**Proceedings Paper:**
Astley, J.R., Reilly, J., Robinson, S. et al. (3 more authors) (2024) 1596: The impact of dosimetric features on interpretable overall survival prediction in NSCLC radiotherapy. In: Radiotherapy and Oncology. ESTRO 2024, 03-07 May 2024, Glasgow, UK. Elsevier BV , S4488-S4492.

https://doi.org/10.1016/s0167-8140(24)01965-0

# The impact of dosimetric features on interpretable overall survival prediction in NSCLC radiotherapy

Joshua R. Astley[1], James Reilly[1], Stephen Robinson[1], Jim M. Wild[1,2], Matthew Q. Hatton[1], Bilal A. Tahir[1,2]*

[1]*Division of Clinical Medicine, The University of Sheffield, Sheffield, UK*
[2]*Insigneo Institute for in silico medicine, The University of Sheffield, Sheffield, UK*

Words: 739/750

## Background and Purpose:

The Radiation Therapy Oncology Group 0617 revealed that a high-dose regimen of 74 Gray over 37 fractions (74Gy/37fr) unexpectedly reduced overall survival (OS) compared to the standard 60Gy/30fr regimen in non-small cell lung cancer (NSCLC) patients [1]. This diminished OS has been attributed to increased radiation exposure to the heart. Additionally, inadvertent radiation to other organs-at-risk (OARs), i.e. the lungs and oesophagus, can give rise to radiation-induced conditions like pneumonitis and oesophagitis that correlate with reduced OS in NSCLC patients receiving radiotherapy [1-3].

Traditionally, Cox proportional hazards (CPH) models have been employed to evaluate OS, but these models may be overly simplistic due to their assumption of a linear relationship between covariates and outcomes. Alternative, non-linear machine learning (ML) methods, including random survival forests (RSFs) and more recently, deep learning (DL), have been suggested [4, 5]. However, these approaches are generally opaque, making interpretation challenging. In our study, we juxtaposed CPH, RSF, and DL for predicting OS in NSCLC patients undergoing radiotherapy, utilising pre-treatment covariates. Established methodologies were used to elucidate the impact of each covariate on OS prediction.

## Materials and Methods:

The dataset comprised clinical, demographic, treatment, and time-to-event survival data for 471 NSCLC patients treated with radical radiotherapy between 2010 and 2016. Patients were treated with hypofractionated accelerated radiotherapy (55Gy/20fr over four weeks) or continuous hyperfractionated accelerated radiotherapy (CHART) (54Gy/36fr over 12 days). For patients undergoing chemotherapy, regimens were platinum-based doublets with gemcitabine, vinorelbine or pemetrexed. Covariates included age, sex, stage (TNM v7), administration of chemotherapy, neutrophil-lymphocyte ratio (NLR) and planning target volume (PTV). OAR dose-volume parameters were calculated for the heart, lungs and oesophagus, including mean and maximum dose, volume receiving 5Gy ($V_{5Gy}$) in 5Gy increments up to $V_{50Gy,}$ and the dose received by 5% of the volume ($D_{5\%}$) in 5% increments up to $D_{50\%}$ and the mean dose to the spinal cord. The date of death or last follow-up was recorded.

Mutual information (MI) is a measure of uncertainty and is calculated between two sets of features, assessing non-linear relationships between them. The MI was calculated between each dose-volume feature and the time-to-event, where the dosimetric feature from each OAR with the highest MI was selected as a covariate. The following covariates were included as 'standard' features: age, sex, stage, administration of chemotherapy, NLR, PTV and spinal cord mean dose. We added OAR dosimetric variables with the highest MI for the heart, lungs and oesophagus.

Three survival prediction models were trained: the conventional CPH, the ML-based RSF and the DL-based DeepSurv [5]. We compared these models with and without dose-volume OAR features. Features were pre-processed prior to OS prediction, including ordinally encoding the participants' stage and normalising all non-categorical features.

10-fold Monte-Carlo cross-validation was employed with a split of 70%:10%:20% for training, validation and testing, respectively. Performance was evaluated using Harrell's concordance index (C-index) and the integrated Brier score (IBS). Local interpretable model-agnostic explanation (LIME) values, adapted for use in survival analysis, were generated using the `survlimepy` library [6] for each model.

Wilcoxon tests were used to assess significances of differences between OS models with and without dose-volume features for each survival prediction method. Friedman tests with *post-hoc* multiple comparisons were used to assess significances of differences between the best-performing feature combinations for each model.

## Results:

For all survival approaches, the inclusion of dosimetric features generated improved performance over approaches which did not integrate dose-volume features. The DeepSurv survival model exhibited the best-performing IBS, achieving an IBS of 0.12, whereas the RSF approach generated the best-performing C-index, achieving a C-index of 0.66 (Table 1). Using the C-index, no significant difference was observed between the CPH, RSF and DeepSurv. However, using the IBS, the DeepSurv approach significantly outperformed the CPH. No significant difference was observed between the RSF and DeepSurv approaches using the IBS.

Table 1. Performance of CPH, RSF and DeepSurv models in terms of mean (95% CI) C-index and IBS, computed via 10-fold Monte-Carlo cross-validation, for each combination of feature inputs. The best testing set values for C-index and IBS are shown in bold.

| Method | | IBS Mean (95% CI) | | C-Index Mean (95% CI) | |
|---|---|---|---|---|---|
| | | **Standard** | **Standard + dosimetric** | **Standard** | **Standard + dosimetric** |
| **CPH** | Train | 0.13 (0.13, 0.14) | 0.13 (0.13, 0.14) | 0.59 (0.58, 0.59) | 0.62 (0.61, 0.62) |
| | Validation | 0.16 (0.15, 0.18) | 0.16 (0.15, 0.17) | 0.60 (0.57, 0.63) | 0.63 (0.60, 0.65) |
| | Test | 0.15 (0.14, 0.15) | 0.14 (0.14, 0.15) | 0.61 (0.59, 0.63) | 0.65 (0.63, 0.66) |
| **RSF** | Train | 0.11 (0.11, 0.11) | 0.11 (0.10, 0.11) | 0.66 (0.65, 0.66) | 0.68 (0.67, 0.69) |
| | Validation | 0.15 (0.14, 0.16) | 0.15 (0.14, 0.16) | 0.61 (0.60, 0.63) | 0.62 (0.60, 0.64) |
| | Test | 0.13 (0.12, 0.14) | 0.13 (0.12, 0.14) | 0.65 (0.63, 0.66) | **0.66 (0.64, 0.68)** |
| **DeepSurv** | Train | 0.12 (0.12, 0.13) | 0.12 (0.11, 0.13) | 0.66 (0.63, 0.66) | 0.66 (0.65, 0.67) |
| | Validation | 0.14 (0.13, 0.15) | 0.14 (0.13, 0.15) | 0.63 (0.61, 0.66) | 0.64 (0.61, 0.67) |
| | Test | 0.13 (0.12, 0.14) | **0.12 (0.11, 0.13)** | 0.63 (0.61, 0.65) | 0.65 (0.64, 0.66) |

Abbreviations: C-Index = Harrell's concordance index, IBS = integrated Brier score, CPH = Cox proportional hazards, RSF = random survival forest, CI = confidence interval

Survival-adapted LIME (SurvLIME) values were calculated for each testing set case with the median weighting reported, providing an overall understanding of feature importance for each survival model. Figure 1 displays median SurvLIME values for the DeepSurv model for all testing set cases.

## Conclusion:

We show that, using pre-treatment covariates, survival analysis approaches which integrate dose-volume information generate improved OS prediction. Additionally, a DL approach demonstrated superior performance over CPH using the IBS for OS prediction. We employ explainable techniques to provide transparency and interpretability.

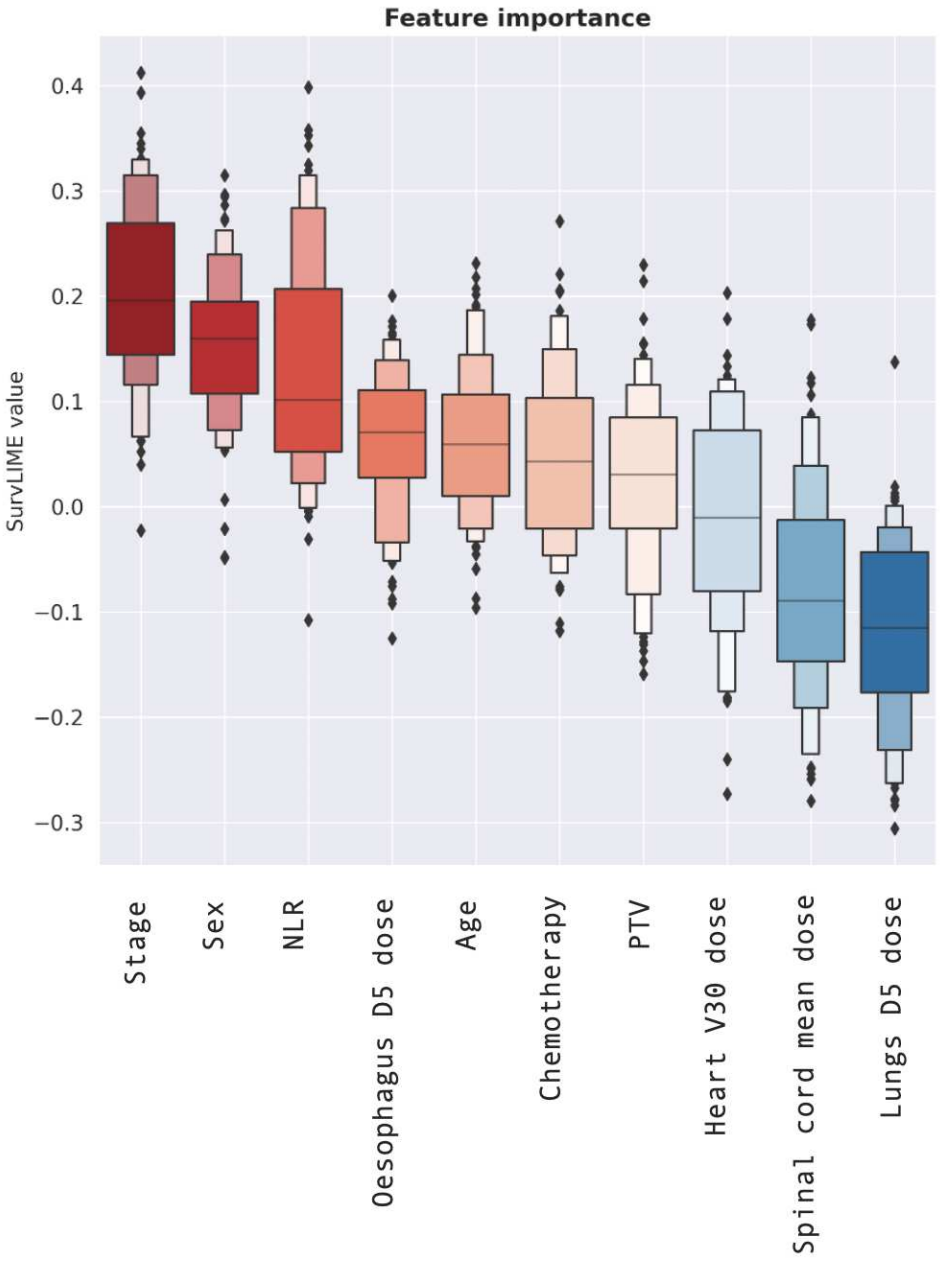


Figure 5. SurvLIME values for the best-performing DeepSurv model.

## References:

[1] Bradley JD, Paulus R, Komaki R, Masters G, Blumenschein G, Schild S, et al. Standard-dose versus high-dose conformal radiotherapy with concurrent and consolidation carboplatin plus paclitaxel with or without cetuximab for patients with stage IIIA or IIIB non-small-cell lung cancer (RTOG 0617): a randomised, two-by-two factorial phase 3 study. Lancet Oncol. 2015;16:187-99.
[2] Guberina M, Eberhardt W, Stuschke M, Gauler T, Heinzelmann F, Cheufou D, et al. Heart dose exposure as prognostic marker after radiotherapy for resectable stage IIIA/B non-small-cell lung cancer: secondary analysis of a randomized trial. Ann Oncol. 2017;28:1084-9.
[3] Shen L, Liu C, Jin J, Han C, Zhou Y, Zheng X, et al. Association of lung and heart dose with survival in patients with non-small cell lung cancer underwent volumetric modulated arc therapy. Cancer Manag Res. 2019;11:6091-8.
[4] Hemant I, Udaya BK, Eugene HB, Michael SL. Random survival forests. The Annals of Applied Statistics. 2008;2:841-60.
[5] Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Med Res Methodol. 2018;18:24.
[6] Pachón-García C, Hernández-Pérez C, Delicado P, Vilaplana V. SurvLIMEpy: A Python package implementing SurvLIME. arXiv preprint arXiv:230210571. 2023.