# Machine Learning Methods in Classification of Prolonged Radiation Therapy in Oropharyngeal Cancer

Matthew I. Saleem BS[1], Seungjun Ahn PhD[2], Eun Jeong Oh PhD[3], Tristan Tham MD[1]
[1]Department of Otolaryngology – Head and Neck Surgery, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell
[2]Department of Population Health Science and Policy, Icahn School of Medicine at Mount Sinai
[3]Institute of Health System Science, Feinstein Institutes for Medical Research

## INTRODUCTION

- Oropharyngeal squamous cell carcinoma (OPSCC) is a subtype of head and neck cancer (HNC) often treated with chemoradiation (CRT) and radiation therapy (RT)[1]
- Delays in RT and CRT leading to prolonged radiation treatment duration (RTD) have been associated with poorer overall survival (OS) in OPSCC patients[2]
- Machine learning (ML) has been used to predict OS and occult nodal metastases in HNC patients, but never prolonged RTD in OPSCC patients[3]
- There is clinical utility in applying ML algorithms to predict treatment delay

## METHODS

- Retrospective analysis of the National Cancer Database (NCDB) queried for OPSCC patients from 2004-2016 who received RT or CRT
- Outcome variable was RTD, dichotomized into prolonged (≥ 50 days) or not prolonged (< 50 days)
- Eight ML algorithms were compared to standard multivariable logistic regression across several classification performance metrics, with a 70%/30% training and test split
- Survival analysis of risk stratified groups was performed using Kaplan-Meier curves, further divided by race, tumor size, or HPV status

## RESULTS

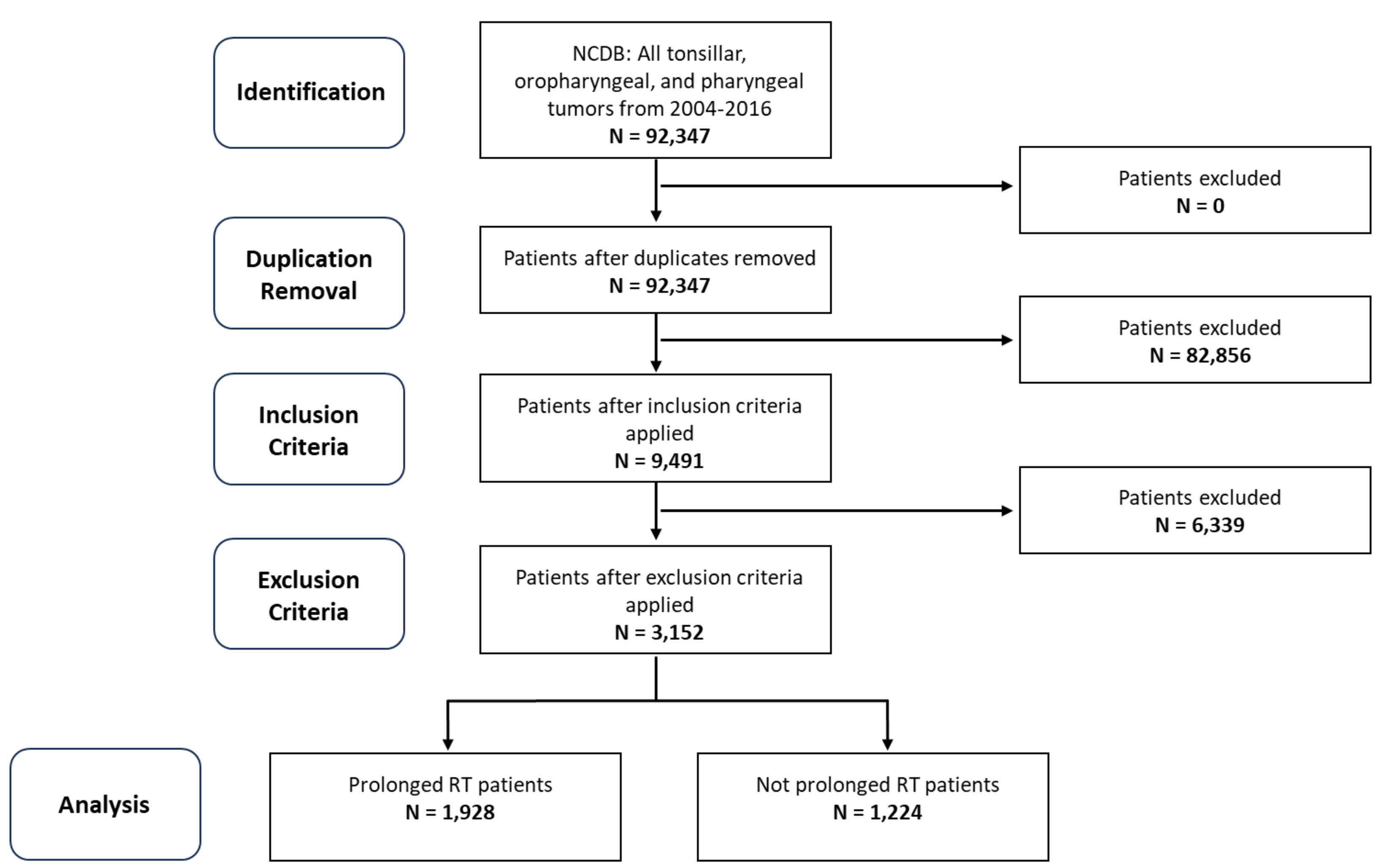


**Figure 1** Strengthening the Reporting of Observational Studies in Epidemiology flowchart

## RESULTS CONT.

| | LASSO | Enet | SCAD | MCP | RF | SVM-l | SVM-nl | XGBoost | Logistic |
|---|---|---|---|---|---|---|---|---|---|
| AUROC | 0.58 (0.02) | 0.58 (0.02) | 0.58 (0.02) | 0.58 (0.02) | **0.65 (0.02)** | 0.52 (0.06) | 0.56 (0.02) | **0.65 (0.02)** | 0.58 (0.02) |
| AUPRC | 0.67 (0.02) | 0.67 (0.02) | 0.67 (0.02) | 0.66 (0.02) | **0.72 (0.02)** | 0.62 (0.05) | 0.66 (0.02) | **0.72 (0.02)** | 0.66 (0.02) |
| CIN | -0.09 (0.21) | -0.11 (0.22) | **0.00 (0.17)** | **0.00 (0.17)** | 0.01 (0.10) | 0.33 (16.7) | -0.04 (0.25) | 0.17 (0.09) | 0.20 (0.11) |
| CSL | 1.20 (0.42) | 1.25 (0.43) | **0.99 (0.31)** | **1.01 (0.32)** | 1.07 (0.14) | 0.26 (36.5) | 1.08 (0.56) | 1.38 (0.20) | 0.55 (0.18) |
| BRS | 0.23 (0.00) | 0.23 (0.00) | 0.23 (0.00) | 0.23 (0.00) | **0.22 (0.00)** | 0.24 (0.00) | 0.24 (0.00) | 0.23 (0.00) | 0.24 (0.00) |
| F1S | 0.63 (0.02) | 0.63 (0.03) | 0.63 (0.02) | 0.63 (0.02) | **0.67 (0.04)** | 0.55 (0.12) | 0.61 (0.10) | 0.66 (0.04) | 0.62 (0.06) |
| MCC | 0.15 (0.03) | 0.15 (0.03) | 0.15 (0.03) | 0.15 (0.03) | 0.24 (0.03) | 0.07 (0.10) | 0.12 (0.03) | **0.25 (0.03)** | 0.15 (0.02) |
| SENS | 0.58 (0.03) | 0.58 (0.04) | 0.58 (0.03) | 0.59 (0.02) | **0.63 (0.09)** | 0.50 (0.18) | 0.56 (0.17) | 0.60 (0.06) | 0.56 (0.10) |
| SPEC | 0.57 (0.04) | 0.58 (0.05) | 0.57 (0.04) | 0.56 (0.03) | 0.61 (0.09) | 0.52 (0.19) | 0.54 (0.17) | **0.65 (0.06)** | 0.59 (0.10) |
| PPV | 0.68 (0.02) | 0.68 (0.02) | 0.68 (0.02) | 0.68 (0.02) | 0.72 (0.02) | 0.63 (0.07) | 0.66 (0.03) | **0.73 (0.02)** | 0.68 (0.02) |
| NPV | 0.47 (0.02) | 0.47 (0.02) | 0.47 (0.02) | 0.47 (0.02) | **0.51 (0.03)** | 0.41 (0.07) | 0.45 (0.03) | **0.51 (0.03)** | 0.46 (0.03) |

**Figure 2** Performance metrics of eight machine learning methods based on complete-case analysis. For each metric, median is reported with mean absolute deviation in parentheses. Best results are bolded. Abbreviations: LASSO, least absolute shrinkage and selection operator; Enet, elastic-net; SCAD, smoothly clipped absolute deviation; MCP, minimax concave penalty; RF, random forest; SVM-l; support vector machine with linear kernel; SVM-nl; support vector machine with non-linear kernel; XGBoost, extreme gradient boosting; Logistic, logistic regression; AUROC, area under the receiver-operating characteristic curve; AUPRC, area under the precision-recall curve; CIN, calibration intercept; CSL, calibration slope; BRS, Brier score; F1S, F1-score; MCC, Matthews correlation coefficient; SENS, sensitivity; SPEC, specificity; PPV, positive predictive value; NPV, negative predictive value

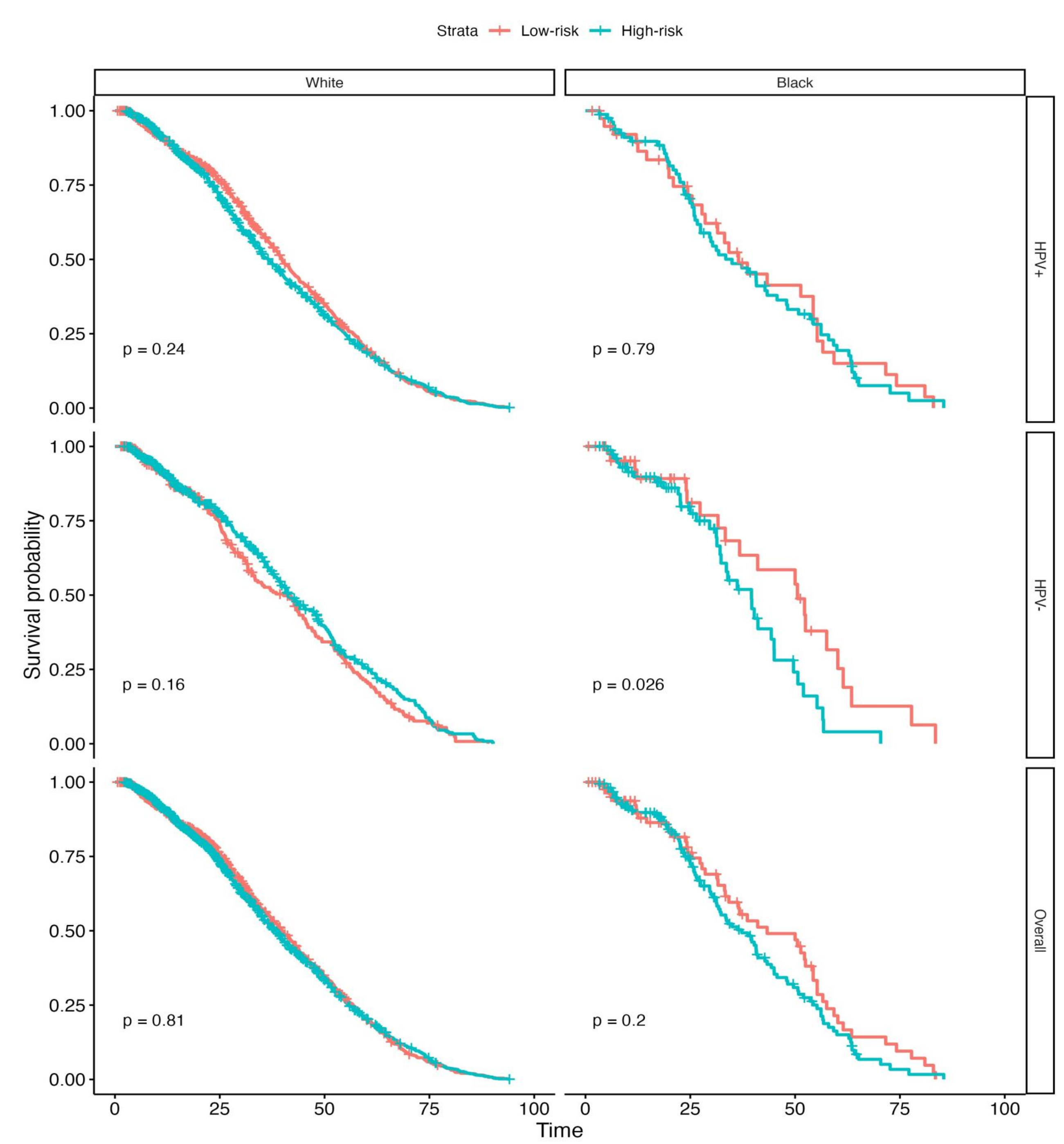


**Figure 3** Kaplan-Meier curves of overall survival stratified by the predicted class of prolonged RT (High-risk vs Low-risk) using the RF model, separately by faceting variables: HPV status and race. The difference in survival between the predicted groups of prolonged RT was significant (P = 0.026) among Black patients in HPV– cases.

## RESULTS CONT.

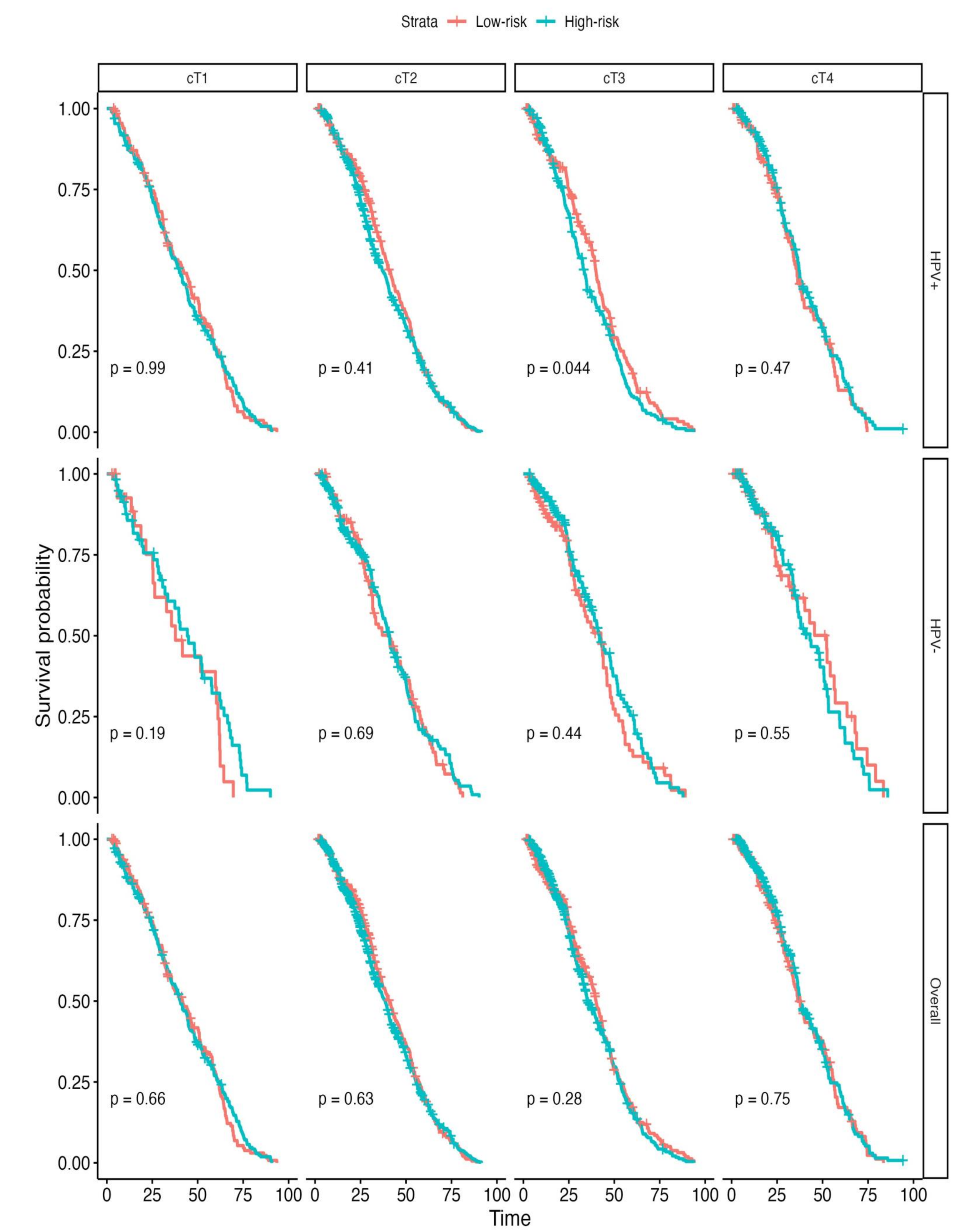


**Figure 3** Kaplan-Meier curves of overall survival stratified by the predicted class of prolonged RT (High-risk vs Low-risk) using the RF model, separately by faceting variables: HPV status and clinical T stage. The difference in survival between the predicted class of prolonged RT was significant (P = 0.044) among cT3 in HPV+ cases.

## CONCLUSIONS

- The RF model is superior to traditional logistic regression when stratifying risk of prolonged RTD in OPSCC patients
- RF risk stratified groups had significant survival differences based on Kaplan-Meier curves
- Classifying patients at high risk of prolonged RTD can potentially facilitate early intervention and improve overall survival

## REFERENCES

1. Hamilton D, Khan MK, O'hara J, Paleri V. The changing landscape of oropharyngeal cancer management. *J Laryngol Otol*. 2017;131(1):3-7. doi:10.1017/S0022215116009178
2. Goel AN, Frangos M, Raghavan G, et al. Survival impact of treatment delays in surgically managed oropharyngeal cancer and the role of human papillomavirus status. *Head & Neck*. 2019;41(6):1756-1769. doi:10.1002/hed.25643
3. Bur AM, Holcomb A, Goodwin S, et al. Machine learning to predict occult nodal metastasis in early oral squamous cell carcinoma. *Oral Oncology*. 2019;92:20-25. doi:10.1016/j.oraloncology.2019.03.011