

## **Precision toxicity correlates of tumor spatial proximity to organs at risk in cancer patients receiving intensity-modulated radiotherapy**

### ***†-Contributing authors:***

Andrew Wentzel<sup>a</sup>, Peter Hanula<sup>a</sup>, Lisanne V. van Dijk, PhD<sup>b</sup>, Baher Elgohari<sup>b,f</sup>, Abdallah S. R. Mohamed, MD, MSc<sup>b</sup>, Carlos E. Cardenas, PhD<sup>c</sup>, Clifton D. Fuller, MD, PhD<sup>b</sup>, David M. Vock<sup>d</sup>, Guadalupe Canahuate<sup>e</sup>, G. Elisabeta Marai<sup>a</sup>

### ***Affiliations:***

<sup>a</sup>Department of Computer Science, The University of Illinois at Chicago, Chicago, IL, 60612, USA.

<sup>b</sup>Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, 77030 USA.

<sup>c</sup>Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, TX, 77030 USA.

<sup>d</sup>Division of Biostatistics, University of Minnesota, Minneapolis, MN, 55455, USA.

<sup>e</sup>Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA, 52242, USA.

<sup>f</sup>Department of Clinical Oncology and Nuclear Medicine, Mansoura University, Mansoura, Dakhalia, 35516, Egypt

### ***\*Co-Corresponding authors:***

Correspondence to Andrew Wentzel (awentze2@uic.edu) and/or G. Elisabeta Marai, PhD (gmarai@uic.edu).

Total number of words = 3419 tables=3, and figures=3

***Running head: Toxicity correlates of tumor location***

### ***Keywords***

Radiation therapy; Oropharynx cancer; Head and Neck cancer; Tumor location; Intensity-modulated radiotherapy; Data Interpretation, Statistical; Medical Informatics; Dysphagia;

### ***†- Co-author specific contributions:***

All listed co-authors performed the following:

- 1. Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work;*

2. *Drafting the work or revising it critically for important intellectual content;*
3. *Final approval of the version to be published;*
4. *Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.*

Specific additional individual cooperative effort contributions to study/manuscript design/execution/interpretation, in addition to all criteria above are listed as follows:

- *AW, PH, GEM - designed and developed similarity measure, data extraction and curation, statistical analysis, and interpretation*
- *LVD, BE, ASRM, CC - direct patient care provision, direct clinical data collection; interpretation and analytic support*
- *GC - supervised statistical analysis, data extraction, graphic construction*
- *DV, CDF - analytic support, guarantor of statistical quality*
- *AW, GC, LVD, ASRM, CDF, GEM - manuscript writing and editing*
- *GC, CDF, GEM - primary investigator(s); conceived, coordinated, and directed all study activities, responsible for data collection, project integrity, manuscript content and editorial oversight and correspondence*

***Prior presentation:***

Preliminary analyses and portions of this data were accepted for a poster presentation at the 2019 American Society of Radiation Oncology (ASTRO) Annual Meeting, September 25 - 28, 2019, Chicago, IL, USA.

A detailed description of the novel dose prediction method used in this work has been published in the IEEE Transactions on Visualization and Computer Graphics journal, and has been orally presented at the IEEE VIS conference, October 21-25, 2019, Vancouver, BC, Canada.

**Abstract**

**Purpose:** Using a 200 Head and Neck cancer (HNC) patient cohort, we employ patient similarity based on tumor location, volume, and proximity to organs at risk to predict radiation-associated dysphagia (RAD) in a new patient receiving intensity modulated radiation therapy (IMRT).

**Material and Methods:** All patients were treated using curative-intent IMRT. Anatomical features were extracted from contrast-enhanced tomography scans acquired pre-treatment. Patient similarity was computed using a topological similarity measure, which allowed for the prediction of normal tissues' mean doses. We performed feature selection and clustering, and used the resulting groups of patients to forecast RAD. We used Logistic Regression (LG) cross-validation to assess the potential toxicity risk of these groupings.

**Results:** Out of 200 patients, 34 patients were recorded as having RAD. Patient clusters were significantly correlated with RAD ( $p < .0001$ ). The area under the receiver-operator curve (AUC) using pre-established, baseline features gave a predictive accuracy of .79, while the addition of our cluster labels improved accuracy to .84.

**Conclusion:** Our results show that spatial information available pre-treatment can be used to robustly identify groups of RAD high-risk patients. We identify feature sets that considerably improve toxicity risk prediction beyond what is possible using baseline features. Our results also suggest that similarity-based predicted mean doses to organs can be used as valid predictors of risk to organs.

## Introduction

Radiation-associated dysphagia (RAD) is one of the severe sequelae of treatment in head and neck (HNC) cancer patients undergoing radiation therapy (RT), with chronic toxicity arising even after acute symptoms have ceased [1]. Chronic RAD is even more relevant in the era of Human Papillomavirus associated (HPV) HNC, where the majority of patients have curable disease with prolonged survival, and thereby endure later toxicities which are otherwise not encountered in patients with aggressive HPV negative disease and relatively shorter survival duration. Proper assessment of the risk of chronic RAD is essential to identify appropriate approaches to prevent and/or early treat patients before the occurrence of advanced, crippling toxicity [2,3,4].

Several studies have demonstrated that risk factors such as patient age and tumor subsite are associated with the development of late RAD [2,5]. Furthermore, the dose administered to anatomical structures such as the swallowing muscles has been used in normal tissue complication probability (NTCP) models [2,3,4,6] to predict the risk of chronic RAD following RT planning. However, these NTCP models require information that is only available after the development of two radiation treatment plans for that

patient, which is extremely time and resource expensive. Other examples of non-dosimetric clinical (surrogate) markers that may have a relationship with the risk of developing RAD are the tumor size or extension (e.g. TNM staging) [6,7], the location of high dose regions, and muscle invasion. Finally, the variations in spatial organization of organs at risk around GTVs may be equally important.

Using these complex tumor and anatomical spatial distributions to identify and categorize similar patients can potentially provide a patient grouping methodology for RAD risk assessment **before** radiotherapy planning. To this end, risk assessment at the initial diagnosis, before radiation plans are available, would be extremely valuable in helping physicians make informed decisions regarding the treatment plan for personalized cancer treatments.

Most studies that look at stratification of patients, such as TNM staging, are centered on overall survival, rather than toxicities. This study proposes a novel HNC toxicity risk criterion based on unsupervised clustering of similar patients using diagnostic imaging data. We hypothesize that spatial characteristics of target volumes and surrounding organs at risk, that are known during the initial diagnosis, can play an important role in determining risk of post-treatment swallowing complications in HNC patients. We further hypothesize that groups derived from unsupervised clustering of patients in the cohort, based on these tumor and organ at risk features, are associated with RAD. Furthermore, these groups act as a staging system for RAD risk. This staging improves risk prediction using TNM staging and demographic information, without requiring dosimetric information.

## **Methods**

Our model segments the cohort into 4 groups using hierarchical agglomerative clustering (HAC), a method commonly used in data-mining. First, HAC considers each patient as a separate group. HAC then selects the two groups that are 'closest' according to a given distance measure, and merges them into a single cluster. This process is repeated until HAC reaches the desired number of groups, chosen here as 4 to align with current TNM staging. Our innovation consists of the use within HAC of a novel distance measure over spatially-aware covariates. This approach allows us to automatically generate clusters that are equivalent to high and low RAD risk groups. To demonstrate the novel value of the clusters generated through this approach, we create a multivariate

regression model for predicting RAD and show that models which include our clusters outperform predictive models that rely only on standard clinical covariates.

### ***Patient Cohort***

Oropharyngeal cancer (OPC) patients who were treated using curative-intent IMRT [8] at MD Anderson Cancer Center between 2005 and 2013 were collected retrospectively using an IRB approved protocol. Demographics, diagnostic categorization and treatment information was retrospectively retrieved from the electronic medical records. Patients prospectively underwent physical and endoscopic examinations, as well as radiological and pathological assessments. RAD was assessed during follow up that occurred 6 months after completion of treatment. Inclusion criteria for our study were: 1) Pathologically proven OPC with at least 1 identified gross tumor volume (GTV), 2) Received IMRT with/without chemotherapy with curative intent, 3) Patient surviving 6 months post-treatment, and 4) Available pre-treatment imaging data for all ROIs as described below. All patients received diagnostic contrast-enhanced computed tomography (CECT) imaging. Imaging data for 245 patients were available over this period. From this set, 45 patients were excluded due to missing contouring data on one or more of the 41 OARs considered. We defined dysphagia as the presence of either a feeding tube insertion or aspiration. Aspiration rate is defined as grade 2+ aspiration per CTCAE guidelines [9]. No patients had feeding tube insertion at the baseline assessment, and three patients had pre-treatment aspiration (Table 1).

After GTVs were manually contoured [10], other tumor spatial characteristics were automatically extracted from CECT imaging data for 41 OARs and all GTVs as described in (Appendix A). Because treatment doses are not available at the time of diagnosis, mean radiation doses to each ROI were estimated using a published predictive model [11] (Appendix B). All candidate ROIs, along with their mean treatment doses, predicted mean doses, and minimum tumor-organ distances are listed in appendix table B.1. Of the spatial and dosimetric characteristics, 5 covariates that were most representative of the anatomical information relevant for predicting RAD were identified using data-mining techniques as described in (Appendix C).

## **Statistical Analysis**

5 covariates were identified for the final model: the predicted treatment doses to the extended oral cavity, mandible, and medial pterygoids; and the minimum euclidean surface distances between the GTV and mandible and medial pharyngeal constrictor. Clustering was performed using hierarchical agglomerative clustering with a weighted linkage distance [12] and the L2-norm as the distance function, and all covariates were normalized to have a mean of 0 and standard-deviation of 1 across the cohort. We report results for  $k = 4$  clusters, to be consistent with existing TMN staging.

To assess how well these clusters discriminate between high and low-risk patients when correcting for existing known clinical covariates [13], we trained logistic regression models on different combinations of existing clinical covariates (see *Existing Clinical Covariates* in Table 1), spatial covariates, and our cluster labels. To prevent overfitting our model, we used leave-one-out cross-validation for generating the prediction for each patient (i.e. when predicting the risk for a given patient, we excluded that patient when training the model) [14]. To assess the predictive power of each model, we report the area under the receiver-operator curve (AUC) for each model, which has been traditionally used to evaluate medical diagnosis tests where there are many more negative cases than positive [15]. AUC score serves as a measure of how well the prediction ranks relative risk in the cohort by comparing the number of true positives against the number of false positives as the sensitivity of the prediction is adjusted, making it a more valuable measurement than accuracy or explained variance in a model than metrics such as accuracy or precision. These results were run using feeding tube toxicity, aspiration, and combined RAD as dependent variables.

Because clusters can be sensitive to changes in the data when the dataset is small, we performed an additional experiment to assess how varying and reducing the cohort size affected our results. Specifically, we randomly removed a number of patients  $P$  from the dataset, such that at least 2 patients with RAD-related toxicity were included in the cohort. We then re-performed the clustering using our 5 selected spatial covariates on the new subset of patients, and calculated the AUC score for logistic regression as before, using the baseline clinical covariates with and without adding in our 4 spatial clusters. We repeated this process 500 times for each  $P$ , for  $0 < P < 150$  (75%) patients. The mean AUC score and 25% confidence intervals were then calculated for each  $P$  with vs. without using spatial clusters. By testing multiple

variations of the data, we can better validate that our results are not due to overfitting the model and can be applied to smaller cohorts.

Additional analysis was performed to compare the performance of our model to other swallowing related muscles as well and an analysis of the relationship between our clusters and relevant swallowing muscles (Appendix D), as well as an analysis of clusters performed on clinical features (Appendix E).

Hierarchical clustering was implemented using the scipy library [16,17]. Fisher's exact test was performed using the R software package [18] using a two-tailed test and a 95% confidence interval. Logistic regression was implemented using the scikit-learn package using L-BFGS solver [19], and an L2 regularization penalty [20] was used for its ability to provide numerical stability when dealing with many correlated variables, as in [21]. After tuning, the regularization penalty coefficient was set to 1 for all models as it yielded the highest validation AUC score for all models.

## Results

Cohort statistics, including demographics, clinical covariates, and spatial features included in later statistical analysis are reported in Table 1. At the 6-month follow-up, 34 patients (17%) required either feeding tube alone (15), aspiration (12), or both (7). Per-cluster summaries of the 5 covariates included in our clusters are detailed in Table 2.

Using these 5 spatial variables, four spatial clusters were identified. Table 2 shows a cluster breakdown of the spatial features' mean values and the percentage of patients experiencing toxicity. Visualizations of the predicted doses and tumor-organ distances for each cluster are shown in Figure 1. Cluster labels are significantly associated with RAD toxicity ( $p < .0001$ ). In particular, Cluster 4 is a high risk group with half of the patients experiencing toxicity. Clusters 1, 2 and 3 have considerably lower risk. Cluster 4 groups patients with the highest predicted doses and a larger tumor spread. Clusters 2 and 3 group patients with lower predicted doses, where Cluster 2 captures patients with lateral tumors and Cluster 3 captures patients with central tumors with considerable overlap with the tongue. Interestingly, Cluster 1 identifies three outlier patients with tumors positioned at the base of the tongue, with relatively high predicted mean doses, but surprisingly low toxicity risk. ‘

Table 3 reports the AUC from the logistic regression model with several different combinations of features included. Scores for the baseline set alone (AUC = 0.79) were markedly lower than the baseline features with spatial cluster labels added (AUC = 0.84). This difference was most pronounced in identifying feeding tube toxicity (AUC increase = .07) compared to aspiration (AUC increase = 0.02). Receiver-operator curves (ROC) for our classifier results when including vs. excluding spatial clusters are shown in Figure 2. Of the clinical covariates, T-stage was the most important predictor of RAD, and overall performance is comparable between using T-stage or the spatial clusters as the independent variables in the model. However, when both T-stage and spatial clusters are combined within a single model, AUC notably improves from .68-.70 to .82. Furthermore, when the spatial clusters are combined with the clinical features, AUC improves to .84 (the maximum observed in our experiments).

From our sensitivity analysis experiment, mean AUC scores across different cohort sizes are shown in Figure 3. Prediction improvement from including spatial information was maintained even after removing up to 75% of the cohort, with a mean AUC improvement across all tests of .028, and a mean improvement of .036 with 75% of the cohort removed. AUC improvement was confirmed to be statistically significant when comparing subsampled populations for all subsample sizes using a dependent t-test ( $p < .001$ ). Overall, our results show that our spatial clusters robustly improve prediction scores, even when there are large perturbations to the data, which supports the hypothesis that our methodology should be beneficial when generalized to similar cohorts.

## **Discussion**

Our results support the hypothesis that relative tumor-OAR positioning has a strong association to the development of late-stage dysphagia. This paper demonstrates a novel classification method based on unsupervised clustering of patient-specific anatomical features and predicted dose parameters in HNC patients receiving (chemo)-radiation therapy. We demonstrate that clustering of anatomical OAR and tumor distribution can meaningfully improve the prediction of radiation-induced toxicity, compared to commonly used non-dosimetric clinical variables, such as T-stage and age (Table 1).



The resulting clusters allow for sophisticated, combined representation of complex three-dimensional proximity of OARs to the tumor location, which is unique per patient due to variations in anatomy and tumor extent. Subsequently, we showed that the identified patient clusters, defined by these proximity features and surrogate OAR dose parameters, are highly associated with the risk of developing dysphagia-related toxicity 6 months following radiotherapy. In particular, Cluster 4 showed the strongest association with development of RAD. These results show that our clustering can identify patients with similar anatomic distribution, and related dose distributions. Despite not using learned parameters, these clusters are notably correlated with RAD.

Our final model used predicted doses to the extended oral cavity, mandible, and combined medial pterygoid muscle doses, along with the minimum distances between a GTV and the mandible and medial pharyngeal constrictor. Predicted doses for the three included OARs were highest for the high-risk cluster (Table 2, Cluster 4), suggesting that they are the most representative of the doses to all the organs around the oral cavity, which as a whole may contribute to RAD. This hypothesis is consistent with the fact that median doses are highest for all OARs for this group compared to any other group (Figure 1). Additional analysis in Appendix E, Table E.1 shows that SPC proximity is the best indicator of membership in the high risk cluster, which further supports the idea that the predicted doses in our clusters encapsulate dosimetric information relevant to swallowing or mastication muscles that are not explicitly encoded in our covariates. The extended oral cavity and mandible represent large ROIs around the mouth, and thus together serve as an indicator for the overall doses to the muscles used for mastication. The addition of the geometric mean of the predicted doses to the medial pterygoid likely further segments out patients with and strong dosing to both sides of the head rather than those with unilaterally biased dosing. The two tumor-organ distances, the mandible and MPC, represent organs central to the mouth and throat, respectively, and are thus indicative of the spread of disease near the organs responsible for mastication and swallowing. In this case, the MPC was likely selected over the SPC as it has less overlap with the throat, which is more consistently captured by the extended oral cavity. One cluster, Cluster 3, had significantly smaller tumor-organ distances for both the Mandible and MPC, as well as overall, than other clusters, despite having below-average predicted doses to relevant ROIs. Cluster 3 then represents a cluster with slightly elevated risk than the baseline due to having a high tumor spread, despite low treatment

dose. Our spatial clustering furthermore identified a surprising group with high predicted-doses but low toxicity (Cluster 1), which featured tumors located at the base of the tongue. This grouping may capture an overlooked phenomenon in estimating the NTCP in standard models where only certain OARs dosimetric criteria are considered. These standard models are likely an over-simplification, as other OARs may contribute to toxicity. Specifically, dysphagia can be induced through many different combinations of muscle, mucosal or glandular damage [22–24].

Of pre-existing clinical variables considered, T-stage was the most significant predictor of dysphagia. This is an intuitive result, considering that T-stage is an indicator of tumor size – thus an indicator of high dose reach – and potential tumor muscle infiltration. The most important aspect was whether the patients were in T-stage 4, with 72.2% of all patients in this category experiencing RAD and 55.5% experiencing aspiration. T-stage alone performed comparably to our spatial cluster labels alone (rows 1 and 2 in Table 3), likely due to the fact that they both consider similar spatial features, such as the extension of the tumor into the pterygoid muscle [25]. T-stage is likely predictive as it acts as an indicator of tumor size and the aggressiveness of the treatment a patient may receive, but doesn't capture all relevant spatial information that affects dose distribution and toxicity. In contrast, our approach captures specific radiation dose distributions across T-stage labels (e.g., T3 and T4 in the high-risk cluster), as well as additional features related to tumor-location that are specifically relevant to feeding tube toxicity or other toxicity that may manifest later. Furthermore, combining our spatial clusters with T-stage improves the toxicity prediction performance of either T-stage alone or spatial cluster alone (Table 3, row 3). Notably, the best AUC scores are observed when the cluster labels are included in the predictive model in combination with T-stage and other clinical features (bold values in Table 3).

Importantly, these results have the potential to impact early-treatment decision making, as these predictions are performed with imaging data alone, do not require time-intensive dose optimization, and can be fully automated after GTV contouring. In contrast to synergistic NTCP models that require actual dose parameters, our results indicate that adequate pre-radiation risk forecasting can be used with diagnostic CT data in combination with tumor annotation and OAR contouring. By providing a granular and continuous risk prediction at the patient level, our approach can be used to identify patients who are in need of exceptional efforts to maintain swallowing function in a

granular way. Low risk patients can be encouraged to maintain oral intake and prophylactic swallowing exercises. Intermediate risk cases can have pre-therapy assessment for short-interval PEG placement in an adaptive manner, as well as more frequent surveillance (e.g. mid-therapy MBS assessment). High-risk patients can be given nutritional support, aggressive swallowing exercises, low-threshold for PEG placement or prophylactic PEG placement, rapid effort to accelerate nutrition in the post-therapy interval (e.g. goal PEG duration <3 months), short-interval post-therapy MBS, and high-frequency post-therapy surveillance for dysphagia symptoms. High risk patients can also be pre-selected for proton therapy referral without doing an elaborate dose comparison [26]. Using our model, “low”, “intermediate”, and “high” risk stratification thresholds can be determined by local resource availability, patient-physician discussion, and clinical considerations.

One limitation of our study is the homogeneity of the cohort. Data was drawn from a single institution, tumor site was limited to oropharynx, and the cohort was largely white and male, similar to previous studies of HNC patients. Our demographic also showed a higher count of patients with HPV-driven tumor and fewer smokers, which is consistent with previous findings showing this trend [27], but makes it difficult to draw any conclusions about the relationship between HPV status and toxicity. Finally, our cohort is also restricted to patients that received IMRT, which was state-of-the-art at the time of treatment. However, more recent studies should consider volume-modulated arc therapy as well.

Most studies that look at stratification of patients, such as TNM staging, are centered on overall survival, rather than toxicities. Thus, many of these studies fail to capture negative outcomes in surviving patients. However, with growing survival rates in HNC patients more work needs to be done to improve post-treatment quality of life for survivors. Importantly, this work is designed to be methodologically rigorous and generalizable, and thus we have eschewed highly simplified single dose/volume thresholds in order to achieve a degree of statistical validity across dose/volume as continuous metrics. Specifically, this work enables us to circumvent single dose/volume metrics in favor of more accurate, individualized risk profiles, rather than population threshold approaches based on high-dimensionality reduction, such as Lyman-Kutcher NTCP models, where the entire dose-volume histogram (DVH) is compressed to a single generalized equivalent uniform dose value (gEUD). The clinical value of the model is

thus targeted less at general dose prescriptions per-se, and more as a granular risk stratification tool for identifying personalized patient-specific risk. Operationalizing conceptually compact predictions as an “app” or API that could integrate within a treatment planning system remains the focus of future work.

In conclusion, using medical imaging information and estimated dosimetric information created at the time of diagnosis, our proposed methodology identifies four groups within a cohort of 200 patients that were significantly correlated ( $p < 0.0001$ ) with dysphagia. Furthermore, our risk-stratification results improve predictive models for dysphagia that already incorporate all possible relevant demographic or clinical information, such as tumor staging, age, and total dose-to-tumor. We believe that our proposed methodology of automatically generating a simple stratified risk score for dysphagia could be applied to identifying high-risk groups of other negative patient outcomes and better guide future treatment recommendations.

#### **Conflict of interest statement**

The authors declare no conflicts of interest

#### **Funding Sources and financial disclosures**

This work was supported by the National Institutes of Health [NCI-R01-CA214825, NCI-R01CA225190] and the National Science Foundation [CNS-1625941, CNS-1828265]. We thank all members of the Electronic Visualization Laboratory, members of the MD Anderson Head and Neck Cancer Quantitative Imaging Collaborative Group, and our collaborators at the University of Iowa and University of Minnesota.

#### **References**

- [1] J.A. Langendijk, P. Doornaert, I.M. Verdonck-de Leeuw, C.R. Leemans, N.K. Aaronson, B.J. Slotman, Impact of late treatment-related toxicity on quality of life among patients with head and neck cancer treated with radiotherapy, *J. Clin. Oncol.* (2008). <https://doi.org/10.1200/JCO.2007.14.6647>.
- [2] MD Anderson Head and Neck Cancer Symptom Working Group, Spatial-Non-spatial Multi-Dimensional Analysis of Radiotherapy Treatment/Toxicity Team (SMART3), Chronic radiation-associated dysphagia in oropharyngeal cancer survivors: Towards age-adjusted dose constraints for deglutitive muscles, *Clin Transl Radiat Oncol.* (2019). <https://doi.org/10.1016/j.ctro.2019.06.005>.

- [3] M.J. Awan, A.S.R. Mohamed, J.S. Lewin, C.A. Baron, G.B. Gunn, D.I. Rosenthal, F.C. Holsinger, D.L. Schwartz, C.D. Fuller, K.A. Hutcheson, Late radiation-associated dysphagia (late-RAD) with lower cranial neuropathy after oropharyngeal radiotherapy: a preliminary dosimetric comparison, *Oral Oncol.* 50 (2014) 746–752. <https://doi.org/10.1016/j.oraloncology.2014.05.003>.
- [4] M.J. Krasin, K.M. Wiese, S.L. Spunt, C.-H. Hua, N. Daw, F. Navid, A.M. Davidoff, L. McGregor, T.E. Merchant, L.E. Kun, L. McCraey, K.A. Hoth, X. Yan, X. Xiong, Jaw Dysfunction Related to Pterygoid and Masseter Muscle Dosimetry After Radiation Therapy in Children and Young Adults With Head-and-Neck Sarcomas, *International Journal of Radiation Oncology\*Biography\*Physics.* (2012). <https://doi.org/10.1016/j.ijrobp.2010.09.031>.
- [5] J.A. Logemann, D.E. Bytell, Swallowing disorders in three types of head and neck surgical patients, *Cancer.* (1979).
- [6] L. van der Molen, W.D. Heemsbergen, R. de Jong, M.A. van Rossum, L.E. Smeele, C.R.N. Rasch, F.J.M. Hilgers, Dysphagia and trismus after concomitant chemo-Intensity-Modulated Radiation Therapy (chemo-IMRT) in advanced head and neck cancer; dose-effect relationships for swallowing and mastication structures, *Radiother. Oncol.* (2013). <https://doi.org/10.1016/j.radonc.2013.03.005>.
- [7] B. O’Sullivan, S.H. Huang, J. Su, A.S. Garden, E.M. Sturgis, K. Dahlstrom, N. Lee, N. Riaz, X. Pei, S.A. Koyfman, D. Adelstein, B.B. Burkey, J. Friborg, C.A. Kristensen, A.B. Gothelf, F. Hoebbers, B. Kremer, E.-J. Speel, D.W. Bowles, D. Raben, S.D. Karam, E. Yu, W. Xu, Development and validation of a staging system for HPV-related oropharyngeal cancer by the International Collaboration on Oropharyngeal cancer Network for Staging (ICON-S): a multicentre cohort study, *Lancet Oncol.* (2016). [https://doi.org/10.1016/S1470-2045\(15\)00560-4](https://doi.org/10.1016/S1470-2045(15)00560-4).
- [8] C.E. Cardenas, A.S.R. Mohamed, R. Tao, A.J.R. Wong, M.J. Awan, S. Kuruvila, M. Aristophanous, G.B. Gunn, J. Phan, B.M. Beadle, S.J. Frank, A.S. Garden, W.H. Morrison, C.D. Fuller, D.I. Rosenthal, Prospective Qualitative and Quantitative Analysis of Real-Time Peer Review Quality Assurance Rounds Incorporating Direct Physical Examination for Head and Neck Cancer Radiation Therapy, *Int. J. Radiat. Oncol. Biol. Phys.* 98 (2017) 532–540. <https://doi.org/10.1016/j.ijrobp.2016.11.019>.
- [9] US Department of Health and Human Services., Modified from Common Terminology Criteria for Adverse Events (CTCAE) (version 4.0), National Institutes of Health, National Cancer Institute. (2009). [https://ctep.cancer.gov/protocolDevelopment/electronic\\_applications/ctc.htm](https://ctep.cancer.gov/protocolDevelopment/electronic_applications/ctc.htm).
- [10] A.S.R. Mohamed, M.-N. Ruangskul, M.J. Awan, C.A. Baron, J. Kalpathy-Cramer, R. Castillo, E. Castillo, T.M. Guerrero, E. Kocak-Uzel, J. Yang, L.E. Court, M.E. Kantor, G. Brandon Gunn, R.R. Colen, S.J. Frank, A.S. Garden, D.I. Rosenthal, C.D. Fuller, Quality Assurance Assessment of Diagnostic and Radiation Therapy–Simulation CT Image Registration for Head and Neck Radiation Therapy: Anatomic Region of Interest–based Comparison of Rigid and Deformable Algorithms, *Radiology.* 274 (2015). <https://doi.org/10.1148/radiol.14132871>.
- [11] A. Wentzel, P. Hanula, T. Luciani, B. Elgohari, H. Elhalawani, G. Canahuate, D. Vock, C.D. Fuller, G.E. Marai, Cohort-based T-SSIM Visual Computing for Radiation Therapy Prediction and Exploration, *IEEE Trans. Vis. Comput. Graph.* (2020). <https://doi.org/10.1109/TVCG.2019.2934546>.
- [12] R.R. Sokal, C.D. Michener, University of Kansas, A Statistical Method for Evaluating Systematic Relationships, 1958.

- [13] A.R. Feinstein, Clinical biostatistics, *Clinical Pharmacology & Therapeutics*. 11 (1970) 282–292. <https://doi.org/10.1002/cpt1970112282>.
- [14] M. Kearns, D. Ron, Algorithmic stability and sanity-check bounds for leave-one-out cross-validation, *Neural Comput.* (1999). <https://doi.org/10.1162/089976699300016304>.
- [15] K. Hajian-Tilaki, Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation, *Caspian J Intern Med*. 4 (2013) 627–635. <https://www.ncbi.nlm.nih.gov/pubmed/24009950>.
- [16] A. Chaturvedi, P.E. Green, J. Douglas Carroll, K-modes Clustering, *Journal of Classification*. (2001). <https://doi.org/10.1007/s00357-001-0004-3>.
- [17] R. Johansson, *Numerical Python: Scientific Computing and Data Science Applications with Numpy, SciPy and Matplotlib*, Apress, 2018.
- [18] R. Core Team, *An Introduction to R*, Samurai Media Limited, 2015.
- [19] G. Varoquaux, L. Buitinck, G. Louppe, O. Grisel, F. Pedregosa, A. Mueller, *Scikit-learn, GetMobile: Mobile Computing and Communications*. 19 (2015) 29–33. <https://doi.org/10.1145/2786984.2786995>.
- [20] S.L. Cessie, S. Le Cessie, J.C. Van Houwelingen, Ridge Estimators in Logistic Regression, *Applied Statistics*. (1992) 1. <https://doi.org/10.2307/2347628>.
- [21] E. Vágó, Logistic Ridge Regression for Clinical Data Analysis (A Case Study), *Applied Ecology and Environmental Research*. (2006). [https://doi.org/10.15666/aeer/0402\\_171179](https://doi.org/10.15666/aeer/0402_171179).
- [22] A. Eisbruch, H.M. Kim, F.Y. Feng, T.H. Lyden, M.J. Haxer, M. Feng, F.P. Worden, C.R. Bradford, M.E. Prince, J.S. Moyer, G.T. Wolf, D.B. Chepeha, R.K. Ten Haken, Chemo-IMRT of oropharyngeal cancer aiming to reduce dysphagia: swallowing organs late complication probabilities and dosimetric correlates, *Int. J. Radiat. Oncol. Biol. Phys.* (2011). <https://doi.org/10.1016/j.ijrobp.2010.12.067>.
- [23] K. Wopken, H.P. Bijl, A. van der Schaaf, H.P. van der Laan, O. Chouvalova, R.J.H.M. Steenbakkens, P. Doornaert, B.J. Slotman, S.F. Oosting, M.E.M.C. Christianen, B.F.A.M. van der Laan, J.L.N. Roodenburg, C.R. Leemans, I.M. Verdonck-de Leeuw, J.A. Langendijk, Development of a multivariable normal tissue complication probability (NTCP) model for tube feeding dependence after curative radiotherapy/chemo-radiotherapy in head and neck cancer, *Radiother. Oncol.* (2014). <https://doi.org/10.1016/j.radonc.2014.09.013>.
- [24] M.E.M.C. Christianen, C. Schilstra, I. Beetz, C.T. Muijs, O. Chouvalova, F.R. Burlage, P. Doornaert, P.W. Koken, C.R. Leemans, R.N.P.M. Rinkel, M.J. de Bruijn, G.H. de Bock, J.L.N. Roodenburg, B.F.A.M. van der Laan, B.J. Slotman, I.M. Verdonck-de Leeuw, H.P. Bijl, J.A. Langendijk, Predictive modelling for swallowing dysfunction after primary (chemo)radiation: results of a prospective observational study, *Radiother. Oncol.* (2012). <https://doi.org/10.1016/j.radonc.2011.08.009>.
- [25] S.H. Huang, B. O'Sullivan, Overview of the 8th Edition TNM Classification for Head and Neck Cancer, *Curr. Treat. Options Oncol.* 18 (2017). <https://doi.org/10.1007/s11864-017-0484-y>.
- [26] J.A. Langendijk, P. Lambin, D. De Ruyscher, J. Widder, M. Bos, M. Verheij, Selection of patients for radiotherapy with protons aiming at reduction of side effects: the model-based approach, *Radiother. Oncol.* (2013). <https://doi.org/10.1016/j.radonc.2013.05.007>.
- [27] K.R. Dahlstrom, G. Calzada, J.D. Hanby, A.S. Garden, B.S. Glisson, G. Li, D.B. Roberts, R.S. Weber, E.M. Sturgis, An evolution in demographics, treatment, and

outcomes of oropharyngeal cancer at a major cancer center, *Cancer*. 119 (2013).  
<https://doi.org/10.1002/cncr.27727>.

## Tables

**Table 1.** Cohort demographics not included during classification

<i>Characteristic</i>	<i>Count (Percent)</i>
<b>General Demographics</b>	
<b>Gender</b>	
Male	172 (86%)
Female	28 (14%)
<b>Race</b>	
White/Caucasian	189 (94.5%)
African American/Black	5 (2.5%)
Hispanic/Latino	3 (1.5%)
Other	3 (1.5%)
<b>Dysphagia</b>	
Pre-Treatment Aspiration	3 (1.5%)
Post-Treatment Aspiration	19 (9.5%)
Post-Treatment Feeding Tube	22 (11%)
<b>Treatment Modality</b>	
One Side of Neck	18 (9%)
Both Sides of Neck	182 (91%)
<b>Existing Clinical Covariates*</b>	
<b>Smoking*</b>	
Never	94 (47%)
Former	69 (34.5%)
Current	37 (18.5%)
<b>T Classification*</b>	
T1	54 (27%)
T2	85 (42.5%)

T3	43 (21.5%)
T4	18 (9%)
<b>N Classification*</b>	
0	7 (3.5%)
1	25 (12.5%)
2	163 (81.5%)
3	5 (2.5%)
<b>Ajcc 8th Edition*</b>	
1	20 (10%)
2	103 (51.5%)
3	7 (3.5%)
4	19 (9.5%)
N/A	51 (25.5%)
<b>HPV Status*</b>	
Positive	130 (65%)
Negative	20 (10%)
Unknown	50 (25%)
<b>Pathological Grade*</b>	
1	1 (0.5%)
2	53 (26.5%)
3	105 (52.5%)
4	2 (1%)
N/A	39 (19.5%)
<b>Tumor Subsites*</b>	
Base of tongue	103 (51.5%)
Tonsil	81 (40.5%)
NOS	11 (5.5%)
Glossopharyngeal sulcus	3 (1.5%)
Soft Palate	2 (1%)
<b>Therapeutic Combination*</b>	
Chemoradiation	115 (57.5%)



Induction Chemotherapy + Chemoradiation	42 (21%)
Radiation Alone	22 (11%)
Induction Chemotherapy + Radiation Alone	21 (10.5%)

**Tumor Laterality\***

Right	102 (51%)
Left	80 (40%)
Bilateral	18 (9%)

**Age\***

Mean (Range)	59.4 (37-82)
--------------	--------------

**Total Dose To Tumor\***

Mean (Range)	68.5 (60-72)
--------------	--------------

**Spatial Features\*\***

**Extended Oral Cavity Predicted Dose (Gy)\*\***

	51.98
Mean (Range)	(44.87-62.66)

**Mandible Predicted Dose (Gy)\*\***

Mean (Range)	39.5 (32.85-51.95)
--------------	--------------------

**Medial Pterygoid Predicted Doses (Combined)(Gy)\*\***

Mean (Range)	77.23 (64.86-92)
--------------	------------------

**Mandible-Tumor Minimum Euclidean Surface Distance (mm)\*\***

Mean (Range)	4.7 (-1.2-16.33)
--------------	------------------

**Medial Pharyngeal Constrictor-Tumor Minimum Euclidean Surface Distance (mm)\*\***

Mean (Range)	8.43 (-2.06-26.61)
--------------	--------------------

\* Features considered for the baseline clinical features

\*\* Features included in the spatial clustering

**Table 2.** Statistics of each spatial cluster. Values include toxicity and 5 features used for generating the spatial clusters. Cluster labels were tested for correlation using Fisher's exact test. Feeding tube alone and Aspiration were both significantly correlated with cluster labels ( $p < .001$ ) as well as overall RAD ( $p < .0001$ ).

<b>Spatial Cluster</b>	<b>Count</b>	<b>% RAD</b>	<b>% Feeding Tube</b>	<b>% Aspiration</b>	<b>Extended Oral Cavity Predicted Dose (Gy)</b>	<b>Mandible Predicted Dose (Gy)</b>	<b>Average Medial Pterygoid Muscle Predicted Dose (Gy)</b>	<b>Mandible-Tumor Distance (mm)</b>	<b>Medial Pharyngeal Constrictor-Tumor Distance (mm)</b>
<b>Spatial Cluster 1</b>	3	0	0	0	51.96 (1.97)	39.92 (2.16)	37.94 (0.5)	13.35 (1.44)	0.29 (1.64)
<b>Spatial Cluster 2</b>	114	5.3	3.5	2.6	50.84 (1.37)	38.21 (1.02)	38.03 (1.1)	7.18 (3.82)	12.23 (4.55)
<b>Spatial Cluster 3</b>	35	11.4	8.6	5.7	48.66 (3.20)	36.33 (2.82)	35.91 (2.1)	0.04 (1.81)	0.28 (0.60)
<b>Spatial Cluster 4</b>	48	50	31.3	29.2	57.11 (2.93)	44.84 (3.69)	42.03 (1.9)	1.64 (2.28)	5.86 (4.23)

**Table 3.** AUC Scores From Logistic Regression Classification using Leave-One-Out Cross-Validation

**Leave-one-out Cross-Validation AUC Scores (Logistic Regression)**

	Feeding		
	Tube	Aspiration	RAD (Either)
Spatial Clusters	0.64	0.66	0.68
T Stage	0.60	0.76	0.70
T Stage + Spatial Clusters	<b>0.76</b>	0.82	0.82
All Clinical Features	0.64	0.85	0.79
All Clinical Features + Spatial Clusters	0.71	<b>0.87</b>	<b>0.84</b>
Spatial Features	0.72	0.80	0.77
Spatial Features + Clinical Features	0.67	0.86	0.80

**Figure Legends**

**Figure 1.** Visual summaries of the tumor proximity and predicted mean doses to each OAR for the 4 identified spatial clusters. Distances are quantile-scaled to give uniform distribution across each axis within the cohort. Dark lines show the mean values for each cluster while darker shading represents the portion of the cluster within a given quantile. Cluster 4 (n = 48) has a higher predicted mean dose and toxicity, while cluster 3 has a high tumor-proximity and a predicted mean dose with low mean values but high variance. Cluster 2 consists of the majority of the low-risk cohort, while cluster 1 consists of a group of 3 patients with more localized dose distributions and no RAD.

**Figure 2.** Receiver-Operator curves (ROC) for 3 outcomes using Clinical Features vs Clinical Features + Spatial Clusters.

**Figure 3.** AUC scores for Logistic Regression cross-validation AUC score when using clinical variables and our spatial clusters, vs using clinical variables alone. For each step the trial was rerun 500 times with patients randomly removed, with at least 2 patients with toxicity included in the remaining set, so that cross-validation could still be run. Spatial clustering was re-calculated at each run and cross-validation AUC scores were calculated with and without clusters in the model on the reduced subset of patients.