主論文

Multi-institutional dose-segmented dosiomic analysis for predicting radiation pneumonitis after lung stereotactic body radiation therapy

Takanori Adachi, M.S.,[1,2] Mitsuhiro Nakamura, Ph.D.,[1,2] Takashi Shintani, M.D., Ph.D.,[2] Takamasa Mitsuyoshi, M.D., Ph.D.,[2,3] Ryo Kakino, M.S.,[1,2] Takashi Ogata, M.D.,[2,3] Tomohiro Ono, Ph.D.,[2] Hiroaki Tanabe, R.T.T.,[3] Masaki Kokubo, M.D., Ph.D.,[3] Takashi Sakamoto, M.D.,[4] Yukinori Matsuo, M.D., Ph.D.,[2] Takashi Mizowaki, M.D., Ph.D.[2]

[1)]Division of Medical Physics, Department of Information Technology and Medical Engineering, Human Health Sciences, Graduate School of Medicine, Kyoto University, Kyoto, Japan

[2)]Department of Radiation Oncology and Image-applied Therapy, Graduate School of Medicine, Kyoto University, Kyoto, Japan

[3)]Department of Radiation Oncology, Kobe City Medical Center General Hospital, Kyoto, Japan

[4)]Department of Radiation Oncology, Kyoto Katsura Hospital, Kyoto, Japan

Corresponding author:

Mitsuhiro Nakamura, Ph.D.

Division of Medical Physics, Department of Information Technology and Medical Engineering, Human Health Sciences, Graduate School of Medicine, Kyoto University

53 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan

Tel: +81-75-751-4176

E-mail: m_nkmr@kuhp.kyoto-u.ac.jp

Running head: Dosiomic analysis for RP after SBRT

Meeting presentation: This work was presented in part as a BLUE RIBBON ePOSTER at the 2020 Joint AAPM|COMP Virtual Meeting, July 12–16, 2020.

Abstract

Purpose: To predict radiation pneumonitis (RP) grade 2 or worse after lung stereotactic body radiation therapy (SBRT) using dose-based radiomic (dosiomic) features.

Methods: This multi-institutional study included 247 early-stage non-small cell lung cancer patients who underwent SBRT with a prescribed dose of 48–70 Gy at an isocenter between June 2009 and March 2016. Ten dose–volume indices (DVIs) were used, including the mean lung dose, internal target volume size, and percentage of entire lung excluding the internal target volume receiving greater than x Gy (x = 5, 10, 15, 20, 25, 30, 35, and 40). A total of 6,808 dose-segmented dosiomic features, such as shape, first order, and texture features, were extracted from the dose distribution. Patients were randomly partitioned into two groups: model training (70%) and test datasets (30%) over 100 times. Dosiomic features were converted to z-scores (standardized values) with a mean of zero and a standard deviation (SD) of one to put different variables on the same scale. The feature dimension was reduced using the following methods: inter-feature correlation based on Spearman's correlation coefficients and feature importance based on a light gradient boosting machine (LightGBM) feature selection function. Three different models were developed using LightGBM as follows: (i) a model with 10 DVIs (DVI model), (ii) a model with the selected dosiomic features (dosiomic model), and (iii) a

model with 10 DVIs and selected dosiomic features (hybrid model). Suitable hyperparameters were determined by searching the largest average area under the curve (AUC) value in the receiver operating characteristic curve (ROC–AUC) via stratified five-fold cross-validation. Each of the final three models with the closest the ROC–AUC value to the average ROC–AUC value was applied to the test datasets. The classification performance was evaluated by calculating the ROC–AUC, AUC in the precision–recall curve (PR–AUC), accuracy, precision, recall, and f1-score. The entire process was repeated 100 times with randomization, and 100 individual models were developed for each of the three models. Then the mean value and SD for the 100 random iterations were calculated for each performance metric.

Results: Thirty-seven (15.0%) patients developed RP after SBRT. The ROC-AUC and PR-AUC values in the DVI, dosiomic, and hybrid models were 0.660 ± 0.054 and 0.272 ± 0.052, 0.837 ± 0.054 and 0.510 ± 0.115, and 0.846 ± 0.049 and 0.531 ± 0.116, respectively. For each performance metric, the dosiomic and hybrid models outperformed the DVI models ($p < 0.05$). Texture-based dosiomic feature was confirmed as an effective indicator for predicting RP.

Conclusions: Our dose-segmented dosiomic approach improved the prediction of the incidence of RP after SBRT.

Keywords: stereotactic body radiation therapy; radiation pneumonitis; dosiomics; machine learning; multi-institutional study

Introduction

Stereotactic body radiation therapy (SBRT) is typically used to treat early-stage non-small cell lung cancer (NSCLC). The standard treatment for early-stage NSCLC is surgery; however, SBRT is an effective treatment method for patients with inoperable tumors or refusing surgical resection (1). Several multi-institutional phase II studies on SBRT in early-stage NSCLC have indicated high local control rates at 3 years (1–4).

Radiation pneumonitis (RP) is the dominant toxicity after lung SBRT. Symptomatic RP predominantly occurs one year after SBRT (5, 6), and the reported incidence rates of RP grade 2 or worse ranged from 2.4% to 28.0% (7–12). Although most RP cases are grade 2 and manageable, some are severe and life-threatening (13). The risks of RPs reduce the benefits of SBRT; hence, it is important to consider the predictive factors of RP incidence. Several studies have been performed to identify dose–volume indices (DVIs), such as the mean lung dose (MLD) and the percentage of the lung volume receiving greater than x Gy ($V_{x\ Gy}$) corresponding to RP after SBRT (7–12). However, these features describe few characteristics of the dose distribution, and no clear consensus is available regarding the appropriate use of DVIs for predicting RP after SBRT.

Recently, machine learning has progressed rapidly in the field of radiation oncology (14). In particular, a radiomic approach that enables the extraction of

89 quantitative medical imaging features for predicting prognostic outcomes was applied to
90 computed tomography (CT) images to predict patient prognosis, such as local recurrence
91 and distant metastasis after lung SBRT (15–18). Similar to the CT-based radiomic
92 approach, some studies have reported the effectiveness of a dose-based radiomic
93 (dosiomic) approach based on three-dimensional (3D) dose distributions to predict
94 toxicities after radiation therapy (19–22). In addition to the limited DVIs, dosiomic
95 features provide quantitative analysis based on multidimensional data, such as the shape,
96 statistics, and texture features of the dose distribution. Although some studies have
97 considered dosiomic features for predicting radiation-induced toxicities, the prediction of
98 RP after SBRT in patients with early-stage NSCLC has not been performed. Furthermore,
99 as a dosiomic approach, no study has focused on using several regions of the dose
100 distribution, for example, inside the region of $V_{x\ Gy}$, as used in the extraction of DVIs.
101 This study aimed to investigate the effectiveness of dose-segmented dosiomic
102 features for predicting the incidence of RP after lung SBRT in patients with early-stage
103 NSCLC. We developed three predictive models with DVIs, dosiomic features, and a
104 combination of DVI and dosiomic features using a light gradient boosting machine
105 (LightGBM) based on a gradient-boosting decision tree algorithm.
106

Materials and methods

Overall workflow

The overall workflow of this study is shown in Figure 1. The study involved the following steps: (I) acquire patient data from Digital Imaging and Communications in Medicine Radiation Therapy (DICOM-RT) files and partitioning patients into training and test datasets, (II) extract 10 DVIs and 6,808 dosiomic features from the dose distribution, (III) select features using inter-feature correlation based on Spearman's correlation coefficients (CCs) and feature importance based on a LightGBM feature selection function, and developing predictive models for RP using LightGBM, and (IV) evaluate the three predictive models. The previous steps were repeated 100 times with randomization.

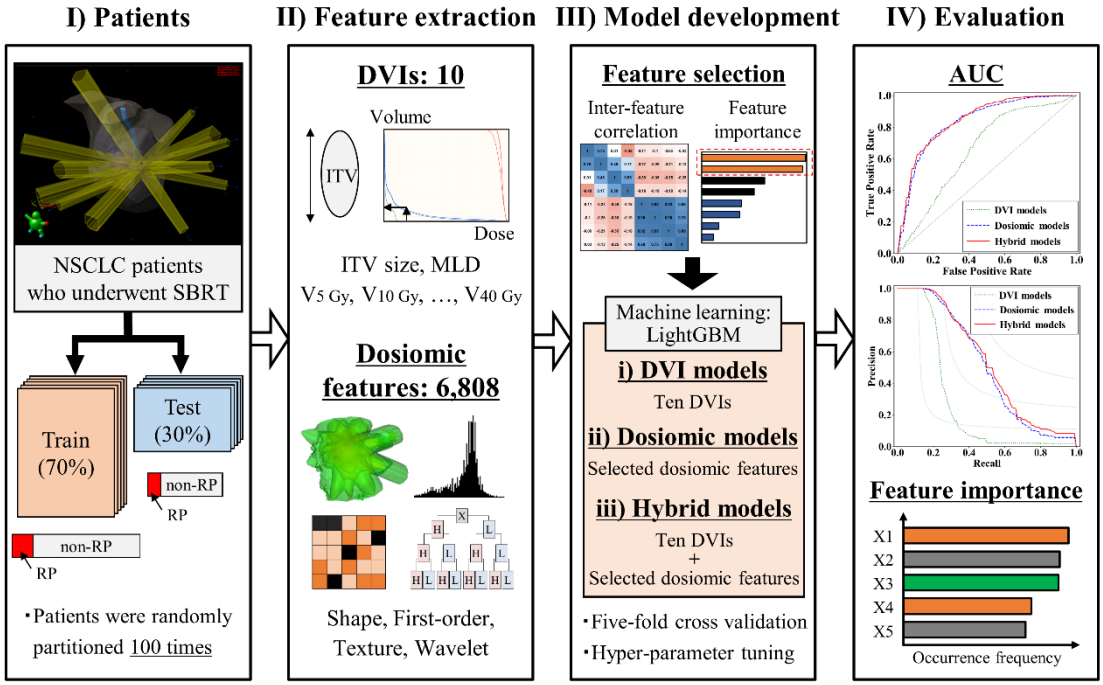


Figure 1: Overall workflow of dosiomic analysis performed in this study. *Abbreviations*: NSCLC = non-small cell lung cancer, SBRT = stereotactic body radiation therapy, RP = radiation pneumonitis, DVI = dose volume index, ITV = internal target volume, MLD = mean lung dose, $V_{x\,Gy}$ = percentage of entire lung excluding ITV volume receiving greater than x Gy, LightGBM = a light gradient boosting machine, AUC = area under the curve.

Patients

Our retrospective study was approved by the Institutional Review Board of Kyoto University Hospital and other institutions (approval number: R1536). We considered 685 early-stage NSCLC patients who underwent lung SBRT with non-

coplanar 3D-conformal radiotherapy at three institutions between June 2009 and March 2016. The patients were treated with 6 MV X-rays; the prescribed dose was 48–70 Gy in 4–8 fractions at an isocenter. All patients received radiation doses greater than 100 Gy of a biologically effective dose with $\alpha/\beta = 10$ at the isocenter. The patient selection criteria are shown in Figure 2. A total of 247 patients with NSCLC were eligible for this study.

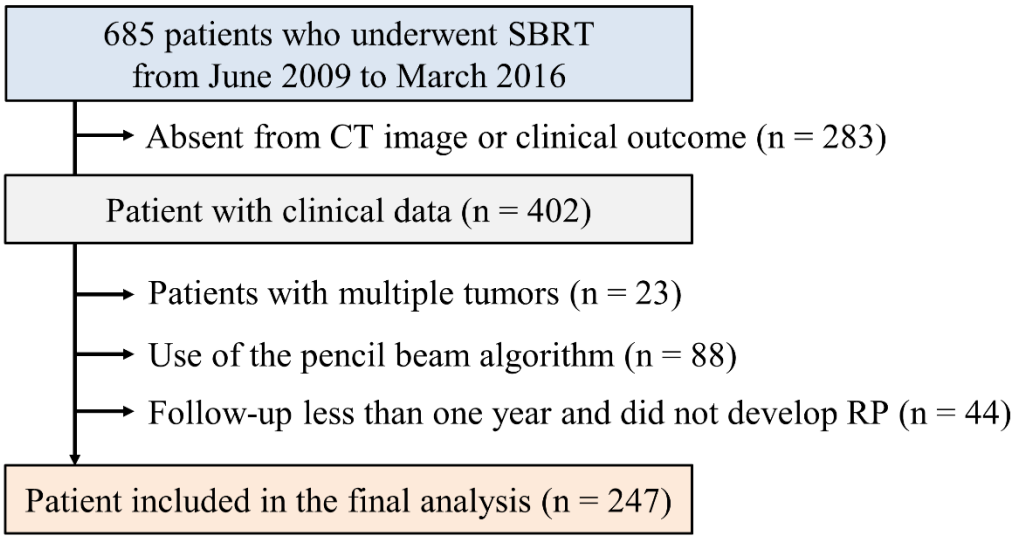


Figure 2: Patient inclusion and exclusion criteria. The exclusion criteria were as follows: the absence of CT and clinical data, multiple tumors, use of pencil beam algorithm, no development of RP, and a follow-up period of less than one year. *Abbreviations*: SBRT = stereotactic body radiation therapy, CT = computed tomography, RP = radiation pneumonitis.

Four-dimensional averaged CT images, and an expiratory breath-hold CT image were used for dose calculation. The dose distribution was calculated using the X-ray voxel Monte Carlo (XVMC) algorithm on an iPlan radiation therapy system (Brainlab, Munich, Germany). The grid size for the dose calculation was 2.3 × 2.3 × 2.5 mm. The dose-volume constraints of organs at risk were based on the Japan Clinical Oncology Group 0403 and 1408 protocols (23).

Symptomatic RP was graded based on the National Cancer Institute Common Terminology Criteria for Adverse Events version 4.0, published by the National Cancer Institute, Cancer Therapy Evaluation Program. In this study, cases of RP grade 2 or worse were considered symptomatic.

The patients were randomly partitioned into two groups: 70% for the training datasets and 30% for the test datasets over 100 times. The RP incidence rate was approximately equal for each dataset. This study was categorized as "Type 1b: Development and validation using resampling" in the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis statement (24).

Feature extraction

From the DICOM-RT plan and dose files, contour information and dose

160 distribution were extracted. Thereafter, DVIs and dosiomic features were extracted from
161 the dose distribution. The 10 DVIs, referred to in previous studies, were as follows: MLD,
162 internal target volume (ITV) size, $V_{5\ Gy}$, $V_{10\ Gy}$, $V_{15\ Gy}$, $V_{20\ Gy}$, $V_{25\ Gy}$, $V_{30\ Gy}$, $V_{35\ Gy}$, and
163 $V_{40\ Gy}$ (12). In this study, the entire lung, excluding the ITV (Lung-ITV), was regarded as
164 a normal lung.

165 Dosiomic features were extracted from the dose distribution using PyRadiomics
166 version 2.2.0, which enables the extraction of a large group of developed features from
167 multiple medical images, including the dose distribution (25). Considered as dosiomic
168 features, Lung-ITVs receiving doses greater than x Gy (x = 5, 10, 15, 20, 25, 30, 35, and
169 40) were used as regions of interest (ROIs) that were also used in the extraction of DVIs
170 (Figure 3). The dosiomic features, extracted from the eight types of ROIs without
171 preprocessing were as follows: 112 (14 × 8) shape features, 144 (18 × 8) first-order
172 features, and 600 (75 × 8) texture features; features extracted with wavelet filters were
173 1,152 (144 × 8) first-order features and 4,800 (600 × 8) texture features (Supplemental
174 Table 1). In the first-order and texture features, wavelet filters were used in eight
175 decompositions (i.e., LLL, LLH, LHL, LHH, HLL, HLH, HHL, and HHH) to extract
176 multiple features from different frequency bands of the original dose distribution. A total
177 of 6,808 dosiomic features were used in this study.

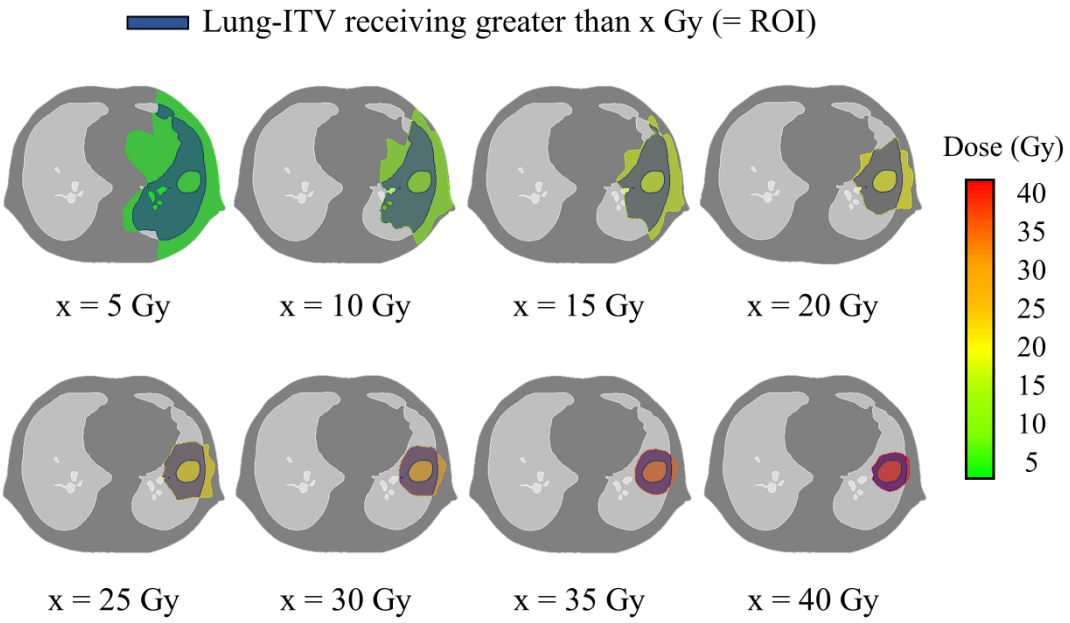


Figure 3: Eight types of ROIs in dose-segmented dosiomic features extraction. Lung-ITV receiving greater than x Gy (x = 5, 10, 15, 20, 25, 30, 35, and 40) were used as the ROIs. Area painted in dark blue shows Lung-ITV receiving greater than x Gy. *Abbreviations*: Lung-ITV = entire lung excluding internal target volume; ROI = region of interest.

Feature selection

To reduce the feature dimensions, dosiomic features were selected based on the redundancy and significance of the features. First, redundant features were eliminated using Spearman's correlation analysis. The Spearman's CCs were calculated after the

dosiomic features were converted to z-scores. The z-scores are the standardized values with a mean of zero and a standard deviation (SD) of one to put different variables on the same scale. If a high correlation exists between two features (CC $\geq$ 0.8), then one of the two features that are highly correlated with the other remaining features is eliminated (18, 26, 27). The threshold value above was determined to eliminate the multicollinearity of the features for the next step. Second, important features were selected using the "Boruta" package based on the LightGBM, which is widely used for binary classifications (28). The LightGBM repeatedly utilizes the learning results of prior weak predictors for subsequent predictors and yields a feature selection method based on a feature significance score. In LightGBM feature selection, features with significantly worse importance than shadow ones are eliminated, whereas features related to the outcome variable are selected.

Model building

For the training datasets, three different models were developed using LightGBM as follows: (i) a model with 10 DVIs (DVI model), (ii) a model with the selected dosiomic features (dosiomic model), and (iii) a model with 10 DVIs and selected dosiomic features (hybrid model). To address the imbalanced sample size in the datasets,

the minority and majority classes were automatically weighted based on their sample size proportion using the LightGBM parameter "scale_pos_weight." To develop significantly accurate models and prevent overfitting to the training datasets, hyperparameters for LightGBM were optimized using stratified five-fold cross-validation with Bayesian optimization (29). Suitable hyperparameters were determined by searching the largest average area under the curve value in the receiver operating characteristic curve (ROC–AUC) of the five models. Furthermore, each of the final three models with the closest ROC–AUC value to the average ROC–AUC value in five-fold cross-validation was applied to the test datasets. The significance of the three models' features was calculated based on the LightGBM by counting the frequency of the top five important features.

Performance evaluation

The ROC curve is a diagnostic tool for evaluating the performance of binary classifiers. However, because class imbalance can introduce bias to the majority class, high predictive performances on ROC curves may not be high on precision–recall (PR) curves (30). To focus on the incidence of RP belonging to the minor class, the AUCs in the precision–recall curves (PR–AUC) were calculated. Considering the performance metrics for the predictive models, the accuracy [= (TP + TN) / (TP + TN + FP + FN)],

recall [= TP / (TP + FN)], precision [= TP / (TP + FP)], and f1-score [= 2TP / (2TP + FP + FN)] were calculated, where TP is true positive, TN is true negative, FP is false positive, and FN is false negative (31). To assess the binary classifications, a default cutoff value of 0.5 was used as the threshold for the predictive probability of each model. The entire process was repeated 100 times with randomization, and 100 individual models were developed for each of the three models. Then the mean value and SD for the 100 random iterations were calculated for the ROC–AUC, PR–AUC, accuracy, recall, precision, and f1-score.

Statistical analysis

The statistical significance of the ROC–AUC and the PR–AUC for the three models was evaluated using the paired t-test and Wilcoxon signed-rank test. Bonferroni correction was applied to adjust the p-value of multiple comparisons. An appropriate test was determined based on normality using the Shapiro–Wilk test. Statistical analyses were performed using R software version 3.6.1 (32). A p-value < 0.05 was considered statistically significant.

Results

Patient characteristics

Table 1 summarizes the characteristics of the study participants in the three institutions. The median follow-up duration was 36.3 months (range: 1.6–119.0 months), and the median duration for the RP incidence was 3.7 months (range: 1.6–8.7 months). A total of 37 patients (15.0%) developed RP grade 2 or worse after SBRT. The number of patients who developed RP after SBRT with the prescribed doses of 48, 50, 56, 60, 64, and 70 Gy were 23 (15.3%), 4 (26.7%), 0 (0%), 4 (13.3%), 0 (0%), and 6 (12%), respectively. In the training and test datasets, the data distribution of the RP class was consistent. Table 2 summarizes the mean value and SD for 10 DVIs. All treatment plans satisfied the dose–volume constraints shown in Table 2.

253 **Table 1: Characteristics of the study participants.**

	All (n = 247)	Institution A (n = 120)	Institution B (n = 91)	Institution C (n = 36)
Age [years]	Median: 78 (range: 41–92)	Median: 78 (range: 56–92)	Median: 80 (range: 58–92)	Median: 79 (range: 41–92)
Sex [Male / Female]	158 / 89	86 / 34	54 / 37	18 / 18
Histology [AC / SqCC / Others / unknown]	93 / 45 / 7 / 102	39 / 23 / 3 / 55	46 / 19 / 3 / 23	8 / 3 / 1 / 24
Tumor location [RUL / RML / RLL / LUL / LLL]	67 / 19 / 66 / 63 / 32	31 / 8 / 37 / 30 / 14	28 / 7 / 18 / 24 / 14	8 / 4 / 11 / 9 / 4
Prescribe dose [48 / 50 / 56 / 60 / 64 / 70 (Gy)]	150 / 15 / 1 / 30 / 1 / 50	69 / 0 / 0 / 1 / 0 / 50	71 / 0 / 0 / 20 / 0 / 0	10 / 15 / 1 / 9 / 1 / 0
Planning CT [Ave-CT / Ex-CT]	156 / 91	120 / 0	0 / 91	36 / 0
RP grade [2 / 3 / 4 / 5]	33 / 3 / 0 / 1	8 / 1 / 0 / 1	16 / 2 / 0 / 0	9 / 0 / 0 / 0
RP rate [Total / training / test (%)]	15.0 / 15.1 ± 0.0 / 14.7 ± 0.0	8.3 / 8.4 ± 1.5 / 8.1 ± 3.5	19.8 / 19.7 ± 2.2 / 20.2 ± 5.0	25.0 / 26.1 ± 4.7 / 22.9 ± 6.2

254

Abbreviations: AC = adenocarcinoma; SqCC = squamous cell carcinoma; RUL = right upper lobe; RML = right middle lobe; RLL = right lower lobe; LUL = left upper lobe; LLL = left lower lobe; CT = computed tomography; Ave-CT = four-dimensional averaged CT images generated from 10 respiratory phase CT images; Ex-CT = expiratory breath-hold CT images; RP = radiation pneumonitis.

Note: The incidence rates of RP in the training and test are shown as mean ± SD because we randomly partitioned the patients into training and test datasets 100 times while maintaining the RP rate for each dataset in three institutions.

Table 2: Mean value and standard deviation for dose-volume indices by the prescribed dose.

Object	DVI	Dose–volume constraints	All (n = 247)	48 Gy (n = 150)	50 Gy (n = 15)	56 Gy (n = 1)	60 Gy (n = 30)	64 Gy (n = 1)	70 Gy (n = 50)
ITV	size (cm^3)		16.7 ± 15.3	15.8 ± 14.6	6.9 ± 7.1	13.4	24.6 ± 19.2	18.9	16.6 ± 12.7
Lung-ITV	Mean lung dose (Gy)	≤ 18 Gy	4.3 ± 1.6	3.9 ± 1.4	4.1 ± 1.6	4.2	5.5 ± 1.9	3.6	4.7 ± 1.4
	$V_{5\ Gy}$ (%)		19.9 ± 6.6	19.2 ± 6.7	18.4 ± 6.9	18.4	21.8 ± 6.7	13.5	20.7 ± 5.4
	$V_{10\ Gy}$ (%)		12.8 ± 5.4	11.9 ± 5.4	13.2 ± 5.4	13.1	15.5 ± 5.9	10.1	13.5 ± 4.2
	$V_{15\ Gy}$ (%)	≤ 25%	8.7 ± 4.4	7.6 ± 4.0	9.9 ± 4.5	8.0	11.5 ± 5.1	7.5	9.5 ± 3.6
	$V_{20\ Gy}$ (%)	≤ 20%	6.1 ± 3.4	5.3 ± 3.0	6.9 ± 3.9	5.8	8.4 ± 4.1	5.3	6.6 ± 2.7
	$V_{25\ Gy}$ (%)		4.5 ± 2.6	3.8 ± 2.2	4.8 ± 2.7	4.5	6.5 ± 3.2	4.1	4.8 ± 2.2
	$V_{30\ Gy}$ (%)		3.4 ± 2.0	2.9 ± 1.7	3.4 ± 1.8	3.5	5.2 ± 2.6	3.2	3.7 ± 1.8
	$V_{35\ Gy}$ (%)		2.6 ± 1.6	2.1 ± 1.3	2.5 ± 1.4	2.8	4.1 ± 2.2	2.4	2.9 ± 1.4
	$V_{40\ Gy}$ (%)		1.9 ± 1.3	1.5 ± 1.0	1.8 ± 1.1	2.1	3.3 ± 1.8	1.7	2.3 ± 1.2

Abbreviations: DVI = dose volume index; ITV = internal target volume; Lung-ITV = entire lung excluding ITV; $V_{x\ Gy}$ = percentage of the Lung-ITV volume receiving greater than x Gy.

Note: dose–volume constraints were based on the Japan Clinical Oncology Group 0403 and 1408 protocols (22).

Model performance

Figure 4 shows the resultant ROC and PR curves for predicting RP in the training and test datasets. The mean ± SD of the ROC-AUC and PR-AUC values for the test datasets in the DVI, dosiomic, and hybrid models were 0.660 ± 0.054 and 0.272 ± 0.052, 0.837 ± 0.054 and 0.510 ± 0.115, and 0.846 ± 0.049 and 0.531 ± 0.116, respectively. In the analyses, the ROC–AUC and PR–AUC values in the dosiomic and hybrid models were significantly higher than those in the DVI models ($p < 0.05$). Furthermore, no significant differences were observed in the ROC–AUC and PR–AUC between the dosiomic and hybrid models.

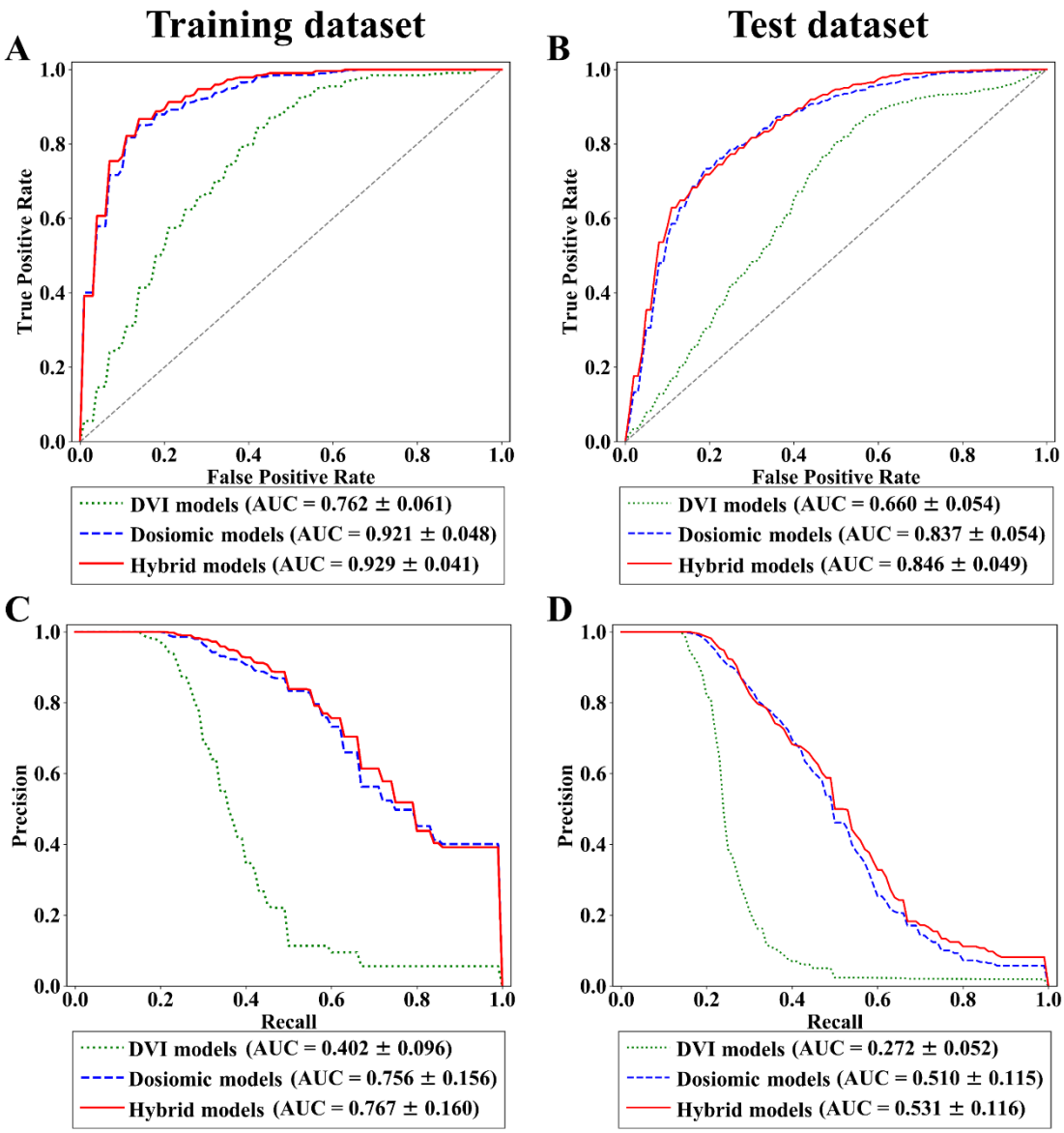


Figure 4: Mean receiver operating characteristic (ROC) and precision–recall (PR) curves with 100 individual models for each of the three models; ROC curves in (A) training and (B) test datasets, and PR curves in (C) training and (D) the test datasets. For the curves, each line indicates the DVI models (green), dosiomic models (blue), and hybrid models (red). *Abbreviations*: DVI = dose volume index; AUC = area under the curve.

Table 3 summarizes the mean ± SD of the accuracy, precision, recall, and f1-score in the training and test datasets. For each performance metric, the dosiomic and hybrid models outperformed the DVI models. Furthermore, for the hybrid models, each performance metric value improved by adding the DVIs to the dosiomic models.

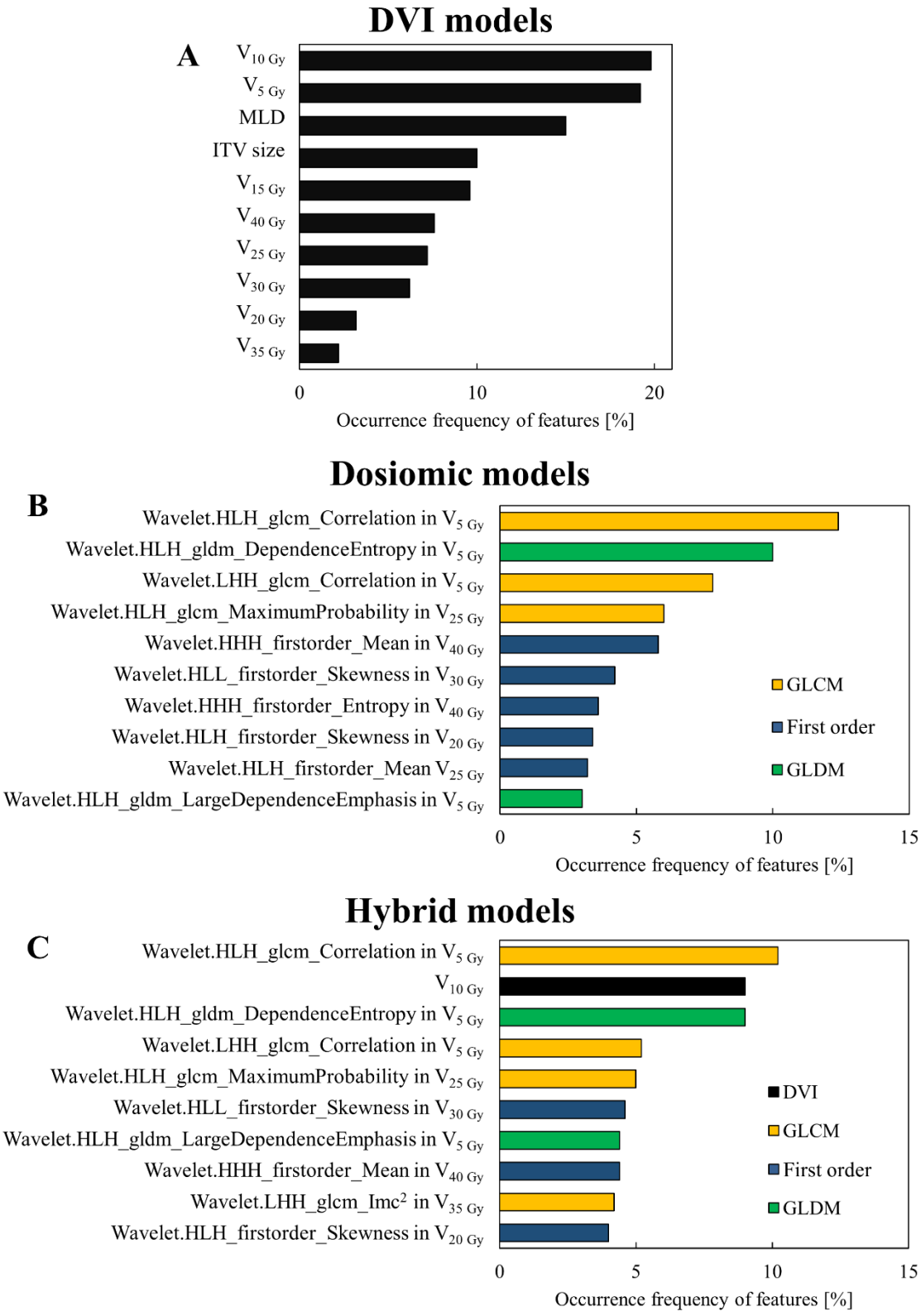
Table 3: Performance metrics of three models.

		Accuracy	Recall	Precision	F1-score
DVI models	Training	0.699 ± 0.067	0.620 ± 0.140	0.285 ± 0.066	0.383 ± 0.070
	Test	0.653 ± 0.062	0.507 ± 0.170	0.214 ± 0.053	0.297 ± 0.073
Dosiomic models	Training	0.899 ± 0.051	0.821 ± 0.133	0.658 ± 0.155	0.718 ± 0.113
	Test	0.843 ± 0.038	0.660 ± 0.145	0.482 ± 0.096	0.551 ± 0.098
Hybrid models	Training	0.900 ± 0.047	0.837 ± 0.128	0.662 ± 0.163	0.725 ± 0.109
	Test	0.846 ± 0.038	0.660 ± 0.138	0.494 ± 0.093	0.556 ± 0.087

Abbreviation: DVI = dose volume index.

Feature importance

In the DVI models, $V_{10\ Gy}$ was the most frequently occurring feature among the 10 DVIs (Figure 5A). In the dosiomic and hybrid models, the gray level co-occurrence matrix (GLCM) was indicated as the most important group (Supplemental Figure 1), whereas "Wavelet.HLH_glcm_Correlation in $V_{5\ Gy}$" was classified as a highly frequent feature (Figures 5B, C).


DVI models
A
V10 Gy
V5 Gy
MLD
ITV size
V15 Gy
V40 Gy
V25 Gy
V30 Gy
V20 Gy
V35 Gy
0
10
20
Occurrence frequency of features [%]
Dosiomic models
B
Wavelet.HLH_glcm_Correlation in V5 Gy
Wavelet.HLH_gldm_DependenceEntropy in V5 Gy
Wavelet.LHH_glcm_Correlation in V5 Gy
Wavelet.HLH_glcm_MaximumProbability in V25 Gy
Wavelet.HHH_firstorder_Mean in V40 Gy
Wavelet.HLL_firstorder_Skewness in V30 Gy
Wavelet.HHH_firstorder_Entropy in V40 Gy
Wavelet.HLH_firstorder_Skewness in V20 Gy
Wavelet.HLH_firstorder_Mean V25 Gy
Wavelet.HLH_gldm_LargeDependenceEmphasis in V5 Gy
GLCM
First order
GLDM
0
5
10
15
Occurrence frequency of features [%]
Hybrid models
C
Wavelet.HLH_glcm_Correlation in V5 Gy
V10 Gy
Wavelet.HLH_gldm_DependenceEntropy in V5 Gy
Wavelet.LHH_glcm_Correlation in V5 Gy
Wavelet.HLH_glcm_MaximumProbability in V25 Gy
Wavelet.HLL_firstorder_Skewness in V30 Gy
Wavelet.HLH_gldm_LargeDependenceEmphasis in V5 Gy
Wavelet.HHH_firstorder_Mean in V40 Gy
Wavelet.LHH_glcm_Imc2 in V35 Gy
Wavelet.HLH_firstorder_Skewness in V20 Gy
DVI
GLCM
First order
GLDM
0
5
10
15
Occurrence frequency of features [%]

295

296

297 **Figure 5**: Importance of features used in (A) DVI, (B) dosiomic, and (C) hybrid models.
298 Vertical axis shows top-10 important features; horizontal axis shows normalized
299 occurrence frequency of features. In the dosiomic and hybrid models,
300 "Wavelet.HLH_glcm_Correlation in $V_{5\ Gy}$" was the most important feature. The feature
301 indicates the glcm correlation with wavelet filter in HLH decomposition inside the region
302 of $V_{5\ Gy}$. *Abbreviation*: DVI = dose volume index; ITV = internal target volume; MLD =
303 mean lung dose; $V_{x\ Gy}$ = percentage of entire lung excluding ITV volume receiving greater
304 than x Gy; GLCM = gray level co-occurrence matrix; GLDM gray level dependence
305 matrix.
306
307 Figure 6 shows examples of the dose distribution, feature map, and follow-up
308 CT images after SBRT comprising RP grade 2 and non-RP cases. The dosiomic feature
309 maps of (C) and (D) were calculated based on planned dose distributions of (A) and (B),
310 respectively. Follow-up CT scans of (E) and (F) were obtained after treatment, and they
311 indicate the same axial slice as those in the dosiomic feature maps of (C) and (D),
312 respectively. Although $V_{10\ Gy}$ was greater than the median value (11.8%) in both cases,
313 the feature map calculated by the "Wavelet.HLH_glcm_Correlation in $V_{5\ Gy}$" feature with
314 RP grade 2 cases used predominantly low values (blue) near the RP position, whereas the

non-RP case used predominantly high values (red) in the ROI. The feature maps calculated from the dose distribution were highly correlated with the RP incidence based on follow-up CT after treatment.

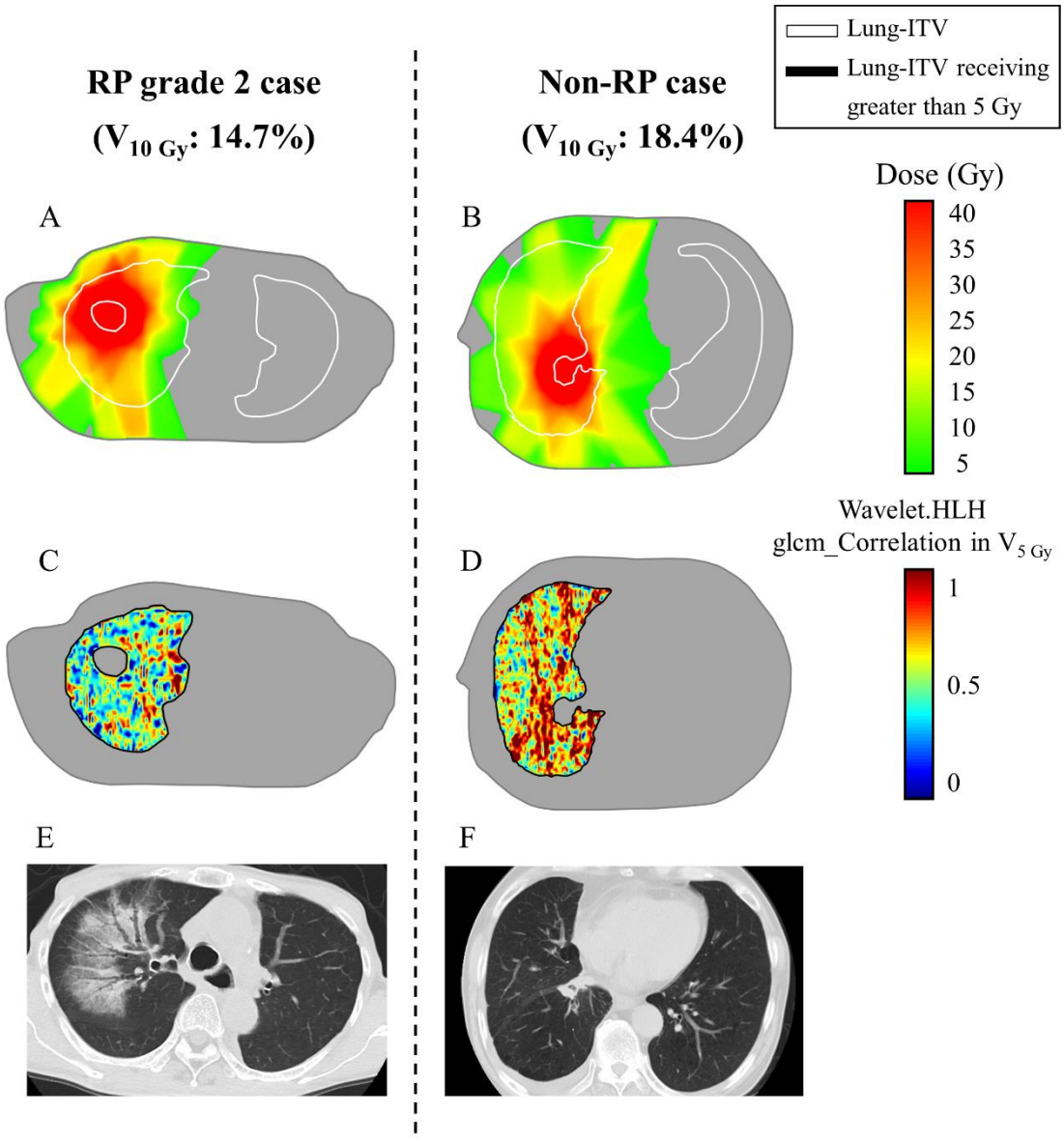


Figure 6: Example of (A, B) dose distribution; (C, D) dosiomic feature map calculated by "Wavelet.HLH_glcm_Correlation in $V_{5\ Gy}$" feature; (E, F) follow-up CT scan after

SBRT. “Wavelet.HLH_glcm_Correlation in $V_{5\ Gy}$” indicates the glcm correlation with wavelet filter in HLH decomposition inside $V_{5\ Gy}$ region. Follow-up CT scans are shown at (E) 271 days and (F) 203 days after SBRT. RP grade 2 case is shown in (A, C, E), and non-RP case is shown in (B, D, F). White contour lines show Lung-ITV in (A) and (B); black contour lines show Lung-ITV receiving greater than x Gy in (C) and (D). *Abbreviations*: RP = radiation pneumonitis; Lung-ITV = entire lung excluding the internal target volume; $V_{x\ Gy}$ = percentage of Lung-ITV volume receiving greater than x Gy.

Discussion

We investigated a novel approach for predicting RP after lung SBRT in patients with early-stage NSCLC using dose-segmented dosiomic features with the machine-learning algorithm in LightGBM. To the best of our knowledge, our study is the first to focus on dose-segmented dosiomic features extracted from the regions of $V_{x\ Gy}$. Subsequently, we developed an accurate model that thoroughly described the characteristics of the dose distribution. Furthermore, we identified that the dosiomic feature, “Wavelet.HLH_glcm_Correlation in $V_{5\ Gy}$” was more significant than the other DVIs. Our dosiomic approach can aid in predicting RP after SBRT.

A previous study reported the effectiveness of dosiomics in predicting the RP

incidence. Liang et al. predicted RP grade 2 or worse after volumetric modulated arc therapy (VMAT) in stage I to IV NSCLC patients (19). They used the ipsilateral, contralateral, and the entire lung as the ROI because the irradiation field for patients with multiple stages was large. Their AUC for predicting RP improved from 0.676 (with a DVI model) and 0.744 (with a normal tissue complication probability model) to 0.782 (with a dosiomic model). Applying this approach to our study in early-stage NSCLC patients will affect the interpretation of the partial-dose distribution that correlates with the RP incidence because the non-irradiated lung area may not be critical for the analysis. To thoroughly analyze the dose distribution, our dose-segmented dosiomic approach focused on whether dose-segmented dosiomic features improve the prediction of RP after SBRT. In this study, we developed accurate predictive models for RP after SBRT using dose-segmented dosiomic features rather than DVIs. We discovered that "Wavelet.HLH_glcm_Correlation in $V_{5\ Gy}$" was the most important feature in the dosiomic and hybrid models. Liang et al. indicated that the "GLCM contrast" feature was a significant feature correlated with the incidence of RP after VMAT, although their irradiation technique and patient characteristics differed from those of our study (19). Therefore, the group of GLCM features may contribute to the incidence of RP regardless of the irradiation technique.

The unbalanced patient dataset can result in a bias to the majority class (non-RP class) in the model building. Although the current study is based on TRIPOD type 1b (24), the patient partition and model development were repeated 100 times to focus on the minority class (RP grade 2 or worse class). Moreover, PR curves were calculated with emphasis on the incidence of RP in the minor class (30). For predicting RP after VMAT, Liang et al. conducted a PR curve analysis using a convolution 3D neural network (33); the result was consistent with our dosiomic approach. In our study, the PR–AUC significantly improved when switching from the DVI models to the dosiomic and hybrid models. As shown in Figure 6, regardless of the similarity of $V_{10\,Gy}$, the dosiomic feature maps can distinguish between the RP and non-RP cases based on the non-uniform quantitative information inside the ROIs. The most selected dosiomic feature was "GLCM correlation" at $V_{5\,Gy}$. The linear dependency of the doses on neighboring voxels mainly contributed to this feature (34), indicating that the non-uniform dose distribution between the connected voxels was correlated with RP after SBRT. Moreover, the predictive performances of the dosiomic and hybrid models were equivalent. This is because most of the DVIs implemented in the hybrid models were commonly regarded as dosiomic features representing the dose distribution shape. Hence, it is important to apply dosiomic features to RP prediction, and it can be concluded that the hybrid models,

including DVIs, does not necessarily improve the predictive performance. Furthermore, our dosiomic approach can be applied to radiation treatment planning. In general, DVIs are used as indicators of RP. In this study, we found that texture features derived from GLCM, representing information regarding the heterogeneity of dose distribution, would be better indicators for RP than DVIs. Although our dosiomic study did not directly reveal clinical application details, the finding can potentially be applied to treatment plan optimization. In addition to the conventional DVI-based optimization, dosiomic-based optimization that can analyze the quantitative information inside the ROI may contribute to reducing the incidence of RP after SBRT.

Multiple DVIs are often considered as dose constraints for reducing the risk of developing RP after SBRT (7–12). Matsuo et al. analyzed 74 patients who underwent SBRT based on a dose distribution calculated using an analytical anisotropic algorithm (AAA) (12). They discovered that $V_{25\ Gy}$ and the planning target volume size were significant factors for RP. Moreover, Ryckman et al. analyzed 93 patients who underwent SBRT using a dose-calculation algorithm, including the AAA, pencil beam, and collapsed cone convolution (CCC) (8). They discovered that $V_{20\ Gy}$ and MLD were associated with the development of RP. Although these studies identified significant factors for RP, dose-calculation algorithms included several types. The dose distributions calculated with

393 “type c” algorithms such as the XVMC are more accurate compared with those calculated
394 with “type a” algorithms such as the pencil beam and “type b” algorithms such as AAA
395 and CCC in inhomogeneous regions such as the lungs (35, 36). Therefore, dose
396 distributions calculated with “type a” and “type b” algorithms overestimate the dose in
397 low-density lung regions compared with those with “type c.” Hence, the values of DVIs
398 and dosiomic features can vary depending on the dose calculation algorithm. Therefore,
399 caution is required when extracting dosiomic features based on different dose calculation
400 algorithms.

401 Several limitations are presented in this study, including the following. First, this
402 study was based on retrospective analysis. Although this study included patients who
403 underwent SBRT at three institutions, they might contain biases associated with treatment
404 protocols, such as planning policy and dose constraints in each institution. Moreover,
405 because patients with multiple tumors were excluded, our dosiomic approach can be
406 applied to patients with a single tumor. Nevertheless, accurate models for predicting RP
407 were developed. Second, various fractionation schemes were included in this study. In
408 this case, treatment expectations are typically quantified using equivalent doses of 2 Gy
409 fractions (37). However, in our analysis, no treatment plans violated the dose–volume
410 constraints, and no bias was observed in the incidence of RP after SBRT among various

prescribed doses; hence, we did not convert physical doses to biological doses. Third, we included only the dose-related features extracted from the dose distribution. Several studies have shown that the incidence of RP is strongly correlated with other clinical factors (7, 38). However, owing to missing clinical factors, such as performance status and smoking history, we did not include them in this study. These features may improve the prediction of RP after SBRT if included in the dosiomic analysis.

Conclusions

We developed predictive models for RP grade 2 or worse after SBRT using dose-segmented dosiomic features and DVIs. The dosiomic features improved the prediction compared with other DVIs. This novel approach for analyzing dose distribution can enhance the practicality of the precision technique for reducing the risk of RP after SBRT.

Conflict of Interest Statement

We have no financial relationships to disclose.

Acknowledgments

This study was partly supported by the Takeda Science Foundation. We sincerely

429 appreciate all the staff members at the Medical Physics Laboratory of Kyoto University
430 Graduate School of Medicine (http://medicalphysics.hs.med.kyoto-u.ac.jp/) for their
431 significant technical support and valuable comments regarding this study.

References

1. Nagata Y, Hiraoka M, Shibata T, et al. Prospective trial of stereotactic body radiation therapy for both operable and inoperable T1N0M0 non-small cell lung cancer: Japan Clinical Oncology Group Study JCOG0403. *Int J Radiat Oncol Biol Phys*. 2018;93(5):989–96.
2. Baumann P, Nyman J, Hoyer M, et al. Outcome in a prospective phase II trial of medically inoperable stage I non-small-cell lung cancer patients treated with stereotactic body radiotherapy. *J Clin Oncol*. 2009;27(20):3290–6.
3. Timmerman R, Paulus R, Galvin J, et al. Stereotactic body radiation therapy for inoperable early stage lung cancer. *JAMA*. 2010;303:1070–6.
4. Fakiris AJ, McGarry RC, Yiannoutsos CT, et al. Stereotactic body radiation therapy for early-stage non-small-cell lung carcinoma: four-year results of a prospective phase II study. *Int J Radiat Oncol Biol Phys*. 2009;75:677–82.
5. Baker R, Han G, Sarangkasiri S, et al. Clinical and dosimetric predictors of radiation pneumonitis in a large series of patients treated with stereotactic body radiation therapy to the lung. *Int J Radiat Oncol Biol Phys*. 2013;85(1):190–5.
6. Chang JY, Liu H, Balter P, et al. Clinical outcome and predictors of survival and pneumonitis after stereotactic ablative radiotherapy for stage I non-small cell lung

cancer. *Radiat Oncol*. 2012;7(1):152.

7. Ueki N, Matsuo Y, Togashi Y, et al. Impact of pretreatment interstitial lung disease on radiation pneumonitis and survival after stereotactic body radiation therapy for lung cancer. *J Thorac Oncol.* 2015;10(1):116–25.

8. Ryckman JM, Baine M, Carmicheal J, et al. Correlation of dosimetric factors with the development of symptomatic radiation pneumonitis in stereotactic body radiotherapy. *Radiat Oncol.* 2020;15(1):33.

9. Zhao J, Yorke ED, Li L, et al. Simple factors associated with radiation-induced lung toxicity after stereotactic body radiation therapy of the thorax: a pooled analysis of 88 studies. *Int J Radiat Oncol Biol Phys*. 2016;95(5):1357–66.

10. Guckenberger M, Baier K, Polat B, et al. Dose-response relationship for radiation-induced pneumonitis after pulmonary stereotactic body radiotherapy. *Radiother Oncol.* 2010;97(1):65–70.

11. Barriger RB, Forquer JA, Brabham JG, et al. A dose–volume analysis of radiation pneumonitis in non–small cell lung cancer patients treated with stereotactic body radiation therapy. *Int J Radiat Oncol Biol Phys*. 2012;82(1):457–62.

12. Matsuo Y, Shibuya K, Nakamura M, et al. Dose–volume metrics associated with radiation pneumonitis after stereotactic body radiation therapy for lung cancer. *Int J*

Radiat Oncol Biol Phys. 2012;83(4):e545–e549.

13. Yamashita H, Nakagawa K, Nakamura N, et al. Exceptionally high incidence of symptomatic grade 2-5 radiation pneumonitis after stereotactic radiation therapy for lung tumors. *Radiat Oncol*. 2007;2:21.

14. El Naqa I, Ruan D, Valdes G, et al. Machine learning and modeling: Data, validation, communication challenges. *Med Phys*. 2018;45(10):834–e840.

15. Li Q, Kim J, Balagurunathan Y, et al. CT imaging features associated with recurrence in non-small cell lung cancer patients after stereotactic body radiotherapy. *Radiat Oncol*. 2017;12(1):158.

16. Lafata KJ, Hong JC, Geng R, et al. Association of pre-treatment radiomic features with lung cancer recurrence following stereotactic body radiation therapy. *Phys Med Biol*. 2019;64(2):025007.

17. Huynh E, Coroller TP, Narayan V, et al. CT-based radiomic analysis of stereotactic body radiation therapy patients with lung cancer. *Radiother Oncol.* 2016;120(2):258–66.

18. Kakino R, Nakamura M, Mitsuyoshi T, et al. Application and limitation of radiomics approach to prognostic prediction for lung stereotactic body radiotherapy using breath-hold CT images with random survival forest: A multi-institutional study. *Med*

Phys. 2020;47(9):4634-4643.

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.

20. Rossi L, Bijman R, Schillemans W, et al. Texture analysis of 3D dose distributions for predictive modelling of toxicity rates in radiotherapy. *Radiother Oncol.* 2018; 29(3):548–53.

21. 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.

22. 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.

23. Kimura T, Nagata Y, Eba J, et al. A randomized phase III trial of comparing two dose-fractionations stereotactic body radiotherapy (SBRT) for medically inoperable stage IA non-small cell lung cancer or small lung lesions clinically diagnosed as primary lung cancer: Japan clinical oncology group study JCOG1408 (J-SBRT Trial). *Jpn J Clin Oncol.* 2017;47(3):277-281.

24. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a

multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Br J Surg*. 2015;102(3):148-158.

25. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. 2017;77(21):e104-e107.

26. Lv W, Yuan Q, Wang Q, et al. Radiomics analysis of PET and CT components of PET/CT imaging integrated with clinical parameters: application to prognosis for nasopharyngeal carcinoma. *Mol Imaging Biol.* 2019;21(5):954-964.

27. Wu G, Woodruff HC, Sanduleanu S, et al. Preoperative CT-based radiomics combined with intraoperative frozen section is predictive of invasive adenocarcinoma in pulmonary nodules: a multicenter study. *Eur Radiol*. 2019;30(5):2680-2691.

28. Ke G, Meng Q, Finely T, et al. LightGBM: a highly efficient gradient boosting decision tree. *Adv Neural Inform Process Syst.* 2017;30:3149-3157.

29. Zhang J, Mucs D, Norinder U, Svensson F. LightGBM: an effective and scalable algorithm for prediction of chemical toxicity-application to the tox21 and mutagenicity data sets. *J Chem Inf Model*. 2019;59(10):4150–4158.

30. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One*.

2015;10(3): e0118432.

31. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. *Information Processing & Management*. 2009;45:427–37.

32. R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation Statistical Computing; 2016.

33. Liang B, Tian Y, Chen X, et al. Prediction of radiation pneumonitis with dose distribution: a convolutional neural network (CNN) based model. *Front Oncol*. 2020;9:1500.

34. Chaddad A, Desrosiers C, Bouridane A, Toews M, Hassan L, Tanougast T. Multi texture analysis of colorectal cancer continuum using multispectral imagery. *PLoS One*. 2016;11(2):e0149893.

35. Knöös T, Wieslander E, Cozzi L, et al. Comparison of dose calculation algorithms for treatment planning in external photon beam therapy for clinical situations. *Phys Med Biol*. 2006;51(22):5785–807.

36. Ojala JJ, Kapanen MK, Hyödynmaa SJ, Wigren TK, Pitkänen MA. Performance of dose calculation algorithms from three generations in lung SBRT: comparison with full Monte Carlo-based dose distributions. *J Appl Clin Med Phys*. 2014;15(2):4–18.

37. Jones B, Dale RG, Deehan C, Hopkins KI, Morgan DA. The role of biologically

effective dose (BED) in clinical oncology. *Clin Oncol (R Coll Radiol)*. 2001; 13:71–81.

38. Kong FM, Wang S. Nondosimetric risk factors for radiation-induced lung toxicity. *Semin Radiat Oncol*. 2015; 25(2):100–9.