



# A comparison of machine learning methods for predicting recurrence and death after curative-intent radiotherapy for non-small cell lung cancer: Development and validation of multivariable clinical prediction models

Sumeet Hindocha,<sup>a,b,c,d,k</sup> Thomas G. Charlton,<sup>e</sup> Kristofer Linton-Reid,<sup>d</sup> Benjamin Hunter,<sup>a,c,d,k</sup> Charleen Chan,<sup>c</sup> Merina Ahmed,<sup>f</sup> Emily J. Robinson,<sup>g</sup> Matthew Orton,<sup>h</sup> Shahreen Ahmad,<sup>e</sup> Fiona McDonald,<sup>a,c</sup> Imogen Locke,<sup>f</sup> Danielle Power,<sup>i</sup> Matthew Blackledge,<sup>j</sup> Richard W. Lee,<sup>a,k,l,\*</sup> and Eric O. Aboagye<sup>c,\*</sup>

<sup>a</sup>Lung Unit, The Royal Marsden NHS Foundation Trust, Fulham Road, London SW36JJ, UK

<sup>b</sup>AI for Healthcare Centre for Doctoral Training, Imperial College London, Exhibition Road, London SW7 2BX, UK

<sup>c</sup>Department of Clinical Oncology, Institute of Cancer Research NIHR Biomedical Research Centre, London, UK

<sup>d</sup>Cancer Imaging Centre, Department of Surgery and Cancer, Imperial College London, Du Cane Road, London W12 0NN, UK

<sup>e</sup>Guy's Cancer Centre, Guy's and St Thomas' NHS Foundation Trust, Great Maze Pond, London SE19RT UK

<sup>f</sup>Lung Unit, The Royal Marsden NHS Foundation Trust, Downs Road, Sutton SM25PT, UK

<sup>g</sup>Clinical Trials Unit, Royal Marsden NHS Foundation Trust, Downs Road, Sutton SM25PT, UK

<sup>h</sup>Artificial Intelligence Imaging Hub, Royal Marsden NHS Foundation Trust, Downs Road, Sutton SM25PT, UK

<sup>i</sup>Department of Clinical Oncology, Charing Cross Hospital, Fulham Palace Road, London W6 8RF, UK

<sup>j</sup>Radiotherapy and Imaging, Institute of Cancer Research, 123 Old Brompton Road, London SW7 3RP, UK

<sup>k</sup>Early Diagnosis and Detection Centre, National Institute for Health Research (NIHR) Biomedical Research Centre at The Royal Marsden NHS Foundation Trust and the Institute of Cancer Research, London

<sup>l</sup>National Heart and Lung Institute, Imperial College, London, UK

## Summary

**Background** Surveillance is universally recommended for non-small cell lung cancer (NSCLC) patients treated with curative-intent radiotherapy. High-quality evidence to inform optimal surveillance strategies is lacking. Machine learning demonstrates promise in accurate outcome prediction for a variety of health conditions. The purpose of this study was to utilise readily available patient, tumour, and treatment data to develop, validate and externally test machine learning models for predicting recurrence, recurrence-free survival (RFS) and overall survival (OS) at 2 years from treatment.

**Methods** A retrospective, multicentre study of patients receiving curative-intent radiotherapy for NSCLC was undertaken. A total of 657 patients from 5 hospitals were eligible for inclusion. Data pre-processing derived 34 features for predictive modelling. Combinations of 8 feature reduction methods and 10 machine learning classification algorithms were compared, producing risk-stratification models for predicting recurrence, RFS and OS. Models were compared with 10-fold cross validation and an external test set and benchmarked against TNM-stage and performance status. Youden Index was derived from validation set ROC curves to distinguish high and low risk groups and Kaplan-Meier analyses performed.

**Findings** Median follow-up time was 852 days. Parameters were well matched across training-validation and external test sets: Mean age was 73 and 71 respectively, and recurrence, RFS and OS rates at 2 years were 43% vs 34%, 54% vs 47% and 54% vs 47% respectively. The respective validation and test set AUCs were as follows: 1) RFS: 0.682 (0.575–0.788) and 0.681 (0.597–0.766), 2) Recurrence: 0.687 (0.582–0.793) and 0.722 (0.635–0.81), and 3) OS: 0.759 (0.663–0.855) and 0.717 (0.634–0.8). Our models were superior to TNM stage and performance status in predicting recurrence and OS.

**Interpretation** This robust and ready to use machine learning method, validated and externally tested, sets the stage for future clinical trials entailing quantitative personalised risk-stratification and surveillance following curative-intent radiotherapy for NSCLC.

\*Corresponding author at: Lung Unit, The Royal Marsden NHS Foundation Trust, Fulham Road, London SW36JJ, UK.  
E-mail addresses: [Richard.Lee@rmh.nhs.uk](mailto:Richard.Lee@rmh.nhs.uk) (R.W. Lee), [eric.aboagye@imperial.ac.uk](mailto:eric.aboagye@imperial.ac.uk) (E.O. Aboagye).

eBioMedicine 2022;77:  
103911

Published online xxx  
<https://doi.org/10.1016/j.ebiom.2022.103911>

**Funding** A full list of funding bodies that contributed to this study can be found in the Acknowledgements section.

**Copyright** © 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

**Keywords:** Non-small cell lung cancer; Radiotherapy; Machine learning; Recurrence; Overall survival; Prediction; Early detection

### Research in context

#### *Evidence before this study*

We searched PubMed Central for papers published from database inception to 6<sup>th</sup> June 2020, with the terms “non-small cell lung cancer”, “radiotherapy”, “machine learning”, “prognosis”, “outcome” and “prediction” with no language restrictions. This returned 809 results which were manually screened for suitability. The majority of returned results either did not involve NSCLC or pertained to imaging (radiomic and deep-learning) or biological models which was not our focus. Six studies related to our work.

To the best of our knowledge, no published studies to date have used routinely available clinical data to compare multiple machine learning methods to build prediction models for recurrence and survival following curative-intent radiotherapy in non-small cell lung cancer.

Previous studies using clinical data for outcome prediction following radiotherapy for NSCLC have explored cox proportional-hazards models and support vector machines and have generally focused on overall survival (OS), with few studies looking at recurrence or recurrence-free survival (RFS). This is likely due to accurate data on death being more readily accessible from national-level registries than recurrence data; however recurrence and RFS may offer more clinical utility than OS in stratifying follow-up, allowing clinicians to potentially intervene earlier to provide further, potentially curable treatment.

#### *Added value of this study*

Our study is the first to compare multiple machine learning algorithms and feature reduction methods using routinely available clinical data and to develop, validate and externally test prediction models for recurrence, RFS and OS following radical radiotherapy for NSCLC. Such approaches may overcome the limitations of traditional statistical methods such as Cox proportional hazards models.

#### *Implications of all the available evidence*

Performance of our models exceed traditional methods and show consistency across validation and external test sets. This robust and readily replicable machine learning method, validated and externally tested, sets the stage for future clinical trials entailing quantitative personalised risk-stratification and surveillance following curative-intent radiotherapy for NSCLC.

### Introduction

Lung cancer is the leading cause of cancer deaths worldwide.<sup>1</sup> Non-small cell lung cancer (NSCLC) accounts for 85% of lung cancers, with approximately 1 in 5 patients alive 5 years after diagnosis.<sup>2</sup> Recurrence is reported in up to 36% of patients receiving curative-intent treatment for NSCLC.<sup>3</sup> Surveillance is recommended across international guidelines for NSCLC patients treated with curative-intent radiotherapy.<sup>3</sup> Surveillance ensures on-going patient support, management of co-morbidities and cancer treatment-related side effects as well as detection of recurrence of the treated cancer or second (metachronous) primary cancers. Curative treatment following local recurrence results in 5-year survival rates of 15%.<sup>4</sup> Earlier detection of recurrence may therefore improve survival and quality of life. Similarly, surveillance-stratification may allow for better resource allocation.

International guidance highlights a lack of high-quality evidence to formulate specific recommendations on the nature and frequency of follow-up after radiotherapy.<sup>3</sup> Most guidance is based on expert consensus and advises follow-up with clinical review and contrast-enhanced chest CT at 3–6 monthly intervals for 2–3 years. Without an agreed framework, follow-up differs according to local policy or even individual clinician opinion.

The UK’s National Institute of Healthcare and Clinical Excellence (NICE) have recommended further research into use of prognostic factors to develop risk-stratification models to inform optimal surveillance.<sup>4</sup> Risk-stratification models may allow for personalised follow-up for NSCLC patients undergoing curative-intent radiotherapy, resulting in potentially earlier detection of recurrence for high-risk patients or avoidance of unnecessary scans and hospital visits for low-risk patients. Such models would improve patient care and healthcare resource management.

Previous studies using clinical data for outcome prediction following radiotherapy for NSCLC have explored cox proportional-hazards models and support vector machines, and have generally focused on OS, achieving validation or external test set AUCs between 0.61 and 0.69 and C-Index scores between 0.58 and 0.69.<sup>3,5–8</sup> It is possible that more advanced machine learning techniques can better handle large clinical datasets leading to superior performance. Comparison of the performance

of several machine learning models has been undertaken for systemic anti-cancer therapy for advanced NSCLC,<sup>9</sup> but is lacking in the radical radiotherapy cohort.

The intent of this study was therefore to utilise a broad range of readily available patient, tumour, and treatment related features in comparing performance of several machine learning methods to develop and validate prediction models for recurrence, recurrence-free survival (RFS) and overall survival (OS). Such models may be used to guide personalised risk-stratification and surveillance following curative-intent radiotherapy for NSCLC in future.

## Methods

### Ethics

This study was approved by the UK Health Research Authority (reference number: 20/HRA/3051), Clinical-Trials.gov identifier: NCT04721444.

### Datasets

We utilised 3 independent, novel datasets of patients receiving primary curative-intent radiotherapy for stage I to III NSCLC, yielding a total of 722 patients across 5 hospitals.

Each dataset was retrospectively collated from electronic patient record (EPR) and radiotherapy treatment planning systems (TPS) at UK National Health Service (NHS) Trusts:

- Dataset X consists of 434 patients with stage I to III disease treated at X Trust with stereotactic or conventional radiotherapy with or without chemotherapy between 26/9/2014 and 23/10/2018.
- Dataset Y consists of 111 patients with stage I to III disease treated at Y Trust with conventional radiotherapy with or without chemotherapy between 3/2/2014 and 10/1/2019.
- Dataset Z consists of 177 patients with stage I to III disease treated at Z Trust with stereotactic or conventional radiotherapy with or without chemotherapy between 21/1/2016 and 18/12/2018.

The datasets with the most (Dataset X) and least (Dataset Y) patients were combined and randomly divided into training and validation sets. The testing set (Dataset Z) was selected by ring-fencing one specific site as a geographically external test set. The data were collected in early 2021, ensuring a minimum of 2 years of follow-up for all patients. Those with no known recurrence or death within 60 days of the 2-year endpoint, or no recurrence within 60 days of death, were taken to have no event. This cut-off was agreed upon by the authors after discussion about how best to reduce bias and factoring in the variable nature of clinical follow-up

appointments. 60 days was taken to be mean timeframe between follow-up appointments and thus an estimated half-way point between the last time a patient was seen and the 2-year endpoint. Therefore, we agreed that if a patient had no known recurrence or death within 60 days of the endpoint, or no recurrence within 60 days of death, it was unlikely that a recurrence/death had occurred.

The following patient demographics and clinical parameters were collected: sex, age, ethnicity, World Health Organisation (WHO) performance status, smoking status, TNM8 T-stage, TNM8 N-stage, TNM8 overall clinical stage, size of primary lesion, FDG PET-CT Standard Uptake Value (SUV) of primary lesion, nodal avidity and maximal nodal SUV, whether nodes were sampled (e.g. with endoscopic bronchial ultrasound, EBUS), whether there was a confirmed pathological diagnosis (e.g. with biopsy) and histological type, body mass index, pre-treatment forced expiratory volume in 1 second (FEV<sub>1</sub>, as percent predicted) and diffusing capacity for carbon monoxide (TLCO, as percent predicted), pre and post-treatment neutrophil and lymphocyte counts, type of radiotherapy treatment received (stereotactic body radiotherapy (SBRT) or conventional radiotherapy with or without chemotherapy), total dose in Gy, number of fractions, biologically effective dose in Gy (assuming an  $\alpha/\beta$  value of 10) radiotherapy gross tumour volume (GTV), radiotherapy planning target volume (PTV) and dates of radiotherapy planning scan and first and last fraction of radiotherapy.

To meet pre-processing requirements for machine learning, categorical data were converted to numeric. One-hot-encoding converted each level of each categorical feature into a new binary feature. To mitigate for resultant increase in data dimensionality, prior to one-hot-encoding, levels of some categorical features were combined, for example, “never”, “ex” and “current” smokers were binarized to “never” and “ever”. Implicit associations between variables were made explicit: dates of planning scan and first and last fraction were replaced with the number of days between them. Missing clinical data were assumed missing at random and non-dependent on outcome. Features with more than 25% of observations missing were removed. Missing data for remaining features was imputed using the multiple imputation with chained equations (MICE) package with default arguments in R. Highly correlated features were removed using the `treatment_corr` function, with a threshold of 0.85, removing one of each pair of correlated features (Pearson correlation for continuous and Spearman correlation for categorical features). Continuous features were standardized.

### Statistics

Patient demographics and clinical parameters are summarised as means and standard deviations for

continuous features, and frequencies and percentages for categorical features. Comparisons between datasets are summarised using Wilcoxon rank sum test for continuous features and Fisher exact test for categorical features. Time to event data for each of the study outcomes (recurrence, RFS and OS) were binarized at 2 years from the first fraction of radiotherapy for classification purposes – cases scored “1” if there was recurrence or death within 2 years, and “0” otherwise. Owing to the nature of clinical follow-up, 340 patients (47%) were not seen at or after the 2-year endpoint. As simply excluding these patients would bias the dataset, those with no known recurrence or death within 60 days of the endpoint, or no recurrence within 60 days of death, were taken to have no event. Those last seen more than 60 days from the endpoint or date of death were excluded ( $n = 65$ , 9%). More information can be found in the Supplementary Material section “Further details on datasets”. To ensure non-biased dataset assignment for training, validation, and testing, the datasets with the most (Dataset X) and least patients (Dataset Y) were combined and then cases randomly assigned as training and validation with an 80:20 ratio, stratified by the binarized outcome. The Z dataset was locked for testing.

#### Feature handling and modelling

Dimensionality reduction may be required prior to modelling to increase prediction accuracy, prevent overfitting, and reduce computational cost. We explored a combination of 10 linear, Bayesian, neural-net and tree-based machine learning algorithms applied to 8 different feature sets (either no feature reduction or following 1 of 7 feature reduction methods, where possible). Feature reduction methods included correlation-based, multivariate linear penalised and recursive approaches). A brief description of these algorithms, feature reduction methods and resultant feature sets can be found in Supplementary Table 2. Hyper-parameter optimisation was performed via grid-search with 3 repeats of 10-fold cross-validation using the caret package in R. Hyper-parameters of the final selected models are listed in Supplementary Table 1.

Receiver-Operator Characteristic (ROC) curves were created for the validation set results of each algorithm-feature set combination and the Area Under the Curve (AUC) calculated. Ensemble prediction models were then explored by averaging the predictions of the 3 algorithms with the highest AUC in the validation set for each particular outcome being predicted. Where the ensemble model was superior, it was selected as the final model for predicting that particular outcome on the external test set. Otherwise, the single algorithm with the highest AUC was selected as the final model.

Caret’s varImp function<sup>10</sup> was used to identify features that contributed most significantly to model

performance. It provides a generic method for calculating variable importance by either utilising a model’s native feature importance ranking method or using ROC curve analysis for each feature.<sup>10</sup> It is not possible to do this for averaged predictions across algorithms and so where an ensemble was selected as the final model, this was performed for the component algorithms instead.

#### Benchmarking

TNM-stage and performance status are known prognostic factors in NSCLC.<sup>11,12</sup> Logistic regression models based on TNM stage and performance status were developed to benchmark our prediction models against. As above, the Caret package was used with hyper-parameter optimisation performed via grid-search with 3 repeats of 10-fold cross-validation. AUC was calculated.

#### Risk-groups

Youden’s Index was calculated from the validation set ROC curve for each final model and used to create a classification threshold for outcome prediction. This threshold was also used to separate groups into high (outcome event occurs within 2 years of first fraction of radiotherapy) or low (outcome does not occur within 2 years of first fraction of radiotherapy) risk groups. Time to event data were then used to create Kaplan Meier curves demonstrating the difference in recurrence/RFS/OS between high and low risk groups. Log-rank test with a significance level of 0.05 was used to determine difference between survival curves. Performance of the risk models was assessed using the external test set. Using the same threshold as for Kaplan-Meier analysis, classification was performed for each outcome. Balanced accuracy and F1 score were recorded, in order to convey performance in the context of imbalanced endpoints. All analyses were carried out using R 3.5.1.

#### Role of the funding source

The funding source played no role in study design, data collection, data analysis, data interpretation, or writing of the report. All authors had full access to all the data in the study and the corresponding authors had final responsibility for the decision to submit for publication.

#### Results

In total, 657 patients were included in the study. Median follow-up was 852 days. The combined training-validation and external test sets included 498 and 159 patients respectively. Patient demographics and clinical parameters are summarised in Table 1.

Datasets were well-matched for most parameters. The external test set had a higher proportion of patients

Parameter	Combined Training & Validation Sets n = 498	External Test Set n = 159	P-value
Age (IQR)	74 (14)	72 (14)	.054
Sex (%)			.907
- Male	273 (54.8)	88 (55.4)	
- Female	225 (45.2)	71 (44.7)	
WHO Performance Status (%)			.023
- 0	82 (16.5)	16 (10.1)	
- 1	282 (56.6)	85 (53.5)	
- 2	118 (23.7)	55 (34.6)	
- Missing	16 (3.2)	3 (1.9)	
Body Mass Index (IQR)	25.1 (6.5)	26.22 (7.1)	.044
- Missing, n (%)	103 (20.7)	7 (4.4)	
Smoking Status (%)			.165
- Never	42 (8.4)	8 (5.0)	
- Ever	437 (87.8)	141 (88.7)	
- Missing	19 (3.8)	10 (6.3)	
TNM8 T stage (%)			.161
- 1	192 (38.6)	77 (48.4)	
- 2	133 (26.7)	36 (22.6)	
- 3	74 (14.9)	17 (10.7)	
- 4	99 (19.9)	29 (18.2)	
TNM8 N stage (%)			.041
- 0	276 (55.4)	106 (66.8)	
- 1	50 (10.0)	7 (4.4)	
- 2	130 (26.1)	35 (22.0)	
- 3	42 (8.4)	11 (6.9)	
FEV1, percent predicted (IQR)	76 (33.2)	68.5 (34.5)	.004
- Missing, n (%)	46 (9.2)	17 (10.7)	
TLCO, percent predicted (IQR)	60 (25)	57 (25.8)	.092
- Missing, n (%)	79 (15.9)	25 (15.7)	
Days from planning scan to first fraction (IQR)	18 (7)	18 (6.0)	.856
Size of primary (IQR)	33 (28)	30 (28.5)	.050
- Missing, n (%)	16 (3.2)	0	
SUV primary (IQR)	10.35 (8.9)	9.3 (9.8)	.829
- Missing, n (%)	54 (10.8)	8 (5.0)	
Max nodal SUV (IQR)	7.35 (6.8)	5.7 (4.7)	<.001
- Missing, n (%)	63 (12.7)	48 (30.2)	
Nodal avidity (%)			.024
- Yes	194 (39.0)	67 (42.1)	
- No	286 (57.4)	92 (57.9)	
- Missing	18 (3.6)	0	
Nodal Sampling (%)			.380
- Yes	159 (31.9)	57 (35.9)	
- No	334 (67.1)	99 (62.3)	
- Missing	5 (1.0)	3 (1.9)	
Histology (%)			<.001
- Adenocarcinoma	223 (44.8)	59 (37.1)	
- Squamous	158 (31.7)	56 (35.2)	
- Other	49 (9.8)	5 (3.1)	
- No pathology	68 (13.7)	39 (24.5)	
Treatment type (%)			.045
- SBRT	174 (34.9)	73 (45.9)	

Table 1 (Continued)

Parameter	Combined Training & Validation Sets <i>n</i> = 498	External Test Set <i>n</i> = 159	<i>P</i> -value
- Conventional RT	125 (25.1)	33 (20.8)	
- Chemo + RT	199 (40.0)	53 (33.3)	
Number of fractions (IQR)	20 (27)	20 (27)	.402
Total Dose, Gy (IQR)	55 (9)	55 (9)	.137
Biologically Effective Dose, Gy (IQR)	76.8 (45.4)	76.8 (38.7)	.023
Planning Target Volume, cm <sup>3</sup> (IQR)	218.39 (343.2)	126.62 (314.3)	.097
Recurrence at 2 years (%)	214 (43.0)	54 (34.0)	.051
Recurrence or death at 2 years (%)	267 (53.6)	74 (46.5)	.122
Death at 2 years (%)	185 (37.2)	54 (34.0)	.508
Median length of follow-up (range)	836 (0–2462)	868 (0–1442)	

**Table 1: Demographic and clinical parameters for combined training-validation and external test sets (prior to imputation). Features not used for modelling are not shown. Categorical data are summarised with means and percentages and *p*-values pertain to Fishers exact test. Continuous data are summarised with median and inter-quartile range (IQR) and *p*-values pertain to Wilcoxon rank sum test.**

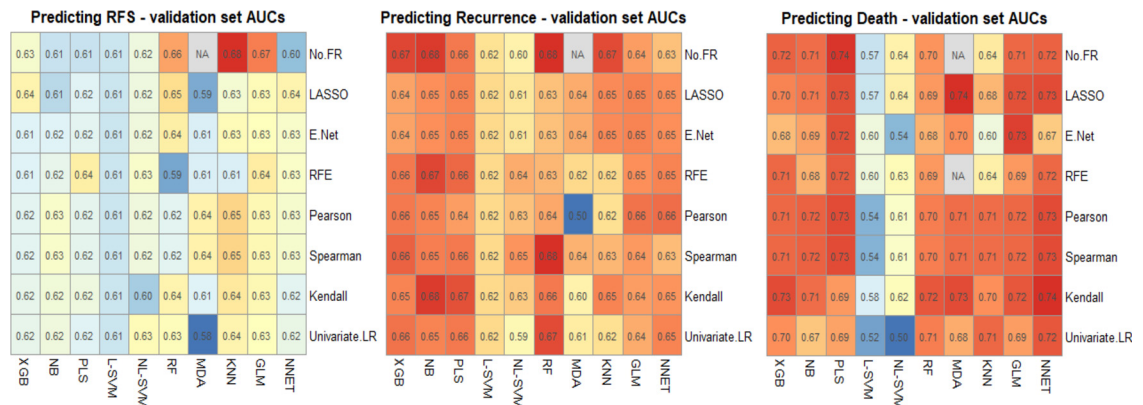
with performance status 2 (34.6% vs 23.7%) and with earlier stage disease (TNM8 T1-stage 1 48.4% vs 38.6%, No-stage 66.7% vs 55.4%). There were fewer cases of adenocarcinoma (37.1% vs 44.8%) and more patients lacking histological diagnosis (24.5% vs 13.7%) and receiving SBRT (45.9% vs 34.9%) in the external test set. The recurrence, RFS and OS endpoints were marginally imbalanced in the validation and external test sets.<sup>13</sup>

Following data pre-processing a total of 12 continuous and 22 discrete features were used for machine learning. These are listed in Supplementary Table 2, together with the features that remained after each feature reduction method for each of the measured outcomes. The results of our experiments with machine learning algorithms and feature set combinations on

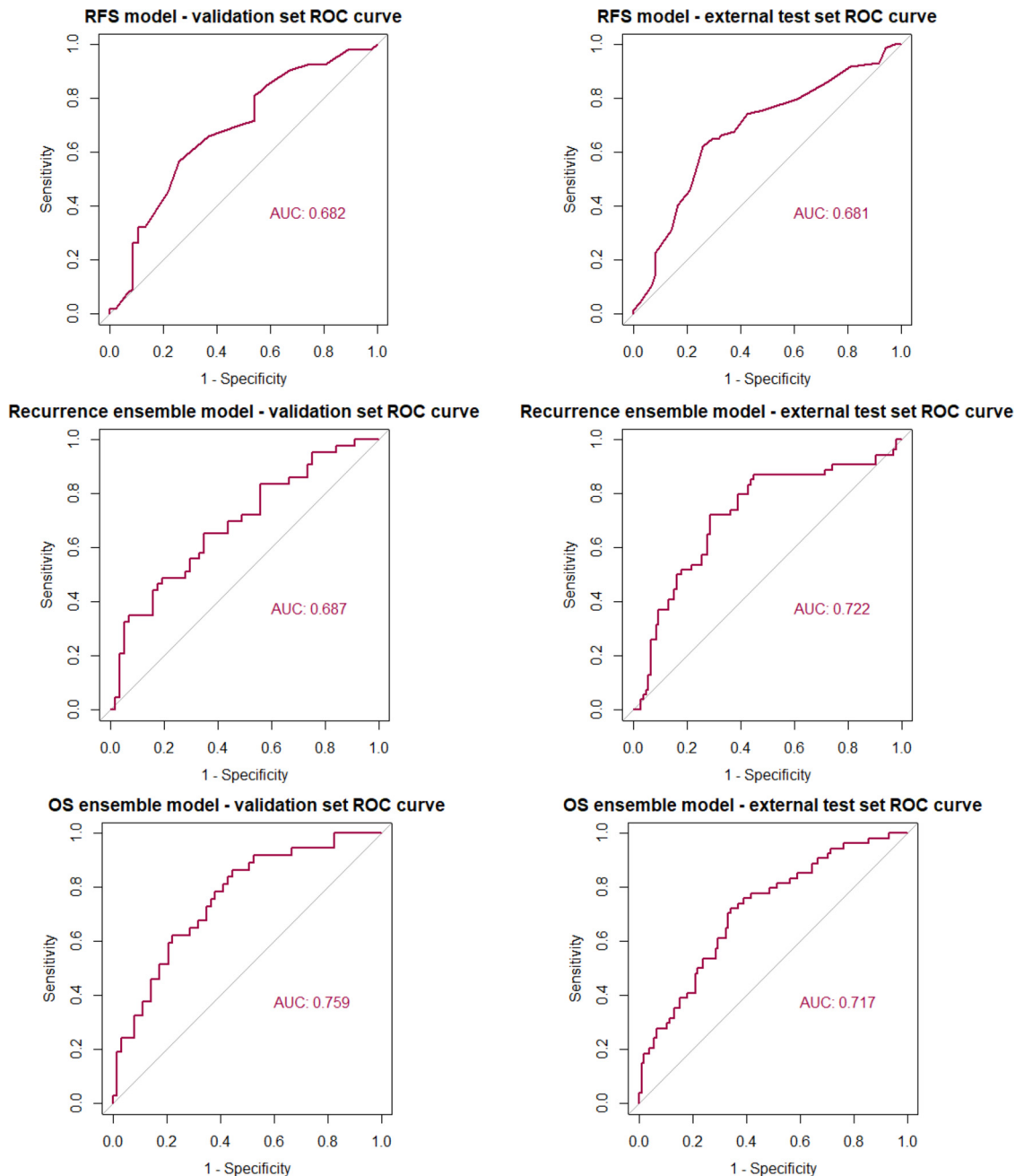
the validation set for each outcome are shown in Figure 1. These are in the form of heatmaps showing the AUC for each combination of machine learning algorithm (rows) with feature selection method (columns).

Averaging across the machine learning algorithms (columns), the best performing feature sets for predicting RFS and recurrence were the full sets with no further feature reduction and for predicting death, the best performing feature set was after Kendall's rank correlation.

The final prediction models chosen were as follows: For RFS, KNN alone, for recurrence, an ensemble of NB, RF and KNN, and for OS, an ensemble of MDA, XGB and NNET. Results of the final models on validation and external test sets are shown in Figure 2 (ROC



**Figure 1.** Heatmaps illustrating the performance of each machine learning algorithm (rows) with each feature reduction method (columns), measured by validation set AUC. No FR: No feature reduction (full feature set used), LASSO: Least Absolute Shrinkage and Selection Operator, E Net: Elastic-Net, RFE: Recursive Feature Elimination, Univariate LR: Univariate Logistic Regression, XGB: Extreme Gradient Boosting machine, NB: Naïve-Bayes, PSL: Partial Least Squares, L-SVM: Linear Support Vector Machine, NL-SVM: Non-linear (radial) SVM, RF: Random Forest, MDA: Mixture Discriminant Analysis, KNN: K-Nearest Neighbours, GLM: Generalised Linear Model, NNET: Neural Network.



**Figure 2.** ROC curves for the validation and external test set for each prediction.

curves) and [Table 2](#) (AUCs with 95% confidence intervals), together with a comparison against both the TNM stage and the performance status-based benchmarking models. Performance of our models was consistent across both validation and external test sets and whilst there was some overlap in confidence intervals with the benchmarking models, absolute AUC was superior for predicting recurrence and OS. For predicting RFS, our model was superior to the performance status-based

model in both validation and external test sets, however, was eclipsed by the TNM stage-based model in external test set performance.

Kaplan Meier survival curves for validation and external test sets for each outcome are shown in [Figure 3](#). Log-rank tests for the difference between survival curves was significant across the validation sets for RFS and recurrence ( $P < 0.01$ ) and approached significance for OS ( $P = 0.0993$ ). Stronger significance was

Outcome		Validation Set		External Test Set	
		AUC	95% CI	AUC	95% CI
RFS	<b>Our prediction model</b>	0.682	0.575–0.788	0.681	0.597–0.766
	<b>TNM based model</b>	0.650	0.541–0.760	0.695	0.616–0.774
	<b>PS based model</b>	0.464	0.363–0.565	0.499	0.418–0.58
Recurrence	<b>Our prediction model</b>	0.687	0.582–0.793	0.722	0.635–0.810
	<b>TNM based model</b>	0.670	0.563–0.777	0.707	0.622–0.791
	<b>PS based model</b>	0.506	0.402–0.609	0.584	0.503–0.665
OS	<b>Our prediction model</b>	0.759	0.663–0.855	0.717	0.634–0.800
	<b>TNM based model</b>	0.649	0.541–0.756	0.665	0.579–0.751
	<b>PS based model</b>	0.459	0.357–0.561	0.531	0.447–0.615

**Table 2: AUC with 95% confidence intervals for the validation and external test set for each prediction model, benchmarked against models based on TNM-stage and performance status.**

demonstrated for all 3 outcomes in the external test set ( $P < 0.001$ ), confirming a statistically significant difference in outcome between the high and low risk groups for prediction of each outcome. Classification metrics including balanced accuracy, F1 score, sensitivity, specificity, positive and negative predictive value along with Brier scores and calibration curves are detailed in Supplementary Tables 3 and 7 and Figure 1.

#### Feature importance (subsection of results)

Supplementary Tables 4a-c show features ranked in order of importance for the KNN model for predicting RFS and for the component algorithms of the final ensemble models for predicting recurrence and OS. For RFS the top 5 features were PTV, size of the primary, the total number of fractions, whether treatment was with SBRT and the BED.

For predicting recurrence, feature importance for KNN and NB were identical, owing to the fact that these models do not have their own corresponding varImp package methods and instead use a filter approach where AUC is used to measure variable importance. For these models the top 5 features were also PTV, primary tumour size, number of fractions, SBRT treatment and BED. The RF model also ranked PTV and size of the primary, as well as the SUV of the primary, patient age and BMI in the top 5 most important features.

For predicting OS, the 8 features remaining after Kendall's rank correlation were PTV, size of the primary, BED, T1 stage, treatment with SBRT, treatment with chemoradiotherapy, whether there was nodal avidity and smoking status. There was a wide distribution of feature importance across features for the XGB and NNET models.

#### Discussion

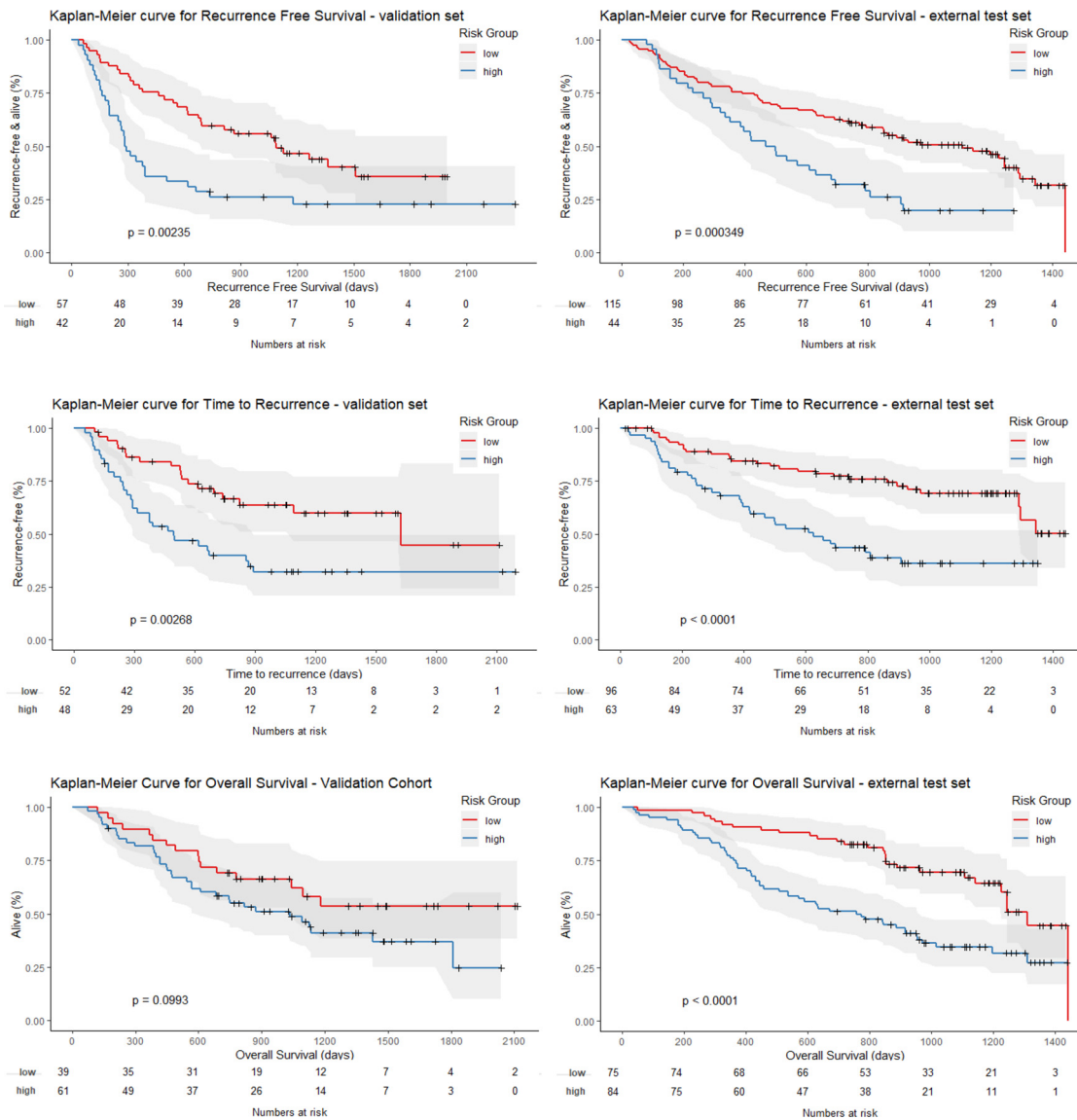
In this multicentre UK study of over 700 patients treated with curative-intent radiotherapy for NSCLC, we

have compared machine learning algorithms and feature selection methods using routinely available clinical data and have developed and externally tested prediction models which are able to categorise patients into low and high risk for recurrence, RFS and OS, two years from the start of treatment. Such models may have future utility in personalised surveillance stratification, whereby only those at greatest risk have the most intensive follow-up.

AUCs for all three models are consistent between validation and external test sets though the recurrence model may demonstrate underfitting and the OS model overfitting in validation sets. Performance of the models is reasonable, demonstrating minor improvement on TNM-based prognostication, but highlight the need for more effective risk-stratification measures following radical radiotherapy for NSCLC – a call that may be better addressed with radiomic or deep-learning based image analysis.

For predicting RFS and recurrence, averaging across the machine learning algorithms, the best performing feature sets were those with no prior feature reduction. Unlike with -omics data where feature sets tend to be very large and are likely to contain noise and highly correlated features, this clinical feature set is comparatively small ( $n = 34$ ) with all features possibly contributing to model performance to some degree, rendering feature reduction unnecessary. This is particularly the case where the models have inherent ability to handle multi-dimensional data and collinearity,<sup>14</sup> for example LASSO and tree-based methods which have their own internal regularisation (L1-norm and number of estimators respectively). This was not the case for predicting OS however where Kendall's rank yielded the best performing feature set. One explanation for this is that OS had a lower event rate compared to recurrence and RFS, and therefore a higher risk of over-fitting to the majority class. As a result, feature reduction may play a more significant role in mitigating against this. Kendall's rank may have emerged as the best performing correlation





**Figure 3.** Kaplan Meier survival curves for low and high-risk groups in both validation and external test sets for each prediction model. *P*-values correspond to log-rank tests.

method because it does not assume a normal distribution, is more robust to outliers and better suited to smaller datasets

Looking at the top 5 features ranked by importance for the final models or their component algorithms, these are features related to: size (PTV and size of the primary), stage (T1 stage and whether there was nodal avidity), treatment (BED, total number of fractions and treatment with SBRT) and also smoking status (ranked top 5 by the XGB and NNET algorithms contributing to the OS model) and SUV of the primary, BMI and age (ranked top 5 by the RF algorithm contributing to recurrence).

Whilst tumour size and stage are known prognostic factors for NSCLC,<sup>12,15,16</sup> despite the high feature importance ranking in our models, strong evidence for tumour volume is lacking.<sup>17–19</sup> The high importance of features related to treatment presumably reflects that our study included patients treated with SBRT as well as conventional radiotherapy and thus likely two cohorts of patients – those with early-stage disease that were older and not fit for surgery (treated with SBRT) and younger patients with more advanced disease (treated with conventional (chemo)radiotherapy). The former is less likely to experience recurrence both due to having early-stage disease but also as they are more likely to die

of other comorbidities prior to recurrence. Smoking is linked to increased comorbidities and smoking status is known to be an independent prognostic factor for survival in NSCLC.<sup>11</sup> A number of studies have demonstrated prognostic utility of PET SUV in NSCLC.<sup>20–23</sup> BMI and age are also likely to reflect the older and frailer cohort of patients described above that were likely to have been treated with SBRT.

A recent study also explored recurrence and OS prediction following curative-intent radiotherapy for NSCLC, using cox-proportional hazards models. Backwards stepwise elimination was used to select significant features resulting in 6 features for predicting recurrence and 5 features for predicting OS. External testing was not performed in this study, however on comparison of validation set AUCs with our external test set, our models demonstrate superior performance for predicting OS and similar for recurrence (0.717 vs 0.607 and 0.722 vs 0.72, respectively). Whilst this study did not rank feature importance, PS and nodal sampling with EBUS were common to both models<sup>3</sup> and are known to hold prognostic value.<sup>11,24</sup> Interestingly these were not considered in the top important features by our models.

Previous studies using clinical (non-imaging) data for outcome prediction following curative intent radiotherapy for NSCLC have generally focused on homogeneous cohorts and OS as the key end-point, with few studies looking at recurrence or RFS.<sup>5–8</sup> This is likely due to accurate data on death being more readily accessible from national-level registries than recurrence data, however recurrence and RFS may offer more clinical utility than OS in stratifying follow-up, allowing clinicians to potentially intervene earlier to provide further, potentially curable treatment.

Earlier studies have focused on traditional statistical and machine learning models such as Cox-proportional hazards and SVMs.<sup>3,5–8</sup> Cox models have limitations in that the proportional hazard assumption and linearity of each variable must be satisfied, which can be difficult to do so with real-world data and may result in an inappropriate model fit. Secondly, in the context of tied samples, approximations are often employed to improve computational efficiency however these may result in significantly different results.<sup>25</sup> Furthermore, survival analysis can experience multiple-testing problems in the context of high-dimensional data and may yield a high number of false positive significant features, which reduces reliability of analysis.<sup>26</sup> Feature reduction based approaches coupled with more advanced machine learning algorithms may mitigate this issue and have been shown to provide more accurate alternatives to traditional Cox proportional-hazards models.<sup>14</sup>

Another potential benefit of exploring more advanced machine learning approaches is the ability of models to highlight “important” features as described above. Until recently prediction models tended to rely

on clinician-determined input features thought to be of significance based on previous research. This approach is at risk of biasing and limiting choice of input features through human assumption. This can be mitigated to some degree by selecting a large number of potential input features and allowing machine learning models to select those that are best performing.<sup>25</sup> These may not necessarily correlate with those thought to be most important by clinicians and may highlight features that were previously not considered.

A previous study compared discriminative performance of 6 classifiers in predicting a range of outcomes following radiotherapy for different tumours.<sup>27</sup> Random forest and elastic net models showed the best overall discrimination, but no single classifier performed well across all datasets. The authors concluded that future investigators should benchmark models against random forest and elastic net models but that overall, informed preselection of a classifier based on specific datasets is advised. We build on this work by exploring a larger number of classifiers (including random forest and elastic net) together with feature reduction and believe that our methodology of applying this to our specific dataset is a robust approach to selecting the most suitable model.

Other strengths of our work are the multicentre design including a large cohort of over 700 patients from expert UK radiotherapy centres which is thus likely representative of UK (and probably European practice). Models were built using readily available clinical data with no requirement for access to, and complex pre-processing of imaging data. A broad set of features was used encompassing patient demographics, fitness, tumour characteristics and treatment parameters. The results are consistent across both validation and external test sets. Furthermore, this study included stage I-III disease treated with SBRT and conventional (chemo) radiotherapy, thus increasing clinical utility.

Whilst previous studies have demonstrated use of clinical data in developing prognostic prediction models following radiotherapy for NSCLC, none of these have compared multiple feature reduction methods and machine learning classification algorithms, as conveyed in our work. Finally, whilst these studies focus on survival analysis using few features, in this work, we have utilised the entire available feature set in our predictive modelling approach. Jochems et al. used a Bayesian network to predict 2-year survival using T-stage, N-stage, age and total tumour dose, with an external test set AUC of 0.66.<sup>28</sup> With our wider range of features, our survival model achieves superior performance of 0.717.

Imaging and biological markers have been used for prognostication in NSCLC, however this requires complex pre-processing of imaging data or biological samples and may necessitate invasive procedures. The focus of our study was to use routinely available electronic health record clinical data to overcome these challenges.

We believe that this methodology may be replicated across health systems, including those in resource-poor settings, using local clinical datasets to benefit surveillance stratification for patients following curative-intent radiotherapy globally. In addition, to demonstrate that there is a true benefit in using models built using imaging and biological markers, these should be compared with state-of-the-art clinical feature-based models. Our methodology has the potential to act as a clinical baseline for such models, as it provides reasonable predictions with clinically interpretable features. Furthermore, it may be combined with imaging and biological markers to enhance future models.

Weaknesses of our study include its retrospective nature, and reliance on data retrieved from the EPR and TPS of participating centres. As some participating sites were tertiary referral centres, some clinical information was not available if omitted at the time of initial referral. Owing to missing data, we had to omit 4 features: pre and post-treatment lymphocyte and neutrophil counts, which are known to have prognostic significance in this cohort.<sup>29–31</sup> On the contrary, this was necessary as to maintain generalisability and utility, our models need to be built using features that are consistently available at other centres.

Treatment heterogeneity is a limitation that may be addressed by developing separate models for each treatment type. It would have been difficult to obtain sufficiently large datasets for a single treatment modality however, as we have curated novel datasets for this work. Additionally, within a single treatment modality there remains variation in dose and fractionation. Treatment modality, dose and number of fractions were features used for modelling to account for this and we therefore believe that our models allow for wider clinical utility. This work, including the novel datasets we have curated, may serve as a foundation for future models which are more specific to radiotherapy dose-fractionation schedules and treatment modalities.

Patients assumed to have NSCLC, without confirmed pathological diagnosis were included in this study as this is reflective of clinical practice, particularly for early-stage disease treated with SBRT. There is therefore potential inclusion of patients with benign disease or small cell lung cancer which may confound recurrence and survival rates, however our models were provided with this information and such patients would still benefit from post-treatment risk-stratification. Our study did not include patients treated with tri-modality therapy (surgery before or after radiotherapy) and we do not have data on patients treated with adjuvant Durvalumab as per the PACIFIC trial<sup>32</sup> or for oligometastatic recurrence which may impact on recurrence or survival.

In developing our ensemble models, we used averaged predictions of the component algorithms, however more complex ensemble architectures such as random forest or majority weighting may yield superior results.

Finally, owing to the patient cohort used for training and validation, whilst our models are reproducible, they are unlikely to be generalisable outside the UK or Europe. Replication of these data in future international prospective clinical trials are warranted.

In this study we have compared multiple machine learning algorithms and feature selection methods to develop prognostic prediction models for recurrence, RFS and OS following curative-intent radiotherapy for NSCLC. Our models demonstrate reasonable performance exceeding traditional methods and show consistency across validation and external test sets. More advanced imaging-based models may surpass this approach but there is potential value in use of routinely available clinical data. This robust and ready to use machine learning method, validated and externally tested, sets the stage for future clinical trials entailing quantitative personalised risk-stratification and surveillance following curative-intent radiotherapy for NSCLC.

#### Contributors

Sumeet Hindocha, Kristofer Linton-Reid, Merina Ahmed, Richard Lee, and Eric Aboagye were involved in the study design.

Sumeet Hindocha, Thomas Charlton, Benjamin Hunter, and Charleen Chan collected the data.

Sumeet Hindocha analysed and interpreted the data.

Sumeet Hindocha undertook literature review and wrote the initial draft.

Sumeet Hindocha and Kristofer Linton-Reid undertook radiomic analysis.

Kristofer Linton-Reid, Emily Robinson, Matthew Orton, and Matthew Blackledge provided statistical and machine learning advice.

Merina Ahmed, Shahreen Ahmad, Fiona McDonald, Imogen Locke, Danielle Power and Richard Lee provided clinical advice.

All authors reviewed, contributed to, and approved the manuscript. All authors had access to the data and Sumeet Hindocha and Richard Lee directly accessed and verified the underlying data reported in the manuscript. Sumeet Hindocha, Richard Lee and Eric Aboagye were responsible for the decision to submit the manuscript.

#### Declaration of interests

Richard Lee receives funding from Cancer Research UK, Innovate UK (co-funded with Roche and Optelium), NIHR and RM Partners outside of this work. He is Joint Clinical Lead for the NHS England Lung Health Checks programme and a National Specialty Lead for the NIHR, and receives funding directly to his institution, outside of this work for these roles. He receives consulting fees from the Royal Marsden Private Care, and honoraria from Cancer Research UK as a member

of the Early Diagnosis grants peer review panel, not related to this work.

### Data sharing

Due to confidentiality, data collected for the study are not publicly available for download, however the corresponding authors can be contacted for academic enquiries. Tools for machine learning are indicated in the methods section and the code for the predictive vectors will be deposited at <https://github.com/orgs/cancer-imaging-group/dashboard>

### Acknowledgements

This study represents independent research funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at The Royal Marsden NHS Foundation Trust and The Institute of Cancer Research, London, and by the Royal Marsden Cancer Charity. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. Sumeet Hindocha is funded by the UKRI CDT in AI for Healthcare <http://ai4health.io> (Grant No. P/S023283/1), by Imperial College London and by the Royal Marsden & Institute of Cancer Research NIHR Biomedical Research Centre. Benjamin Hunter is funded by Cancer Research UK (grant reference C309/A31316) and Royal Marsden Partners. Richard Lee is funded by the Royal Marsden Cancer Charity and by the Royal Marsden & Institute of Cancer Research NIHR Biomedical Research Centre. Eric Aboagye receives funding from the Imperial College Biomedical Research Centre and Experimental Cancer Medicines Centre, paid to his institution. Richard Lee and Eric Aboagye provided joint supervision for this work.

### Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:[10.1016/j.ebiom.2022.103911](https://doi.org/10.1016/j.ebiom.2022.103911).

### References

- Arshad MA, Thornton A, Lu H, et al. Discovery of pre-therapy 2-deoxy-2-<sup>18</sup>F-fluoro-D-glucose positron emission tomography-based radiomics classifiers of survival outcome in non-small-cell lung cancer patients. *Eur J Nucl Med Mol Imaging*. 2019;46(2):455–466. Springer Berlin Heidelberg;Feb 1.
- Baek S, He Y, Allen BG, et al. Deep segmentation networks predict survival of non-small cell lung cancer. *Sci Rep*. 2019;9(1):17286.. [Internet]Dec 21 [cited 2019 Dec 5]Available from: <http://www.nature.com/articles/s41598-019-53461-2>.
- Evison M, Barrett E, Cheng A, et al. Predicting the Risk of Disease recurrence and death following curative-intent radiotherapy for non-small cell lung cancer: the development and validation of two scoring systems from a large multicentre UK cohort. *Clin Oncol*. 2021;33(3):145–154. W.B. Saunders;Mar 1.
- The diagnosis and treatment of lung cancer (update) Full Guideline. 2019. National Institute for Health and Care Excellence. [cited 2020 July 30]. Available from: <https://www.nice.org.uk/guidance/ng122/evidence/full-guideline-pdf-6722113502>.
- Sher A, Medavaram S, Nemesure B, Clouston S, Keresztes R. Risk stratification of locally advanced non-small cell lung cancer (Nscl) patients treated with chemo-radiotherapy: an institutional analysis. *Cancer Manag Res*. 2020;12:7165–7171. [Internet]. Cancer Manag Res[cited 2021 Jul 27]Available from: <https://pubmed.ncbi.nlm.nih.gov/32848470/>.
- Oberije C, De Ruyscher D, Houben R, et al. A validated prediction model for overall survival from stage III non-small cell lung cancer: Toward survival prediction for individual patients. *Int J Radiat Oncol Biol Phys*. 2015;92(4):935–944. [Internet]. Elsevier Inc.;Jul 15 [cited 2020 Jul 30]Available from: [/pmc/articles/PMC4786012/?report=abstract](https://pmc/articles/PMC4786012/?report=abstract).
- Oberije C, Nalbantov G, Dekker A, et al. A prospective study comparing the predictions of doctors versus models for treatment outcome of lung cancer patients: a step toward individualized care and shared decision making. *Radiother Oncol*. 2014;112(1):37–43. Elsevier Ireland Ltd;Jul 1.
- Dekker A, Vinod S, Holloway L, et al. Rapid learning in practice: a lung cancer survival decision support system in routine patient care data. *Radiother Oncol*. 2014;113(1):47–53. [Internet]. Elsevier Ireland Ltd[cited 2020 Jul 30]Available from: <https://pubmed.ncbi.nlm.nih.gov/25241994/>.
- Siah KW, Khozin S, Wong CH, Lo AW. Machine-learning and stochastic tumor growth models for predicting outcomes in patients with advanced non-small-cell lung cancer. *JCO Clin Cancer Inform*. 2019;3(3):1–11. [Internet]. American Society of Clinical Oncology (ASCO)Nov [cited 2020 Jul 30]Available from: <https://pubmed.ncbi.nlm.nih.gov/31539267/>.
- M. Kuhn 15 variable importance the caret package [Internet]. 2019 [cited 2021 Aug 12]. Available from: <https://topepo.github.io/caret/variable-importance.html>.
- Kawaguchi T, Takada M, Kubo A, et al. Performance status and smoking status are independent favorable prognostic factors for survival in non-small cell lung cancer: a comprehensive analysis of 26,957 patients with NSCLC. *J Thorac Oncol*. 2010;5(5):620–630. Elsevier;May 1.
- Ball D, Mitchell A, Giroux D, Rami-Porta R. Effect of tumor size on prognosis in patients treated with radical radiotherapy or chemoradiotherapy for non-small cell lung cancer: An analysis of the staging project database of the international association for the study of lung cancer. *J Thorac Oncol*. 2013;8(3):315–321. [Internet]. J Thorac Oncol[cited 2021 Jul 27]Available from: <https://pubmed.ncbi.nlm.nih.gov/23334092/>.
- He H, Ma Y. *Imbalanced Learning: Foundations, Algorithms, and Applications*. Wiley; 2013:1–210. Imbalanced Learning: Foundations, Algorithms, and Applications.
- Spooner A, Chen E, Sowmya A, et al. A comparison of machine learning methods for survival analysis of high-dimensional clinical data for dementia prediction. *Sci Rep*. 2020;10(1):1–10. [Internet]. Nature Publishing GroupNov 23 [cited 2021 Jul 27]Available from: <https://www.nature.com/articles/s41598-020-77220-w>.
- Asamura H, Chansky K, Crowley J, et al. The international association for the study of lung cancer lung cancer staging project: proposals for the revision of the N descriptors in the forthcoming 8th edition of the TNM classification for lung cancer. *J Thorac Oncol*. 2015;10:1675–1684. J Thorac Oncol[cited 2021 Jul 27]Available from: <https://pubmed.ncbi.nlm.nih.gov/26709477/>.
- Rami-Porta R, Bolejack V, Crowley J, et al. The IASLC lung cancer staging project: proposals for the revisions of the T descriptors in the forthcoming eighth edition of the TNM classification for lung cancer. *J Thorac Oncol*. 2015;10(7):990–1003. [Internet]. J Thorac Oncol;Jul 4 [cited 2021 Jul 27]Available from: <https://pubmed.ncbi.nlm.nih.gov/26134221/>.
- Ball DL, Fisher RJ, Burmeister BH, et al. The complex relationship between lung tumor volume and survival in patients with non-small cell lung cancer treated by definitive radiotherapy: a prospective, observational prognostic factor study of the trans-tasman radiation oncology group (TROG 99.05). *Radiother Oncol*. 2013;106(3):305–311. [Internet]. Radiother Oncol;Mar [cited 2021 Jul 27]Available from: <https://pubmed.ncbi.nlm.nih.gov/23333017/>.
- Käsmann L, Niyazi M, Blanck O, et al. Predictive and prognostic value of tumor volume and its changes during radical radiotherapy of stage III non-small cell lung cancer: a systematic review. *Strahlenther Onkol*. 2018;194:79–90. Strahlenther Onkol[cited 2021 Jul 27]Available from: <https://pubmed.ncbi.nlm.nih.gov/29030654/>.

- 19 Yu Y, Guan H, Xing LG, Xiang YB. Role of gross tumor volume in the prognosis of non-small cell lung cancer treated with 3D conformal radiotherapy: a meta-analysis. *Clin Ther*. 2015;37(10):2256–2266. [Internet]. *Clin Ther* 1 [cited 2021 Jul 27] Available from: <https://pubmed.ncbi.nlm.nih.gov/26293808/>.
- 20 Chen F, Yao Y, Ma C, et al. Ratio of maximum standardized uptake value to primary tumor size is a prognostic factor in patients with advanced non-small cell lung cancer. *Transl Lung Cancer Res*. 2015;4(1):18–26. [Internet]. *Transl Lung Cancer Res* [cited 2021 Jul 27] Available from: <https://pubmed.ncbi.nlm.nih.gov/25806343/>.
- 21 Paesmans M, Garcia C, Wong CYO, et al. Primary tumour standardised uptake value is prognostic in nonsmall cell lung cancer: a multivariate pooled analysis of individual data. *Eur Respir J*. 2015;46(6):1751–1761. [Internet]. European Respiratory Society Dec 1 [cited 2021 Jul 27] Available from: <https://erj.ersjournals.com/content/46/6/1751>.
- 22 Kohutek ZA, Wu AJ, Zhang Z, et al. FDG-PET maximum standardized uptake value is prognostic for recurrence and survival after stereotactic body radiotherapy for non-small cell lung cancer. LUNG CANCER [Internet]. *Lung Cancer*. 2015;89(2):115–120. Aug 1 [cited 2021 Jul 27] Available from: <https://pubmed.ncbi.nlm.nih.gov/26078260/>.
- 23 Horne ZD, Clump DA, Vargo JA, et al. Pretreatment SUV max predicts progression-free survival in early-stage non-small cell lung cancer treated with stereotactic body radiation therapy. *Radiat Oncol*. 2014;9(1):1–6. 2014 91 [Internet]. BioMed Central Jan 30 [cited 2021 Jul 27] Available from: <https://ro-journal.biomedcentral.com/articles/10.1186/1748-717X-9-41>.
- 24 Moore S, Leung B, Wu J, Ho C. Real-world treatment of stage III NSCLC: the role of trimodality treatment in the era of immunotherapy. *J Thorac Oncol*. 2019;14(8):1430–1439. [Internet]. J Thorac Oncol Aug 1 [cited 2021 Aug 12] Available from: <https://pubmed.ncbi.nlm.nih.gov/31002953/>.
- 25 Lee B, Chun SH, Hong JH, et al. DeepBTS: prediction of recurrence-free survival of non-small cell lung cancer using a time-binned deep neural network. *Sci Rep*. 2020;10(1):145–154. [Internet]. Nature Research Dec 1 [cited 2020 Jul 30]. Available from: <https://pmc/articles/PMC7005286/?report=abstract>.
- 26 Zhang Y, Oikonomou A, Wong A, Haider MA, Khalvati F. Radio-mics-based prognosis analysis for non-small cell lung cancer. *Sci Rep*. 2017;7(1):1–8. [Internet]. Nature Publishing Group Apr 18 [cited 2021 Jul 27] Available from: <https://www.nature.com/articles/srep46349>.
- 27 Deist TM, Dankers F, Valdes G, et al. Machine learning algorithms for outcome prediction in (chemo)radiotherapy: an empirical comparison of classifiers. *Med Phys*. 2018;45(7):3449–3459. [Internet]. 2018/06/13 Jul Available from: <https://pubmed.ncbi.nlm.nih.gov/29763967>.
- 28 Jochems A, Deist TM, El Naqa I, et al. Developing and validating a survival prediction model for NSCLC patients through distributed learning across 3 countries. *Int J Radiat Oncol Biol Phys*. 2017;99(2):344–352. [Internet]. 2017/04/24. Elsevier Science Inc Oct 1 Available from: <https://pubmed.ncbi.nlm.nih.gov/28871984>.
- 29 Dong X, Zhou ZM, Bi N, et al. Prognostic value of pretreatment neutrophil-to-lymphocyte ratio (NLR) in locally advanced non-small cell lung cancer patients treated with thoracic radiation. *Zhonghua Zhong Liu Za Zhi*. 2018;40(6):446–451. [Internet]. *Zhonghua Zhong Liu Za Zhi* Jun 23 [cited 2021 Jul 27] Available from: <https://pubmed.ncbi.nlm.nih.gov/29936771/>.
- 30 Wang D, Guo D, Li A, Wang P, Teng F, Yu J. The post-treatment neutrophil-to-lymphocyte ratio and changes in this ratio predict survival after treatment of stage III non-small-cell lung cancer with conventionally fractionated radiotherapy. *Futur Oncol*. 2020;16(9):439–449. [Internet]. *Futur Oncol* Mar 1 [cited 2021 Jul 27] Available from: <https://pubmed.ncbi.nlm.nih.gov/32141321/>.
- 31 Palomar-Abril V, Soria-Comes T, Campos ST, Ureste MM, Bosch VG, Maiques ICM. Dynamic evaluation of neutrophil-to-lymphocyte ratio as prognostic factor in stage III non-small cell lung cancer treated with chemoradiotherapy. *Clin Transl Oncol*. 2020;22(12):2333–2340. [Internet]. *Clin Transl Oncol* Dec 1 [cited 2021 Jul 27] Available from: <https://pubmed.ncbi.nlm.nih.gov/32449125/>.
- 32 Antonia SJ, Villegas A, Daniel D, et al. Overall survival with durvalumab after chemoradiotherapy in stage III NSCLC. *N Engl J Med*. 2018;379(24):2342–2350. [Internet]. Massachusetts Medical Society Sep 25 [cited 2021 Jul 27] Available from: <https://www.nejm.org/doi/full/10.1056/nejmoa1809697>.