









## A MACHINE LEARNING-BASED MODEL FOR THE PREDICTION OF XEROSTOMIA AFTER HEAD AND NECK RADIOTHERAPY

YEDEKCI, Yağız <sup>1</sup>, YILMAZ, Melek Tugce <sup>1</sup>, SARI, Sezin Yuce <sup>1</sup>, CENGİZ, Mustafa <sup>1</sup>, OZYIGIT, Gokhan <sup>1</sup>, YAZICI, Gozde <sup>1</sup>

<sup>1</sup> Department of Radiation Oncology, Faculty of Medicine, Hacettepe University, Ankara 06100, Turkey

### Correspondence:

Yagiz Yedekci Department of Radiation Oncology, Faculty of Medicine, Hacettepe University, Ankara 06100, Turkey  
[yagiz.yedekci@hacettepe.edu.tr](mailto:yagiz.yedekci@hacettepe.edu.tr)

Received: 6 February 2026

Revised: 10 March 2026

Accepted: 30 March 2026

### ABSTRACT

**Objective:** The aim of this study is to create an ML model to predict xerostomia in patients undergoing RT.

**Methods and Materials:** Data from patients with LAHNC treated in our department between 2012 and 2017 were used in this retrospective study. The patients were selected with an aim of a minimum follow-up period of longer than two years. RIX was assessed with the XI which includes 11 questions. The ML method of random forest (RF) was used for the prediction of RIX. We developed the model using dosimetric parameters as input and RIX scores as output.

**Findings:** The model worked with an accuracy rate of 96.6% to generate the importance of the survey questions on the total RIX score. The question "I have difficulties swallowing certain foods" was the most effectual and "I suck sweets or cough lollies to relieve dry mouth" was the least effectual question, respectively. Also, our ML model predicted the RIX scores with an accuracy of 81%.

**Conclusion:** The questions in the XI have different impacts on the total score. Our ML model showed that the XI items could be reduced from 11 to nine, costing an only 0.3% decrease in accuracy. The ML model offers to delineate the parotid stem cells and record the doses to these cells for a more accurate assessment of RIX.

**Keywords:** Machine learning, Xerostomia, head and neck radiotherapy

## INTRODUCTION

Radiation-induced xerostomia (RIX) is one of the most common late side effects of head and neck radiotherapy (RT). Xerostomia is a subjective feeling of dry mouth which significantly affects the quality of life (QoL). Clinical symptoms include not only pain, tooth decay, tooth loss, and oral infections, but also swallowing and speech disorders, and a decreased sense of taste [1]. Since the main goal of the treatment of locally advanced head and neck cancer (LAHNC) is organ preservation, anatomically sparing the organ with a dysfunction contradicts the basic philosophy of the treatment. Therefore, protecting the function of the salivary glands should be one of the main goals of the RT treatment plan.

RIX may develop early during the course of radiotherapy and frequently persists after treatment completion. The saliva flow decreases by 50-60% in the first week of RT and to 20% by the 7th week [2]. Depending on the dose received by the relevant

salivary gland and the volume irradiated, recovery may occur 12-18 months after RT; however, permanent functional impairment is frequently observed.[3]. Although xerostomia rates in LAHNC have decreased with modern RT techniques and parotid sparing treatment approaches, the rate of RIX is still reported approximately 40% [4]. For this reason, the search for a preventive or repairing treatment approach to salivary gland damage continues. These approaches include parotid stem cell sparing treatment, salivary gland transfer techniques, and gene therapies [5-8]. Since the most effective management of radiation-induced xerostomia is its prevention, reliable predictive tools are critically needed prior to the initiation of RT. Accurate pre-treatment prediction is essential to support individualized treatment planning and to optimize organ-sparing strategies in modern radiotherapy.

In recent years, there has been a great deal of interest in using machine learning (ML) algorithms to predict the outcomes of RT [9,10]. ML approaches offer significant advantages in radiotherapy toxicity

prediction, particularly due to their ability to model complex relationships among multiple clinical and dosimetric variables. ML algorithms receive and analyze existing data to recognize the embedded pattern in data and predict output values in an acceptable range. Modeling the relationship between treatment-induced side effects and related covariates is an ideal application area for ML approaches. Although the mean dose to the parotid gland is a good predictor of RIX [11], a number of studies revealed that various patient and treatment-related factors also have an effect on RIX such as the doses to the submandibular glands, oral cavity and parotid stem cells, and patient related factors (e.g. age, educational status, smoking) [12,13]. While previous ML models have demonstrated promising results in predicting radiotherapy-induced toxicities using these clinical and dosimetric features, many rely solely on physician-graded objective scales and overlook nuanced, patient-reported symptoms. Integrating detailed subjective symptom evaluations into ML frameworks remains a relatively unexplored frontier. In this context, developing a novel ML algorithm that synergizes conventional dosimetric parameters with specific, patient-reported outcome measures is crucial for advancing existing work and achieving a more precise RIX prediction.

Being a subjective feeling, the measurement of the presence and grading of xerostomia is not easy. Collecting the saliva and measuring its amount is one objective but onerous way. A more practical and clinically feasible alternative for evaluating xerostomia involves the use of validated patient-reported questionnaires. One widely utilized instrument is the Xerostomia Inventory (XI), which systematically evaluates the severity of xerostomia symptoms during daily activities [14].

The aims of this study are to create an ML model using dosimetric parameters and the XI scores to predict xerostomia in patients undergoing RT, and to define the importance of the items in the XI.

## MATERIAL AND METHODS

### Clinical Data

Data from patients with LAHNC treated in our department between 2012 and 2017 were used in this retrospective study. Patients <70 years of age without distant metastases that received RT with or without induction and/or concurrent chemotherapy were included in the study. Exclusion criteria were a histopathology other than squamous cell carcinoma, a

comorbid disease or medication that may cause xerostomia, a total dental prosthesis, and a prior, postoperative and/or unilateral RT to the neck and/or primary tumor site. The patients were selected with an aim of a minimum follow-up period of longer than two years.

All treatment plans were generated using the Eclipse treatment planning system (version 8, Varian). Radiotherapy was delivered with the intensity-modulated radiotherapy (IMRT) technique via Clinac DHX linear accelerators. The RT protocol was as follows: 70 Gy to the gross tumor volume (both in the primary tumor site and for gross lymph nodes), 60 Gy to the whole organ of the primary tumor site and involved nodal neck area, and 54 Gy to the elective nodal area in 33 fractions using a simultaneous integrated boost (SIB) technique. Routine adaptive re-planning during the treatment course was not performed. Dose constraints for the parotid glands (PG) and other organs at risk (OARs) where minor salivary glands are widely distributed are given in Table 1. No dose constraints were defined for the parotid stem cells or submandibular glands during optimization, but the dose-volume histogram (DVH) parameters for these structures were systematically recorded. To ensure standardized delineation, the parotid stem cell regions were specifically contoured in accordance with the previously published institutional guidelines detailed by Sari et al.[7]. RIX was assessed with the XI, which includes 11 questions. The Turkish version of the XI questionnaire was administered to all patients via telephone at a cross-sectional point with a minimum of 2 years of follow-up (median follow-up 55 months, range: 24-94.3 months) to allow sufficient time for the recovery of the salivary glands. Each question has an answer on a scale of 1 to 5 (never=1, hardly ever=2, occasionally=3, fairly often=4, and very often=5), and the overall score is between 11 and 55. A higher score indicates a more severe RIX. This study was approved by the Institutional Review Board (Project number: GO 20/1061)

### Machine learning

In this study, the ML method of random forest (RF) was used for the prediction of RIX. RF is an ensemble learning method which creates decision trees on randomly selected data samples. The prediction from each tree is evaluated and the best solution is selected by means of voting. RF can be used both for classification and regression. Also, it allows defining the relative feature importance which helps selecting the most contributing features for the target value. The RF algorithm was carried out using

the Python programming language (version 3.8) and the Scikit-Learn library (Python Software Foundation). To ensure reproducibility, the key hyperparameters of the RF model were explicitly defined: the number of trees in the forest (*n\_estimators*) was set to 200, the maximum depth of the trees (*max\_depth*) was set to 10, and the minimum number of samples required to split an internal node (*min\_samples\_split*) was 2. A random state of 42 was assigned to maintain consistent results across multiple runs.

The RF algorithm was applied to two different data;

-To the XI, to obtain the importance of the survey questions on the total RIX score, and

-To the dosiomics, to predict the XI scores.

First, the model was implemented to the XI data. The survey questions were selected as the features to obtain the importance of the variables in the XI. Secondly, we investigated the relation between the dosimetric parameters and RIX by using the RF algorithm. For this purpose, the mean doses to the lips, pharynx, submandibular glands, oral cavity, parotid gland and stem cells, and the maximum doses to the parotid gland and stem cells were evaluated. The V26Gy (the volume receiving 26 Gy or higher) and V30Gy values for the parotid glands were also recorded. The above mentioned dosiomics were selected as features. The ML model was developed to predict the total score of the XI. The data was split into two parts; 75% to train the algorithm and 25% to test the model. The prediction output by the model was compared to the actual scores in order to obtain the final results. The accuracy was used as the model performance indicator. Correctly classified instances are known as accuracy which is defined in the following equation;

$$\text{Accuracy} = (Tp+Tn)/(Tp+Tn+Fp+Fn)$$

In the context of predicting radiation-induced xerostomia (RIX), these variables are explicitly defined as follows:

True Positive (TP): The model correctly identifies a patient as having severe RIX.

True Negative (TN): The model correctly identifies a patient as not having severe RIX.

False Positive (FP): The model incorrectly predicts that a patient has severe RIX when they actually do not.

False Negative (FN): The model incorrectly predicts that a patient does not have severe RIX when they actually do. The mean squared error (mse) and the Pearson correlation coefficient *r* were calculated for the predicted and true values of the total XI score.



Figure 1: The importance of XI questions on total score.

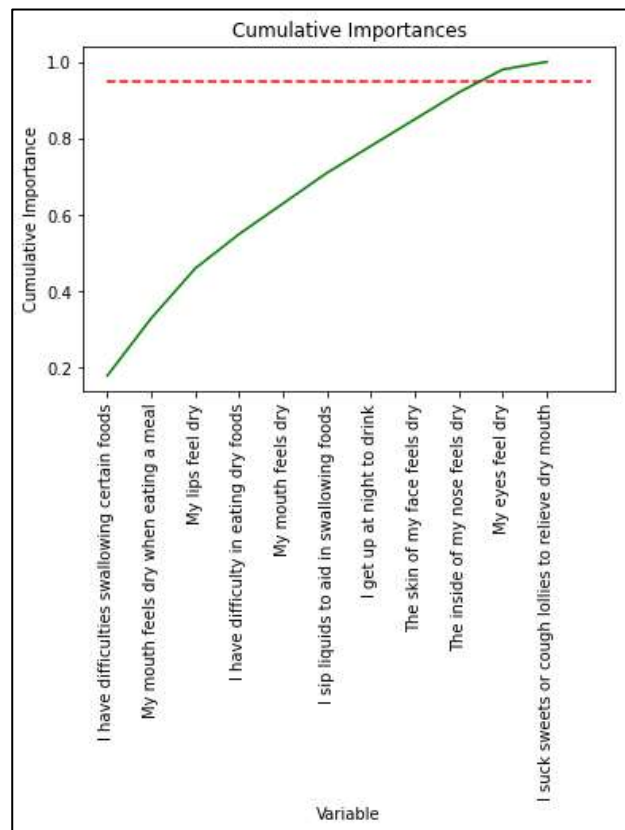


Figure 2: Cumulative importance of variables, dashed line: 95% of the cumulative importance.

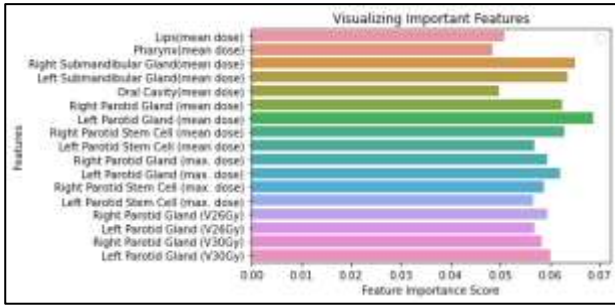


Figure 3: The importance of dose volume histogram (DVH) parameters on XI scores were evaluated using the random forest model.

Table 1. Dose constraints for the structures

Organ at risk	Constraint
Lips	mean <20 Gy
Pharynx	mean <60 Gy
Oral cavity	mean <35 Gy
Parotid gland	mean <26 Gy OR V <sub>30Gy</sub> * <50%

\*V30Gy: Volume receiving ≥30 Gy

## RESULTS

A total of 880 patients treated with IMRT for LAHNC were initially evaluated. Following the application of strict exclusion criteria—which eliminated patients with a follow-up of less than 2 years, age over 70, metastatic disease, postoperative or unilateral RT, non-squamous cell carcinoma histology, and factors intrinsically affecting salivary function (e.g., total dental prosthesis, specific comorbidities, or medications)—87 patients were deemed eligible for the final analysis. The characteristics of all patients are given in Table 2. All patients were administrated the XI and the results were evaluated with our ML model which aimed to define the importance of the survey questions over the XI total score.

The model worked with an accuracy rate of 96.6%. The relative importance of the questions is given in

Fig. 1. The question “I have difficulties swallowing certain foods” was the most effectual and “I suck sweets or cough lollies to relieve dry mouth” was the least effectual question, respectively. Fig. 2 shows the cumulative importance weights of the variables. The dashed line shows 95% of the total importance. The total importance impact of “My eyes feel dry” and “I suck sweets or cough lollies to relieve dry mouth” on the cumulative weight was less than 5%. When we omitted these questions from the questionnaire and ran the model again, the accuracy decreased to 96.3%. In the next stage, we selected the most effectual five items, and the model worked with an accuracy of 86%.

Table 3 shows the mean doses obtained from the patients’ data. Our ML model predicted the RIX scores with an accuracy of 81%. Fig. 3 shows the importance of the scores of the DVH parameters on RIX. The importance of the scores ranged between 0.07 and 0.04, meaning the impact of the related DVH parameters on the total score was between 7% and 4%. Fig. 4 shows the prediction and true values of the total XI scores with 7.1 mse and 0.91 r value. After omitting the least effectual questions from the data, we recalculated the model in which the accuracy was 80.5%, and the mse and r values were 7.4 and 0.89.

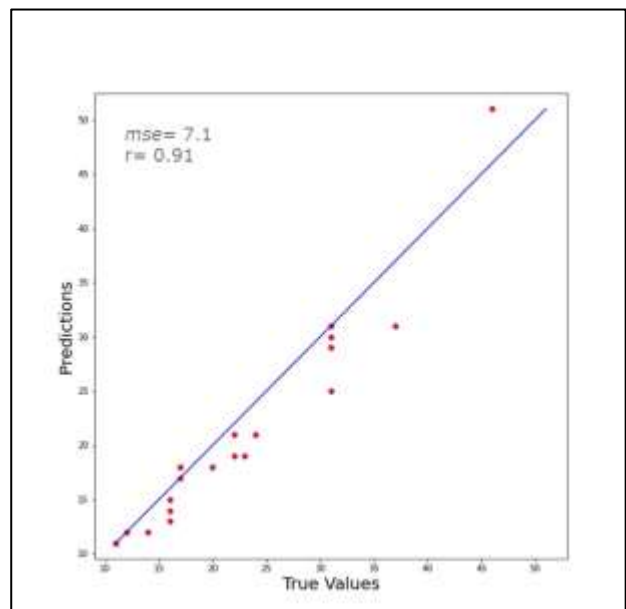


Figure 4: The predicted values using the random forest model and true values of XI scores are plotted, together with the mean square error and Pearson correlation coefficient.

## DISCUSSION

The primary objective of this study was to evaluate the predictive value of comprehensive dosimetric parameters on patient-reported xerostomia using a machine learning approach. While previous studies have utilized ML for toxicity prediction, to the best of our knowledge, this is the first study to employ an ML regression framework specifically designed to predict continuous, patient-reported XI scores using a combination of whole-gland and parotid stem cell DVH parameters. After calculating the feature importance of each question regarding the overall score, the results demonstrated that not all collected survey data are strictly necessary for robust predictions. Notably, it was found that the XI could be streamlined from 11 to 9 items, as two specific questions accounted for a combined relative weight of less than 5%. Eliminating these low-impact features resulted in a marginal decrease of approximately 0.3% in the model's overall predictive performance, a reduction that is considered statistically and clinically negligible.

Previously, the XI was shortened in a Dutch version [15]. In the Summated XI-Dutch version, the five original items were used (i.e. my mouth feels dry when eating a meal, my mouth feels dry, I have difficulty in eating dry foods, I have difficulties swallowing certain foods, and my lips feel dry). Our results with ML also showed that these questions were the five most effectual questions on the total XI score. However, the model accuracy decreased from 96.6% to 86% in the summated form in our study. In terms of ML, reducing the number of the inputs will provide a significant calculation time advantage in an RF ML model, particularly in situations where large data is used. The Dutch version of the XI can be used in an ML model but for more accurate results, we offer the XI with nine items as summated version.

In this study, we developed an ML model using the RF algorithm. The RF algorithm is frequently used for predicting outcomes and side effects. Previously, Deist et al. [9] tested the performance of six common algorithms including decision tree, RF, neural network, support vector machine, elastic net logistic regression, and Logit Boost to predict the outcomes and toxicity of chemoradiotherapy. They reported that the RF and elastic net logistic regression provided higher performance compared to the other algorithms. There are also disease site-based studies that showed RF has the best performance to predict toxicity [16,17]. The results indicate that the RF algorithm was suitable for predicting xerostomia related to head and neck RT.

The ML model learns the parameters from the available data which makes the characteristics of the dataset absolutely crucial. We used dosimetric parameters and the total XI score to train the model and predict RIX. The most important advantage of our model is that it predicts RIX by considering the doses of all structures that can cause xerostomia. Experimental results have demonstrated that the proposed model can accurately predict the total XI score. Accurately predicting RIX will facilitate clinical decision making for personalized treatment. Despite the promising predictive performance of the proposed ML regression framework, several methodological limitations must be acknowledged. First, the current model relies exclusively on dosimetric parameters and does not incorporate the potential impact of confounding patient-specific clinical factors, such as age, comorbidities, smoking status, or concurrent medications. For a more comprehensive and advanced level of modeling, future algorithms should integrate these clinical variables alongside dosimetric features. Second, the analysis was based on a relatively small cohort of 87 patients treated at a single institution, which may introduce selection bias and limit the generalizability of the findings. Finally, the lack of an external validation dataset means that the robustness and reproducibility of the model across different clinical settings remain unconfirmed. Future multicenter studies with larger, diverse cohorts are warranted to externally validate and refine the proposed predictive tool.

Previously, it was shown that the total XI score was correlated with the mean doses to the parotid gland, parotid stem cell niches, oral cavity, lips, and submandibular glands [7]. However, there was no correlation with the mean dose to the pharynx [7]. Our ML model showed that all dosimetric parameters have an impact on the total XI score. Salivary glands in the oral cavity and pharynx are minor contributors, secreting less than 10% of the saliva [18]. The clinical findings were compatible with our results. The impact of the mean dose to the oral cavity and pharynx was relatively lower than other features.

We selected the mean and maximum doses to the parotid stem cells as features. As expected, the dose to these cells was one of the most effectual features for RIX prediction. Recently, Steenbakkers et al. [19] showed that the dose to the stem cell region was more predictive for the development of parotid gland function-related xerostomia compared to the dose to the entire parotid gland in a double-blind randomized controlled trial. Since the adequate dose reduction to the parotid glands is not always possible, the parotid

stem cell-sparing treatment approach gains more importance. The results obtained in this study also showed that sparing the parotid stem cells was crucial for preventing xerostomia.

More than one critical structure plays a role in the occurrence of RIX and different dose constraints have been defined in the literature for each of these structures. The dose given to each structure may vary depending on the location of the primary tumor and in some cases, it is not even possible to meet the criteria. Therefore, estimating the multi-factor RIX is quite challenging. Our proposed model considers the doses of all structures associated with xerostomia and allows us to predict the optimal probability for RIX development.

## CONCLUSION

The questions in the XI have different impacts on the total score. Our ML model showed that the XI items could be reduced from 11 to nine, costing an only 0.3% decrease in accuracy. Our proposed ML model can predict the XI score. This prediction considers the doses absorbed by the structures associated with xerostomia. Our results also showed that the doses to the parotid stem cells have an impact on the total XI score. Therefore, we offer to delineate the parotid stem cells and record the doses to these cells for a more accurate assessment of RIX.

## Conflict of Interest

There are no conflicts of interest and no acknowledgements.

## References

1. Wijers OB, Levendag PC, Braaksmma MM, Boonzaaijer M, Visch LL, Schmitz PI: Patients with head and neck cancer cured by radiation therapy: a survey of the dry mouth syndrome in long-term survivors. *Head Neck*. 2002, 24:737-747. 10.1002/hed.10129
2. Franzén L, Funegård U, Ericson T, Henriksson R: Parotid gland function during and following radiotherapy of malignancies in the head and neck. A consecutive study of salivary flow and patient discomfort. *Eur J Cancer*. 1992, 28:457-462. 10.1016/s0959-8049(05)80076-0
3. Roesink JM, Moerland MA, Battermann JJ, Hordijk GJ, Terhaard CH: Quantitative dose-volume response analysis of changes in parotid gland function after radiotherapy in the head-and-neck region. *Int J Radiat Oncol Biol Phys*. 2001, 51:938-946. 10.1016/s0360-3016(01)01717-5
4. Nutting CM, Morden JP, Harrington KJ, et al.: Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): a phase 3 multicentre randomised controlled trial. *Lancet Oncol*. 2011, 12:127-136. 10.1016/s1470-2045(10)70290-4
5. Delporte C, O'Connell BC, He X, et al.: Increased fluid secretion after adenoviral-mediated transfer of the aquaporin-1 cDNA to irradiated rat salivary glands. *Proc Natl Acad Sci U S A*. 1997, 94:3268-3273. 10.1073/pnas.94.7.3268
6. Spiegel JH, Deschler DG, Cheney ML: Microvascular Transplantation and Replantation of the Rabbit Submandibular Gland. *Archives of Otolaryngology-Head & Neck Surgery*. 2001, 127:991-996. 10.1001/archotol.127.8.991
7. Sari SY, Yilmaz MT, Elmali A, et al.: Parotid gland stem cells: Mini yet mighty. *Head Neck*. 2021, 43:1122-1127. 10.1002/hed.26556
8. Steenbakkens R, van Rijn-Dekker MI, Stokman MA, et al.: Parotid Gland Stem Cell Sparing Radiation Therapy for Patients With Head and Neck Cancer: A Double-Blind Randomized Controlled Trial. *Int J Radiat Oncol Biol Phys*. 2022, 112:306-316. 10.1016/j.ijrobp.2021.09.023
9. Deist TM, Dankers FJ, Valdes G, et al.: Machine learning algorithms for outcome prediction in (chemo) radiotherapy: An empirical comparison of classifiers. *Medical physics*. 2018, 45:3449-3459.
10. Isaksson LJ, Pepa M, Zaffaroni M, et al.: Machine Learning-Based Models for Prediction of Toxicity Outcomes in Radiotherapy. *Frontiers in Oncology*. 2020, 10. 10.3389/fonc.2020.00790
11. Deasy JO, Moiseenko V, Marks L, Chao KSC, Nam J, Eisbruch A: Radiotherapy dose-volume effects on salivary gland function. *Int J Radiat Oncol Biol Phys*. 2010, 76:S58-S63. 10.1016/j.ijrobp.2009.06.090
12. Guchelaar HJ, Vermes A, Meerwaldt JH: Radiation-induced xerostomia: pathophysiology, clinical course and supportive treatment. *Support Care Cancer*. 1997, 5:281-288. 10.1007/s005200050075
13. Dirix P, Nuyts S, Van den Bogaert W: Radiation-induced xerostomia in patients with head and neck cancer. *Cancer*. 2006, 107:2525-2534. <https://doi.org/10.1002/cncr.22302>
14. Thomson WM, Chalmers JM, Spencer AJ, Williams SM: The Xerostomia Inventory: a multi-item approach to measuring dry mouth. *Community Dent Health*. 1999, 16:12-17.

15. van der Putten G-J, Brand HS, Schols JM, de Baat C: The diagnostic suitability of a xerostomia questionnaire and the association between xerostomia, hyposalivation and medication use in a group of nursing home residents. *Clinical oral investigations*. 2011, 15:185-192.

16. Dean JA, Wong KH, Welsh LC, et al.: Normal tissue complication probability (NTCP) modelling using spatial dose metrics and machine learning methods for severe acute oral mucositis resulting from head and neck radiotherapy. *Radiotherapy and Oncology*. 2016, 120:21-27.

17. Saednia K, Tabbarah S, Lagree A, et al.: Quantitative Thermal Imaging Biomarkers to Detect Acute Skin Toxicity From Breast Radiation Therapy Using Supervised Machine Learning. *International Journal of Radiation Oncology\*Biography\*Physics*. 2020,106:1071-1083.  
<https://doi.org/10.1016/j.ijrobp.2019.12.032>

18. Humphrey SP, Williamson RT: A review of saliva: normal composition, flow, and function. *The Journal of prosthetic dentistry*. 2001, 85:162-169.

19. Steenbakkens RJHM, van Rijn–Dekker MI, Stokman MA, et al.: Parotid Gland Stem Cell Sparing Radiation Therapy for Patients With Head and Neck Cancer: A Double-Blind Randomized Controlled Trial. *International Journal of Radiation Oncology\*Biography\*Physics*.2022,112:306-316.  
<https://doi.org/10.1016/j.ijrobp.2021.09.023>