Radiotherapy and Oncology 133 (2019) 106–112

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Predicting radiation pneumonitis in locally advanced stage II–III non-small cell lung cancer using machine learning

José Marcio Luna [a,1,*], Hann-Hsiang Chao [a,1], Eric S. Diffenderfer [a], Gilmer Valdes [b], Chidambaram Chinniah [c], Grace Ma [a], Keith A. Cengel [a], Timothy D. Solberg [b], Abigail T. Berman [a], Charles B. Simone II [d]

[a] Department of Radiation Oncology, University of Pennsylvania, Philadelphia; [b] Department of Radiation Oncology, University of California San Francisco; [c] Albany Medical College; and [d] Department of Radiation Oncology, University of Maryland School of Medicine, Baltimore, United States

ARTICLE INFO

Article history:
Received 22 May 2018
Received in revised form 4 January 2019
Accepted 7 January 2019

Keywords:
Radiation pneumonitis
Non-small cell lung cancer
Machine learning
Random forest
RUSBoost
CART
Support vector machines
Logistic regression

ABSTRACT

Background and purpose: Radiation pneumonitis (RP) is a radiotherapy dose-limiting toxicity for locally advanced non-small cell lung cancer (LA-NSCLC). Prior studies have proposed relevant dosimetric constraints to limit this toxicity. Using machine learning algorithms, we performed analyses of contributing factors in the development of RP to uncover previously unidentified criteria and elucidate the relative importance of individual factors.

Materials and methods: We evaluated 32 clinical features per patient in a cohort of 203 stage II–III LA-NSCLC patients treated with definitive chemoradiation to a median dose of 66.6 Gy in 1.8 Gy daily fractions at our institution from 2008 to 2016. Of this cohort, 17.7% of patients developed grade $\geq$2 RP. Univariate analysis was performed using trained decision stumps to individually analyze statistically significant predictors of RP and perform feature selection. Applying Random Forest, we performed multivariate analysis to assess the combined performance of important predictors of RP.

Results: On univariate analysis, lung V20, lung mean, lung V10 and lung V5 were found to be significant RP predictors with the greatest balance of specificity and sensitivity. On multivariate analysis, Random Forest (AUC = 0.66, p = 0.0005) identified esophagus max (20.5%), lung V20 (16.4%), lung mean (15.7%) and pack-year (14.9%) as the most common primary differentiators of RP.

Conclusions: We highlight Random Forest as an accurate machine learning method to identify known and new predictors of symptomatic RP. Furthermore, this analysis confirms the importance of lung V20, lung mean and pack-year as predictors of RP while also introducing esophagus max as an important RP predictor.

Radiation pneumonitis (RP) is a common, potentially dose-limiting, and clinically significant toxicity associated with thoracic irradiation [1]. Increased use of concurrent chemotherapy has made this toxicity particularly relevant, with reported incidence in patients receiving definitive chemoradiation for locally advanced non-small cell lung cancer (LA-NSCLC) ranging from 15 to 40% [2–6], with an even greater proportion of patients developing subclinical changes reflective of pneumonitis [7]. The impact of this toxicity can be associated with significant morbidity and limitation of quality of life, necessitating interventions such as steroid initiation/dependence and oxygen support; resulting in radiotherapy dose limitations which negatively affect clinical outcomes [6]. Therefore, it is imperative to be cognizant of means to mitigate this toxicity.

Prior attempts at identifying predictors of RP have shown the importance of factors such as chemotherapy use and agents, dose fractionation, tumor location, age, and dosimetric/volumetric factors related to the lungs and even heart [6,8–12]. However, there currently is a lack of consensus on the optimal thresholds and cut-offs of dosimetric criteria, as well as the comparative importance of associated predictors. Limited prior analyses have attempted to generate predictive models with varying degrees of success and clinical utility [8,12–15] and have suffered from issues such as heterogeneity in the patient population and treatment techniques. This study aims to add to the existing literature on predictors of RP using machine learning. We examined a cohort of 203 consecutive LA-NSCLC patients treated at our institution using modern treatment planning and delivery technology, reflective of current techniques applicable to modern practice. Our method to analyze clinical and demographic structured data [16–18], incorporates

* Corresponding author at: Department of Radiation Oncology, University of Pennsylvania, 3400 Civic Center Blvd, TRC 8-130, Philadelphia, PA 19104, United States.
E-mail address: Jose.Luna@uphs.upenn.edu (J.M. Luna).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.radonc.2019.01.003
0167-8140/

machine learning algorithms allowing us to identify important features and optimal thresholds predictive for RP development. This study demonstrates the potential power and applicability of machine learning to the modeling of RP predictors and describes prioritization of clinical criteria to consider in clinical practice.

Materials and methods

Patient cohort

With institutional review board approval, we identified a cohort of 203 consecutive patients with histologically confirmed AJCC 7th Edition Stage II–III LA-NSCLC treated between 2008 and 2016 at our institution with sequential or concurrent chemoradiation with platinum-containing regimens. Patients received treatment using either intensity-modulated radiation treatment (IMRT) or proton beam therapy (PBT). RP was graded according to the Common Terminology Criteria for Adverse Events v4.0.

Feature definition

In this study, we analyzed 32 continuous and categorical features. The continuous features were smoking pack-years (pack-year), body mass index (BMI) and age at diagnosis, primary tumor size, PFT pre-bronchodilator, PFT DLCO (% predicted), FEV1 (L), radiation total dose (total dose), radiation fraction size, number of radiation fractions (nr. fractions), mean heart dose (heart mean), heart V5, heart V30, heart V50, heart V60, mean lung dose (lung mean), lung V5, lung V10, lung V20, mean esophagus dose (eso mean), and maximum esophagus dose (eso max). The categorical features were ethnicity, pre-treatment ECOG, 3 month post-RT ECOG, AJCC stage, T Stage, N Stage, radiation treatment modality (photon or proton), concurrent vs. sequential chemotherapy, chemotherapy agent used, tumor grade, and gender. All dosimetric indices were calculated with heterogeneity corrections, using the Analytical Anisotropic Algorithm (photons) or the Proton Convolution Superposition algorithm (Varian Medical Systems, Palo Alto, CA, USA). This set of dose parameters and clinical features, was thoroughly discussed and selected by two highly experienced board-certified thoracic radiation oncologists at our institution based on their expertise, best clinical practice, and current national treatment guidelines.

Univariate analysis

The independent effect of each feature in the prediction of RP was studied using optimally trained decision trees with only one internal node directly connected to the terminal nodes known as *decision stumps*. The missing values were imputed with the mean in the case of continuous features and with the mode in the case of categorical features. Additionally, the feature vector was extended with additionally binary indicator features with value *true* if there were missing values and *false* otherwise, thus increasing the length of the feature vector from 32 to 60 features. The *p*-value associated with the classification performance of the individual features conditioning the stumps was calculated *in-sample*, (*i.e.,* no validation). Moreover, the in-sample sensitivity and specificity of each feature, as well as Pearson's correlation coefficients among features were calculated.

Multivariate analysis

To assess the combined capacity of prediction of the features, we used Random Forest [19], an ensemble algorithm based on compounds of decision trees for classification and regression. The hyperparameter to build the Random Forest model was the *maximum tree-depth* which is an integer value that defines the maximum number of *branch nodes* in the longest path from a root to a terminal node. Once the maximum tree depth has been defined, the observations as well as the class labels are fed into the Random Forest algorithm, to produce an ensemble of decision trees.

Given the limited size of the cohort and the imbalance of the classes, resampling through 5-fold cross-validation was carried out using the entire dataset to perform hyperparameter tuning as well as to evaluate the predictive performance of the Random Forest ensembles (see Section S1 for more details). This analysis corresponds to the development and validation of a predictive model using resampling or analysis type 1b as specified in [20]. The stages of cross-validation in the workflow shown in Fig. 1, are performed separately, however using the same cohort of 203 patients. The hyperparameter optimization was carried out using 5-fold cross-validation 20 times per algorithm to maximize the balanced accuracy (BACC) on the validation group of each fold, thus delivering $N_{DS} = 20$ optimized instances of each model. After the hyperparameters had been tuned, the BACCs of all optimized model instances of all classifiers were computed performing 5-fold cross-validation a number of $N_{PT} = 100$ times over the same data with randomly chosen, but stratified, data partitions. Then the hyperparameters of the single best performing model instance were chosen to estimate the average predictive performance of the validation group of each fold through $N_{AV} = 100$ tests using 5-fold cross-validation, and the average performance indexes, specifically, BACC, the Receiver Operating Characteristic (ROC), the Area Under the ROC Curve (AUC), as well as its respective confidence intervals (95% confidence) and significance were calculated. Furthermore, the classification performance of additional decision tree-based algorithms used in medical applications, namely, CART [21] and RUSBoost [22] was assessed. Additionally, the long-existing methods logistic regression [23] and linear Support Vector Machines (SVM) [24] were implemented for further performance comparison. The hyperparameters of each classifier were tuned to maximize BACC using the Generalized Pattern Search (GPS) algorithm [25].

In this analysis, we trained 10 Random Forest ensembles of 500 trees each (5000 trees total) using the features automatically selected by the univariate analysis with significance level $\alpha = 0.05$ adjusted with Holm–Bonferroni correction for multiple comparison. Subsequently, the frequency of selection of the features at the root and the first level of the trees in the ensembles was calculated to determine their importance. The ranges of the cutoff thresholds used to split the nodes were selected to maximize the impurity gain over all splitting candidates. All the analyses were implemented in Matlab R2017b ® [26].

Results

Patient characteristics

Characteristics of the 203 patients in the study are provided in Tables 1 and 2. The patients' median age was 63 years (range 58–74). Patients received a median dose of 66.6 Gy, at 1.8 Gy per fraction (range 5.4–80 Gy at 1.8–10 Gy per fraction). Within our cohort, 36 patients (17.7%) developed grade $\geq$2 RP. Median follow-up time was 22.6 months (1–88 month range), with a median overall survival (OS) of 23.5 months, 1-year OS of 75.0%, 2-year OS of 49.0%, and 5-year OS of 12.0%. Overall, 88.0% of patients received concurrent chemotherapy, with a carboplatin-based doublet combination (51.2%) being the most common regimen, followed by cisplatin-based doublet (35.0%). Radiation was largely delivered with photon beam treatments (89.7%), with a smaller subset receiving proton beam treatment (10.3%).

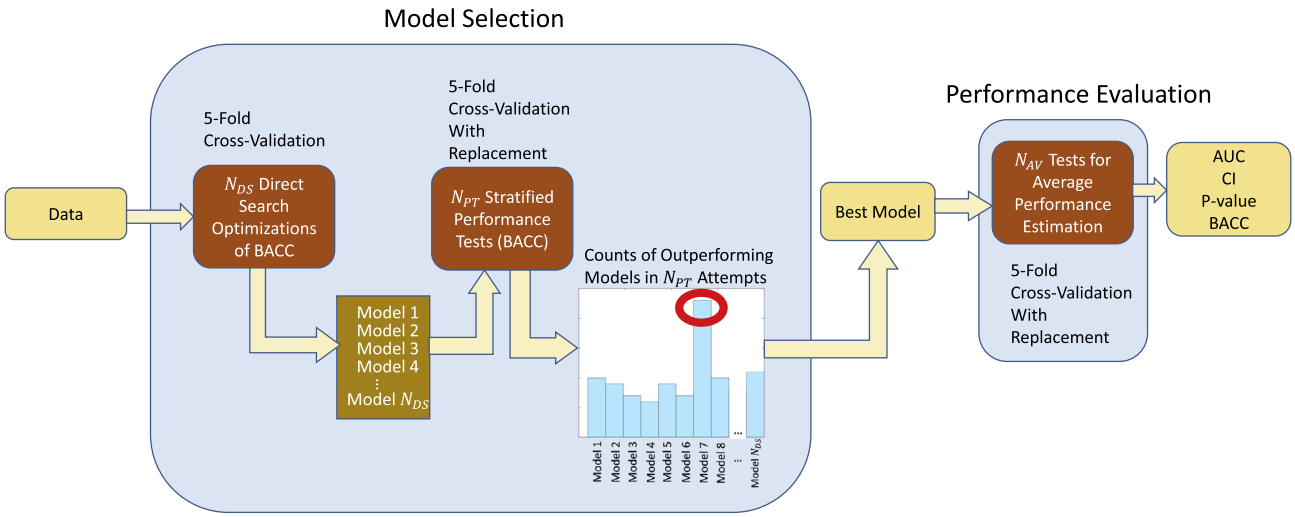


Fig. 1. Multivariate Analysis Workflow. The hyperparameters are tuned to maximize the BACC producing a total of $N_{DS} = 20$ model instances. After that, the optimized instances are subject to $N_{PT} = 100$ performance tests and the single best performing instance (best model) will be used to calculate indexes of average performance prediction, namely, AUC, 95% confidence interval, p-value and BACC using $N_{AV} = 100$ tests for average performance estimation.

Table 1
Summary of categorical patient characteristics. Description of clinical characteristics of the cohort with their respective categorization and percentages.

Categorical features	Classes	Nr. of patients	(%)
Sex	Male	91	44.8
	Female	112	55.2
Smoking history	Former	137	67.5
	Current	26	12.8
	Never	17	8.4
	Not Available	23	11.3
Ethnicity	White	138	68.0
	Black	47	23.2
	Asian	4	2.0
	Other	14	6.9
Pre treatment ECOG	0	78	38.4
Perform. Status	1	55	27.1
	2	14	6.9
	3	2	1.0
	4	2	1.0
	Not recorded	52	25.6
Stage grouping	IIB	1	0.5
	IIIA	120	59.1
	IIIB	82	40.4
Tumor stage	Tx	15	7.4
	T1	51	25.1
	T2	63	31.0
	T3	32	15.8
	T4	42	20.7
Nodal stage	Nx	7	3.4
	N0	9	4.4
	N1	13	6.4
	N2	126	62.1
	N3	48	23.6
Histology	Adenocarcinoma	203	100.0
Radiation modality	Linac	182	89.7
	Proton	21	10.3
Chemotherapy	Concurrent	178	87.7
	Sequential	20	9.9
	None	5	2.5
Chemotherapy agents	Carboplatin-based doublet	104	51.2
	Cisplatin-based doublet	71	35.0
	Platinum-based triplet	6	3.0
	Single agent	2	1.0
	Other	20	9.9

Table 2
Summary of numerical patient characteristics. Description of numerical characteristics of the cohort with their respective median and range.

Continuous features	Median	Range
Age (yr)	63	(58–74)
Pack-Year (current/former smokers)	40	(3–105)
BMI (kg/m^2)	26.00	(14.0–50.0)
Radiation dose delivered (Gy)	66.60	(27.0–80.0)
Dose per fraction (Gy)	1.80	(1.8–4.0)
Lung V5 (%)	44.50	(6.0–77.4)
Lung V20 (%)	28.60	(1.2–43.6)
Lung mean (Gy)	16.44	(1.3–25.0)

Univariate analysis

The univariate analysis showed that lung V20 > 27.4% ($p = 1.1 \times 10^{-4}$), lung mean >15.4 Gy ($p = 4.4 \times 10^{-4}$), lung V10 > 36.3% ($p = 0.001$), and lung V5 > 43.6% ($p = 0.002$) consistently predict the presence of RP. Additionally, among the list of all significant features identified, these features represent the most balanced scores of sensitivity and specificity (≥0.5) (Table 3). Moreover, these four features were also found to be strongly and significantly correlated ($r > 0.88$, $p < 0.01$) (Fig. 2). Additional significant features on univariate analysis, after applying Holm–Bonferroni correction for multiple comparison with $\alpha = 0.05$, were eso max, pack-year, nr. fractions, total dose and heart mean (Table 3). Heart V50, age at diagnosis, heart V60, heart V5, heart V30, primary tumor size and T-stage did not pass the significance test with the Holm–Bonferroni correction. Dosimetric parameters of the same organ at risk (e.g. heart, lung, and esophagus) were found to be strongly positively correlated to each other and also showed a weaker positive correlation with neighboring anatomic organs (e.g. lung-heart, lung-esophagus) (Fig. 2).

Multivariate analysis

A maximum tree depth of 2 was found to be the optimal hyperparameter for Random Forest. The optimal hyperparameters for the remaining classifiers tested for this classification task are shown in Table S1. The predictive power using AUCs, as well as the associated p-values and the optimal BACC for all classifiers are summarized in Table S3. The data illustrate that the tree ensembles Random Forest and RUSBoost display superior performance than single trees. Random Forest (AUC = 0.66, $p = 4.4 \times 10^{-4}$), RUSBoost (AUC = 0.63, $p = 0.03$) and CART (AUC = 0.63, $p = 5.3 \times 10^{-5}$) showed significant classification

Table 3
Univariate feature selection. Features with in-sample significance level $\alpha \leq 0.05$ corrected with Holm–Bonferroni method for multiple comparison.

In-sample					
Features	Threshold	*p*-Value	Significance level α (Holm–Bonferroni Correction)	Specificity	Sensitivity
Lung V20	**>27.4%**	**<0.003**	**0.0031**	**0.66**	**0.69**
Lung Mean	**>15.4 Gy**	**<0.003**	**0.0033**	**0.63**	**0.69**
Lung V10	**>36.3%**	**<0.003**	**0.0035**	**0.68**	**0.61**
Lung V5	**>43.6%**	**<0.003**	**0.0038**	**0.67**	**0.61**
Eso Max	>72.2 Gy	<0.003	0.0032	0.86	0.42
Pack-Year	>4.5	<0.003	0.0034	0.93	0.25
Nr. Fractions	>35.5	<0.003	0.0037	0.45	0.83
Total Dose	>63.9 Gy	<0.003	0.0040	0.44	0.83
Heart Mean	>13.4 Gy	0.0034	0.0042	0.74	0.50
Heart V50	>8.4%	0.0083	0.0043	0.77	0.44
Age at Diagnosis	>79.5 yr	0.0088	0.0045	0.94	0.19
Heart V60	>7.5%	0.0155	0.0048	0.86	0.31
Heart V5	>42.4%	0.0174	0.0050	0.75	0.44
Heart V30	>19.2%	0.0174	0.0053	0.76	0.33
Prim Tumor Size	>1.9 cm	0.0431	0.0059	0.27	0.88
T-Stage	{1,2,3}or{4}	0.0498	0.0063	0.80	0.36

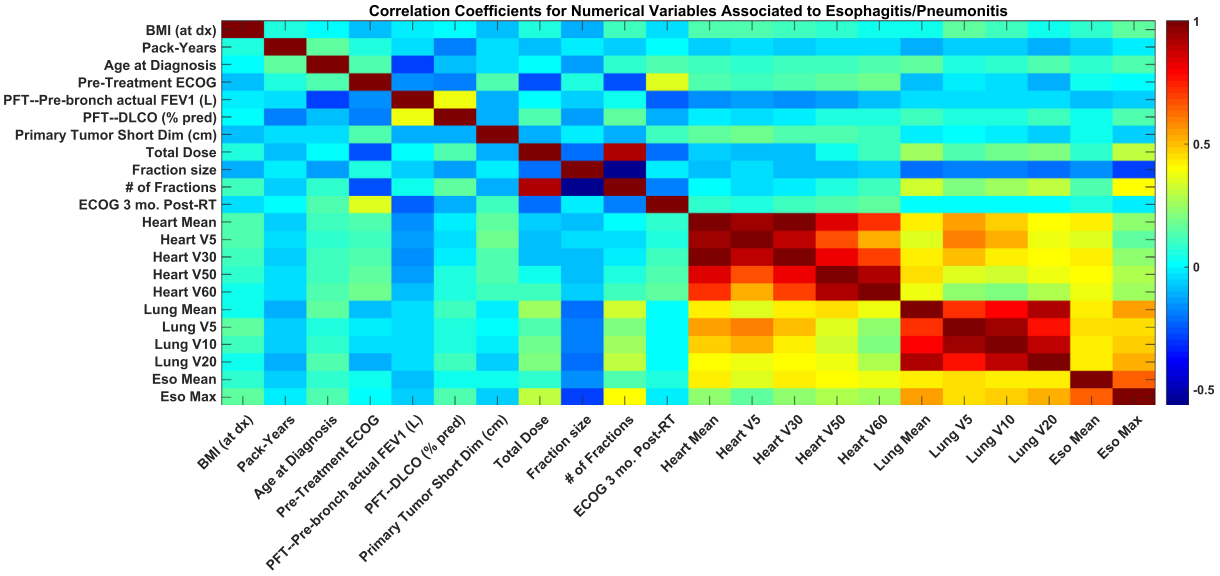


Fig. 2. Feature Correlation Heat Map. Heat map, illustrating the correlation between the continuous features under study.

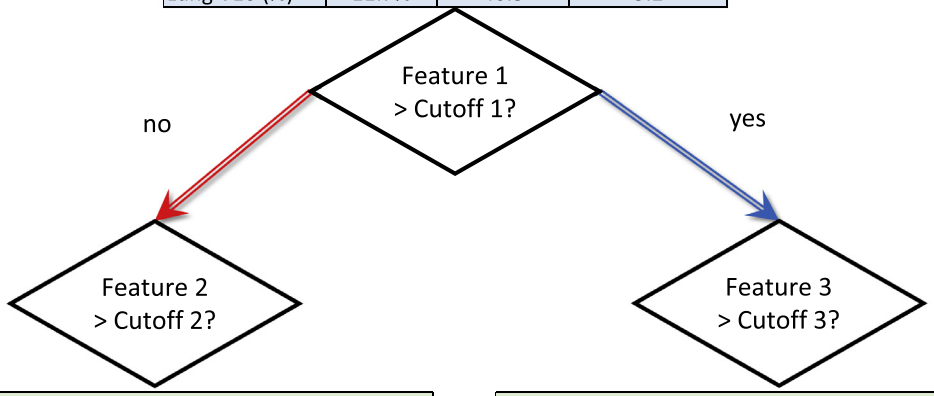
power and significant improvement (min improvement rate = 2.7%, $p = 9.5 \times 10^{-5}$) over the cross-validated performance of the best single predictors, Lung V20 (AUC = 0.61, $p = 5.5 \times 10^{-4}$) (Table S4). Logistic regression exhibits lower performance than Random Forest (AUC = 0.64, $p = 1.9 \times 10^{-3}$), and its BACC = 0.54 is the lowest among all the classifiers (Table S3). Linear SVM (AUC = 0.65, $p = 2.4 \times 10^{-3}$) also showed slightly lower performance than Random Forest; however, general SVMs do not allow for the analysis of interactions between features and complicates the estimation of cut-offs, which is the main purpose of this study.

The most frequently chosen features (selection frequency >10%) in the first two levels of the Random Forest ensemble along with the average optimal cutoff thresholds and their respective standard deviations used to split the nodes are presented in Fig. 3. Eso max was the most commonly selected root feature (20.5%) to separate RP events, followed by lung V20 (16.4%), lung mean (15.7%) and pack-year (14.9%). At the 2nd tree level, the most common left child feature (i.e. those falling below the root feature threshold) selected to separate out RP events was total dose (15.0%) followed by pack-year (14.9%) and Lung V5 (14.4%). Similarly, commonly selected right child features were also total dose (15.3%), lung V20 (14.1%) and eso max (13.0%).

Discussion

Owing to the significant morbidity and potential for mortality associated with RP, there have been prior efforts to establish predictive variables to guide clinicians in mitigating this toxicity [2,4,28,5,6,9–12,14,27]. These prior studies utilize a wide range of methodologies, including small single-institution retrospective multivariate analyses [13,15,28–31], systematic reviews [6], international meta-analysis [12], and Bayesian network analysis [14]. Although some of the criteria from those studies have been used in clinical practice, many of those studies are now less generalizable since they were conducted in a pre-IMRT (and pre-proton therapy) era. Additionally, these studies have reported conflicting data regarding the relative importance of factors associated with RP. Our work significantly contributes to the existing body of knowledge in several ways. First, we provide a modern evaluation of the RP question by analyzing a large, and well-curated patient

Tree Root			
Feature 1	**Frequency**	**Cutoff Mean**	**Cutoff Std Dev**
Eso Max (Gy)	20.5%	73.0	1.8
Lung V20 (%)	16.4%	31.1	4.1
Lung Mean (Gy)	15.7%	17.9	2.1
Pack-Year	14.9%	11.9	11.8
Lung V5 (%)	12.2%	50.4	7.1
Lung V10 (%)	11.7%	40.5	5.2



Left Child			
Feature 2	**Frequency**	**Cutoff Mean**	**Cutoff Std Dev**
Total Dose (Gy)	15.0%	69.5	8.8
Pack-Year	14.9%	19.0	17.3
Lung V5 (%)	14.4%	36.9	18.4
Lung Mean (Gy)	13.6%	13.4	7.0
Eso Max (Gy)	13.1%	69.3	11.2
Lung V20 (%)	12.3%	27.1	11.2
Lung V10 (%)	11.1%	28.0	17.0

Right Child			
Feature 3	**Frequency**	**Cutoff Mean**	**Cutoff Std Dev**
Total Dose (Gy)	15.3%	64.9	6.7
Lung V20 (%)	14.1%	32.2	4.7
Eso Max (Gy)	13.0%	72.3	4.9
Pack-Year	12.5%	26.3	20.0
Luna Mean (Gy)	11.7%	18.2	2.7
Luna V5 (%)	11.4%	53.2	8.8
Lung V10 (%)	11.3%	42.3	6.4
Nr. Fractions	10.7%	34.8	3.8

Fig. 3. Feature Frequency using Random Forest. Most frequently chosen predictors (>10%) with their respective frequency and the means and standard deviations of the cutoffs.

cohort treated in a relatively uniform manner in the IMRT/proton era. Second, we highlight the power of a machine learning methodology to simultaneously assess multiple features to uncover potentially important factors for RP and provide guidance regarding comparative importance for optimization.

Prior studies have established the importance of lung dosimetric parameters in RP development, including lung mean [6,9], lung V20 [5,12,32,33], and agents for concurrent chemotherapy [12]. Our patient cohort included a largely uniform set of patients, of whom the vast majority are treated with concurrent chemoradiation, which allows us to identify RP predictors within a population of patients receiving modern treatment regimens. In this context, our findings are congruent with the existing literature, and we confirm the importance of lung V20 (Fig. 3), as it was found to be the second most commonly selected tree root with the Random Forest method (16.4%), as well as lung mean (15.7%). Since BMI could affect total lung volume, there could be bias on the interpretation of lung V20; however, in our cohort, 96% of the patients had BMI $\leq$ 37, which reassures the accuracy of our results.

Additionally, we extend the existing literature by highlighting the potential importance of eso max, as it is the most common tree root Random Forest feature (20.5%), which has not been previously well established as a dosimetric feature to consider in the existing literature. Eso max, although a significant predictor, it is not causative of RP. It rather reflects the distribution and volume of pulmonary disease surrounding it, which eventually will be reflected in the pulmonary dosimetric outcomes. This is evidenced in Fig. 2, where esophagus dosimetric indices are positively correlated with lung dosimetric indices.

Increasing age was found in our univariate analysis to predict for RP risk, which is in keeping with prior reports [12,34]. Our univariate feature selection confirms greater age at diagnosis as a significant feature with increasing sensitivity for RP at higher age cutoffs (Table 3).

Random Forest was chosen to carry out the main analyses presented in this paper given its high AUC with respect to the other tested classifiers and its ability of providing variable importance and cutoff thresholds that could be clinically interpreted by physicians. Furthermore, Random Forest is intrinsically robust to *multicollinearity* effects given the randomness induced in the selection of the subset of features for each tree in the ensemble. This comes at the expense of the *interpretability* [35,36] that less accurate algorithms such as CART and MediBoost [18,37] offer, which is of interest in areas such as medicine where making informed decisions require clear understanding of the model. In addition, Random Forest also has the important ability to identify cutoff thresholds in continuous variables such as dosimetric criteria. For example, the algorithm most often finds average cutoff thresholds such as 73 Gy (SD = 1.8) for eso max, 31.1% (SD = 4.1) for lung V20, 17.9 Gy (SD = 2.1) for lung mean, 11.9 (SD = 11.8) for pack-year, 50.4% (SD = 7.1) for lung V5 and 40.5% (SD = 5.2) for lung V10 (Fig. 3). This information provides the clinician valuable guidance

to accept dosimetric cutoffs in plan evaluation in order to possibly mitigate RP development.

In an attempt to minimize improper utilization of the available algorithms, we tested and validated several multivariate analysis techniques (RUSBoost, Random Forest, CART, logistic regression and linear SVM) to confirm and verify our results [17] (Section S3). In order to improve exploration of the hyperparameter space with respect to a grid search, the hyperparameters of each classifier were tuned to maximize BACC using the GPS algorithm [25]. Through rigorous use of random subsampling, decision tree modeling and multiple permutations, we demonstrated the robustness of the Random Forest tool and used it to identify RP predictors. Furthermore, this study has followed the guidelines provided in [38] for predicting radiation therapy outcomes using machine learning from a clinician's perspective.

Although general deficiencies have been identified in applying models between populations with poor outcomes even when using well-established models [8,38,39], there has been success with cross-validated machine learning methods generalizing to independent data from the Radiation Therapy Oncology Group (RTOG) [40]. Therefore, given the limited size of the cohort and the lack of balance of the classes, 5-fold cross-validation was implemented to perform hyperparameter tuning and to evaluate the predictive performance of the Random Forest ensembles.

The limitations of this study lie with the modestly sized patient cohort from a single institution, which is reflected in the low sensitivity and specificity of the predictive features. Also notice that even though Random Forest could be used to predict multiple outcomes, our model predicts the presence or absence of RP ≥ 2 but not the grade, because the diversity on the RP grades present in the current cohort is rather poor to be able to classify the grade of RP with sufficient statistical power. Therefore, increased patient numbers would lend itself to increased statistical power which would complement the proven ability to identify significant predictors of our machine learning technique.

Future analysis will be aimed at using toxicity predictors to predict treatment response to radiotherapy for LA-NSCLC using both clinical data and PET/CT radiomic features in larger cohorts, as well as the analysis of detailed combinations of features using tree ensembles. It is worth mentioning that patients analyzed in the current report represent a very characterized and highly detail annotated dataset of homogenously treated patients. As such, we have invested valuable effort in carrying out the rigorous analysis presented here in order to minimize the risk of statistical bias in our results, therefore obtaining reliable and conclusive outcomes. Our long-term goal is to produce a tool to assist oncologists and dosimetrists to personalize radiotherapy treatment by minimizing radiation-induced toxicities for LA-NSCLC.

The strength of this study, one of the first of its type applied to RP, is in the application of a machine learning analysis based on accurate ensembles of decision trees (Random Forest), to identify factors predicting both for and against the development of symptomatic RP by simultaneously analyzing multiple heterogeneous patient- and tumor-related variables and dosimetric factors. We highlight and reinforce the importance of lung V20, lung mean, eso max, pack-year, lung V5 and total dose, and we also identify the relative importance and potential cutoffs of other contributing dosimetric and clinical factors. Confirming previous results obtained through different methods is a crucial step to reliably extend our machine learning implementations, which are rigorously supported by mathematical principles, to more complex analyses regarding the prediction of radiation pneumonitis. These metrics can serve to provide a guide for clinicians in the evaluation of treatment plans for LA-NSCLC patients receiving concurrent chemoradiation and highlight the power of the machine learning techniques for future studies.

Acknowledgements

This work was partially supported by the Abramson Cancer Center of the University of Pennsylvania through award granted by the Emerson Collective.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.radonc.2019.01.003.

References

[1] Simone 2nd CB. Thoracic Radiation Normal Tissue Injury. Semin Radiat Oncol 2017;27:370–7. https://doi.org/10.1016/j.semradonc.2017.04.009.
[2] Barriger RB, Fakiris AJ, Hanna N, Yu M, Mantravadi P, McGarry RC. Dose-volume analysis of radiation pneumonitis in non-small-cell lung cancer patients treated with concurrent cisplatinum and etoposide with or without consolidation docetaxel. Int J Radiat Oncol Biol Phys 2010;78:1381–6. https://doi.org/10.1016/j.ijrobp.2009.09.030.
[3] Graves PR, Siddiqui F, Anscher MS, Movsas B. Radiation pulmonary toxicity: from mechanisms to management. Semin Radiat Oncol 2010;20:201–7. https://doi.org/10.1016/j.semradonc.2010.01.010.
[4] Hernando ML, Marks LB, Bentel GC, Zhou SM, Hollis D, Das SK, et al. Radiation-induced pulmonary toxicity: a dose-volume histogram analysis in 201 patients with lung cancer. Int J Radiat Oncol Biol Phys 2001;51:650–9. https://doi.org/10.1016/S0360-3016(01)01685-6.
[5] Ramella S, Trodella L, Mineo TC, Pompeo E, Stimato G, Gaudino D, et al. Adding ipsilateral V20 and V30 to conventional dosimetric constraints predicts radiation pneumonitis in stage IIIA-B NSCLC treated with combined-modality therapy. Int J Radiat Oncol Biol Phys 2010;76:110–5. https://doi.org/10.1016/j.ijrobp.2009.01.036.
[6] Rodrigues G, Lock M, D'Souza D, Yu E, Van Dyk J. Prediction of radiation pneumonitis by dose-volume histogram parameters in lung cancer – A systematic review. Radiother Oncol 2004;71:127–38. https://doi.org/10.1016/j.radonc.2004.02.015.
[7] McDonald S, Rubin P, Phillips TL, Marks LB. Injury to the lung from cancer therapy: clinical syndromes, measurable endpoints, and potential scoring systems. Int J Radiat Oncol Biol Phys 1995;31:1187–203. https://doi.org/10.1016/0360-3016(94)00429-O.
[8] Bradley JD, Hope A, El Naqa I, Apte A, Lindsay PE, Bosch W, et al. A nomogram to predict radiation pneumonitis, derived from a combined analysis of RTOG 9311 and institutional data. Int J Radiat Oncol Biol Phys 2007;69:985–92. https://doi.org/10.1016/j.ijrobp.2007.04.077.
[9] Kocak Z, Borst GR, Zeng J, Zhou S, Hollis DR, Zhang J, et al. Prospective assessment of dosimetric/physiologic-based models for predicting radiation pneumonitis. Int J Radiat Oncol Biol Phys 2007;67:178–86. https://doi.org/10.1016/j.ijrobp.2006.09.031.
[10] Marks LB, Bentzen SM, Deasy JO, Kong F-M, (Spring), Bradley JD, Vogelius IS, et al. Radiation dose-volume effects in the lung. Int J Radiat Oncol Biol Phys 2015;76:S20–7. https://doi.org/10.1016/j.ijrobp.2009.06.091.
[11] Marks LB, Munley MT, Bentel GC, Zhou SM, Hollis D, Scarfone C, et al. Physical and biological predictors of changes in whole-lung function following thoracic irradiation. Int J Radiat Oncol Biol Phys 1997;39:563–70. https://doi.org/10.1016/S0360-3016(97)00343-X.
[12] Palma DA, Senan S, Tsujino K, Barriger RB, Rengan R, Moreno M, et al. Predicting radiation pneumonitis after chemoradiation therapy for lung cancer: an international individual patient data meta-analysis. Int J Radiat Oncol Biol Phys 2013;85:444–50. https://doi.org/10.1016/j.ijrobp.2012.04.043.
[13] Kim M, Lee J, Ha B, Lee R, Lee K-J, Suh HS. Factors predicting radiation pneumonitis in locally advanced non-small cell lung cancer. Radiat Oncol J 2011;29:181–90. https://doi.org/10.3857/roj.2011.29.3.181.
[14] Luo Y, El Naqa I, McShan DL, Ray D, Lohse I, Matuszak MM, et al. Unraveling biophysical interactions of radiation pneumonitis in non-small-cell lung cancer via Bayesian network analysis. Radiother Oncol 2017;123:85–92. https://doi.org/10.1016/j.radonc.2017.02.004.
[15] Wang D, Shi J, Liang S, Lu S, Qi X, Wang Q, et al. Dose-volume histogram parameters for predicting radiation pneumonitis using receiver operating characteristic curve. Clin Transl Oncol 2013;15:364–9. https://doi.org/10.1007/s12094-012-0931-y.
[16] Darcy AM, Louie AK, Roberts LW. Machine learning and the profession of medicine. JAMA – J Am Med Assoc 2016;315:551–2. https://doi.org/10.1001/jama.2015.18421.
[17] Valdes G, Solberg TD, Heskel M, Ungar L, Simone 2nd CB. Using machine learning to predict radiation pneumonitis in patients with stage I non-small cell lung cancer treated with stereotactic body radiation therapy. Phys Med Biol 2016;61:6105–20. https://doi.org/10.1088/0031-9155/61/16/6105.
[18] Valdes G, Luna JM, Eaton E, Simone 2nd CB, Ungar LH, Solberg TD. MediBoost: a patient stratification tool for interpretable decision making in the era of precision medicine. Sci Rep 2016;6:1–8. https://doi.org/10.1038/srep37854.

[19] Svetnik V, Liaw A, Tong C, Christopher Culberson J, Sheridan RP, Feuston BP. Random forest: a classification and regression tool for compound classification and QSAR modeling. J Chem Inf Comput Sci 2003;43:1947–58. https://doi.org/10.1021/ci034160g.
[20] Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. BMC Med 2015;13:1–10. https://doi.org/10.1186/s12916-014-0241-z.
[21] Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees, vol. 19, 1984. doi:10.1371/journal.pone.0015807.
[22] Seiffert C, Khoshgoftaar TM, Van Hulse J, Napolitano A. RUSBoost: A hybrid approach to alleviating class imbalance. IEEE Trans Syst Man Cybern Part ASystems Humans 2010;40:185–97. https://doi.org/10.1109/TSMCA.2009.2029559.
[23] Cox DR. The regression analysis of binary sequences. J R Stat Soc Ser B 1958;20:215–42. https://doi.org/10.1007/BF03180993.
[24] Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20:273–97. https://doi.org/10.1023/A:1022627411411.
[25] Kolda TG, Lewis RM, Torczon V. Optimization by direct search: new perspectives on some classical and modern methods. SIAM Rev 2003;45:385–482. https://doi.org/10.1137/S003614450242889.
[26] MathWorks. Statistics and machine learning toolbox release notes. MatLab 2015:118.
[27] Rancati T, Ceresoli GL, Gagliardi G, Schipani S, Cattaneo GM. Factors predicting radiation pneumonitis in lung cancer patients: a retrospective study. Radiother Oncol 2003;67:275–83. https://doi.org/10.1016/S0167-8140(03)00119-1.
[28] Park YH, Kim JS. Predictors of radiation pneumonitis and pulmonary function changes after concurrent chemoradiotherapy of non-small cell lung cancer. Radiat Oncol J 2013;31:34–40. https://doi.org/10.3857/roj.2013.31.1.34 [doi].
[29] Dang J, Li G, Ma L, Diao R, Zang S, Han C, et al. Predictors of grade $\geq$ 2 and grade $\geq$ 3 radiation pneumonitis in patients with locally advanced non-small cell lung cancer treated with three-dimensional conformal radiotherapy. Acta Oncol (Madr) 2013;52:1175–80. https://doi.org/10.3109/0284186X.2012.747696.
[30] Dang J, Li G, Zang S, Zhang S, Yao L. Comparison of risk and predictors for early radiation pneumonitis in patients with locally advanced non-small cell lung cancer treated with radiotherapy with or without surgery. Lung Cancer 2014;86:329–33. https://doi.org/10.1016/j.lungcan.2014.10.005.
[31] Dang J, Li G, Zang S, Zhang S, Yao L. Risk and predictors for early radiation pneumonitis in patients with stage III non-small cell lung cancer treated with concurrent or sequential chemoradiotherapy. Radiat Oncol 2014;9.. https://doi.org/10.1186/1748-717X-9-172.
[32] Graham MV, Purdy JA, Emami B, Harms W, Bosch W, Lockett MA, et al. Clinical dose-volume histogram analysis for pneumonitis after 3D treatment for non-small cell lung cancer (NSCLC). Int J Radiat Oncol Biol Phys 1999;45:323–9. https://doi.org/10.1016/S0360-3016(99)00183-2.
[33] Bledsoe TJ, Nath SK, Decker RH. Radiation pneumonitis. Clin Chest Med 2017;38:201–8. https://doi.org/10.1016/j.ccm.2016.12.004.
[34] Parashar B, Edwards A, Mehta R, Pasmantier M, Wernicke AG, Sabbas A, et al. Chemotherapy significantly increases the risk of radiation pneumonitis in radiation therapy of advanced lung cancer. Am J Clin Oncol 2010;34:1. https://doi.org/10.1097/COC.0b013e3181d6b40f.
[35] Balu A, Nguyen T V., Kokate A, Hegde C, Sarkar S. A forward-backward approach for visualizing information flow in deep networks; 2017.
[36] Lipton ZC. The Mythos of Model Interpretability; 2016.
[37] Luna JM, Eaton E, Ungar LH, Diffenderfer E, Jensen ST, Gennatas ED, et al. Tree-Structured Boosting: Connections Between Gradient Boosted Stumps and Full Decision Trees. Symp. Interpret. Mach. Learn. NIPS 2017, Long Beach, CA: 2017, p. 1–8.
[38] Kang J, Schwartz R, Flickinger J, Beriwal S. Machine learning approaches for predicting radiation therapy outcomes: a clinician's perspective. Int J Radiat Oncol Biol Phys 2015;93:1127–35. https://doi.org/10.1016/j.ijrobp.2015.07.2286.
[39] Bentzen SM, Constine LS, Deasy JO, Eisbruch A, Jackson A, Marks LB, et al. Quantitative analyses of normal tissue effects in the clinic (QUANTEC): an introduction to the scientific issues. Int J Radiat Oncol Biol Phys 2010;76:3–9. https://doi.org/10.1016/j.ijrobp.2009.09.040.
[40] El Naqa I, Bradley JD, Lindsay PE, Hope AJ, Deasy JO. Predicting radiotherapy outcomes using statistical learning techniques. Phys Med Biol 2009;54. https://doi.org/10.1088/0031-9155/54/18/S02.