European Radiology (2022) 32:8716–8725
https://doi.org/10.1007/s00330-022-08869-2

IMAGING INFORMATICS AND ARTIFICIAL INTELLIGENCE

Importance of CT image normalization in radiomics analysis: prediction of 3-year recurrence-free survival in non-small cell lung cancer

Doohyun Park[1] · Daejoong Oh[1,2] · MyungHoon Lee[2] · Shin Yup Lee[3,4] · Kyung Min Shin[5] · Johnson SG Jun[2] · Dosik Hwang[1,6,7,8]

Received: 8 February 2022 / Revised: 9 April 2022 / Accepted: 9 May 2022 / Published online: 31 May 2022

Abstract

Objectives To analyze whether CT image normalization can improve 3-year recurrence-free survival (RFS) prediction performance in patients with non-small cell lung cancer (NSCLC) relative to the use of unnormalized CT images.

Methods A total of 106 patients with NSCLC were included in the training set. For each patient, 851 radiomic features were extracted from the normalized and the unnormalized CT images, respectively. After the feature selection, random forest models were constructed with selected radiomic features and clinical features. The models were then externally validated in the test set consisting of 79 patients with NSCLC.

Results The model using normalized CT images yielded better performance than the model using unnormalized CT images (with an area under the receiver operating characteristic curve of 0.802 vs 0.702, $p = 0.01$), with the model performing especially well among patients with adenocarcinoma (with an area under the receiver operating characteristic curve of 0.880 vs 0.720, $p < 0.01$).

Conclusions CT image normalization may improve prediction performance among patients with NSCLC, especially for patients with adenocarcinoma.

Key Points

- *After CT image normalization, more radiomic features were able to be identified.*
- *Prognostic performance in patients was improved significantly after CT image normalization compared with before the CT image normalization.*
- *The improvement in prognostic performance following CT image normalization was superior in patients with adenocarcinoma.*

Keywords Radiomics · Prognosis · Computed tomography · Non-small cell lung cancer

Abbreviations

AC	Adenocarcinoma
AUC	Area under the receiver operating characteristic curve
CT	Computed tomography
GGO	Ground-glass opacity
GLCM	Gray level co-occurrence matrix
GLSZM	Gray level size zone matrix

✉ Dosik Hwang
dosik.hwang@yonsei.ac.kr

[1] School of Electrical and Electronic Engineering, Yonsei University, Seoul, Republic of Korea

[2] D&P BIOTECH Inc., Seoul, Republic of Korea

[3] Department of Internal Medicine, School of Medicine, Kyungpook National University, Daegu, Republic of Korea

[4] Lung Cancer Center, Kyungpook National University Chilgok Hospital, Daegu, Republic of Korea

[5] Department of Radiology, School of Medicine, Kyungpook National University, Kyungpook National University Chilgok Hospital, Daegu, Republic of Korea

[6] Center for Healthcare Robotics, Korea Institute of Science and Technology, 5, Hwarang-ro 14-gil, Seongbuk-gu, Seoul 02792, Republic of Korea

[7] Department of Oral and Maxillofacial Radiology, Yonsei University College of Dentistry, Seoul, Republic of Korea

[8] Department of Radiology and Center for Clinical Imaging Data Science (CCIDS), Yonsei University College of Medicine, Seoul, Republic of Korea

HR	Hazard ratio
NSCLC	Non-small cell lung cancer
RF	Random forest
SqCC	Squamous cell carcinoma
TNM	Tumor-node-metastasis

Introduction

Worldwide, lung cancer is a leading cause of cancer-related death [1]. Despite complete resection as a potentially curative treatment in early-stage non-small cell lung cancer (NSCLC), 30–55% of patients experience recurrence and die of the disease [2]. Because even patients who are in the same pathologic stage can have different risks of recurrence and death [3], it is difficult to make accurate prognoses and to ensure proper selection of candidates who are suitable for post-operative adjuvant therapy. Although pathologic stage is the most important predictor of prognosis after surgical resection of NSCLC [4, 5], it has been shown that tumor morphology could also be related to patient prognosis [6, 7].

Recent studies in radiomics have sought to predict the prognosis using quantitative radiomic features from computed tomography (CT) images [8, 9]. However, since the quality of medical images varies depending on the reconstruction algorithm, the quantitative radiomic features extracted from those images also vary [10–14]. Especially for the CT images, the radiomic features vary depending on the acquisition parameters such as voxel size and reconstruction kernel [15–20]. Shafiq-ul-Hassan et al [15] demonstrated that the radiomic features are dependent on the voxel size and that they can be stabilized following feature normalization. In addition, Choe et al [17] showed that the radiomic features are highly dependent on the reconstruction kernel of the CT image, especially for texture and wavelet features. Radiomics studies should therefore be performed using CT images obtained with the same acquisition parameters. However, since it is difficult to obtain CT images using the same acquisition parameters in a retrospective study, it is also difficult to show the importance of normalization prognostically.

Thus, the purpose of this study was to evaluate whether CT image normalization can improve prognostic prediction performance.

Materials and methods

Patient population

We collected data from 320 lung cancer patients who underwent surgical resection between March 2011 and September 2016. None of these patients had received prior radiotherapy or chemotherapy. Among them, 26 patients who had clinical missing data and one patient who did not have images in digital imaging and communications in medicine format were removed. In addition, we used two inclusion criteria: (1) early-stage NSCLC (tumor-node-metastasis [TNM] stage ≤ II), which comprises 88% (280/320) of our collected data, and (2) major histological types, namely adenocarcinoma (AC) and squamous cell carcinoma (SqCC), which comprise 95% (303/320) of our collected data [21]. As a result, a total of 47 patients were excluded. Lastly, 61 patients who had incomplete 3-year follow-up were removed to predict the 3-year recurrence or death. A total of 185 patients were ultimately included in our study. The training set was made up of 106 patients whose CT images were obtained from referring hospitals and the test set was made up of 79 patients whose CT images were externally obtained from Kyungpook National University Chilgok Hospital (KNUCH). In this study, we defined time to recurrence as the interval from surgery to the first evidence of disease recurrence or last evaluation, while duration of cancer-specific survival was calculated from the date of surgery until the date of cancer-related death or final follow-up. Data flow diagram and characteristics of patients are shown in Fig. 1 and Table 1, respectively.

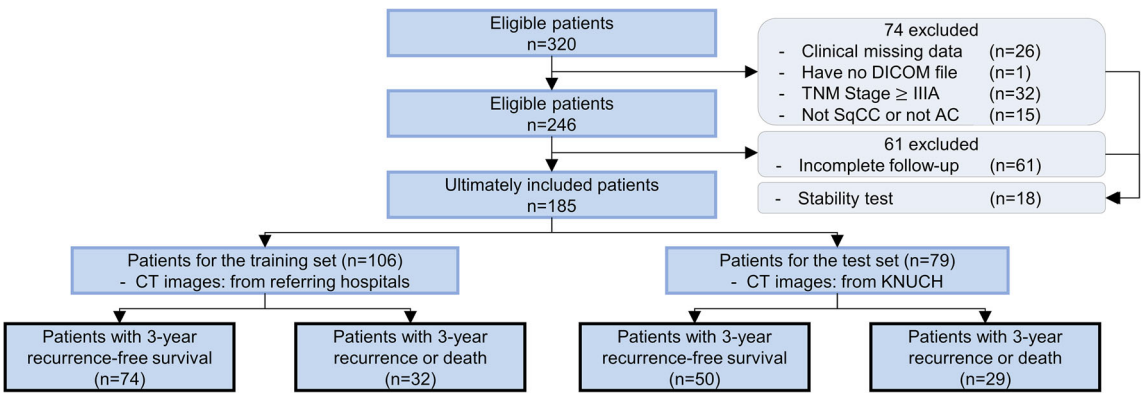

Fig. 1 Data flow diagram of patients. DICOM, digital imaging and communications in medicine; SqCC, squamous cell carcinoma; AC, adenocarcinoma; KNUCH, Kyungpook National University Chilgok Hospital

Table 1 Demographic and pathological characteristics

Characteristic	Training set (n = 106)	Test set (n = 79)	p value*
Mean age (year)†	65.5 ± 8.1	65.4 ± 10.1	0.93
Sex			0.33
Men	69 (65%)	45 (57%)	
Women	37 (35%)	34 (43%)	
Smoking history			0.73
Never smokers	39 (37%)	32 (41%)	
Pack-years of ever smokers†	33.9 ± 19.4	38.1 ± 22.6	
TNM stage			0.14
IA	43 (41%)	41 (52%)	
IB	34 (32%)	20 (25%)	
IIA	23 (22%)	11 (14%)	
IIB	6 (5%)	7 (9%)	
Histology			0.15
Adenocarcinoma	80 (75%)	51 (65%)	
Squamous cell carcinoma	26 (25%)	28 (35%)	
Recurrence			0.44
3-year recurrence-free survival	74 (70%)	50 (63%)	
3-year recurrence or death	32 (30%)	29 (37%)	

*Calculated by the chi-squared test and t test

†Mean ± standard deviation

If standard deviation is not specified, data represents the number of patients

CT acquisition parameters

All CT scans were obtained with intravenous iodine contrast media using various scanners and reconstruction kernels. The CT protocol was as follows: 100–130 kVp, 72–679 mAs, 0.70–5.00-mm slice thickness, 0.51–0.88 mm pixel size. More detailed information is available in Supplement A.

CT image normalization

Our proposed radiomics process to predict the patient's prognosis is shown in Fig. 2. First, we normalized the voxel sizes to 1-mm iso-voxels through cubic interpolation. Second, we normalized the reconstruction kernels using the algorithm developed by Gallardo-Estrella et al [22]. The CT image normalization was performed using MATLAB R2019a. A detailed explanation of the reconstruction kernel normalization is provided in Supplement B. In this study, we defined the radiomic features obtained from the voxel size and reconstruction kernel–normalized CT images as F_{norm}, and those obtained from the unnormalized CT images as F_{unnorm}.

Segmentation and radiomic feature extraction

After a radiologist (with 15 years of chest radiology experience) assigned the tumor location in CT images, an engineer (with 6 years of CT research experienced) segmented the

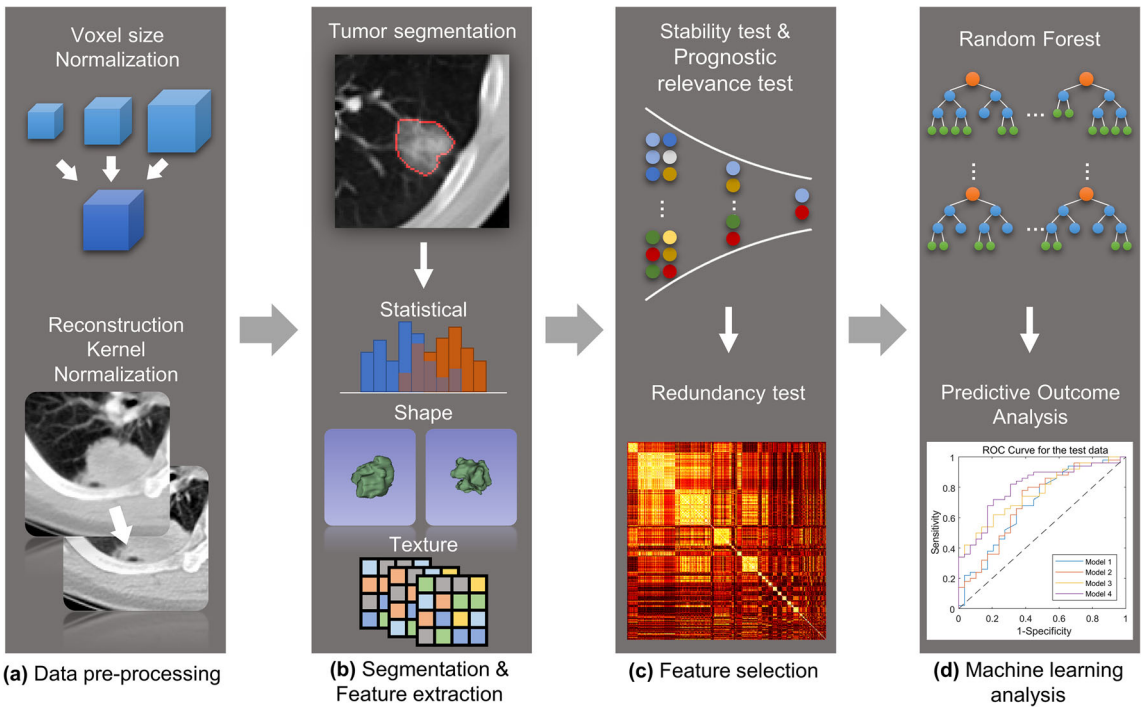

Fig. 2 Proposed radiomics process to predict patient prognosis. **a** Voxel size normalization into 1-mm iso-voxels. Reconstruction kernel normalization into General Electric's BONE kernel. **b** From the region of segmented tumor, a total of 851 radiomic features were extracted. **c** Stability and Wilcoxon rank-sum tests were performed to select the prognosis-relevant radiomic features. After that, a correlation test was performed to eliminate redundant features. **d** Using the selected features, a random forest model was constructed to predict the 3-year recurrence-free survival rate

tumors semi-automatically using the Watershed algorithm in 3D Slicer 4.9.0 [23]. Overall, 851 radiomic features including first-order, shape, gray level dependence matrix, gray level run length matrix, gray level size zone matrix (GLSZM), neighborhood gray tone difference matrix, gray level co-occurrence matrix (GLCM), and wavelet features were extracted by PyRadiomics from the segmented region [24]. The extracted features are described in Supplement C.

Feature selection—stability test

Radiomic features are calculated from the region of segmented tumor. However, one problem is that the segmented region depends on which software tool is used and on the proficiency of radiologists [25–28]. If a radiomic feature value is highly dependent on how and who segmented the tumor region, it is not suitable as a prognostic factor. Therefore, we selected radiomic features that do not vary significantly within the different regions of segmented tumor according to the stability test. The stability test was performed using an open data set that includes two CT scans taken within 15 min [29] and the data of the 18 excluded patients. For the excluded patient data, we manually segmented the tumor twice and then calculated radiomic features within both of the regions of segmented tumor. After that, we selected the features in each dataset with symmetric mean absolute percentage errors of less than 2.5% (Supplement D). The stability test was performed using MATLAB R2019a.

Feature selection—prognostic relevance test

There are various feature selection methods used to find prognostically relevant features, including filter, wrapper, and embedded methods. Parmer et al [30] showed that a prognostic model trained using features that were selected by the Wilcoxon rank-sum test showed the best performance in predicting 2-year survival rates of patients with NSCLC [31]. Therefore, we used the Wilcoxon rank-sum test to select prognostically relevant features.

The TNM stage is one of the most significant factors in predicting the prognosis of patients [32]. Also, one of the goals of this study is to improve performance of the prognostic model by adding radiomic features to the clinicopathological features. Therefore, in order to select non-redundant features with the TNM stage, we performed prognostic relevance test on patients with TNM stage I (n = 77). If the prognostic relevance test is performed on all patients, model performance could deteriorate due to overfitting because radiomic features correlated with the TNM stage may be selected. The prognostic relevance test was performed using MATLAB R2019a.

Feature selection—redundancy test

We eliminated redundant features to prevent overfitting by our proposed method. The detailed algorithm is shown in Supplement E.

Prognostic model

We used a random forest (RF) algorithm to construct a 3-year RFS prediction model with clinicopathological features and the selected radiomic features [33]. The prognostic model construction was performed using MATLAB R2019a. The clinicopathological features used for the model training were TNM stage, age, and sex, which showed the highest 5-fold cross-validation area under the receiver operating characteristic curve (AUC) for all combinations of five clinicopathological features (TNM stage, age, sex, pack-year, and histology).

In order to prevent overfitting, we bound the patients with TNM stages IIA and IIB as TNM stage II, and we augmented the training data 10 times. A detailed explanation is given in Supplement F. The RF parameters were determined by the nested 5-fold cross-validation.

Statistical analysis

We used the following commercially available software: data pre-processing and RF model construction (MATLAB. (2019). 9.7.0.1190202 (R2019a): The MathWorks Inc.), statistical analysis (R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, URL https://www.R-project.org/), tumor segmentation (3DSlicer 4.9.0, https://slicer.readthedocs.io/), and radiomic feature extraction (PyRadiomics, https://pyradiomics.readthedocs.io/) [23, 24]. The p values of the AUCs were calculated using Delong's method [34]. In this study, p values lower than 0.05 were considered statistically significant. We stratified the test set into low-risk and high-risk groups. Thresholds for stratifying the two groups were set as median values of output of the RF models for the training set. We calculated the hazard ratios (HRs) by using the Cox proportional hazards model, and plotted Kaplan-Meier curves using the R 3.6.2 [35, 36]. Statistical significance between low-risk and high-risk groups was calculated by log-rank test [37].

In this study, because we selected the radiomic features that satisfied the stability, prognostic relevance, and redundancy tests, one problem was that only one feature was selected for F_{norm}. Accordingly, we lowered the statistical significance criterion to $p < 0.10$ for the prognostic relevance test in order to identify more radiomic features for the prognostic model construction.

Results

Selected features

Among the extracted 851 radiomic features, 78 satisfied the stability test. Among those 78 features, four satisfied the prognostic relevance test for both of F_{unnorm} and F_{norm}. Lastly, after the redundancy test, 1 and 3 features were ultimately selected for F_{unnorm} and F_{norm}, respectively. In F_{unnorm}, waveletHLL - GLCM - InverseDifferenceNormalized (f_{IDN}) was selected. In F_{norm}, Original-FirstOrder-90Percentile (f_{90p}), Original-GLSZM-SmallAreaEmphasis (f_{SAE}), and the f_{IDN} were selected. Histograms for the selected radiomic features are shown in Fig. 3.

Prediction performance of the RF models—AUC

In this study, we used the RF as the prognostic model, and the experimental results described in this section are summarized in Table 2 and Fig. 4. In this study, we defined the RF model trained by three clinicopathological features as RF_{clinic}. The test AUC (± standard deviation, p value) of the RF_{clinic} was 0.690 (0.626–0.754, $p = 0.029$, compared with RF_{norm}). We analyzed the importance of CT image normalization with two comparisons. First, we defined the RF models trained by the selected radiomic features in F_{unnorm} and F_{norm}, respectively, with three clinicopathological features as $RF_{unnorm1}$ and RF_{norm}. The test AUCs for $RF_{unnorm1}$ and RF_{norm} were 0.702 (0.639–0.765, $p = 0.010$, compared with RF_{norm}) and 0.802 (0.752–0.853, reference), respectively. Second, we defined the RF models trained by the f_{IDN}, f_{90p}, and f_{SAE} in F_{unnorm} with three clinicopathological features as $RF_{unnorm2}$. The $RF_{unnorm2}$ had lower AUC of 0.767 (0.714–0.820, $p = 0.208$, compared with RF_{norm}) compared to the AUC of RF_{norm}.

The efficiency of the CT image normalization was superior to that of patients with AC. The test AUCs of RF_{clinic}, $RF_{unnorm1}$, $RF_{unnorm2}$, and RF_{norm} for the patients with AC were 0.723 (0.640–0.806, $p = 0.038$, compared with RF_{norm}), 0.720 (0.632–0.808, $p = 0.008$, compared with RF_{norm}), 0.834 (0.776–0.892, $p = 0.176$, compared with RF_{norm}), and 0.880 (0.829–0.931, reference), respectively. Figure 5 shows the receiver operating characteristic (ROC) curves for the 3-year RFS predictions with RF models.

Prediction performance of the RF models—Kaplan-Meier analysis

We performed Kaplan-Meier analysis by stratifying the test set into the low-risk and the high-risk groups. The Kaplan-Meier curves are shown in Fig. 6, and the HRs described in this section are shown in Table 2. The HRs (95% CI, p value) of RF_{clinic}, $RF_{unnorm1}$, $RF_{unnorm2}$, and RF_{norm} were 2.638 (1.271–5.474, $p < 0.01$), 3.471 (1.635–7.369, $p < 0.001$), 2.928 (1.381–6.208, $p < 0.01$), and 5.950 (2.264–15.635, $p < 0.0001$), respectively. Although all the models could stratify the low-risk and high-risk groups significantly, the results were more significant following the CT image normalization.

Similar to the AUC, the efficiency of the CT image normalization was superior to that of patients with AC. The HRs of the RF_{clinic}, $RF_{unnorm1}$, $RF_{unnorm2}$, and RF_{norm} for the patients with AC were 3.290 (1.101–9.831, $p<0.05$), 2.749 (0.920–8.210, $p = 0.059$), 4.100 (1.419–11.845, $p < 0.01$), and 10.296 (2.296–46.175, $p < 0.001$), respectively.

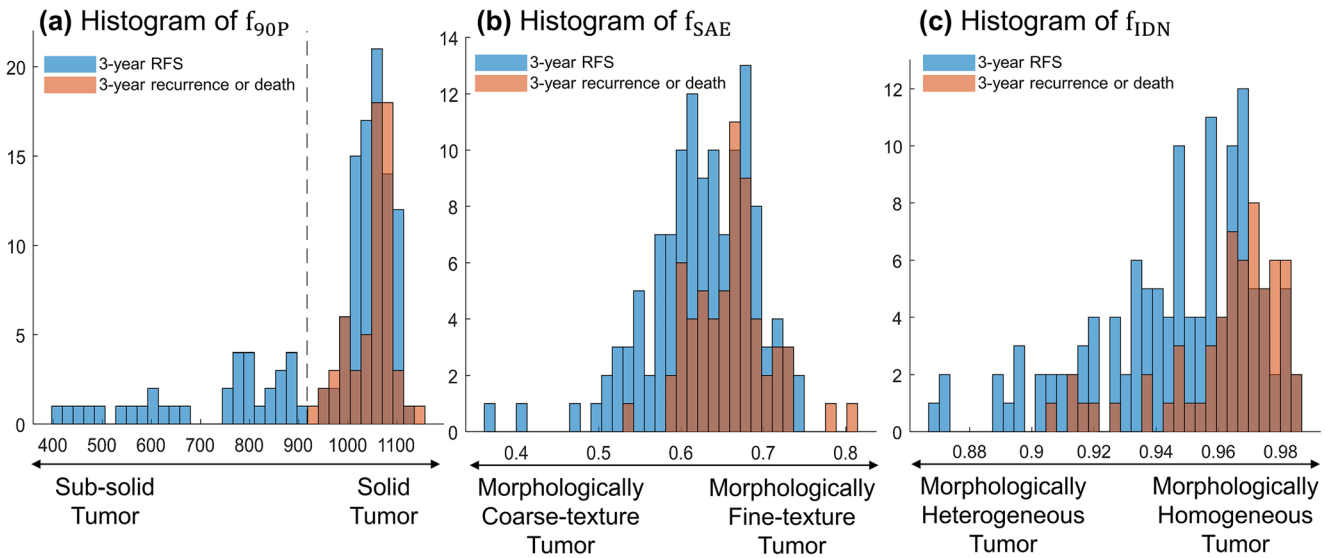

Fig. 3 Histograms of selected three features with normalized CT image. The blue bar represents patients with 3-year recurrence-free survival (RFS). The red bar represents patients with 3-year recurrence or death. In **a**, the black dashed line is plotted on the values of 917

Table 2 Prediction performance of 3-year recurrence-free survival for the test set according to different random forest (RF) models

Model	Patients with adenocarcinoma and squamous cell carcinoma (n = 79)				Patients with adenocarcinoma (n = 51)			
	AUC (± standard deviation)	p value*	HR (95% CI)	p value†	AUC (± standard deviation)	p value*	HR (HR, 95% CI)	p value†
RF_{clinic}	0.690 (0.626–0.754)	0.029	2.638 (1.271–5.474)	< 0.01	0.723 (0.640–0.806)	0.038	3.290 (1.101–9.831)	< 0.05
$RF_{unnorm1}$	0.702 (0.639–0.765)	0.010	3.471 (1.635–7.369)	< 0.001	0.720 (0.632–0.808)	0.008	2.749 (0.920–8.210)	0.059
$RF_{unnorm2}$	0.767 (0.714–0.820)	0.208	2.928 (1.381–6.208)	< 0.01	0.834 (0.776–0.892)	0.176	4.100 (1.419–11.845)	< 0.01
RF_{norm}	0.802 (0.752–0.853)	Reference	5.950 (2.264–15.635)	< 0.0001	0.880 (0.829–0.931)	Reference	10.296 (2.296–46.175)	< 0.001

* p values were calculated using Delong's method

† p values were calculated by log-rank test

AUC area under the receiver operating characteristic curve, *HR* hazard ratio, *CI* confidence interval

Discussion

In our study, the prognostic model using the normalized CT images showed promising results for predicting the prognosis of patients with NSCLC. By normalizing the CT images, we were able to select three radiomic features using our proposed feature selection process. The RF model trained using the selected radiomic features with clinicopathological features

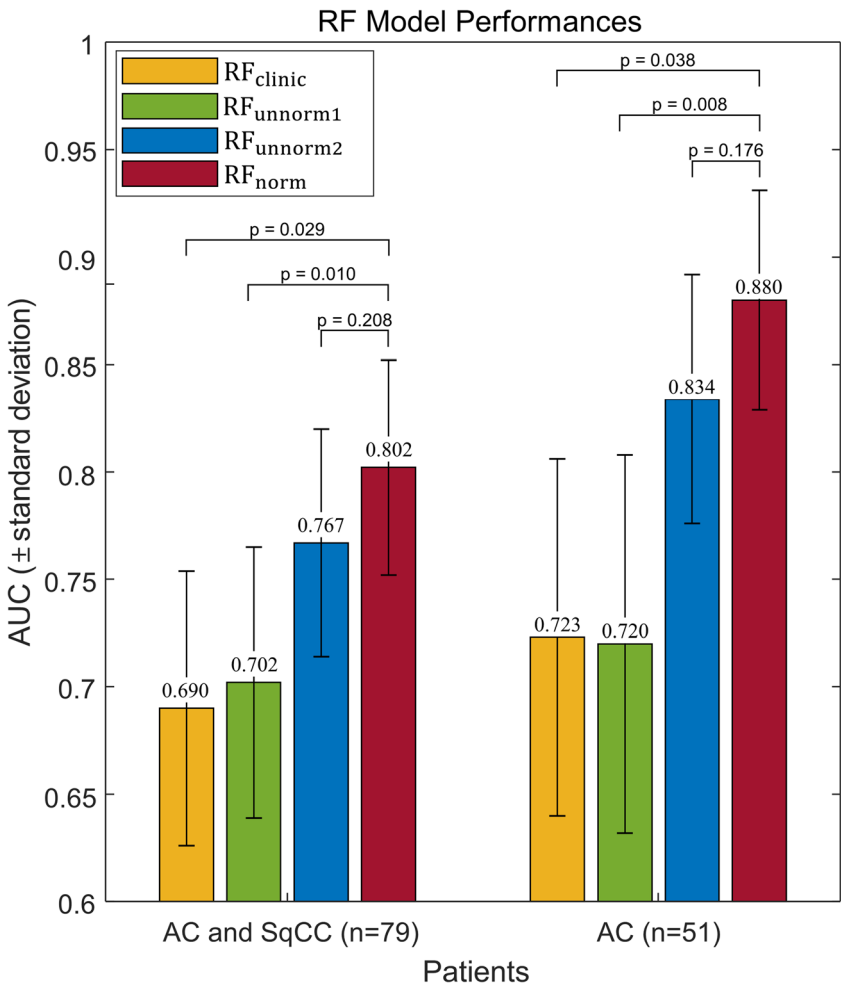

Fig. 4 Random forest (RF) model performances for the prediction of 3-year recurrence-free survival. AUC, area under the receiver operating characteristic curve; AC, adenocarcinoma; SqCC, squamous cell carcinoma. RF_{clinic}, RF model trained by clinicopathological features. $RF_{unnorm1}$, RF model trained by the selected radiomic features in F_{unnorm} with clinicopathological features. $RF_{unnorm2}$, RF model trained by the f_{IDN}, f_{90p}, and f_{SAE} in F_{unnorm} with clinicopathological features. RF_{norm}, RF model trained by the f_{IDN}, f_{90p}, and f_{SAE} in F_{norm} with clinicopathological features. The features f_{IDN}, f_{90p}, and f_{SAE} are the selected radiomic features after the CT image normalization. The F_{unnorm} and the F_{norm} are the radiomic features obtained from unnormalized and normalized CT images, respectively

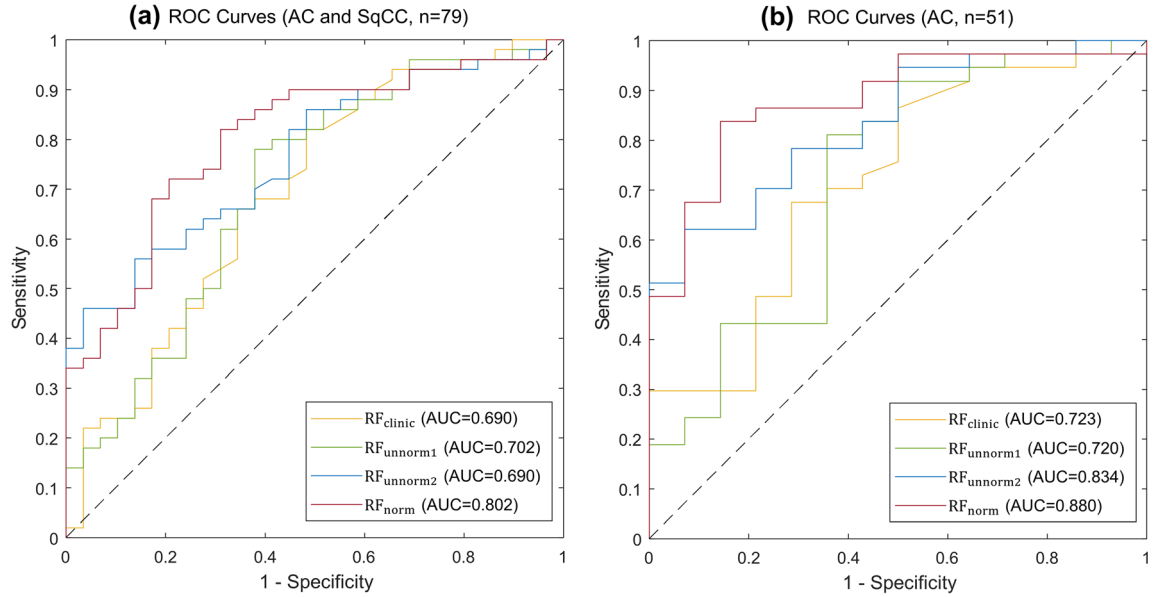

Fig. 5 Receiver operating characteristic (ROC) curves for the prediction of 3-year recurrence-free survival using random forest (RF) models. **a** Patients with adenocarcinoma (AC) and squamous cell carcinoma (SqCC). **b** Patients with AC. AUC, area under the receiver operating characteristic curve. RF_{clinic}, RF model trained by clinicopathological features. $RF_{unnorm1}$, RF model trained by the selected radiomic features in F_{unnorm} with clinicopathological features. $RF_{unnorm2}$, RF model trained by the f_{IDN}, f_{90p}, and f_{SAE} in F_{unnorm} with clinicopathological features. RF_{norm}, RF model trained by the f_{IDN}, f_{90p}, and f_{SAE} in F_{norm} with clinicopathological features. The features f_{IDN}, f_{90p}, and f_{SAE} are the selected radiomic features after the CT image normalization. The F_{unnorm} and the F_{norm} are the radiomic features obtained from unnormalized and normalized CT images, respectively

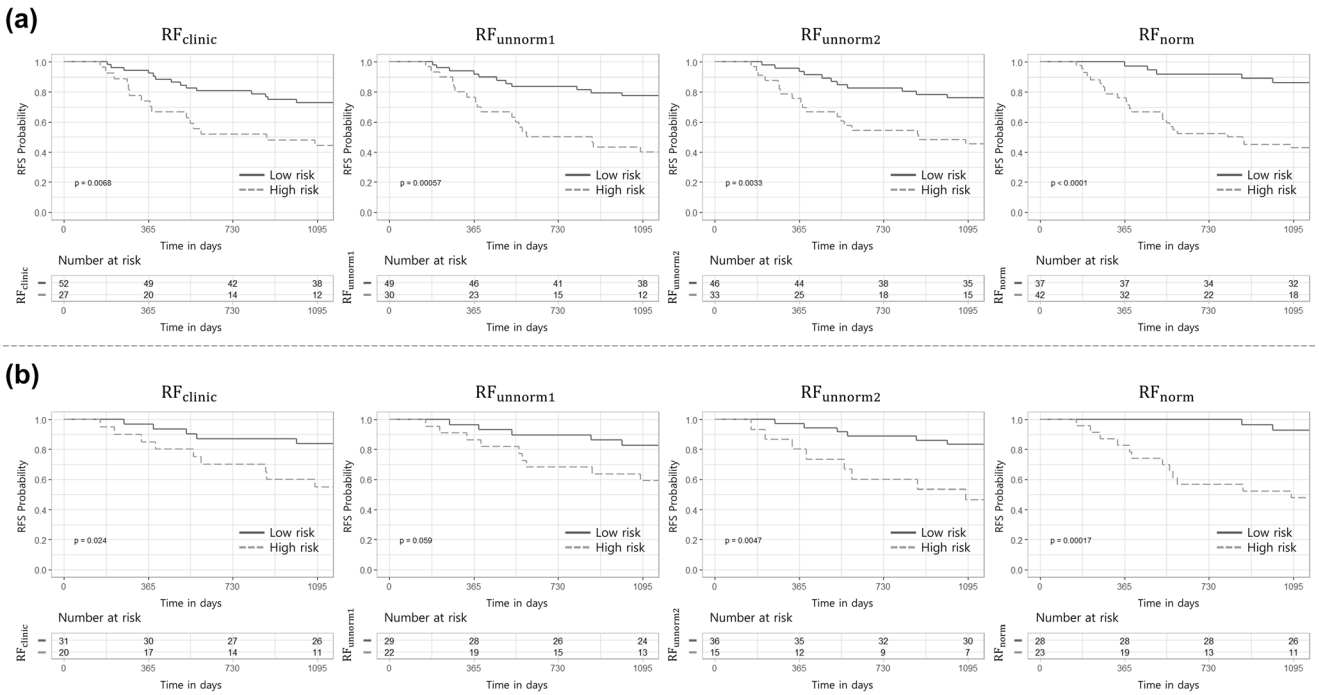

Fig. 6 Kaplan-Meier curves for low-risk and high-risk groups. **a** Patients with adenocarcinoma (AC) and squamous cell carcinoma (SqCC). **b** Patients with AC. RF_{clinic}, RF model trained by clinicopathological features. $RF_{unnorm1}$, RF model trained by the selected radiomic features in F_{unnorm} with clinicopathological features. $RF_{unnorm2}$, RF model trained by the f_{IDN}, f_{90p}, and f_{SAE} in F_{unnorm} with clinicopathological features. RF_{norm}, RF model trained by the f_{IDN}, f_{90p}, and f_{SAE} in F_{norm} with clinicopathological features. The features f_{IDN}, f_{90p}, and f_{SAE} are the selected radiomic features after the CT image normalization. The F_{unnorm} and the F_{norm} are the radiomic features obtained from unnormalized and normalized CT images, respectively. Thresholds for stratifying the two groups were set as median values of output of the RF models for the training set

was able to predict the 3-year RFS of patients with an AUC of 0.802. Our model showed promising result for predicting the 3-year RFS among patients with AC, with an AUC of 0.880. Kaplan-Meier analysis of the stratified low-risk and high-risk groups yielded an HR of 5.950 for patients with AC and SqCC, and an HR of 10.296 for patients with AC.

Our study has three major findings. First, as a result of the proposed feature selection process to select stable, prognostically relevant, and non-redundant features, three radiomic features were identified following CT image normalization, whereas only one radiomic feature was identified prior to normalization. Second, the RF model using the normalized CT images yielded better performance than the model using the unnormalized CT images, with an AUC of 0.702 vs 0.802 ($p < 0.05$), respectively. Third, the performance improvement after CT image normalization was superior in patients with AC, with an AUC of 0.720 vs 0.880 ($p < 0.01$), respectively.

Most of the radiomics studies select radiomic features using a specific method such as the filter, wrapper, or embedded method [9]. However, radiomic features are calculated in a segmented region of tumor, and the segmented region depends on what software tool is used and on the proficiency of radiologists [25–28]. We therefore selected radiomic features that do not vary significantly when the region of segmented tumor is slightly changed. As a result, only 78 (< 10%) out of a total of 851 radiomic features satisfied the stability test in this study. Among the features that satisfied the stability test, three radiomic features were selected from the normalized CT images following the prognostic relevance and redundancy tests. After the CT image normalization, the p values of prognostic relevance test for the selected three features (f_{90p}, f_{SAE}, and f_{IDN}) for the training and test set ($n = 185$) were lowered from 0.003 to < 0.001, from 0.012 to 0.001, and from 0.116 (×10e −3) to 0.003 (×10e−3), respectively. This implies that the value of radiomic features can more accurately represent the pathological characteristics of a tumor after the CT image normalization.

By using the selected radiomic features with clinicopathological features in F_{norm}, the RF_{norm} was able to predict the 3-year RFS of patients with an AUC of 0.802. This is significantly improved performance compared to the RF_{clinic} (AUC of 0.690). The improvement in performance is because the radiomic features can quantitatively calculate tumor phenotypes that cannot be observed in clinical features.

The RF_{norm} was also superior to the $RF_{unnorm1}$ and the $RF_{unnorm2}$, respectively. This result implies two things. First, the performance of the prognostic model can be improved by using more radiomic features, which can only be selected after the CT image normalization. Second, even if the same features were used for the prognostic model, those features became much more effective in terms of prognosis prediction following the CT image normalization.

The model's performance showed particular improvement among patients with AC. With the subgroup of patients with AC, the AUCs of the $RF_{unnorm1}$, $RF_{unnorm2}$, and RF_{norm} were 0.720, 0.834, and 0.880, respectively. This is because f_{90p}, one of the selected radiomic features, is particularly efficient at predicting the prognosis of patients with AC. The mathematical meaning of the f_{90p} is the 90th percentile value of the segmented region [38]. For the F_{norm}, we found that there was no recurrence for 18% (34/185) of patients ($f_{90p} \leq 917$), and all those cases were patients with AC. In this study, the f_{90p} was only able to be selected following the CT image normalization. Therefore, the performance of RF_{norm} might be greatly improved, as it more accurately predicts the prognosis of patients with AC.

Clinically, the f_{90p} is related to the degree of ground-glass opacity (GGO) in the volume of tumor. It is well known that patients with lung cancer who had ground-glass nodules are considered to have less invasive tumors and less risk of recurrence [39, 40]. To reproduce these findings, a radiologist (with 15 years of chest radiology experience) classified the degree of GGO into five levels for all the patients following the semantic annotations of Bakr et al [41]. As a result, we found that the lower the f_{90p} value, the greater the proportion of GGO (correlation coefficient: −0.656). In other words, f_{90p} can be used to quantify the proportion of GGO.

The performance improvement of the RF model after CT image normalization also affected by the f_{SAE} and the f_{IDN}. The f_{SAE} and f_{IDN} represent texture and local homogeneity of the tumor, respectively [38, 42]. Since both features are related to the texture of the tumor, the feature values after normalization better reflect the morphological characteristics of the tumor than before normalization.

Radiomics study has a complex process from data collection to model evaluation. Since the processes of the radiomics studies are not consistent with all studies, Lambin et al [43] proposed the Radiomics Quality Score (RQS), which can quantitatively evaluate the quality of the radiomics study. Chen et al [44] analyzed the RQS of 15 radiomics studies, and the average RQS was 29.6%. Also, the RQS of this study was 38.9%, which is the fourth highest among the analyzed studies in [44].

This study demonstrates the importance of the CT image normalization: more radiomic features were able to be identified following CT image normalization, which consequently improved prognostic model performance. In addition, even if the same features had been used for the prognostic model, the prognosis prediction performance would have improved after normalization.

Nevertheless, this study has a few limitations. First, the main limitation of this study is that the performance of the normalization has not been quantitatively verified. Because it is a retrospective study, we could not obtain the reference image used as the normalization target, and thus, we could not

quantitatively compare the normalized CT images with the ground-truth CT images. Second, patients with complete follow-up after 3 years of surgical resection were used in this study. Therefore, it is necessary to be aware that bias may occur in the results due to loss to follow-up patients. Despite these limitations, we found that CT image normalization made the radiomic features more prognostically relevant for patients with NSCLC, which contributes to improving the predictive performance of the prognostic model.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00330-022-08869-2.

Acknowledgements This research was supported by D&P BIOTECH Inc. and partially supported by the Yonsei Signature Research Cluster Program of 2022 (2022-22-0002), the KIST Institutional Program(Project No.2E31051-21-204), the Institute of Information and Communications Technology Planning and Evaluation (IITP) Grant funded by the Korean Government (MSIT) Artificial Intelligence Graduate School Program, Yonsei University (2020-0-01361), and the Graduate School of YONSEI University Research Scholarship Grants in 2018. The authors sincerely thank In Yong Park for his diligent proofreading of this paper.

Funding This research was funded by D&P BIOTECH Inc.

Declarations

Guarantor The scientific guarantor of this publication is Prof. Dosik Hwang.

Conflict of interest The authors of this manuscript declare relationships with the following companies: D&P BIOTECH Inc. Mr. Park, Mr. Oh, Dr. Lee, Dr. Jun, and Dr. Hwang have a patent "METHOD FOR PREDICTING PROGNOSIS IN CANCER PATIENT USING CLINICAL INFORMATION AND RADIOMIC FEATURE" pending. Dr. Shin and Dr. Lee have nothing to disclose.

Statistics and biometry One of the authors has significant statistical expertise.

Informed consent Written informed consent was waived by the Institutional Review Board.

Ethical approval Institutional Review Board approval was obtained.

Methodology
- retrospective
- diagnostic or prognostic study
- multicenter study

References

1. Siegel RL, Miller KD, Jemal A (2019) Cancer statistics, 2019. CA Cancer J Clin 69:7–34. https://doi.org/10.3322/caac.21551
2. Crosbie PA, Shah R, Summers Y, Dive C, Blackhall F (2013) Prognostic and predictive biomarkers in early stage NSCLC: CTCs and serum/plasma markers. Transl Lung Cancer Res 2:382. https://doi.org/10.3978/j.issn.2218-6751.2013.09.02
3. Detterbeck FC, Boffa DJ, Tanoue LT (2009) The new lung cancer staging system. Chest 136:260–271. https://doi.org/10.1378/chest.08-0978
4. Lee SY, Jung DK, Choi JE et al (2017) Functional polymorphisms in PD-L1 gene are associated with the prognosis of patients with early stage non-small cell lung cancer. Gene 599:28–35. https://doi.org/10.1016/j.gene.2016.11.007
5. Lee SY, Jin CC, Choi JE et al (2016) Genetic polymorphisms in glycolytic pathway are associated with the prognosis of patients with early stage non-small cell lung cancer. Sci Rep 6:35603. https://doi.org/10.1038/srep35603
6. Aoki T, Hanamiya M, Uramoto H, Hisaoka M, Yamashita Y, Korogi Y (2012) Adenocarcinomas with predominant ground-glass opacity: correlation of morphology and molecular biomarkers. Radiology 264:590–596. https://doi.org/10.1148/radiol.12111337
7. Lee HY, Lee SW, Lee KS et al (2015) Role of CT and PET imaging in predicting tumor recurrence and survival in patients with lung adenocarcinoma: a comparison with the International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society Classification of Lung Adenocarcinoma. J Thorac Oncol 10:1785–1794. https://doi.org/10.1097/JTO.0000000000000689
8. Aerts HJWL, Velazquez ER, Leijenaar RT et al (2014) Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 5:1–9. https://doi.org/10.1038/ncomms5006
9. Liu Z, Wang S, Dong D et al (2019) The applications of radiomics in precision diagnosis and treatment of oncology: opportunities and challenges. Theranostics 9:1303. https://doi.org/10.7150/thno.30309
10. Oh D, Kim S, Park D et al (2018) Correction of severe beam-hardening artifacts via a high-order linearization function using a prior-image-based parameter selection method. Med Phys 45:4133–4144. https://doi.org/10.1002/mp.13072
11. Kim Y, Oh D, Hwang D (2017) Small-scale noise-like moiré pattern caused by detector sensitivity inhomogeneity in computed tomography. Opt Express 25:27127–27145. https://doi.org/10.1364/OE.25.027127
12. Kim Y, Baek J, Hwang D (2014) Ring artifact correction using detector line-ratios in computed tomography. Opt Express 22:13380–13392. https://doi.org/10.1364/OE.22.013380
13. Eo T, Jun Y, Kim T, Jang J, Lee HJ, Hwang D (2018) KIKI-net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images. Magn Reson Med 80:2188–2201. https://doi.org/10.1002/mrm.27201
14. Eo T, Shin H, Jun Y, Kim T, Hwang D (2020) Accelerating Cartesian MRI by domain-transform manifold learning in phase-encoding direction. Med Image Anal 63:101689. https://doi.org/10.1016/j.media.2020.101689
15. Shafiq-ul-Hassan M, Latifi K, Zhang G, Ullah G, Gillies R, Moros E (2018) Voxel size and gray level normalization of CT radiomic features in lung cancer. Sci Rep 8:1–9. https://doi.org/10.1038/s41598-018-28895-9
16. Traverso A, Wee L, Dekker A, Gillies R (2018) Repeatability and reproducibility of radiomic features: a systematic review. Int J Radiat Oncol Biol Phys 102:1143–1158. https://doi.org/10.1016/j.ijrobp.2018.05.053
17. Choe J, Lee S, Do K et al (2019) Deep learning–based image conversion of CT reconstruction kernels improves radiomics reproducibility for pulmonary nodules or masses. Radiology 292:365–373. https://doi.org/10.1148/radiol.2019181960
18. Berenguer R, Pastor-Juan MR, Canales-Vázquez J et al (2018) Radiomics of CT features may be nonreproducible and redundant: influence of CT acquisition parameters. Radiology 288:407–415. https://doi.org/10.1148/radiol.2018172361

19. Orlhac F, Frouin F, Nioche C, Ayache N, Buvat I (2019) Validation of a method to compensate multicenter effects affecting CT radiomics. Radiology 291:53–59. https://doi.org/10.1148/radiol.2019182023
20. Park BW, Kim JK, Heo C, Park KJ (2020) Reliability of CT radiomic features reflecting tumour heterogeneity according to image quality and image processing parameters. Sci Rep 10:1–13. https://doi.org/10.1038/s41598-020-60868-9
21. Kawase A, Yoshida J, Ishii G et al (2011) Differences between squamous cell carcinoma and adenocarcinoma of the lung: are adenocarcinoma and squamous cell carcinoma prognostically equal? Jpn J Clin Oncol 42:189–195. https://doi.org/10.1093/jjco/hyr188
22. Gallardo-Estrella L, Lynch DA, Prokop M et al (2016) Normalizing computed tomography data reconstructed with different filter kernels: effect on emphysema quantification. Eur Radiol 26:478–486. https://doi.org/10.1007/s00330-015-3824-y
23. Fedorov A, Beichel R, Kalpathy-Cramer J et al (2012) 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magn Reson Imaging 30:1323–1341. https://doi.org/10.1016/j.mri.2012.05.001
24. Griethuysen JJ, Fedorov A, Parmar C et al (2017) Computational radiomics system to decode the radiographic phenotype. Cancer Res 77:e104–e107. https://doi.org/10.1158/0008-5472.CAN-17-0339
25. Parmar C, Velazquez ER, Leijenaar R et al (2014) Robust radiomics feature quantification using semiautomatic volumetric segmentation. PLoS One 9:e102107. https://doi.org/10.1371/journal.pone.0102107
26. Owens CA, Peterson CB, Tang C et al (2018) Lung tumor segmentation methods: impact on the uncertainty of radiomics features for non-small cell lung cancer. PLoS One 13:e0205003. https://doi.org/10.1371/journal.pone.0205003
27. Kim S, Bae WC, Masuda K, Chung CB, Hwang D (2018) Fine-grain segmentation of the intervertebral discs from MR spine images using deep convolutional neural networks: BSU-Net. Appl Sci Basel 8:1656. https://doi.org/10.3390/app8091656
28. Kim S, Bae WC, Masuda K, Chung CB, Hwang D (2018) Semi-automatic segmentation of vertebral bodies in MR images of human lumbar spines. Appl Sci Basel 8:1586. https://doi.org/10.3390/app8091586
29. Zhao B, James LP, Moskowitz CS et al (2009) Evaluating variability in tumor measurements from same-day repeat CT scans of patients with non–small cell lung cancer. Radiology 252:263–272. https://doi.org/10.1148/radiol.2522081593
30. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL (2015) Machine learning methods for quantitative radiomic biomarkers. Sci Rep 5:13087. https://doi.org/10.1038/srep13087
31. Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. Ann Math Stat 18:50–60
32. Woodard GA, Jones KD, Jablons DM (2016) Lung cancer staging and prognosis. Lung Cancer 170:47–75. https://doi.org/10.1007/978-3-319-40389-2_3
33. Breiman L (2001) Random forests. Mach Learn 45:5–32. https://doi.org/10.1023/A:1010933404324
34. DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44:837–845. https://doi.org/10.2307/2531595
35. Kaplan EL, Meier P (1958) Nonparametric estimation from incomplete observations. J Am Stat Assoc 53:457–481. https://doi.org/10.1080/01621459.1958.10501452
36. Cox DR (1972) Regression models and life-tables. J R Stat Soc Ser B Stat Methodol 34:187–202. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x
37. Mantel N (1966) Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemother Rep 50:163–170
38. Clausi DA (2002) An analysis of co-occurrence texture statistics as a function of grey level quantization. Can Aeronaut Space J 28:45–62. https://doi.org/10.5589/m02-004
39. Moon Y, Sung SW, Moon SW, Park JK (2016) Risk factors for recurrence after sublobar resection in patients with small (2 cm or less) non-small cell lung cancer presenting as a solid-predominant tumor on chest computed tomography. J Thorac Dis 8:2018. https://doi.org/10.21037/jtd.2016.07.90
40. Hattori A, Matsunaga T, Takamochi K, Oh S, Suzuki K (2017) Importance of ground glass opacity component in clinical stage IA radiologic invasive lung cancer. Ann Thorac Surg 104:313–320. https://doi.org/10.1016/j.athoracsur.2017.01.076
41. Bakr S, Gevaert O, Echegaray S et al (2018) A radiogenomic dataset of non-small cell lung cancer. Sci Data 5:1–9. https://doi.org/10.1038/sdata.2018.202
42. Zwanenburg A, Leger S, Vallières M, Löck S (2016) Image biomarker standardisation initiative. arXiv 1612:07003. https://doi.org/10.48550/arXiv.1612.07003
43. Lambin P, Leijenaar RT, Deist TM et al (2017) Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 14:749–762. https://doi.org/10.1038/nrclinonc.2017.141
44. Chen Q, Zhang L, Mo X et al (2021) Current status and quality of radiomic studies for predicting immunotherapy response and outcome in patients with non-small cell lung cancer: a systematic review and meta-analysis. Eur J Nucl Med Mol Imaging 49:345–360. https://doi.org/10.1007/s00259-021-05509-7