T4N1M0 IV	Yes	1	2
11	48	M	T3N1aM0 III	Yes	1	2
12	72	M	T2N1M0 II	No	0	1
13	57	M	T2N1M0 II	No	1	2
14	56	M	T4N2M0 IV	Yes	1	2
15	40	M	T3N1M0 III	Yes	1	2
16	37	F	T2N2M0 III	Yes	2	2
17	70	M	T2N1M0 II	No	0	1
18	48	M	T3N1M0 III	Yes	1	2
19	43	M	T2N0M0 II	No	1	2
20	56	M	T4N2M0 IV	Yes	1	1
21	59	M	T3N0M0 III	Yes	2	2
22	32	F	T3N1M0 III	Yes	1	2
23	32	M	T2N2M0 III	Yes	0	2
24	47	F	T4N2M0 IV	Yes	1	2
25	51	F	T4N1M0 IV	Yes	1	2
26	62	F	T4N2M0 IV	Yes	1	2
27	62	M	T3N0M0 III	Yes	1	2
28	35	M	T1N2M0 III	Yes	1	2
29	30	M	T3N3M0 IV	Yes	1	2
30	29	F	T3N3bM0 IV	Yes	1	2
31	48	M	T3N0M0 III	Yes	1	2
32	53	M	T3N1M0 III	Yes	0	1
33	50	M	T3N0M0 III	Yes	1	2
34	35	M	T3N2M0 III	Yes	1	2
35	53	F	T2N1M0 II	No	1	2

Table 2 Feature score calculation, selection and predication model for the predicted saliva amount (PSA)

Feature score calculation	Feature selection	Saliva amount prediction
f_regression	SelectKBest	SVR
mutual_info_regression		DecisionTreeRegressor
DecisionTreeRegressor	RecursiveFeatureElimination (RFE)	ExtraTreesRegressor
ExtraTreesRegressor		Lasso
Lasso		LinearRegression
LinearRegression		RandomForestRegressor
RandomForestRegressor		Ridge
Ridge		RidgeCV
RidgeCV		

$$f = \frac{r_j^2}{1-r_j^2} * (n-2) \quad [4]$$

where X_j represents the vector of all patients on the feature of j, Y is the vector of SA label. And *std*() indicates standard deviation function. r_j is the sample correlation coefficient. Hence, the larger the f value, the greater the correlation between the j-feature and the dependent variable Y.

(II) mutual_info_regression: this function is able to estimate the mutual information $I(X,Y)$ (31,32) of two target variables, which can measure the dependency between two samples. The formula is defined as:

$$I(X,Y) = \sum_{X_j \in X} p(X_j, Y) \log \frac{p(X_j, Y)}{p(X_j)\,p(Y)} \quad [5]$$

where $X=\{X_1, X_2,..., X_m\}$ is the set of all features and m represents the number of features, $p(X_j, Y)$ is joint probability distribution function of X_j and Y. And $p(X_j)$ and $p(Y)$ are marginal probability function. Higher $I(X,Y)$ implies higher dependency between two variables.

(III) DecisionTreeRegressor: decision tree regressor is an unsupervised regression method, and it learns simple decision rules from features to predict the target variable (33-35).

(IV) RandomForestRegressor (36): in the integrated model of the random forest, the samples of each tree are constructed from the training set after the put-back sampling. In addition, the selected segmentation point is not the best segmentation point for all features, but the best segmentation point in a random subset of features (37).

(V) ExtraTreesRegressor: in the extra tree, the randomness in the method of calculating the segmentation point is further enhanced. Instead of finding the most discriminating threshold, the threshold here is randomly generated for each candidate feature, and selecting the best one of these thresholds as the segmentation rule (37,38).

(VI) LinearRegression: this function is a least squares linear regression (36,39), which minimizes the residual sum of squares between X and Y by fitting a linear model with the coefficient $w = (w_1, w_2, \ldots, w_p)$:

$$\min_w \|Xw - Y\|_2^2 \quad [6]$$

However, this method relies on the mutual independence of the model terms. The least squares estimate will be very sensitive to random errors and produce a large variance when the columns (features) of X are approximately linearly dependent and terms are correlated (40).

(VII) Ridge: ridge regression solves the problem of ordinary least squares by imposing punishment on the size of the coefficient (40), which minimizes the residual sum of squared with penalty:

$$\min_w \|Xw - Y\|_2^2 + \alpha \|w\|_2^2 \quad [7]$$

where $\alpha \geq 0$, and the larger the α, the greater the shrinkage and the stronger the robustness of the coefficient w to the collinearity (41).

(VIII) RidgeCV: RidgeCV implements Ridge regression

Table 3 Experimental data set for prediction input

SA label	Feature
SA_{10F} or NSA_{10F}	FV_{0F} or NFV_{0F}
SA_{30F} or NSA_{30F}	FV_{0F} or NFV_{0F}
	FV_{10F} or NFV_{10F}
	{FV_{0F}, FV_{10F}} or {NFV_{0F}, NFV_{10F}}
	ΔFV_{10F-0F} or ΔNFV_{10F-0F}
	{FV_{0F}, FV_{10F}, ΔFV_{10F-0F}} or {NFV_{0F}, NFV_{10F}, ΔNFV_{10F-0F}}

SA, saliva amount; NSA, normalized saliva amount; FV, feature value; NFV, normalized feature value.

with cross-validation (CV) of built-in Alpha parameters (40).

(IX) Lasso: this function is a linear model for estimating sparse coefficients. It tends to use fewer variables or parameters, which can effectively reduce the number of variables that the given solution depends on (40,42). Its minimized objective function is as follows:

$$\min_{w} \frac{1}{2n} \left\| Xw - Y \right\|_2^2 + \alpha \left\| w \right\|_1 \quad [8]$$

where α is a constant, n is the number of samples and $\left\| w \right\|_1$ is the $\ell 1$-norm of the parameter vector.

Feature selection

The score calculation function (*Figure 1C*) needs to combine with the feature selection function (*Figure 1D*) to filter out the features. This paper employed the SelectKBest and recursive feature elimination (RFE) (43) to select the features (*Table 2*).

(I) SelectKBest is able to remove all features except the K features with the highest score, and the retained features are the target features, where the K value can be customized.

(II) RFE selects the specified number K of features by recursively considering smaller and smaller sets of features.

SA prediction

The features, filtered by feature selection, were used to predict the SA (*Figure 1E*). Eight prediction functions were used, and most of them were the same as the score calculation function, except for the support vector regression (SVR) (40) (in *Table 2* column 3). The output of this step was the predicted value of the SA (PSA), which could be used as a predictor of the acute xerostomia.

Results

Model selection

Various feature selection and SA prediction function combinations could be considered. There were nine score calculation functions (in *Table 2* column 1) and two feature selection functions (in *Table 2* column 2), which could produce nine combinations for selecting features. And there were also eight methods for SA prediction (in *Table 2* column 3). Therefore, a total 72 combinations of feature selection and SA prediction function were tested in all (*Table 2*). Besides, the experimental data can be preprocessed by various ways, such as whether to normalize, whether to take the difference. Here we considered total 24 forms (*Table 3*). Therefore, the best model and data preprocessing method were finally chosen based on the experimental results.

We designed a set of experiments according to the input data form (as shown in *Table 3*) and the functions (as shown in *Table 2*) to predict SA. The number of selected feature quantity K, which was difficult to be set at very beginning, could be determined by experiments. And range of candidate K was from 1 to 16. And the CV coefficient of the prediction function was initialized to 3.

The training stage was composed of three parts: score calculation, feature selection and SA prediction. It was worth noting that all the 35 patients' data were used in score calculation and feature selection stage, while the CV technique mentioned was only used for SA prediction function in training stage.

The optimal result for PSA_{10F} and PSA_{30F} is shown in *Table 4*. The model is evaluated by MSE (mean square error) (39), which is shown as follows:

$$\mathrm{MSE} = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2 \quad [9]$$

Table 4 The optimal result for PSA_{10F} and PSA_{30F}

PSA	Feature	Score calculation	Feature selection	K	Prediction	CV	MSE	R_MSE
PSA_{10F}	NFV_{0F}	RidgeCV	RFE	8	LinearRegression	3	0.9042	3.6886
PSA_{30F}	{NFV_{0F}, NFV_{10F}, ΔNFV_{10F-0F}}†	RidgeCV	RFE	14	LinearRegression	3	0.0801	0.7297
	ΔNFV_{10F-0F}†	LinearRegression	RFE	14	LinearRegression	3	0.0754	

†, the method to be further compared, and the results were listed in *Table 5*. RFE, Recursive Feature Elimination; MSE, mean square error; CV, cross-validation.

Table 5 Prediction results of PSA_{30F} under different CV coefficients

Feature	Score calculation	Feature selection	K	Prediction	CV	MSE
{NFV_{0F}, NFV_{10F}, ΔNFV_{10F-0F}}	RidgeCV	RFE	14	LinearRegression	4	0.0729
					6	0.0696
ΔNFV_{10F-0F}	LinearRegression	RFE	14	LinearRegression	4	0.0754
					6	0.0704

PSA, predicted saliva amount; RFE, Recursive Feature Elimination; MSE, mean square error; CV, cross-validation.

where n is the number of samples, y_i is the SA true value of the i[th] patient, and $\hat{y}_i$ is predictive value. Reference MSE (R_MSE) is defined as $\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\overline{y_i}\right)^2$. Hence, the smaller the MSE, the more accurate the prediction.

Experimental results showed that the best prediction model for PSA_{10F} was RidgeCV_RFE_LinearRegression. Experimental results also showed that the best two prediction models for PSA_{30F} was RidgeCV_RFE_LinearRegression with {NFV_{0F}, NFV_{10F}, ΔNFV_{10F-0F}} datasets and LinearRegression_RFE_LinearRegression with NFV_{10F-0F} datasets, respectively. They had the similar MSE as shown in *Table 4* (0.0801 *vs.* 0.0754).

In order to determine the final model of predicting PSA_{30F}, the pros and cons of the methods for PSA_{30F} in *Table 4* rows 2 and 3 were further discussed by changing the CV coefficient, and the result was shown in *Table 5*.

Table 5 shows that as the CV of the prediction function increases, the advantages of using {NFV_{0F}, NFV_{10F}, ΔNFV_{10F-0F}} datasets with RidgeCV to calculate feature scores become apparent. Hence, the best predictive model of PSA_{30F} was RidgeCV_RFE_LinearRegression with {NFV_{0F}, NFV_{10F}, ΔNFV_{10F-0F}} datasets.

A large number of experiments demonstrated that NFV prediction results were better than FV, the FV results were not given in the paper. For detailed experimental results, please refer to *Table S1*.

We also did a related experiment to predict NSA ($PNSA_{10F}$ and $PNSA_{30F}$), however, because predicting PSA is more in line with clinical needs, *Table 4* focuses on the result of PSA. The best predictions for $PNSA_{10F}$ and $PNSA_{30F}$ were detailed in *Table S1*.

SA prediction

Predicting $PSA1_{0F}$

We used RidgeCV_RFE_LinearRegression model with NFV_{0F} datasets and SA_{0F} to predict PSA_{10F}. The experimental result with K=8 gave the best result (the smallest MSE of 0.9042). However, prediction results between K=6 (*Table S1*, an MSE of 0.9548) and K=8 were close, which K=6 had fewer features. Therefore, both sets of features could be considered. In this paper, the eight features and their weights were given in *Table 6*. Please note that the "h" means High-pass filter and the "l" means Low-pass filter in "hhl" or "lhh" of the feature names, and the last "r" means Right and the last "l" means Left. Ranking indicated the sequential order of the features' selection. The features were selected by score calculation and feature selection, and the weights were yielded in SA prediction stage by specific function based on the data from all 35 patients. The related weights of the six features were given in *Table S2*.

Predicting SA_{10F} with the LinearRegression function,

Table 6 The eight features and weights of PSA_{10F}

ID	Ranking	Feature	Weights
1	No. 6	0F_wavelet-hhh_firstorder_median_parotid_r	–4.4827
2	No. 3	0F_wavelet-hhh_glcm_correlation_parotid_l	–4.5370
3	No. 8	0F_wavelet-hhl_glcm_clustershade_parotid_l	–2.7172
4	No. 1	0F_wavelet-hll_glszm_lowgraylevelzoneemphasis_parotid_l	8.0633
5	No. 5	0F_wavelet-lhh_firstorder_mean_parotid_r	3.6109
6	No. 2	0F_wavelet-lhh_glrlm_longrunlowgraylevelemphasis_parotid_r	2.6036
7	No. 7	0F_wavelet-lhh_glrlm_shortrunlowgraylevelemphasis_parotid_r	3.6465
8	No. 4	0F_saliva amount	–5.4791
		Bias	5.2181

PSA, predicted saliva amount.

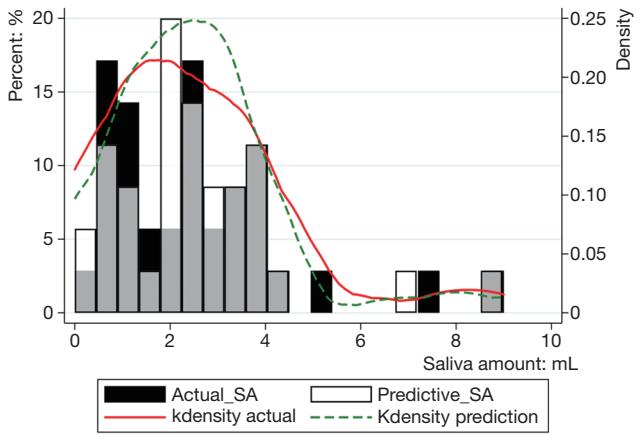

Figure 2 The result of PSA_{10F}. PSA, predicted saliva amount; SA, saliva amount.

MSE was 0.9042 and R^2 score was 0.7406 in the verification phase while CV was 3 (*Table 4*). The R^2 score is shown as:

$$R^2 = 1 - \frac{\sum_1^n (y_i - \hat{y}_i)^2}{\sum_1^n (y_i - \bar{y}_i)^2} \quad [10]$$

where $\bar{y}_i$ is mean SA. The denominator is the true data dispersion degree, and the numerator is the prediction error. The division of the two intends to eliminate the influence of the true data dispersion degree. When the R^2 is higher, the model fits the data better.

Figure 2 shows the prediction results of SA_{10F}.White histogram and green solid line represented the predicted result, black histogram and red dash line were the distribution of true SA and gray part indicated the overlap between reality and prediction. The *x*-axis represented the SA, and the left *y*-axis represented probability for the histogram, while the right *y*-axis represented density for the fitting curve. The histogram showed the distribution of SA and the fitting curve was kernel density estimation, which gave the distribution of the data more explicitly. The true and corresponding predicted values of PSA_{10F} were given in *Table 7*, and the P value between SA_{10F} and PSA_{10F} was 0.9888.

Predicting $PSA3_{0F}$

The RidgeCV_RFE_LinearRegression model with {NFV_{0F}, NFV_{10F}, $\Delta NFV_{10F\text{-}0F}$} datasets were applied to predict PSA_{30F}, K=14 was selected by experiment result. The weights of the fourteen features were given in *Table 8*.

LinearRegression function with 35 CV (leave one out protocol) was utilized to get the best prediction results for PSA_{30F} in the verification phase. The best MSE was 0.0569 and R^2 score was 0.9220. The results showed that the error was small enough by using the above 14 features to predict PSA_{30F}, while the model fitted the data well. Please note that ten delta radiomics features out of the fourteen features played an important role in the PSA_{30F}.

Figure 3 showed the prediction results of PSA_{30F}. The prediction curve was basically consistent with the true curve. The specific values of PSA_{30F} were also given in *Table 7*, and the P value between SA_{30F} and PSA_{30F} was 0.8845.

Independent testing

In general, our model tried to predict the SA based on

Table 7 The true SA and the prediction SA

Patient No.	SA_{10F} (mL)	PSA_{10F} (mL)	SA_{30F} (mL)	PSA_{30F} (mL)
1	2.6	2.36	1.7	1.45
2	0.5	0.00	1.5	1.22
3	9.0	7.08	1.0	1.06
4	3.0	3.22	0.0	0.44
5	4.0	3.13	0.9	0.60
6	3.5	2.66	1.8	1.79
7	7.5	8.97	2.5	2.61
8	4.0	1.95	1.2	1.55
9	3.2	3.41	0.3	0.24
10	0.8	1.36	0.0	0.04
11	2.3	2.48	1.0	0.96
12	1.0	2.01	1.5	1.90
13	0.7	1.17	2.5	3.09
14	0.5	0.80	0.3	0.29
15	0.0	0.47	0.0	0.00
16	2.5	3.04	1.0	1.25
17	1.5	2.19	2.0	2.10
18	2.5	2.48	2.1	2.15
19	1.0	2.06	0.0	0.00
20	3.7	3.43	2.5	2.14
21	3.1	3.65	0.5	0.68
22	1.5	0.73	0.5	0.85
23	1.0	0.64	0.5	0.70
24	2.1	2.26	0.2	0.20
25	4.4	3.67	2.1	1.91
26	2.5	2.99	2.6	2.53
27	1.2	2.12	0.0	0.08
28	5.0	3.70	0.5	0.40
29	1.0	0.99	0.2	0.16
30	3.7	4.27	2.0	1.75
31	0.5	2.01	0.5	0.49
32	2.0	3.95	2.0	1.68
33	3.5	0.99	1.5	1.93
34	0.5	0.00	0.2	0.25
35	2.5	1.85	1.2	0.88

SA, saliva amount; PSA, predicted saliva amount.

Table 8 The fourteen features and weights of PSA_{30F}

ID	Ranking	Feature	Weights
1	No. 1	0F_wavelet-hhl_firstorder_skewness_parotid_l	−1.0103
2	No. 11	0F_wavelet-hll_firstorder_median_parotid_l	−1.1899
3	No. 9	0F_wavelet-hll_ngtdm_strength_parotid_l	−1.3942
4	No. 14	10F_wavelet-hll_ngtdm_busyness_parotid_r	0.6046
5	No. 2	Δoriginal_glszm_sizezonenonuniformity_parotid_l	−1.1343
6	No. 7	Δoriginal_glszm_zonepercentage_parotid_r	1.8854
7	No. 8	Δoriginal_shape_maximum3ddiameter_parotid_l	−1.5198
8	No. 5	Δwavelet-hhh_glszm_sizezonenonuniformitynormalized_parotid_l	1.1849
9	No. 13	Δwavelet-hlh_firstorder_skewness_parotid_l	−0.6800
10	No. 3	Δwavelet-hlh_glrlm_shortrunlowgraylevelemphasis_parotid_l	−1.5883
11	No. 10	Δwavelet-lhh_glcm_differenceentropy_parotid_l	1.2062
12	No. 6	Δwavelet-lhh_gldm_largedependencehighgraylevelemphasis_parotid_l	1.1405
13	No. 12	Δwavelet-llh_glcm_imc2_parotid_l	−1.0097
14	No. 4	Δwavelet-lll_glcm_correlation_parotid_r	−0.5980
		Bias	0.8049

PSA, predicted saliva amount.

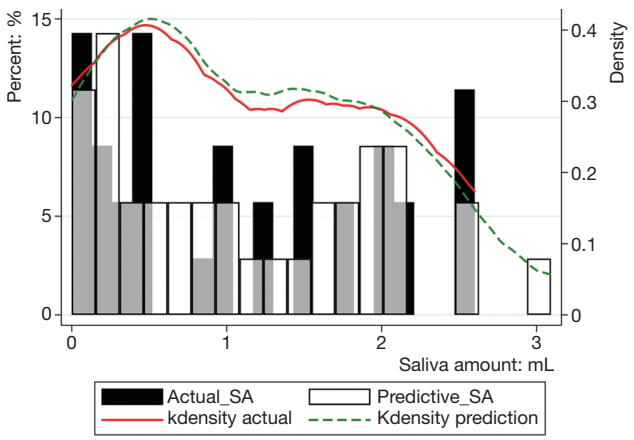

Figure 3 The result of PSA_{30F}. PSA, predicted saliva amount; SA, saliva amount.

specific features (8 features in *Table 6* for PSA_{10F}, 14 features in *Table 8* for PSA_{30F}) and the LinearRegression was used as the SA prediction function. In order to prove the model and the selected features were also applicable to other datasets, four unseen patients were used for testing. The *Table 9* showed that the PSA_{10F} (*Table 9* column 4) were not close to the true value (*Table 9* column 3), while PSA_{30F} were accurate enough with the true value (*Table 9* column 6). The MSE of PSA_{30F} was 0.0233 and R^2 score was 0.8680. The reason might be that most of the selected features (10 out of 14) used for predicting SA_{30F} were delta radiomics, which can effectively represent features' changes during RT.

Discussion

Predicting acute xerostomia through the changes in the PGs had been proposed in previous literatures. However, most of the studies focused on merely a few common features [e.g., volume and Hounsfield units (HU) (6)], which could not comprehensively extract and analyze such large number of features for PGs. In this paper, the method was implemented by using radiomics and delta radiomics. It can be divided into following stages: image collection, ROI segmentation, feature extraction, statistical analysis, classification and prediction. It showed that radiomics can automatically extract a large number of features from the ROI of medical images, and identifies key features based on statistical analysis. Finally, these features were applied in the auxiliary diagnosis and treatment through detailed analysis (8,9,11).

This article quantified the SA by systematically analyzing

Table 9 Four individual patients PSA results

Patient No.	SA_{0F} (mL)	SA_{10F} (mL)	PSA_{10F} (mL)	SA_{30F} (mL)	PSA_{30F} (mL)
36	7.0	3.8	1.70	0.3	0.23
37	6.3	4.4	3.90	0.7	0.42
38	6.9	2.0	1.00	1.2	1.21
39	6.4	2.5	0.00	0.1	0.00

SA, saliva amount; PSA, predicted saliva amount.

a large number of features extracted from the PGs to achieve predicting acute xerostomia. Finally, eight (*Table 6*) and fourteen (*Table 8*) features to applied to predict SA_{10F} and SA_{30F}, and the precision were 0.7406 and 0.9220, respectively. The sensitivity reached 100% because every patient received a PSA. On the other hand, the P values for the 10F and 30F were both greater than 0.05 (0.9888 and 0.8845), which indicated no significant statistical difference between the true and the prediction values, and the experimental results expressed clinical significance.

We noted that delta radiomics (ΔFV_{10F-0F}) held a larger proportion (10/14) when predicting SA_{30F}. It indicated that the variation of features contributed significantly to the quantitative prediction of SA. However, volume and HU, which played an important role in other studies (6), were not selected out in this article. We speculated the reason was that the wavelet features had a large ratio in all the features exacted in our study. Radiomics data contain first-, second-, and higher-order statistics, and the high-order features might overcome the low-order (first-order) grey features in this PGs study.

Moreover, our prediction accuracy for SA_{30F} at the 10^{th} of RT reached 0.9220, in other words, the system could accurately predict the acute xerostomia of patients after 30^{th} fractions of RT as soon as possible, which was a key step for early xerostomia prediction.

Conclusions

We observed and quantified the radiomics changes in the PGs during the fractionated RT for NPC to predict SA_{10F} and SA_{30F}. The RidgeCV was used to calculate the features score, RFE was used to select feature, and Linear Regression was used to predict SA both at 10F and 30F. The optimal result of PSA_{10F} was achieved using 8 features (*Table 6*) with an MSE of 0.9042 and R^2 score of 0.7406. Meanwhile 14 features (*Table 8*) performed the best at PSA_{30F} with an MSE of 0.0569 and R^2 score of 0.9220. This result indicated that the proposed method was able to accurately predict patients' SA at early stage and prevent the xerostomia symptom in advance.

In this paper, only the un-stimulated SA was used to build the model, and the stimulated saliva data can be added into the model to increase the robustness of the method in the following work. In addition, how to accurately predict the late xerostomia for the patients after RT is still a difficult problem. We would work on it in the near future, since the PGs surface/volume reduction might be associated with late xerostomia. The early post treatment model with delta PG-surface and acute xerostomia scores can be considered as a surrogate marker for late xerostomia (44). What is more, we will increase training data and test data gradually and further improve the model to improve the accuracy of SA prediction. Other diseases could also benefit from this model with moderate adjustment.

Acknowledgments

Discussions and inputs from Dr. Allen Li and Dr. Yalan Tao are greatly appreciated. The authors would like to thank the anonymous reviewers, and the guest editor for their constructive comments and suggestions, which greatly improved the quality of the manuscript.

Funding: This research received the support from Natural Science Foundation of Guangdong Province (No. 2017A030310217); Pearl River S&T Nova Program of Guangzhou (No. 201710010162).

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

Ethical Statement: The study was approved by the

institutional review board and the ethical review office from the institution, and the data had been submitted to a public Research Data Deposit platform (www.researchdata.org.cn), with an approval RDD number as RDDB2018000256. Written informed consent was obtained from all individual participants involved in this trial, which was registered on the clinicaltrials.gov (ID: NCT01762514).

References

1. Tribius S, Sommer J, Prosch C, Bajrovic A, Muenscher A, Blessmann M, Kruell A, Petersen C, Todorovic M, Tennstedt P. Xerostomia after radiotherapy. Strahlenther Onkol 2013;189:216-22.
2. Bjordal K, Kaasa S, Mastekaasa A. Quality of life in patients treated for head and neck cancer: A follow-up study 7 to 11 years after radiotherapy. Int J Radiat Oncol Biol Phys 1994;28:847-56.
3. Cox JD, Stetz J, Pajak TF. Toxicity criteria of the Radiation Therapy Oncology Group (RTOG) and the European organization for research and treatment of cancer (EORTC). Int J Radiat Oncol Biol Phys 1995;31:1341-6.
4. Vissink A, Mitchell JB, Baum BJ, Limesand KH, Jensen SB, Fox PC, Elting LS, Langendijk JA, Coppes RP, Reyland ME. Clinical Management of Salivary Gland Hypofunction and Xerostomia in Head-and-Neck Cancer Patients: Successes and Barriers. Int J Radiat Oncol Biol Phys 2010;78:983-91.
5. Vissink A, van Luijk P, Langendijk J, Coppes R. Current ideas to reduce or salvage radiation damage to salivary glands. Oral Diseases 2015;21:e1-10.
6. Wu H, Chen X, Yang X, Tao Y, Xia Y, Deng X, Zheng C, Robbins J, Schultz C, Li XA. Early Prediction of Acute Xerostomia During Radiation Therapy for Head and Neck Cancer Based on Texture Analysis of Daily CT. Int J Radiat Oncol Biol Phys 2018;102:1308-18.
7. van Dijk LV, Brouwer CL, van der Schaaf A, Burgerhof JGM, Beukinga RJ, Langendijk JA, Sijtsema NM, Steenbakkers RJHM. CT image biomarkers to improve patient-specific prediction of radiation-induced xerostomia and sticky saliva. Radiother Oncol 2017;122:185-91.
8. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D, Hoebers F, Rietbergen MM, Leemans CR, Dekker A, Quackenbush J, Gillies RJ, Lambin P. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5:4006.
9. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology 2016;278:563-77.
10. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, Forster K, Aerts HJ, Dekker A, Fenstermacher D, Goldgof DB, Hall LO, Lambin P, Balagurunathan Y, Gatenby RA, Gillies RJ. Radiomics: the process and the challenges. Magn Reson Imaging 2012;30:1234-48.
11. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, Zegers CM, Gillies R, Boellard R, Dekker A, Aerts HJ. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer 2012;48:441-6.
12. Pota M, Scalco E, Sanguineti G, Farneti A, Cattaneo GM, Rizzo G, Esposito M. Early prediction of radiotherapy-induced parotid shrinkage and toxicity based on CT radiomics and fuzzy classification. Artif Intell Med 2017;81:41-53.
13. Chen X, Oshima K, Schott D, Wu H, Hall W, Song Y, Tao Y, Li D, Zheng C, Knechtges P, Erickson B, Li XA. Assessment of treatment response during chemoradiation therapy for pancreatic cancer based on quantitative radiomic analysis of daily CTs: An exploratory study. PLoS One 2017;12:e0178961.
14. Abdollahi H, Mahdavi SR, Mofid B, Bakhshandeh M, Razzaghdoust A, Saadipoor A, Tanha K. Rectal wall MRI radiomics in prostate cancer patients: prediction of and correlation with early rectal toxicity. Int J Radiat Biol 2018;94:829-37.
15. Zhang Z, Yang J, Ho A, Jiang W, Logan J, Wang X, Brown PD, McGovern SL, Guha-Thakurta N, Ferguson SD, Fave X, Zhang L, Mackin D, Court LE, Li J. A predictive model for distinguishing radiation necrosis from tumour progression after gamma knife radiosurgery based on radiomic features from MR images. Eur Radiol 2018;28:2255-63.
16. Zhu T, Das S, Wong TZ. Integration of PET/MR Hybrid Imaging into Radiation Therapy Treatment. Magn Reson Imaging Clin N Am 2017;25:377-430.
17. Liao KY, Chiu CC, Chiang WC, Chiou YR, Zhang G, Yang SN, Huang TC. Radiomics features analysis of PET images in oropharyngeal and hypopharyngeal cancer. Medicine (Baltimore) 2019;98:e15446.
18. Rosen BS, Hawkins PG, Polan DF, Balter JM, Brock KK, Kamp JD, Lockhart CM, Eisbruch A, Mierzwa ML, Ten

Haken RK, El Naqa I. Early Changes in Serial CBCT-Measured Parotid Gland Biomarkers Predict Chronic Xerostomia After Head and Neck Radiation Therapy. Int J Radiat Oncol Biol Phys 2018;102:1319-29.
19. van Timmeren JE, Leijenaar RTH, van Elmpt W, Reymen B, Lambin P. Feature selection methodology for longitudinal cone-beam CT radiomics. Acta Oncologica 2017;56:1537-43.
20. Xu S, Wu Z, Yang C, Ma L, Qu B, Chen G, Yao W, Wang S, Liu Y, Li XA. Radiation-induced CT number changes in GTV and parotid glands during the course of radiation therapy for nasopharyngeal cancer. Br J Radiol 2016;89:20140819.
21. Feng M, Yang C, Chen X, Xu S, Moraru I, Lang J, Schultz C, Li XA. Computed Tomography Number Changes Observed During Computed Tomography–Guided Radiation Therapy for Head and Neck Cancer. Int J Radiat Oncol Biol Phys 2015;91:1041-7.
22. Jha N, Harris J, Seikaly H, Jacobs JR, McEwan AJB, Robbins KT, Grecula J, Sharma AK, Ang KK. A Phase II Study of Submandibular Gland Transfer Prior to Radiation for Prevention of Radiation-induced Xerostomia in Head-and-Neck Cancer (RTOG 0244). Int J Radiat Oncol Biol Phys 2012;84:437-42.
23. Yang WF, Liao GQ, Hakim SG, Ouyang DQ, Ringash J, Su YX. Is Pilocarpine Effective in Preventing Radiation-Induced Xerostomia? A Systematic Review and Meta-analysis. Int J Radiat Oncol Biol Phys 2016;94:503-11.
24. Meirovitz A, Murdoch-Kinch CA, Schipper M, Pan C, Eisbruch A. Grading xerostomia by physicians or by patients after intensity-modulated radiotherapy of head-and-neck cancer. Int J Radiat Oncol Biol Phys 2006;66:445-53.
25. Navazesh M. Methods for Collecting Saliva. Ann N Y Acad Sci 1993;694:72-7.
26. Pieper S, Halle M, Kikinis R. editors. 3D Slicer. 2004 2nd IEEE International Symposium on Biomedical Imaging: Nano to Macro (IEEE Cat No. 04EX821); 2004 18-18 April 2004.
27. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, Bauer C, Jennings D, Fennessy F, Sonka M, Buatti J, Aylward S, Miller JV, Pieper S, Kikinis R. 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magn Reson Imaging 2012;30:1323-41.
28. Pinter C, Lasso A, Wang A, Jaffray D, Fichtinger G. SlicerRT: Radiation therapy research toolkit for 3D Slicer. Med Phys 2012;39:6332-8.
29. Aksoy S, Haralick RM. Feature normalization and likelihood-based similarity measures for image retrieval. Pattern Recognit Lett 2001;22:563-82.
30. Hazelton ML. Univariate Linear Regression. International Encyclopedia of Education 2010:482-8.
31. Huang J, Cai Y, Xu X. editors. A Filter Approach to Feature Selection Based on Mutual Information. IEEE International Conference on Cognitive Informatics; 2007.
32. Nagpal A, Gaur D, Gaur S. editors. Feature selection using mutual information for high- dimensional data sets. IEEE International Advance Computing Conference; 2014.
33. Fredman ML. editor. On the decision tree complexity of the shortest path problems. Symposium on Foundations of Computer Science; 1975.
34. Cervantes B, Monroy R, Medina-Pérez MA, Gonzalez-Mendoza M, Ramirez-Marquez J. Some features speak loud, but together they all speak louder: A study on the correlation between classification error and feature usage in decision-tree classification ensembles. Eng Appl Artif Intell 2018;67:270-82.
35. Shah N, Ziauddin S, Shahid AR. editors. Brain tumor segmentation and classification using cascaded random decision forests. International Conference on Electrical Engineering/electronics; 2017.
36. Grömping U. Variable Importance Assessment in Regression: Linear Regression versus Random Forest. Am Stat 2009;63:308-19.
37. Mishra G, Sehgal D, Valadi JK. Quantitative Structure Activity Relationship study of the Anti-Hepatitis Peptides employing Random Forests and Extra-trees regressors. Bioinformation 2017;13:60-2.
38. Pisane J, Maree R, Ries P, Wehenkel L, Verly JG. editors. An extra-trees-based automatic target recognition algorithm. International Radar Conference-surveillance for A Safer World; 2009.
39. Arsad PM, Buniyamin N, Manan JLA. editors. Neural Network and Linear Regression methods for prediction of students' academic achievement. Global Engineering Education Conference; 2014.
40. Franklin J. The elements of statistical learning: data mining, inference, and prediction. J Am Stat Assoc 2010;99:567.
41. Arpit D, Wu S, Natarajan P, Prasad R, Natarajan P. editors. Ridge Regression based classifiers for large scale class imbalanced datasets. Applications of Computer Vision; 2013.
42. Zhang Z, Hong Y. Development of a novel score for the prediction of hospital mortality in patients with severe

sepsis: the use of electronic healthcare records with LASSO regression. Oncotarget 2017;8:49637-45.
43. Guyon I, Weston J, Barnhill S, Vapnik V. Gene Selection for Cancer Classification using Support Vector Machines. J Mach Learn 2002;46:389-422.
44. van Dijk LV, Brouwer CL, van der Laan HP, Burgerhof JG, Langendijk JA, Steenbakkers RJ, Sijtsema NM. Geometric Image Biomarker Changes of the Parotid Gland Are Associated With Late Xerostomia. Int J Radiat Oncol Biol Phys 2017;99:1101-10.

Cite this article as: Liu Y, Shi H, Huang S, Chen X, Zhou H, Chang H, Xia Y, Wang G, Yang X. Early prediction of acute xerostomia during radiation therapy for nasopharyngeal cancer based on delta radiomics from CT images. Quant Imaging Med Surg 2019;9(7):1288-1302. doi: 10.21037/qims.2019.07.08

Supplementary

Table S1 The optimal result for PSA_{10F}, $PNSA_{10F}$, PSA_{30F} and $PNSA_{30F}$

PSA	Feature	Score calculation	Feature selection	K	DecisionTreeRegressor	ExtraTreesRegressor	Lasso	LinearRegression	RandomForestRegressor	Ridge	RidgeCV	SVR	R_MSE
PSA_{10F}	FV_{0F}	f_regression	SelectKBest	8	2.6147	2.8418	2.2806	3.7410	2.6248	2.1036*	2.1297	4.6501	3.6886
	NFV_{0F}	RidgeCV	RFE	6	3.6999	3.0042	4.4033	0.9548*	2.9245	3.3750	1.7392	4.1900	
	NFV_{0F}	RidgeCV	RFE	8	3.3218	3.2805	4.4033	0.9042*#	3.5698	3.4684	1.6890	4.2368	
$PNSA_{10F}$	FV_{0F}	RandomForestRegressor	RFE	11	0.1133	0.0785	0.0706	0.0604	0.0778	0.0321*	0.0373	0.0575	0.0496
	NFV_{0F}	RidgeCV	RFE	12	0.0201	0.0157	0.0563	0.0059*	0.0182	0.0301	0.0100	0.0352	
PSA_{30F}	FV_{0F}	mutual_info_regression	SelectKBest	11	0.9863	0.4195*	0.7308	1.0910	0.5824	0.7310	71.58	0.7369	0.7297
	FV_{10F}	RandomForestRegressor	RFE	2	0.4460	0.2612*	0.4938	0.5413	0.2653	0.4633	0.4749	0.6175	
	{FV_{0F}, FV_{10F}}	RandomForestRegressor	RFE	2	0.5187	0.2910*	0.4729	0.6013	0.3823	0.4635	0.4782	0.6419	
	ΔFV_{10F-0F}	ExtraTreesRegressor	RFE	8	1.4424	0.7342	0.6158	0.4229*	0.8592	0.5485	4.7127	0.7369	
	{FV_{0F}, FV_{10F}, ΔFV_{10F-0F}}	RandomForestRegressor	RFE	7	0.3826*	0.6701	1.0878	2.0266	0.5011	0.9635	1.0891	0.7369	
	NFV_{0F}	RidgeCV	RFE	13	0.7356	0.6071	0.7349	0.4737	0.6027	0.4141	0.1625*	0.7000	
	NFV_{10F}	RidgeCV	RFE	7	1.1659	0.6228	0.7349	0.1738*	0.6056	0.5724	0.2391	0.7403	
	{NFV_{0F}, NFV_{10F}}	RidgeCV	RFE	15	1.0848	0.6059	0.7349	0.6979	0.6803	0.3775	0.1270*	0.6998	
	ΔNFV_{10F-0F}	LinearRegression	RFE	14	1.3734	0.5227	0.7349	0.0754*#	0.6997	0.3747	0.0990	0.6791	
	{NFV_{0F}, NFV_{10F}, ΔNFV_{10F-0F}}	RidgeCV	RFE	14	1.2432	0.6850	0.7349	0.0801*#	0.7759	0.3675	0.1076	0.6781	
$PNSA_{30F}$	FV_{0F}	DecisionTreeRegressor	RFE	2	0.0155	0.0139	0.0244	0.0260	0.0126*	0.0260	0.0260	0.0267	0.0262
	FV_{10F}	RandomForestRegressor	RFE	7	0.0288	0.0152	0.0236	0.0121*	0.0189	0.0158	30.41	0.0267	
	{FV_{0F}, FV_{10F}}	DecisionTreeRegressor	RFE	8	0.0225	0.0129*	0.0208	0.0335	0.0191	0.0272	0.8601	0.0267	
	ΔFV_{10F-0F}	DecisionTreeRegressor	RFE	3	0.0255	0.0122*	0.0268	0.0157	0.0174	0.0245	0.0273	0.0267	
	{FV_{0F}, FV_{10F}, ΔFV_{10F-0F}}	DecisionTreeRegressor	RFE	9	0.0125*	0.0225	0.0249	209544	0.0184	0.0322	0.8942	0.0267	
	NFV_{0F}	RidgeCV	RFE	14	0.0508	0.0194	0.0263	0.0119	0.0230	0.0141	0.0050*	0.0154	
	NFV_{10F}	RidgeCV	RFE	16	0.0351	0.0176	0.0263	0.0078	0.0187	0.0150	0.0048*	0.0176	
	{NFV_{0F}, NFV_{10F}}	LinearRegression	RFE	15	0.0316	0.0172	0.0263	0.0161	0.0214	0.0123	0.0034*	0.0144	
	ΔNFV_{10F-0F}	RidgeCV	RFE	15	0.0276	0.0167	0.0263	0.0045	0.0240	0.0114	0.0029*	0.0161	
	{NFV_{0F}, NFV_{10F}, ΔNFV_{10F-0F}}	RidgeCV	RFE	16	0.0320	0.0166	0.0263	0.0013*	0.0183	0.0100	0.0018	0.0144	

This table shows the optimal result for PSA_{10F}, $PNSA_{10F}$, PSA_{30F} and $PNSA_{30F}$, where SA means saliva amount, PSA means prediction saliva amount, PNSA means prediction normalized saliva amount, PSA_{10F} means predicting saliva amount at 10th fraction, FV means feature value, NFV means normalized feature value and ΔFV_{10F-0F} means the feature difference between FV_{10F} and FV_{0F}. K represents the number of features selected by the feature selection function. Columns 6 to 13 were saliva amount prediction methods. * was the optimal MSE of one row. R_MSE (reference mean square error) was a reference value. # means the selected models presented in our papers.

Table S2 The six features and weights of predicting PSA_{10F}

ID	Ranking	Feature	Weights
1	No.6	0F_wavelet-hhh_firstorder_median_parotid_r	–3.5515
2	No.3	0F_wavelet-hhh_glcm_correlation_parotid_l	–4.461
3	No.1	0F_wavelet-hll_glszm_lowgraylevelzoneemphasis_parotid_l	6.568
4	No.5	0F_wavelet-lhh_firstorder_mean_parotid_r	3.8897
5	No.2	0F_wavelet-lhh_glrlm_longrunlowgraylevelemphasis_parotid_r	6.1709
6	No.4	0F_saliva amount	–5.1888
		Bias	5.2693

Using RidgeCV_RFE_LinearRegression model with NFV_{0F} datasets to predict SA_{10F}, and the experimental result with K=6 is second best. The related weights of the six features were given in this table, whose MSE was 0.9548 and R^2 score was 0.7149. SA, saliva amount; PSA, predicted saliva amount.