

## Multi-region Multiomic Random Forest Toxicity Modeling of Radiation Pneumonitis

Saurabh S. Nair, MSc,<sup>1</sup> Ramon M. Salazar, PhD,<sup>1</sup> Alexandra O. Leone, MBS,<sup>1</sup> Ting Xu, PhD,<sup>2</sup> Zhongxing Liao, MD,<sup>2</sup> Laurence E. Court, PhD,<sup>1</sup> Joshua S. Niedzielski, PhD<sup>1</sup>, Departments of <sup>1</sup>Radiation Physics and <sup>2</sup>Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

### Abstract

**Purpose** – Radiation pneumonitis (RP) is a major dose-limiting toxicity resulting from non-small-cell lung cancer (NSCLC) radiotherapy. Multiomic features (radiomics and dosiomics) could provide additional predictive information as compared to traditionally used clinical and dose-volume histogram (DVH) parameters. We aimed to investigate the utility of multiomic features to improve RP toxicity models.

**Methods** – Out of 329 NSCLC patients considered, 85 patients (25.84%) were found to have toxicity  $\geq$  grade 2 RP per CTCAE v5.0. A total of 422 radiomic and dosiomic features were extracted. Four toxicity prediction model types were created using clinical factors together with respective features from one of the following groups: (a) DVH (base model), (b) whole lung radiomics and dosiomics (WL-RD), (c) multi-region radiomics and dosiomics (MR - RD) and (d) multi-region DVH, radiomics and dosiomics (MR-DVHRD). Toxicity models were created using a random forest classifier with a Monte Carlo cross-validation approach of 100 iterations, and a training/test split of 80%/20%, respectively. Model predictive performance was evaluated by area under the receiver operating characteristic curve (AUC) and area under the precision-recall curve (AUPRC).

**Results** –The AUC and AUPRC values (mean  $\pm$  standard deviation) for the 4 model types were  $0.81 \pm 0.04 / 0.70 \pm 0.06$  (base model),  $0.82 \pm 0.05 / 0.73 \pm 0.08$  (WL-RD,  $p < 0.05$ ),  $0.83 \pm 0.06 / 0.75 \pm 0.08$  (MR-RD,  $p < 0.05$ ), and  $0.82 \pm 0.05 / 0.72 \pm 0.08$  (MR-DVHRD,  $p < 0.05$ ), wherein a paired test compared the performance metrics of omic models with the base model built on each iteration of cross validation.

**Conclusions** – All multiomic model types outperformed the base DVH model. MR-RD model had the best performance among all model types.

## Introduction

Lung cancer is the leading cause of cancer-related mortality worldwide.<sup>1</sup> Non-small-cell lung cancer (NSCLC) accounts for 85% of all lung cancer cases.<sup>2</sup> Radiation therapy (RT) plays a crucial role in the management of unresectable and locally advanced NSCLC, as RT is the primary treatment modality for inoperable NSCLC. Radiation pneumonitis (RP) remains a dose-limiting toxicity that degrades patient quality-of-life, with symptomatic RP occurring in about 15–40% of lung RT patients.<sup>3</sup> RP can lead to irreversible fibrosis and dyspnea, with possible mortality in severe cases.<sup>3</sup> Non-effective therapeutics and persistent RP remains one of the top dose-limiting toxicities that influences RT planning.<sup>4,5</sup> Various clinical factors like age, gender, tumor location and smoking status have been studied in order to predict RP<sup>6</sup>, but the findings have been inconsistent due to sample size<sup>7</sup> and the heterogeneity of patient populations.<sup>8</sup> Metrics derived from dose-volume histograms (DVHs) have also been used to predict the incidence of RP with mean lung dose (MLD) and volume of lung receiving  $\geq 20\text{Gy}$  (V20) being the most common metrics related to RP.<sup>9–12</sup> However, clinical and DVH feature-based toxicity models have been shown to have suboptimal predictive performance.<sup>13</sup>

Recently, machine learning has emerged in the field of radiation oncology for radiation-induced toxicity outcome prediction modeling (OPM). A random forest machine learning classification approach has been shown to outperform other algorithms for outcome predictions with clinical data among some of the widely used machine learning classification approaches.<sup>14</sup> A process termed radiomics involves extracting quantitative features from medical images, such as computed tomography (CT) scans, and with the aid of a machine learning approach, these features are used to predict therapeutic responses<sup>15–17</sup>. Recently, dosiomics, which is similar to radiomic analyses, but instead uses 3D dose distributions as the feature space, has been introduced into OPM analyses. These omic features are derived from the extracted radiotherapy

dose distributions of patients to predict therapeutic responses. It is possible that dosiomic variables can better predict RP than DVH metrics since they retain the 3D information of the radiation dose distribution.<sup>18,19</sup> Recent studies have also demonstrated the effectiveness of incorporating additional radiomic features derived from a very small region surrounding the tumor into OPM analyses, resulting in improved prediction accuracy<sup>20–22</sup>. Combining multi-region radiomics and dosiomics features into a multiomic OPM approach, based on shape, statistical, and textural patterns from CT Imaging and 3D dose distributions, could provide additional data that improves outcome prediction models.

The aims of this work were to investigate the utility of radiomic and dosiomic features over traditional DVH features, as well as to examine the impact of using multiomic features from a subregion of normal lung to improve RP OPMs.

## **Materials and Methods**

### *Patient Data*

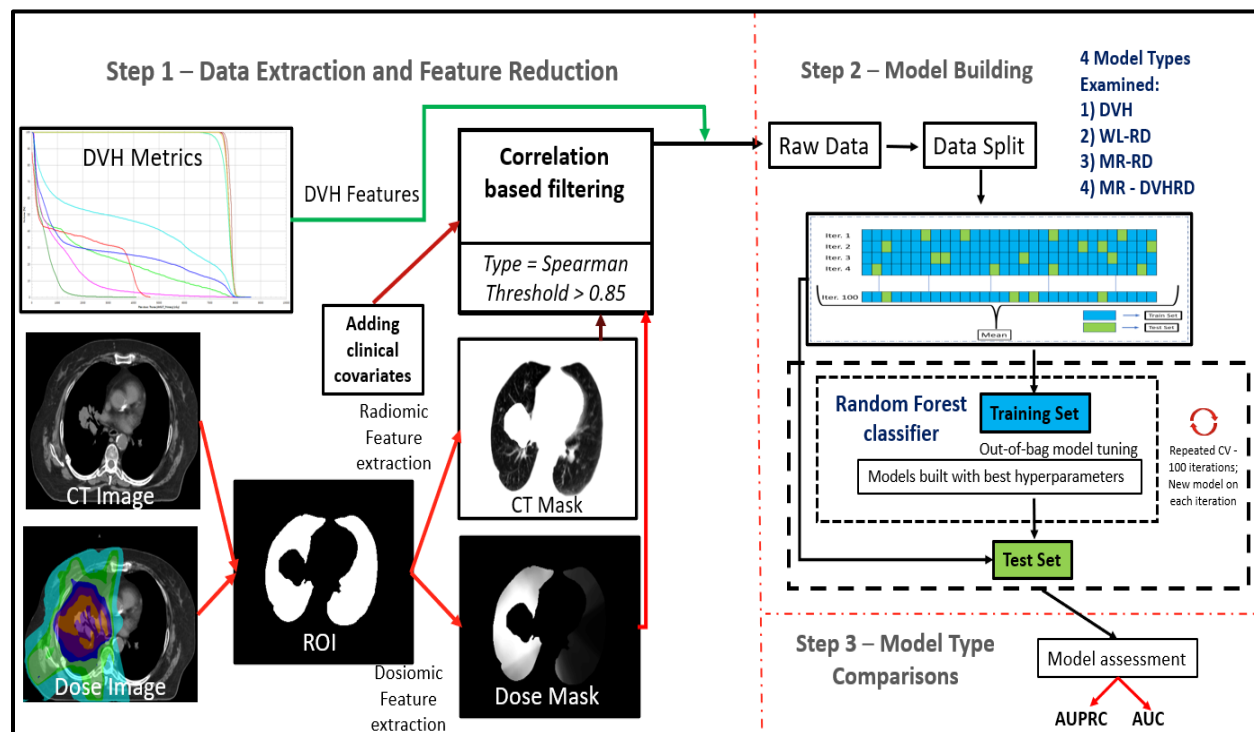
We included 329 NSCLC patients treated at our institution between October 2006 to March 2019 with either intensity modulated radiation therapy/volume-modulated arc therapy (IMRT/VMAT, n=190), passive-scatter proton therapy (PSPT, n=92), or intensity-modulated proton therapy (IMPT, n=47). The CT and dose grid spacing were 2.5 mm and 3×3×3 mm, respectively. All patients received fractionated doses between 1.8–3.0 Gy per fraction for total doses between 60–74 Gy. The study exclusion criteria were: (a) presence of acute lung infection, (b) prior thoracic surgery, (c) patients having <6-month follow-up period unless toxicity noted, (d) patients with a previous history of thoracic radiation therapy. All patients were treated with either induction, concurrent, or adjuvant chemotherapy in addition to RT.

### *Toxicity Evaluation*

Radiation Pneumonitis was graded from 0 to 5 according to the Common Terminology Criteria for Adverse Events (CTCAE v5.0).<sup>23</sup> The approximate time for RP to develop in a patient after radiation therapy is between 1–6 months.<sup>24,25</sup> A patient with an RP grade  $\geq 2$  was considered symptomatic for RP and used as the binary endpoint for this study.

### *Data Analysis Workflow*

The data analysis workflow is illustrated in **Fig 1**. Radiomic and dosiomic features were calculated using CT and dose images, respectively, focusing on two regions of interest (ROI). First was the total lung minus the gross tumor volume (GTV). To study the utility of incorporating additional information about tumor microenvironment in enhancing efficiency of RP OPMs, multomic features were also extracted from a 20mm ring structure created from normal lung surrounding the GTV (termed, “multi-region multiomic features”). Once all features were extracted, we conducted a Spearman correlation analysis to remove heavily correlated features with a correlation coefficient threshold of  $>0.85$  for redundant features<sup>18</sup>. Next, we used a random forest classifier to create predictive toxicity models of 4 different types using the selected feature categories: (a) base DVH model, (b) whole lung based radiomic & dosiomic (WL-RD) model, (c) whole lung and ring (multi-region) based radiomic & dosiomic (MR-RD) model, and (d) multi-region DVH + radiomic + dosiomic (MR-DVHRD) model. The performance metrics between the base and multitomic models were then evaluated. Further details on each component of the study analysis are presented in the following subsections.



**Figure 1** – Overall Study Workflow. DVH, radiomic and dosiomic features are extracted, followed by feature reductions for omic features in the first step. 4 OPM types are built and performance metrics between the DVH and omic model types are compared in the second step. *Abbreviations: DVH, dose volume histograms; ROI, region of interest; AUC, area under the curve; AUPRC, area under precision recall curve; CV, cross validation; WL, whole lung; RD, radiomic & dosiomic; MR, multi-region.*

### Feature Extraction and Selection

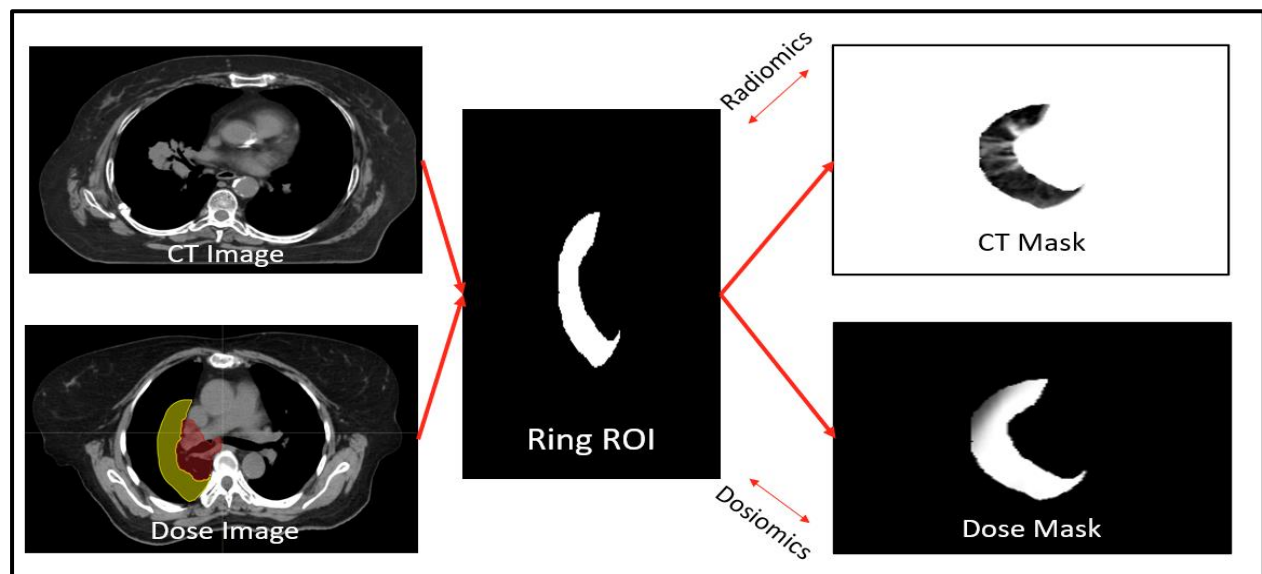
A total of 422 data features were used for this study (13 clinical, 18 dosimetric, 182 dosiomic and 210 radiomic). The clinical features were gender, age, smoking status, cancer stage, histology, type of chemotherapy, tumor position, gross tumor volume (GTV), and performance status. The description of the clinical features is provided in the **Supplemental Table S1**. The

dosimetric, radiomic, and dosiomic features were first extracted from the whole-lung ROI. The dosimetric covariates include mean lung dose, max lung dose, the relative volume receiving at least a given dose (i.e. the value obtained for VxGy[%] refers to the percent volume of a structure receiving at least xGy) from rV5 to rV60 (in increments of 5 Gy). These metrics were extracted from the Raystation treatment planning system (RaySearch Laboratories; Stockholm, Sweden). In order to extract features, the DICOM files were first converted into nearly raw raster data (nrrd) using the 3D Slicer software.<sup>26</sup> Radiomic and dosiomic features were then extracted based on the Imaging Biomarker Standardization Initiative (IBSI)<sup>27</sup> using the pretreatment CT and dose image using Pyradiomics (v3.0.1) software. The extracted features included: 14 shape, 18 first-order, and 73 second-order features respectively for both ROI's.<sup>28</sup> All multiomic features are described in the **Supplemental Materials**. For feature selection, a previously reported filter-based correlation approach was implemented.<sup>18</sup> First, Spearman correlation coefficients (CC) were calculated for all features, with features showing a CC of >0.85 being excluded. The reduction criteria were that if two variables had a higher correlation coefficient than the cutoff, the variable with the highest mean absolute correlation was removed.

### *Multi-region multiomics*

We assume that the greatest effect for normal tissue toxicities are observed closer to the tumor volume and therefore hypothesize that using a sub-region near the GTV could provide more informative multiomic measures. Unique characteristics captured by these features based on tumor-normal tissue interactions and spatial information could be imperative in enhancing the performance of OPMs. Thus, in this study, in addition to whole-lung multiomic features, we created a 20-mm ring subregion ROI of the normal lung around the GTV and extracted multiomic features to examine the added potential of the features close to the tumor in predicting RP. **Figure**

2 shows the sub-region multiomic feature delineation. Once the ring ROI multiomic features are extracted, the data analysis process is the same as depicted in **Fig 1**.



**Figure 2**– Sub-region (ring) multiomic feature extraction. Radiomic and dosiomic features are extracted for a 20-mm ring around the GTV overlapping the whole-lung ROI, followed by feature reductions and analysis as described in **Fig 1**.

### *Model Building*

The process of constructing the model and performance metric evaluation is illustrated in **Fig 1 and Fig 2**. For this study, we created 4 toxicity prediction model types using the clinical factors, along with one of the following feature sets: (a) base DVH, (b) whole-lung based radiomic & dosiomic (WL-RD), (c) whole-lung and ring (multi-region) based radiomic & dosiomic (MR-RD), and (d) multi-region DVH + radiomic + dosiomic (MR-DVHRD). The random forest (RF) classification model has been shown to yield the best discriminative performance in radiotherapy outcomes and toxicity predictions and hence has been used in this study.<sup>14</sup> Hence, to build our toxicity models, we used a random forest classifier. The hyperparameters for the random forest

modeling like *mtry* (controls how much randomness is added to the decision tree creation process), *maxnodes* (maximum number of terminal nodes allowed in a tree) and *ntrees* (number of decision trees that are combined to create the final prediction), were set depending on the number of covariates in each of the 4 models.<sup>29</sup>

The toxicity models were created using a repeated cross-validation approach with 100 iterations. This type of cross-validation is a common method used to approximate the generalizability of the modeling process.<sup>30–32</sup> Each iteration of cross-validation creates a unique model, with the overall process creating 100 unique models for a given model type. This results in a paired comparison of individual models, where statistical significance of performance metrics can be derived from the distribution of models for each model type. For each iteration, the cohort was randomly split into a training set (80%) and test set (20%). Correlation-based feature reductions were performed for all features on the training set to find out optimal features. The remaining features were then used for RF modeling for the specific model type.

### *Model Evaluation*

After selecting the optimal hyperparameters using the training set, the performance of the model was tested on the test set. The following metrics were used to assess each model's performance on the held-out test set: accuracy, area under the receiver operating characteristic curve (AUC), area under the precision-recall curve (AUPRC), precision, and recall. Standard deviation was also computed for performance metric distributions from the repeated cross-validation model building on both the training and test sets. All omic model types were compared to the base model to test the significance of each performance metrics distribution from repeated cross-validation using the paired Wilcoxon sign-rank test, with a p-value of <0.05 being considered as statistically significant. All statistical analyses were performed using R Studio



(4.2.1) (Posit, PBC; Boston, Massachusetts). The open-source R package “caret” with randomForest library was used to build all machine learning outcome prediction models.

## Results

### *Patient Characteristics and Univariate Analysis*

A total of 329 patients were used for this study. The patients who developed a grade 2 or higher RP were 85/329 (25.84% patients positive for RP). The patient characteristics are summarized in **Supplemental Table S1**. The results for the Fisher’s exact test and Mann Whitney U-test for categorial, continuous, and DVH variables are provided in **Supplemental Table S1**.

**Table 1.** Results of test set evaluation for different model types reported as mean of each metric with one standard deviation as the error. Model type comparison with the base model type (DVH metric) for WL-RD, MR-RD, and MR-DVHRD models is conducted with a one-sided Wilcoxon signed-rank test. Models with multiomic features had a better AUC and AUCPR as compared to the base model with MR-RD model type outperforming all the other model types.

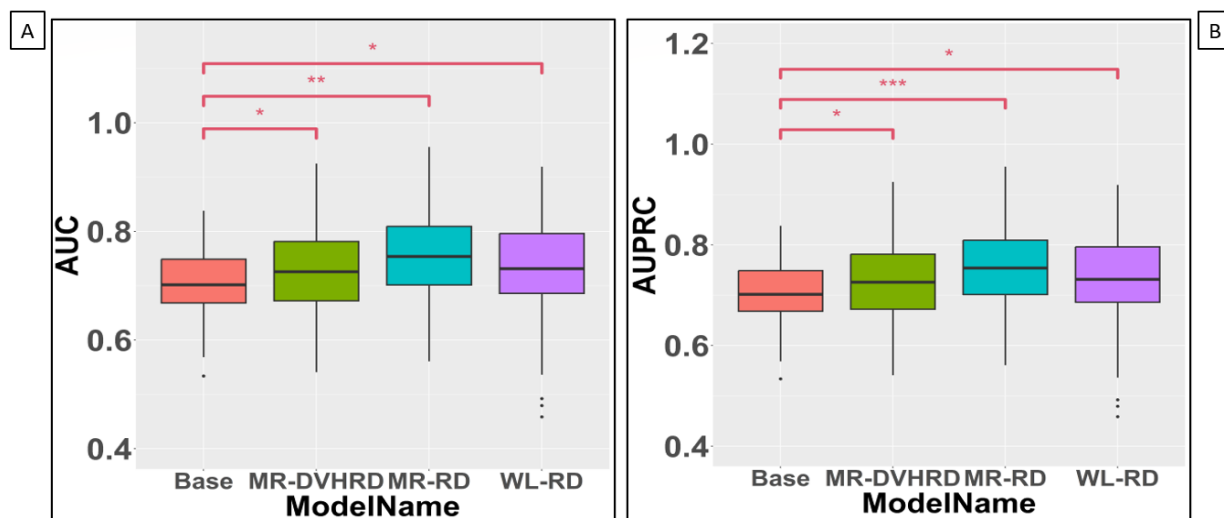
<b>Model Name</b>	<b>Accuracy</b>	<b>AUC</b>	<b>AUPRC</b>	<b>Precision</b>	<b>Recall</b>
<b>Base (DVH)</b>	0.86±0.02	0.81±0.04	0.70±0.06	0.74±0.08	0.69±0.08
<b>WL-RD</b>	0.86±0.03	0.82±0.05*	0.73±0.08*	0.73±0.12	0.69±0.09
<b>MR-RD</b>	0.86±0.03	0.83±0.06**	0.75±0.08***	0.72±0.12	0.73±0.10*
<b>MR-DVHRD</b>	0.84±0.03	0.82±0.05*	0.72±0.08*	0.69±0.11	0.73±0.11*

\*p<0.05; \*\*p<0.001; \*\*\*p<0.00001 – Significance level as compared to the base model

*Abbreviations: DVH, dose volume histogram; AUC, area under the curve; AUPRC, area under precision-recall curve; WL, whole lung; RD, radiomic & dosiomic; MR, multi-region.*

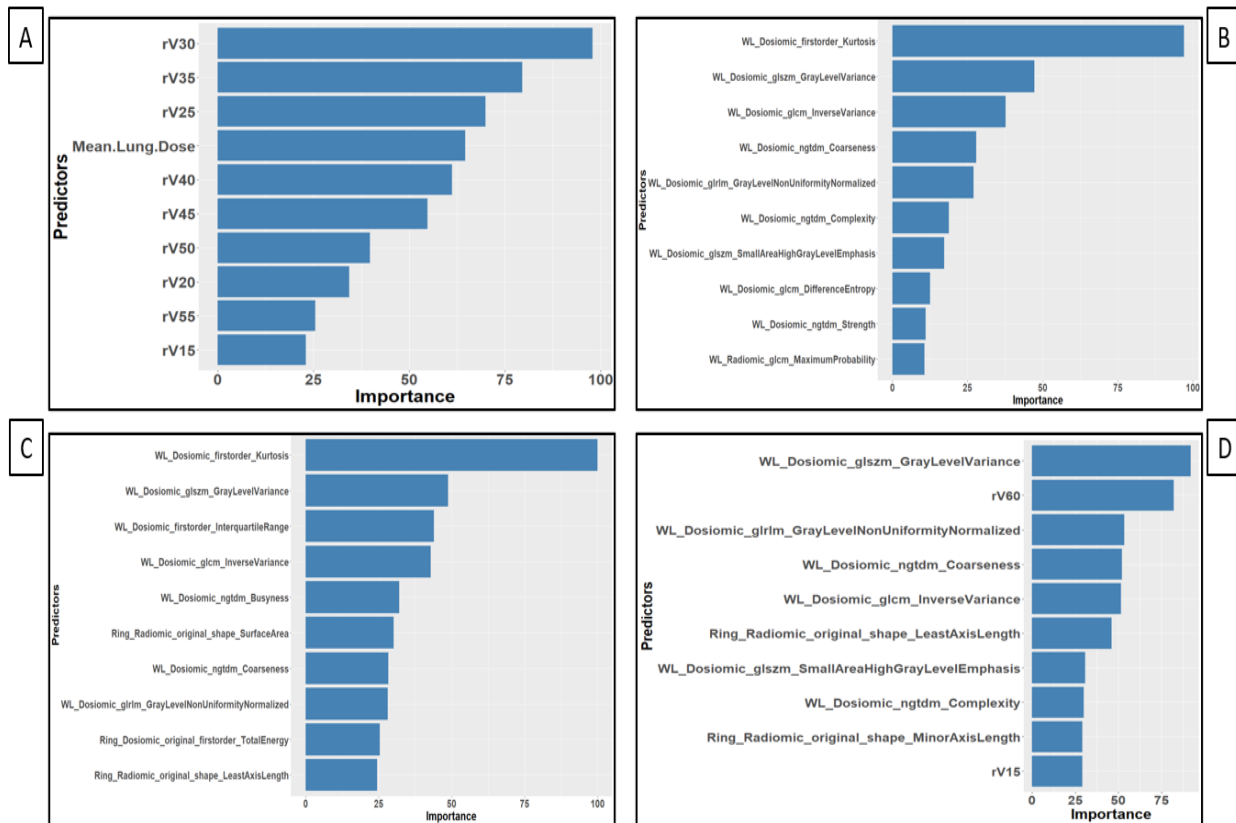
### Multivariate analysis

The results of all the performance metrics for all models (test set) for predicting RP $\geq$ 2 is reported in **Table 1**. The results for the training set are reported in **Supplementary Table S3**. All multiomic models outperformed the base model. Also, the MR-RD model type was the best performing model type among all model types. Boxplots comparing AUC and AUPRC for all the 4 model types are also shown in **Figure 2**. These results show that model types with omic features are better than the base model and the MR-RD model type is the best performing model among all models.



**Figure 3** – Boxplots comparing performance metrics of all 4 model types. A, Area under the curve (AUC) comparison between model types; B, area under precision-recall curve (AUPRC) comparison between model types. In both comparisons, the performance metrics for the multi-omic models are statistically significant as compared to the base model with the MR-RD being the best performing model. A one-sided Wilcoxon ranked significance test was used to check if the distribution of the other models is greater than that of the base model.

The most important (top 10) features are plotted as per rank for the base, WL-RD, MR-RD and MR-DVHRD model types in **Figure 3**. The plots show that for the base model type (**Fig 3A**), the relative V30 and V35 metrics are extremely important followed by relative V25 and mean lung dose. In the WL-RD model type (**Fig. 3B**), Kurtosis was found to be an important model feature, followed by dosiomic GLSZM Gray Level Variance, and dosiomic GLCM Inverse Variance. In the MR-RD model type (**Fig. 3C**), Kurtosis was once again shown to be the most important covariate followed by dosiomic GLSZM Gray Level Variance and Interquartile Range. Also, in the MR-DVHRD model type (**Fig. 3D**), dosiomic GLSZM Gray Level Variance was seen to be the most important covariate followed by rV60, dosiomic GLRLM Gray Level Non-Uniformity Normalized and others. Descriptions for all features are mentioned in the **Supplemental Materials**.



**Figure 4** – *Most Important features for each model type. A, DVH model type top features; B, WL-RD model type top features; C, MR-RD model type top features; D, MR-DVHRD model type top features.*

## Discussion

In this study, we compared the performance metrics (AUC, AUPRC, accuracy, precision, and recall) of the base model with the 3 omic model types (**Table 1**) to determine if multiomic features provided improved predictive value for OPMs of RP over using DVH metrics alone. A comparison of the base model with the WL-RD, MR-RD, and MR-DVHRD model types showed that all the multiomic model types outperformed base DVH models. Also, the MR-RD model type had the highest performance of the 4 model types. This shows that multiomic features, especially those extracted from a restricted area surrounding the GTV, has valuable information for predicting RP. To the best of our knowledge, this is the first study to use multi-region multiomic features to predict RP  $\geq$  grade 2 in NSCLC patients.

Several studies have analyzed radiomic features from CT for predicting RP. A study by Krafft et al.<sup>33</sup> demonstrated that when radiomic features were added to clinical and dosimetric features, the resultant model was able to better predict RP signifying the importance of radiomic features. Similarly, another study by Du et al.<sup>34</sup> used a nomogram-based radiomics approach with esophageal cancer patients and found that adding radiomic features to clinical and DVH metrics enhances the model capability for predicting RP. Dosiomics has been used in various studies to predict RT related treatment efficacy<sup>35</sup> and prognosis<sup>36</sup> but few toxicity prediction-based studies exist. A study conducted by Liang et al.<sup>19</sup> showed that the predictive power of dosiomic features were superior to DVH features; this study indicated that dosiomic features could possess more useful dose information than DVH metrics for predicting toxicity. A combined radiomic and dosiomic study by Zhang et al.<sup>37</sup> used logistic regression based risk scores to predict RP using clinical, DVH, radiomic and dosiomic features. They found that the prediction model had the best

performance when combining, clinical, radiomic, and dosiomic features. However, the paper does not mention which feature set plays the most important role in enhancing model efficacy. In our study as well, the combined clinical, radiomic and dosiomic whole lung model outperformed the base DVH model. However, in our case, the performance of the combined radiomic+dosiomic model type relied mostly on the contributions of the dosiomic features in the model (as illustrated by the top 10 most important features in **Fig. 3B**).

Prior studies have also shown the utility of incorporating additional radiomic features derived from the tumor's outer region, resulting in improved prediction accuracy<sup>20,21</sup>. A study by Kawahara et al.<sup>22</sup> also used multi-region radiomic features to predict RP in NSCLC. However, this study did not include two very important aspects: clinicopathologic factors and dosiomic features which have been proven to be vital in predicting RP. In our study, we built two multiomic multi-region model types: (a) MR-RD (including clinical, and multi-region radiomic and dosiomic features) and (b) MR-DVHRD (including DVH, clinical, and multi-region radiomic and dosiomic features), and both these models outperformed the base model. The MR-RD was found to be the best model type among all models. This shows that additional 3-dimensional dose information especially from a finite area surrounding the GTV in terms of ROI geometry, first-order statistics (individual voxel based) and second-order statistics (based on relationship between voxels) have a lot more predictive power as conventional DVH metrics. Features extracted from a sub-region around GTV allows for a more nuanced assessment of normal tissue radiosensitivity and our results suggests that is vital to take these aspects into consideration in the OPMs while predicting RP. This does confirm our hypothesis that incorporating multi-region features into OPMs can identify patients who are at a greater risk of developing RP  $\geq$  grade 2, thus mitigating potential adverse effects for future patients. The best performing model types based on our results could be used to run toxicity predictions once the treatment is planned for a specific patient. This would provide the physician with a data-driven prediction of the patient developing grade 2 or above RP before the treatment

begins, allowing the physician either to amend the treatment plan or to discuss post-treatment RP risks more accurately with the patient.

The figures for the top 10 most important features, as per Gini index, for the base, WL-RD, ML-RD, and MR-DVHRD model types are shown in **Fig 3**. In the WL-RD and MR-RD model types, it is seen that one first-order feature, kurtosis, is high on the importance scale. Statistically, kurtosis is used to describe the sharpness of the distribution. A study by Wang et al.<sup>38</sup> discusses the potential kurtosis has for quantifying tumor heterogeneity and the progression of lung disease. Hence, it would be of interest to study the relationship between the occurrence of RP and kurtosis to better understand how this feature plays a role in predicting RP. In all the multiomic models, it was the dosiomic features that were dominant emphasizing the need to have dosiomic features in OPMs when predicting RP.

There are a few limitations to our study. This study is a single institution, retrospective analysis. For this study, we have only used the original pretreatment CT and dose images for extracting features. There are studies that have examined the utility of various pre-processed (filtered) CT images, in addition to original CT image, to predict RP.<sup>39,40</sup> A future scope of this work could be to test if features extracted from filtered images help in enhancing the models based out of original omic features; this is motivated by Demircioglu et al. whose study suggests that preprocessing filters can influence radiomic analysis, and they can improve predictive performance of models.<sup>41</sup> Another limitation of this study is that we only use pretreatment CT for our omic analysis. A future scope of this study could be to determine if we can further enhance the performance for dosiomic-based toxicity models by introducing pretreatment PET radiomic features into our modeling scheme, similar to a study by Anthony et al.<sup>42</sup> Due to the inclusion of many radiomic and dosiomic features in our analysis, a feature reduction technique was used to avoid overfitting, thereby potentially excluding features that could be having predictive value. However, given the data-driven nature of our feature reduction process, we were able to find

predictive features. In addition to radiomics and dosiomics, including other omic data, such as genomics and proteomics, could also provide insightful information and as a future scope, they could be introduced into the modeling scheme to improve model robustness. An efficient and effective way of utilizing these features during the treatment planning phase remains to be one goal of our future study.

## **Conclusion**

In this work, we discovered that multiomic models are more efficient than DVH-only models at predicting  $\geq$ grade 2 RP. The best performing OPM included both whole-lung and multi-region multiomic features, thereby demonstrating that a restricted region extracted proximal to the GTV has utility for improving RP-based OPMs. Dosiomic features were the most important in all multiomic RP toxicity model types, thereby demonstrating that these features add significant value to predictive models of RP toxicity. Future outcome prediction models of radiation pneumonitis should include both multi-region multiomic and dosiomic features for the highest performing OPMs.

## References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 2021;71(3):209-249. doi:10.3322/caac.21660
2. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. *CA Cancer J Clin.* 2023;73(1):17-48. doi:10.3322/caac.21763
3. Jain V, Berman A. Radiation Pneumonitis: Old Problem, New Tricks. *Cancers.* 2018;10(7):222. doi:10.3390/cancers10070222
4. Ullah T, Patel H, Pena GM, Shah R, Fein AM. A contemporary review of radiation pneumonitis. *Curr Opin Pulm Med.* 2020;26(4):321. doi:10.1097/MCP.0000000000000682
5. Käsmann L, Dietrich A, Staab-Weijnitz CA, et al. Radiation-induced lung toxicity – cellular and molecular mechanisms of pathogenesis, management, and literature review. *Radiat Oncol Lond Engl.* 2020;15:214. doi:10.1186/s13014-020-01654-9
6. Robnett TJ, Machtay M, Vines EF, McKenna MG, Algazy KM, McKenna WG. Factors predicting severe radiation pneumonitis in patients receiving definitive chemoradiation for lung cancer. *Int J Radiat Oncol Biol Phys.* 2000;48(1):89-94. doi:10.1016/s0360-3016(00)00648-9
7. VOGELIUS IR, BENTZEN SM. A literature-based meta-analysis of clinical risk factors for development of radiation induced pneumonitis. *Acta Oncol Stockh Swed.* 2012;51(8):975-983. doi:10.3109/0284186X.2012.718093
8. Kong FM (Spring), Wang S. Nondosimetric Risk Factors for Radiation-Induced Lung Toxicity. *Semin Radiat Oncol.* 2015;25(2):100-109. doi:10.1016/j.semradonc.2014.12.003
9. Katsui K, Ogata T, Watanabe K, et al. Dose-volume parameters predict radiation pneumonitis after induction chemoradiotherapy followed by surgery for non-small cell lung cancer: a retrospective analysis. *BMC Cancer.* 2019;19(1):1144. doi:10.1186/s12885-019-6359-9
10. Tsujino K, Hirota S, Endo M, et al. Predictive value of dose-volume histogram parameters for predicting radiation pneumonitis after concurrent chemoradiation for lung cancer. *Int J Radiat Oncol.* 2003;55(1):110-115. doi:10.1016/S0360-3016(02)03807-5
11. Graham MV, Purdy JA, Emami B, et al. Clinical dose–volume histogram analysis for pneumonitis after 3D treatment for non-small cell lung cancer (NSCLC). *Int J Radiat Oncol.* 1999;45(2):323-329. doi:10.1016/S0360-3016(99)00183-2
12. Fay M, Tan A, Fisher R, Mac Manus M, Wirth A, Ball D. Dose-volume histogram analysis as predictor of radiation pneumonitis in primary lung cancer patients treated with radiotherapy. *Int J Radiat Oncol.* 2005;61(5):1355-1363. doi:10.1016/j.ijrobp.2004.08.025
13. Bourbonne V, Da-ano R, Jaouen V, et al. Radiomics analysis of 3D dose distributions to predict toxicity of radiotherapy for lung cancer. *Radiother Oncol.* 2021;155:144-150. doi:10.1016/j.radonc.2020.10.040



14. Deist TM, Dankers FJWM, 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. doi:10.1002/mp.12967
15. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology*. 2016;278(2):563-577. doi:10.1148/radiol.2015151169
16. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48(4):441-446. doi:10.1016/j.ejca.2011.11.036
17. Mayerhoefer ME, Materka A, Langs G, et al. Introduction to Radiomics. *J Nucl Med*. 2020;61(4):488-495. doi:10.2967/jnumed.118.222893
18. Adachi T, Nakamura M, Shintani T, et al. Multi-institutional dose-segmented dosiomic analysis for predicting radiation pneumonitis after lung stereotactic body radiation therapy. *Med Phys*. 2021;48(4):1781-1791. doi:10.1002/mp.14769
19. Liang B, Yan H, Tian Y, et al. Dosiomics: Extracting 3D Spatial Features From Dose Distribution to Predict Incidence of Radiation Pneumonitis. *Front Oncol*. 2019;9:269. doi:10.3389/fonc.2019.00269
20. Hao H, Zhou Z, Li S, et al. Shell feature: a new radiomics descriptor for predicting distant failure after radiotherapy in non-small cell lung cancer and cervix cancer. *Phys Med Biol*. 2018;63(9):095007. doi:10.1088/1361-6560/aabb5e
21. Xie C, Yang P, Zhang X, et al. Sub-region based radiomics analysis for survival prediction in oesophageal tumours treated by definitive concurrent chemoradiotherapy. *EBioMedicine*. 2019;44:289-297. doi:10.1016/j.ebiom.2019.05.023
22. Kawahara D, Imano N, Nishioka R, et al. Prediction of radiation pneumonitis after definitive radiotherapy for locally advanced non-small cell lung cancer using multi-region radiomics analysis. *Sci Rep*. 2021;11(1):16232. doi:10.1038/s41598-021-95643-x
23. Common Terminology Criteria for Adverse Events (CTCAE). Published online 2017.
24. Ghaye B, Wanet M, El Hajjam M. Imaging after radiation therapy of thoracic tumors. *Diagn Interv Imaging*. 2016;97(10):1037-1052. doi:10.1016/j.diii.2016.06.019
25. Arroyo-Hernández M, Maldonado F, Lozano-Ruiz F, Muñoz-Montaño W, Nuñez-Baez M, Arrieta O. Radiation-induced lung injury: current evidence. *BMC Pulm Med*. 2021;21(1):9. doi:10.1186/s12890-020-01376-4
26. Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn Reson Imaging*. 2012;30(9):1323-1341. doi:10.1016/j.mri.2012.05.001
27. Zwanenburg A, Vallières M, Abdalah MA, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology*. 2020;295(2):328-338. doi:10.1148/radiol.2020191145

28. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* 2017;77(21):e104-e107. doi:10.1158/0008-5472.CAN-17-0339
29. Scornet E. Tuning parameters in random forests. Coeurjolly JF, Leclercq-Samson A, eds. *ESAIM Proc Surv.* 2017;60:144-162. doi:10.1051/proc/201760144
30. Niedzielski JS, Yang J, Stingo F, et al. A Novel Methodology using CT Imaging Biomarkers to Quantify Radiation Sensitivity in the Esophagus with Application to Clinical Trials. *Sci Rep.* 2017;7(1):6034. doi:10.1038/s41598-017-05003-x
31. Salazar RM, Duryea JD, Leone AO, et al. Random Forest Modeling of Acute Toxicity in Anal Cancer: Effects of Peritoneal Cavity Contouring Approaches on Model Performance. *Int J Radiat Oncol.* 2024;118(2):554-564. doi:10.1016/j.ijrobp.2023.08.042
32. Niedzielski JS, Wei X, Xu T, et al. Development and application of an elastic net logistic regression model to investigate the impact of cardiac substructure dose on radiation-induced pericardial effusion in patients with NSCLC. *Acta Oncol.* 2020;59(10):1193-1200. doi:10.1080/0284186X.2020.1794034
33. Krafft SP, Rao A, Stingo F, et al. The utility of quantitative CT radiomics features for improved prediction of radiation pneumonitis. *Med Phys.* 2018;45(11):5317-5324. doi:10.1002/mp.13150
34. Du F, Tang N, Cui Y, et al. A Novel Nomogram Model Based on Cone-Beam CT Radiomics Analysis Technology for Predicting Radiation Pneumonitis in Esophageal Cancer Patients Undergoing Radiotherapy. *Front Oncol.* 2020;10:596013. doi:10.3389/fonc.2020.596013
35. Wu A, Li Y, Qi M, et al. Dosiomics improves prediction of locoregional recurrence for intensity modulated radiotherapy treated head and neck cancer cases. *Oral Oncol.* 2020;104:104625. doi:10.1016/j.oraloncology.2020.104625
36. Gabryś HS, Buettner F, Sterzing F, Hauswald H, Bangert M. Design and Selection of Machine Learning Methods Using Radiomics and Dosiomics for Normal Tissue Complication Probability Modeling of Xerostomia. *Front Oncol.* 2018;8:35. doi:10.3389/fonc.2018.00035
37. Zhang Z, Wang Z, Yan M, et al. Radiomics and Dosiomics Signature From Whole Lung Predicts Radiation Pneumonitis: A Model Development Study With Prospective External Validation and Decision-curve Analysis. *Int J Radiat Oncol.* 2023;115(3):746-758. doi:10.1016/j.ijrobp.2022.08.047
38. Wang Z, Yang C, Han W, et al. Quantifying lung cancer heterogeneity using novel CT features: a cross-institute study. *Insights Imaging.* 2022;13(1):82. doi:10.1186/s13244-022-01204-9
39. Yang S, Huang S, Ye X, Xiong K, Zeng B, Shi Y. Risk analysis of grade  $\geq 2$  radiation pneumonitis based on radiotherapy timeline in stage III/IV non-small cell lung cancer treated with volumetric modulated arc therapy: a retrospective study. *BMC Pulm Med.* 2022;22(1):402. doi:10.1186/s12890-022-02211-8
40. Hirose T aki, Arimura H, Ninomiya K, Yoshitake T, Fukunaga J ichi, Shioyama Y. Radiomic prediction of radiation pneumonitis on pretreatment planning computed tomography images prior to lung

cancer stereotactic body radiation therapy. *Sci Rep.* 2020;10(1):20424. doi:10.1038/s41598-020-77552-7

41. Demircioğlu A. The effect of preprocessing filters on predictive performance in radiomics. *Eur Radiol Exp.* 2022;6:40. doi:10.1186/s41747-022-00294-w
42. Anthony GJ, Cunliffe A, Castillo R, et al. Incorporation of pre-therapy 18F-FDG uptake data with CT texture features into a radiomics model for radiation pneumonitis diagnosis. *Med Phys.* 2017;44(7):3686-3694. doi:10.1002/mp.12282

## Figure and Table Legend

**Table 1.** Results of test set evaluation for different model types reported as mean of each metric with one standard deviation as the error. Model type comparison with the base model type (DVH metric) for WL-RD, MR-RD, and MR-DVHRD models is conducted with a one-sided Wilcoxon signed-rank test. Models with multiomic features had a better AUC and AUCPR as compared to the base model with MR-RD model type outperforming all the other model types.

**Figure 2 – Overall Study Workflow.** DVH, radiomic and dosiomic features are extracted, followed by feature reductions for omic features in the first step. 4 OPM types are built and performance metrics between the DVH and omic model types are compared in the second step.

**Figure 2– Sub-region (ring) multiomic feature extraction.** Radiomic and dosiomic features are extracted for a 20mm ring around the GTV overlapping the whole lung ROI, followed by feature reductions and analysis as described in **Fig 1**.

**Figure 3 - Boxplots comparing performance metrics of all 4 model types.** A, Area under the curve (AUC) comparison between model types; B, area under precision-recall curve (AUPRC) comparison between model types. In both comparisons, the performance metrics for the multi-omic models are statistically significant as compared to the base model with the MR-RD being the best performing model. A one-sided Wilcoxon ranked significance test was used to check if the distribution of the other models is greater than that of the base model.

**Figure 4** – Most Important features for each model type. A, DVH model type top features; B, WL-RD model type top features; C, MR-RD model type top features; D, MR-DVHRD model type top features.