# Machine Learning Models Using Extended Clinical Variables to Predict Real-World Head and Neck Cancer Radiotherapy Toxicity

Dr Tom Young[1,2], Victoria Butterworth[1,2], Dr Laia Humbert-Vidan[1,2,3], Dr Delali Adjogatse[1,2], Dr Anthony Kong[1,2], Dr Miguel Reis Ferreira[1,2], Dr Mary Lei[1,2], Dr Imran Petkar[1,2], Professor Andy King[2], Dr Teresa Guerrero Urbano[1,2]. [1]Guy's and St Thomas' NHS Foundation Trust, London, UK. [2]King's College London, London, UK. [3]The University of Texas MD Anderson Cancer Center, Houston, USA.

## Background and Aims

- Radiotherapy is an effective treatment for Head and Neck Cancer (HNC) but may cause long-term toxicity significantly impacting quality of life [1]. Normal Tissue Complication Probability (NTCP) models predict radiotherapy (RT)-induced toxicity and guide treatment decision-making [2].
- Traditional Lyman-Kutcher-Burman (LKB) NTCP models are limited, based on single organ-at-risk uniform equivalent dose [3]. Recently proposed NTCP models outperform LKB models but still exclude potentially important clinical variables [4]. Machine learning (ML) techniques may handle more input variables than traditional statistical models [5] but require large accurate datasets for development.
- At our centre we have built a large multi-modal retrospective dataset for almost 3000 HNC patients seen by the HNC Oncology Team at our centre since 2010. This dataset aligns extensive clinical data with diagnostic/radiotherapy imaging data.
- We aimed to leverage this large real-world HNC clinical dataset to develop new RT toxicity prediction models (specifically focusing on xerostomia) incorporating ML techniques.

## Methods

- From the HNC retrospective dataset, data was extracted for oropharynx, larynx and hypopharynx cancer patients treated with radical radiotherapy (with or without induction and/or concomitant chemotherapy) with known 1-year post-RT xerostomia grading (CTCAE v5 toxicity grade 0-1 versus ≥2) (n=312).
- Relevant clinical variables (n=243) were selected: Demographic (n=40), disease (n=6), treatment (n=24), bloods (n=24), baseline toxicity (n=10), and RT dosimetric (parotid glands, Planning Target Volume (PTV) (n=137) data. Univariable and multivariable logistic regression (LR) was performed to assess variable significance, and correlation analysis.
- Multivariable LR (MVLR), Support Vector Machine (SVM), Fully Connected Network (FCN), Gradient Boosting Machine (GBM), Decision Tree (DT), and Random Forest (RF) predictive models were developed using two different input data:
    1. Parotid mean doses alone (Parotid Dose ('PD') models, in keeping with current HNC RT clinical practice)
    2. Significant ($p<0.05$) uncorrelated (Spearman's rank <0.7) clinical variables from LR ('Extended Clinical Variable' (ECV) models)
- Models with no feature selection ('All Variable' as input) were also assessed. Models were trained and tested using 5-fold cross-validation. Area Under Curve (AUC), Brier score and Expected Calibration Error (ECE) were calculated to assess performance.

## Results

- Univariable and correlation analysis showed 27 uncorrelated significant variables for predicting xerostomia. For all models, mean AUC ranged from 0.64 to 0.68 for PD models, and 0.61 to 0.73 for ECV models (Table 1). Mean AUC for ECV models outperformed PD models for all model types except DT. Mean AUC was highest for multivariable LR compared to other model types for all data inputs (Figure 1) (no significant difference between model types with best performing data input type on Friedman test). Mean AUC for ECV models was higher than for 'All Variable' models for all but DT (Figure 2), but this trend did not reach statistical significance on Friedman test comparing different data inputs for all model types.

| | Mean AUC (95% Confidence Interval) | | Brier Score | | ECE | |
|---|---|---|---|---|---|---|
| | Parotid Mean Dose Only (PD) | Extended Clinical Variables (ECV) | Parotid Mean Dose Only (PD) | Extended Clinical Variables (ECV) | Parotid Mean Dose Only (PD) | Extended Clinical Variables (ECV) |
| MVLR | 0.68 (0.64-0.73) | **0.73 (0.69-0.78)** | 0.233 | **0.209** | **0.120** | 0.130 |
| DT | **0.64 (0.59-0.70)** | 0.61 (0.54-0.67) | **0.272** | 0.319 | **0.234** | 0.269 |
| RF | 0.67 (0.65-0.71) | **0.70 (0.66-0.73)** | 0.228 | **0.214** | 0.165 | **0.153** |
| SVM | 0.67 (0.64-0.71) | **0.70 (0.66-0.74)** | 0.228 | **0.216** | 0.119 | **0.119** |
| FCN | 0.65 (0.60-0.71) | **0.71 (0.64-0.80)** | 0.305 | **0.216** | **0.164** | 0.223 |
| GBM | 0.66 (0.62-0.70) | **0.70 (0.67-0.73)** | 0.23 | **0.216** | **0.146** | 0.153 |

Table 1. Model performance evaluation metrics (95% confidence intervals in brackets, bold denotes best data input for each ML model type for each metric) for PD and ECV models.

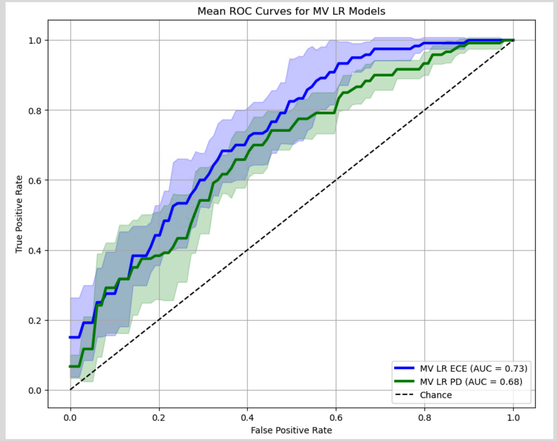



Figure 1. Mean ROC curves for MV LR models: ECE model (blue) versus PD only model (green). Shading denotes upper and lower cross-validation fold results.

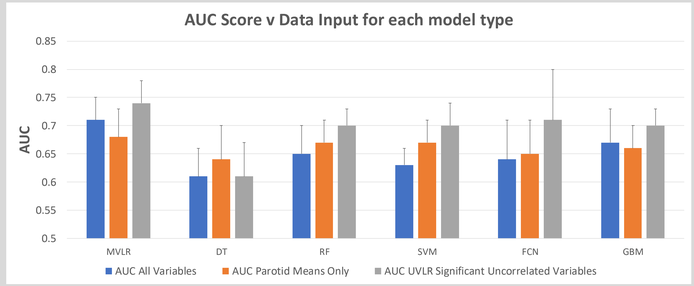



Figure 2. Mean AUC values for all models with different input data: All Variables (Blue), PD (Orange), ECV (Grey).

- Feature importance was assessed for the best performing model (MVLR with ECV). This showed the most important variables included established predictors such as parotid mean dose. However, of note less considered non-dosimetric variables contributing to model performance included higher socioeconomic deprivation indices ($p=0.027$) and platelet count ($p=0.013$).

## Conclusions

Incorporating additional clinical variables increased mean AUC for all model types predicting 1-year xerostomia except DT, suggesting potential benefit of additional clinical variable inclusion for comprehensive NTCP model development.

Ongoing work modelling xerostomia and other HNC RT toxicities will:

- Incorporate additional relevant clinical variables (e.g. submandibular dosimetric data)
- Explore alternative feature selection techniques, model architectures and recalibration techniques
- Incorporate imaging data to develop spatially-aware multi-modal ML predictive models
- Externally validate future models

**References**

1. van der Laan HP, Van den Bosch L, Schuit E, Steenbakkers RJHM, van der Schaaf A, Langendijk JA. Impact of radiation-induced toxicities on quality of life of patients treated for head and neck cancer. Radiother Oncol. 2021 Jul;160:47-53.
2. Joiner, M.C., & van der Kogel, A.J. (Eds.). (2018). Basic Clinical Radiobiology (5th ed.). CRC Press.
3. Appelt AL, Elhaminia B, Gooya A, Gilbert A, Nix M. Deep Learning for Radiotherapy Outcome Prediction Using Dose Data - A Review. Clin Oncol (R Coll Radiol). 2022 Feb;34(2):e87-e96.
4. Van den Bosch L, van der Schaaf A, van der Laan HP, Hoebers FJP, Wijers OB, van den Hoek JGM, Moons KGM, Reitsma JB, Steenbakkers RJHM, Schuit E, Langendijk JA. Comprehensive toxicity risk profiling in radiation therapy for head and neck cancer: A new concept for individually optimised treatment. Radiother Oncol. 2021 Apr;157:147-154. doi: 10.1016/j.radonc.2021.01.024. Epub 2021 Feb 3. PMID: 33545258.
5. H.S. Gabrys, et al. Design and selection of machine learning methods using radiomics and dosiomics for normal tissue complication probability modeling of xerostomia. Front Oncol, 8:35 2018.