

## RESEARCH ARTICLE

# Robust radiogenomics approach to the identification of *EGFR* mutations among patients with NSCLC from three different countries using topologically invariant Betti numbers

Kenta Ninomiya<sup>1</sup>, Hidetaka Arimura<sup>2\*</sup>, Wai Yee Chan<sup>3\*</sup>, Kentaro Tanaka<sup>4</sup>, Shinichi Mizuno<sup>5</sup>, Nadia Fareeda Muhammad Gowdh<sup>3</sup>, Nur Adura Yaakup<sup>3</sup>, Chong-Kin Liam<sup>6</sup>, Chee-Shee Chai<sup>7</sup>, Kwan Hoong Ng<sup>3</sup>

**1** Division of Medical Quantum Science, Department of Health Sciences, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan, **2** Faculty of Medical Sciences, Division of Medical Quantum Science, Department of Health Sciences, Kyushu University, Fukuoka, Japan, **3** Faculty of Medicine, Department of Biomedical Imaging, University of Malaya, Kuala Lumpur, Malaysia, **4** Department of Respiratory Medicine, Kyushu University Hospital, Fukuoka, Japan, **5** Division of Medical Sciences and Technology, Department of Health Sciences, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan, **6** Faculty of Medicine, Department of Medicine, University of Malaya, Kuala Lumpur, Malaysia, **7** Faculty of Medicine and Health Science, Department of Medicine, University Malaysia Sarawak, Kota Samarahan, Sarawak, Malaysia

\* [arimurah@med.kyushu-u.ac.jp](mailto:arimurah@med.kyushu-u.ac.jp) (HA); [waiyec@ummc.edu.my](mailto:waiyec@ummc.edu.my) (WYC)



## OPEN ACCESS

**Citation:** Ninomiya K, Arimura H, Chan WY, Tanaka K, Mizuno S, Muhammad Gowdh NF, et al. (2021) Robust radiogenomics approach to the identification of *EGFR* mutations among patients with NSCLC from three different countries using topologically invariant Betti numbers. PLoS ONE 16(1): e0244354. <https://doi.org/10.1371/journal.pone.0244354>

**Editor:** Francesco Bianconi, Università degli Studi di Perugia, ITALY

**Received:** September 27, 2020

**Accepted:** December 9, 2020

**Published:** January 11, 2021

**Copyright:** © 2021 Ninomiya et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its [Supporting information files](#).

**Funding:** This study was supported by a grant from Center for Clinical and Translational Research of Kyushu University Hospital and JSPS KAKENHI Grant Number 20K08084. The funders had no role in study design, data collection and analysis,

## Abstract

### Objectives

To propose a novel robust radiogenomics approach to the identification of epidermal growth factor receptor (*EGFR*) mutations among patients with non-small cell lung cancer (NSCLC) using Betti numbers (BNs).

### Materials and methods

Contrast enhanced computed tomography (CT) images of 194 multi-racial NSCLC patients (79 *EGFR* mutants and 115 wildtypes) were collected from three different countries using 5 manufacturers' scanners with a variety of scanning parameters. Ninety-nine cases obtained from the University of Malaya Medical Centre (UMMC) in Malaysia were used for training and validation procedures. Forty-one cases collected from the Kyushu University Hospital (KUH) in Japan and fifty-four cases obtained from The Cancer Imaging Archive (TCIA) in America were used for a test procedure. Radiomic features were obtained from BN maps, which represent topologically invariant heterogeneous characteristics of lung cancer on CT images, by applying histogram- and texture-based feature computations. A BN-based signature was determined using support vector machine (SVM) models with the best combination of features that maximized a robustness index (RI) which defined a higher total area under receiver operating characteristics curves (AUCs) and lower difference of AUCs

decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

between the training and the validation. The SVM model was built using the signature and optimized in a five-fold cross validation. The BN-based model was compared to conventional original image (OI)- and wavelet-decomposition (WD)-based models with respect to the RI between the validation and the test.

## Results

The BN-based model showed a higher RI of 1.51 compared with the models based on the OI (RI: 1.33) and the WD (RI: 1.29).

## Conclusion

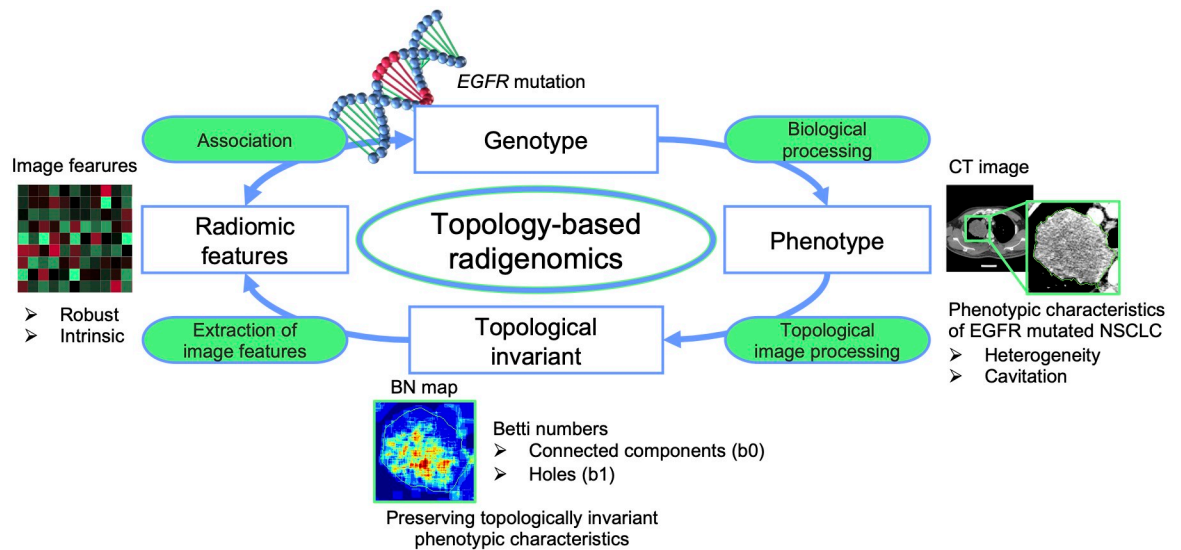
The proposed model showed higher robustness than the conventional models in the identification of *EGFR* mutations among NSCLC patients. The results suggested the robustness of the BN-based approach against variations in image scanner/scanning parameters.

## 1. Introduction

Lung cancer is the leading cause of cancer-related deaths worldwide [1]. Approximately 85% of lung cancer lesions are of the non-small cell lung cancer (NSCLC) subtype [2]. The 5-year survival rates for stages I, II, III, and IV NSCLC are approximately 80, 57, 25, and 5%, respectively [3]. The treatment of patients with late-stage epidermal growth factor receptor (*EGFR*) sensitizing mutations positive NSCLC using tyrosine kinase inhibitors (TKIs) is exemplified as precision medicine which takes into account the individual variability in the genes, environment, and lifestyle of each patient.

In a rapidly developing field of radiomics, researchers have been investigating the associations between medical images and patients' prognostic information (including the *EGFR* mutation status) in a non-invasive manner under the assumption that particular somatic mutations of cancer lead remarkable phenotypes appearing on the medical images (Fig 1) [4–7]. Previous studies demonstrated underlying associations between the *EGFR* mutation status and intra-tumor heterogeneity on computed tomography (CT) images quantified using conventional radiomic features such as shape-, original image (OI)-, wavelet-decomposition (WD)-, and deep learning-based features [8–10]. However, past studies have not investigated the feasibility of their model using datasets with wide variety of imaging parameters or patient populations [8,9], although it was reported that conventional radiomic features can struggle to extract intrinsic image features that are robust to variations in CT scanner or scanning parameters [11,12].

Betti numbers (BNs), which are topological invariant in the homology, has been applied to quantify tumor traits in several medical images such as CT, magnetic resonance, and pathological images [13–17]. The topologically invariants in the BNs indicates unchangeable property of objects under continuous deformation [18]. In two-dimensional images, two types of BNs can be defined, that are zero- and one-dimensional BNs representing the number of connected components ( $b_0$ ) and holes ( $b_1$ ), respectively. In fact, several studies reported that *EGFR* mutated NSCLC appeared to have intra-tumor heterogeneity on contrast enhanced (CE) CT images [19] and hazy areas in tumor regions which cause low-intensity holes on the CT images [20]. Meanwhile, in our previous study, we found that  $b_0$  and  $b_1$  could robustly characterize the intra-tumor heterogeneities and the low-intensity holes associated with prognoses of lung



**Fig 1. Assumption and concept of the present study.** The assumption of this study is that epidermal growth factor receptor (*EGFR*) mutation in non-small cell lung cancer (NSCLC) leads remarkable phenotypes appearing on computed tomography (CT) images and that could be characterized by using Betti numbers (BN) derived from topology theory.

<https://doi.org/10.1371/journal.pone.0244354.g001>

cancer patients, although those features were extracted from the CT images acquired under several scanner/scanning parameters [5]. Therefore, in the present study, we assumed that the BN-based image features could also successfully quantify the intra-tumor heterogeneity and the low-intensity holes related with the *EGFR* mutations in the robust manner.

Here, we present our work on the robust radiogenomics approach based on topologically invariant BNs to the identification of the *EGFR* mutations among patients with NSCLC in three databases acquired from three different countries using 5 manufacturers' scanners with a variety of scanning parameters.

## 2. Methods

### 2.1 Clinical cases

The study protocol was approved by the institutional review boards of Kyushu University Hospital (KUH) and University of Malaya Medical Centre (UMMC). Table 1 and S1–S3 Tables summarize the demographic/clinical characteristics and significant differences between patients with the *EGFR* mutant and wildtype tumors in the datasets obtained from UMMC (Malaysia), KUH (Japan) and The Cancer Imaging Archive (TCIA) (America) [21]. The CE CT images of 194 NSCLC patients (79 *EGFR* mutants and 115 wildtypes) were analyzed. Ninety-nine cases obtained from UMMC were used for training and validation procedures. Forty-one cases collected from KUH and fifty-four cases obtained from TCIA were used for a test procedure. The case numbers selected from TCIA database were listed in S4 Table [21]. These CT images were acquired using several scanners with a variety of scanning parameters (Tables 2 and 3) in which there were statistically significant differences in slice thickness and in-plane pixel size between the training (validation) and the test datasets (Mann–Whitney *U* test,  $p < 0.05$ ) (Table 3). The matrix sizes after reconstruction were  $512 \times 512 \times 46$ –750 (median, 346),  $512 \times 512 \times 38$ –415 (median, 114) and  $512 \times 512 \times 115$ –636 (median, 250) for the datasets obtained from UMMC, KUH and TCIA, respectively.

**Table 1. Distributions and significant differences in demographic/clinical characteristics between patients with sensitizing epidermal growth factor receptor (*EGFR*) mutants and wildtypes in datasets obtained from University of Malaya Medical Centre, Kyushu University Hospital and The Cancer Imaging Archive.**

	<i>EGFR</i> mutant	<i>EGFR</i> wildtype	p value (testing method)
Total number of cases	79	115	
Age (y, min-max (median))	28–87 (66)	27–89 (68)	0.24 (Mann-Whitney U-test)
Sex			$3.79 \times 10^{-6}$ (Chi-squared test)
Male	33	87	
Female	46	28	
Stage			$8.67 \times 10^{-3}$ (Mann-Whitney U-test)
I	13	24	
II	6	13	
III	5	25	
IV	55	53	
Smoking status			$1.73 \times 10^{-9}$ (Mann-Whitney U-test)
Non-smoker	54	26	
Former-smoker	16	50	
Current-smoker	9	39	
Ethnicity			$2.62 \times 10^{-4}$ (Chi-squared test)
Asian	70	70	
Caucasian	8	43	
Hispanic/Latino	1	1	
Native Hawaiian/Pacific Islander	0	1	

<https://doi.org/10.1371/journal.pone.0244354.t001>

The *EGFR* mutation status was examined among lung tumor specimens using the QIAGEN *EGFR* RGQ PCR Kit (QIAGEN, Manchester Ltd, UK), Cobas<sup>®</sup> *EGFR* Mutation Test (Cobas<sup>®</sup>, Roche Molecular System Inc., USA), PNAclamp<sup>™</sup> *EGFR* Mutation Detection Kit (PANAGEN, Daejeon, Korea), or the PNA-LNA PCR clamp (LSI Medience, Japan). The tumor specimens were obtained via image-guided biopsy, endobronchial biopsy, or percutaneous Tru-Cut<sup>®</sup> needle biopsy as clinically indicated. For the detailed description of TCIA dataset, see ref. [22].

Solid tumor regions were defined for each patient as regions of interest (ROIs) to be analyzed using ITK-SNAP [23] and a 3D slicer [24]. All segmentations were performed and/or verified by a radiologist or respiratory physician with more than eight years of experience (WYC and KT).

Anisotropic CT images and ROIs were transformed into isotropic images with an isovoxel size of 0.77 mm (mode pixel size in the training dataset), using cubic and shape-based interpolations [25], respectively. In the present study, an axial plane of each CT image, which contained the maximum axial plane area of the ROIs, was selected for the calculation of radiomic features [26,27]. Since annotations of the ROIs are not created as a routine work in the clinical practice, we considered our algorithm should work with minimal labor. Aside from that, several past studies reported that two-dimensional radiomic features based on the CT images with maximum solid tumor areas showed comparable performance to the three-dimensional features in characterization of tumors [26,27]. Therefore, the two-dimensional feature extraction algorithms were applied in the present study.

A multiple-segmentation dataset, which was used to evaluate the robustness of the signatures against inter-observer variability, consisted of CT images of patients with NSCLC (n = 30) from the Quantitative Imaging Network multisite collection of lung CT data, with

**Table 2. Image acquisition parameters in datasets obtained from University of Malaya Medical Centre (UMMC), Kyushu University Hospital (KUH) and The Cancer Imaging Archive (TCIA).**

Dataset	Manufacturer	No. of cases	Model	No. of cases	Reconstruction Kernel	No. of cases	
UMMC	SIEMENS	92	Definition	39	B20f	71	
			Definition AS+	17	B30f	5	
			Sensation 16	18	B31f	15	
			SOMATOM Definition AS	1	D20f	1	
	SOMATOM Definition AS+	17					
	GE	5	LightSpeed16	5	Standard	5	
	Philips	1	Ingenuity CT	1	B	1	
KUH	TOSHIBA	1	Aquilion ONE	1	FC08	1	
	TOSHIBA	22	Aquilion	1	EC	2	
			Aquilion 64_3	4	FC13	1	
			Aquilion 64_4	10	FC53	1	
			Aquilion ONE	5	FC81	1	
			ID_station	2	FC86	16	
			LARGE			1	
	Philips	16	Brilliance 64	4	B	2	
			iCT256	10	C	1	
			Ingenuity CT	1	YB	1	
			PHILIPS-95201	1	YC	12	
			SOMATOM Definition AS+	2	B60f	1	
	TCIA	CANON	1	Robust	1	9	1
		GE	48	Discovery CT750 HD	33	BONE	1
				LightSpeed VCT	1	BONEPLUS	9
Unavailable				14	LUNG	34	
STANDARD						4	
SIEMENS		5	Unavailable	5	B45f	3	
			Unavailable		B50f	1	
			Unavailable		B70f	1	
TOSHIBA	1	Unavailable	1	FC52	1		

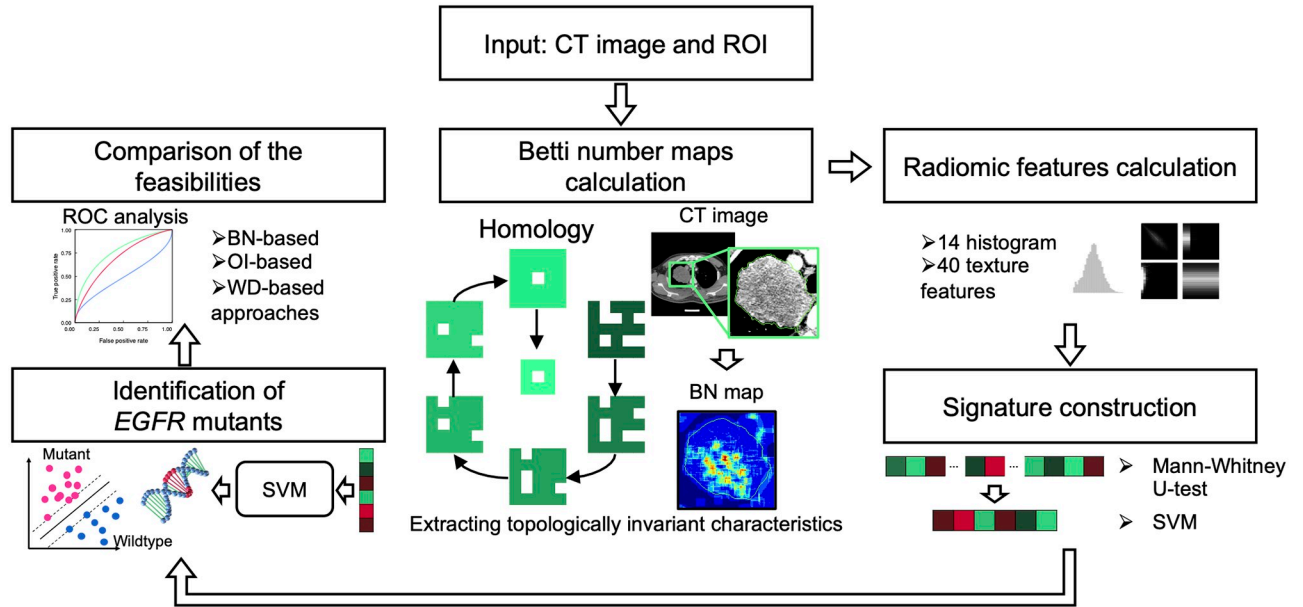
<https://doi.org/10.1371/journal.pone.0244354.t002>

nodule segmentations from TCIA (detailed in ref. [5]) [28–30]. The ROI in each image was independently segmented by three different institutions: Columbia University Medical Center, Stanford University, and Moffitt Cancer Center/University of South Florida. Each institution performed segmentation using their own custom segmentation algorithms under three different sets of initial conditions. These configurations resulted in nine segmentations of each

**Table 3. Comparison of voxel sizes and slice thicknesses in datasets obtained from University of Malaya Medical Centre (UMMC), Kyushu University Hospital (KUH) and The Cancer Imaging Archive (TCIA).**

		UMMC	KUH	TCIA
Voxel size (mm)		0.53–0.98 (0.73)	0.37–0.74 (0.39)	0.60–0.98 (0.79)
pvalue (Mann-Whitney U-test)	v.s. KUH	$1.07 \times 10^{-17}$	–	–
	v.s. TCIA	$1.49 \times 10^{-7}$	$4.62 \times 10^{-16}$	–
Slice thickness (mm)		0.50–5.00 (0.75)	1.00–5.00 (2.00)	0.63–2.50 (1.25)
pvalue (Mann-Whitney U-test)	v.s. KUH	$1.70 \times 10^{-20}$	–	–
	v.s. TCIA	$4.71 \times 10^{-18}$	$3.31 \times 10^{-10}$	–

<https://doi.org/10.1371/journal.pone.0244354.t003>



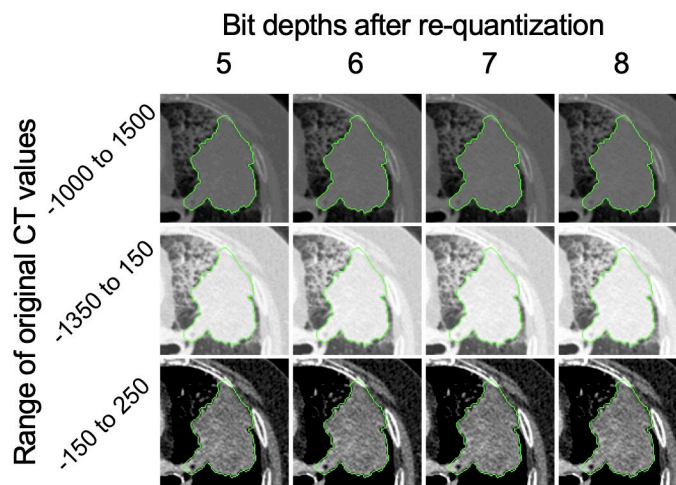
**Fig 2. Overall workflow for the present study.** CT: Computed tomography, ROI: Region of interest, BN: Betti number, SVM: Support vector machine, EGFR: Epidermal growth factor receptor, OI: Original image, WD: Wavelet-decomposition.

<https://doi.org/10.1371/journal.pone.0244354.g002>

tumor, for a total of 270 segmentations. The case numbers for constructing the multiple-segmentation dataset were listed in S5 Table.

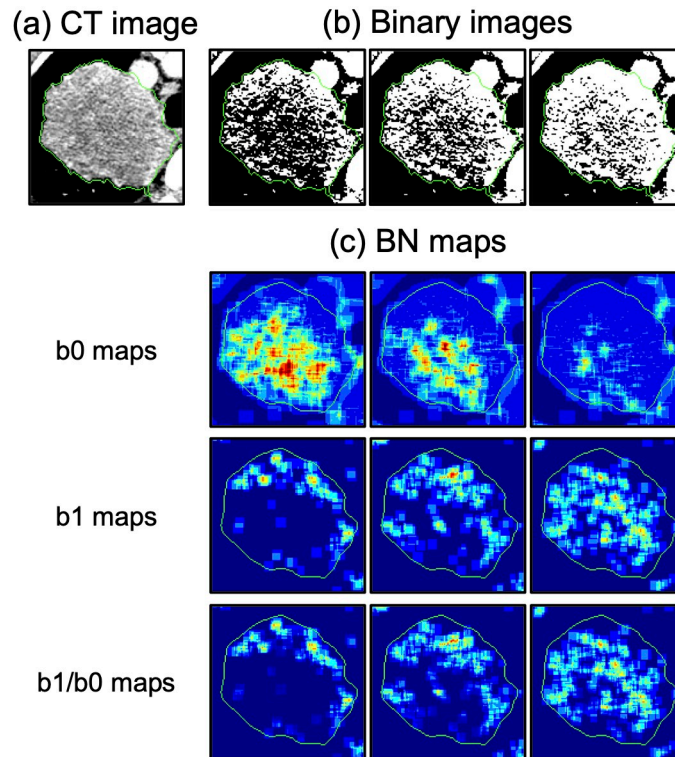
### 2.2 Betti number maps computation

Fig 2 shows an overall workflow of the present study. The BN maps were computed from a  $q$ -bit CT image (Fig 3) by counting the number of connected components ( $b_0$ ), holes ( $b_1$ ), and holes per connected component ( $b_1/b_0$ ) through thresholding (Fig 4) and a convolutional computation procedure [5]. A total of  $2^q \times 3$  BN maps ( $b_0$ ,  $b_1$ , and  $b_1/b_0$  maps) were



**Fig 3.  $q$ -bit images generated from the computed tomography (CT) images using four ranges of Hounsfield units (HU) of CT images (-1000 to 1500 HU, -1350 to 150 HU [lung range window], and -150 to 250 HU [mediastinal range window]) with four bit-depths after re-quantization (5, 6, 7, and 8 bits).**

<https://doi.org/10.1371/journal.pone.0244354.g003>



**Fig 4. Representative images of (a) computed tomography (CT) images, (b) binary images, and (c) Betti number (BN) maps.**

<https://doi.org/10.1371/journal.pone.0244354.g004>

calculated from binary images which were derived from the  $q$ -bit CT images by thresholding the images with values ranging from 0 to  $2^q - 1$ . The  $q$ -bit CT images were obtained by re-quantizing the original CT images into  $q$  bits. The optimal re-quantization level was explored among twelve types of  $q$ -bit CT images generated from the CT images using three ranges of Hounsfield units (HU) of the CT images (-1000 to 1500 HU [5], -1350 to 150 HU [lung range window], and -150 to 250 HU [mediastinal range window]) with four bit-depths after re-quantization (5, 6, 7, and 8 bits) (Fig 3). The kernel sizes and shifting pixels in the calculation of BNs using the convolutional computation were optimized from four kernels (5, 7, 9, and 11 pixels squared) and five shifting pixels (1, 2, 3, 4, and 5 pixels). The detailed algorithms used in the computation of the BN maps are shown in ref. [5].

### 2.3 Extraction of radiomic features

Fifty-four features were computed from the BN maps within the ROI by applying fourteen histogram- and forty texture-based (nine from a gray level co-occurrence matrix (GLCM), thirteen from a gray level run-length matrix (GLRLM), thirteen from a gray level size-zone matrix (GLSZM), and five from a neighborhood gray-tone difference matrix (NGTDM)) feature calculations. The total number of features could be calculated as  $2^q \times 3 \times 54$  (the number of BN maps  $\times$  the number of image features). Therefore, a total of 41,472, 20,736, 10,368, and 5,184 BN features were extracted from the BN maps based on the eight-, seven-, six-, and five-bit CT images, respectively. S6 Table shows a list of the features computed in the present study (see Supporting information).

## 2.4 Construction of radiomic signatures

Radiomic signatures were determined using the best combination of features that maximized a robustness index (RI) evaluating the feasibility and robustness of the support vector machine (SVM) models with a set of features in the identification of the *EGFR* mutants. The best combination was searched among representative features, which were the features showing statistically significant differences (Mann–Whitney *U* test,  $p < 0.05$ ) between median feature values for the *EGFR* mutants and wildtypes. Based on a review paper published by Chalkidou et al. [31], it was not necessary to correct the threshold values for statistical significance ( $p < 0.05$ ) in the present study, as independent tests were performed. The number of constituent features in the signature was determined from one to four using the methodology put forth by Vallières et al. [32] to search for the best combination of the features by maximizing the RI for training and validation defined as:

$$RI = \frac{AUC_{train} - AUC_{valid}}{1 + |AUC_{train} + AUC_{valid}|},$$

where  $AUC_{train}$  and  $AUC_{valid}$  indicate the AUCs for the identification of the *EGFR* mutants in the training and the validation procedures, respectively. These AUCs were obtained in the five-fold cross validation using the SVMs, which were constructed using a Gaussian kernel with a soft margin parameter  $C$  of 1 and gamma of  $1/N$  ( $N$ : the number of features) [33]. Since we applied the five-fold cross validation, the  $AUC_{train}$  was calculated by averaging the AUCs obtained in the five training procedures.

## 2.5 Construction of classifiers for the identification of the *EGFR* mutants

The SVMs were used to build the identification models of the *EGFR* mutants using the BN-based signature. The SVM was implemented with linear, Gaussian, and sigmoid kernels and several soft margin parameters  $C$  ranging from 0.1 to 10 with an interval of 0.3. For the Gaussian and the sigmoid kernels, gamma was also optimized using the values from 0.1 to 10 with an interval of 0.3. These parameters were optimized by maximizing the RI for training and validation using a grid search strategy.

## 2.6 Construction of conventional models

OI- and WD-based models were built to compare the feasibility of the BN-based model. The OI-based features were extracted from original CT images using 54 feature calculation methods described in subsection 2.3. A total of 216 WD features were obtained by applying the 54 feature calculation methods to four WD images. These WD images were obtained from  $q$ -bit CT images by applying either a low-pass filter (L) or a high-pass filter (H), which was derived from a Coiflet 1 mother wavelet, along the  $x$  and  $y$  axes. Therefore, the number of WD images resulted in four [LL, LH, HL and HH (First character L or H represents a low- or a high-pass filter applied along the  $x$  axis and second character represents the same for  $y$  axis)] The radiomic signatures and the SVMs for OI- and WD-based models were constructed in the same way as the BN-based models described in subsections 2.3–2.5.

## 2.7 Evaluation of the identification performance and robustness

The robustness of radiomic signatures against inter-planner variabilities of the ROIs was evaluated using mean intra-class correlation coefficients (ICCs) calculated in the multiple-segmentation dataset (refer to subsection 2.1) for the constituent features in the BN- and the WD-based signatures. Their robustness was evaluated using the following criteria for the ICCs [34]:



poor robustness,  $ICC < 0.5$ ; moderate robustness,  $0.5 \leq ICC < 0.75$ ; good robustness,  $0.75 \leq ICC < 0.9$ ; and excellent robustness,  $ICC \geq 0.9$ .

The BN-based model was compared to the conventional OI- and WD-based models in terms of accuracy and robustness of *EGFR* mutant identification. The models were evaluated using the RI based on the AUCs from the validation and the test procedures. A higher value of the RI determined the best model which could accurately and/or stably identify the *EGFR* mutants in both the validation and the test procedures.

The 95% confidence intervals (CIs) of the AUCs were estimated via bootstrapping (2000 times).

## 2.8 Statistical analysis

In the demographic/clinical characteristics, Mann-Whitney U-tests were applied to assess significant differences of age, stage and smoking status between the *EGFR* mutant and the wild-type. On the other hand, chi-squared tests were applied for sex and ethnicity. Stages from I-IV were used as numerical values to apply the Mann-Whitney U-test. In comparison of smoking status, numerical values were assigned for each category; non-smoker: 1, former-smoker: 2, current-smoker: 3.

Mann-Whitney U-test was applied to assess statistically significant differences of the CT scanning parameters of slice thickness and in-plane pixel size among the UMMC, the KUH and the TCIA datasets.

Statistical differences among the AUCs obtained from the BN- and the OI- or the WD-based approaches were evaluated using a Delong's test (significance threshold;  $p < 0.05$ ) [35]. All analyses were performed using R-4.0.3 (available at <http://www.r-project.org/>).

## 3. Results

Table 4 shows a summary of mean ICCs, the AUCs with CIs, accuracies, sensitivities, specificities and the RIs from the BN-, the OI- and the WD-based models to the identification of the *EGFR* mutants.

The mean ICCs for the BN-, the OI- and the WD-based signatures were 0.84 (good robustness), 0.75 (good robustness) and 0.070 (poor robustness), respectively. Four (*b0\_GLCM\_Energy\_45*, *b1/b0\_GLSZM\_ZSN\_104*, *b1\_GLCM\_SumAverage\_122*, *b0\_GLRLM\_LRLGE\_97*), three (*GLRLM\_SRLGE*, *GLSZM\_LGZE*, *GLSZM\_SZLGE*) and one (*GLSZM\_LGZE\_LL*) features were selected in the signatures for the BN, the OI and the WD, respectively. Digits in the

**Table 4. Summary of mean intra-class correlation coefficients (ICCs), areas under the receiver operating characteristic curves (AUCs) with 95% confidence intervals (CIs), accuracies, sensitivities, specificities and robustness indices from Betti number (BN)-, original image (OI)-, and wavelet decomposition (WD)-based models to the identification of epidermal growth factor receptor mutants.**

		BN	OI	WD
Validation	Mean ICC	0.84	0.75	0.070
	AUC (95% CI)	0.86 (0.78–0.93)	0.69 (0.58–0.79)	0.65 (0.54–0.76)
	Accuracy	0.84	0.67	0.62
	Sensitivity	0.78	0.58	0.75
	Specificity	0.88	0.73	0.53
Test	AUC (95% CI)	0.77 (0.67–0.86)	0.54 (0.42–0.66)	0.71 (0.59–0.81)
	Accuracy	0.71	0.66	0.74
	Sensitivity	0.72	0.38	0.69
	Specificity	0.70	0.86	0.77
	Robustness index	1.51	1.33	1.29

<https://doi.org/10.1371/journal.pone.0244354.t004>

range of 0–255 following the names of the BN features correspond to the threshold values for obtaining BN maps. The optimal parameters of the BN map computations were the kernel size of 7, the shifting pixel of 1, and the 8-bit CT image obtained using the mediastinal window (-150 to 250 HU). The optimal parameters for the BN-based SVM model were the Gaussian kernel with the soft margin parameter  $C$  of 0.4 and the gamma of 0.4. The optimal SVM parameters for the OI-based model were the Gaussian kernel, the soft margin parameter  $C$  of 7.3 and the gamma of 2.5. The optimal parameters for the WD-based model were the 7-bit CT image obtained using the lung window (-1350 to 150 HU) and the SVM using the Gaussian kernel with the soft margin parameter  $C$  of 7.3 and the gamma of 2.5.

Fig 5 shows the ROC curves for the identification of the *EGFR* mutants using the BN-, the OI- and the WD -based models. In the BN-based model, the AUCs for the validation and the test were 0.86 (CI: 0.78–0.93) and 0.77 (CI: 0.67–0.86), respectively. On the other hand, the AUCs in the OI-based model for the validation and the test were 0.69 (0.58–0.79) and 0.54 (0.42–0.66), respectively. Further, the AUCs in the WD-based model for the validation and the test were 0.65 (CI: 0.54–0.76) and 0.71 (CI: 0.59–0.81), respectively (Table 4). The  $p$  values between two AUCs obtained between the BN- and the OI-based models in the validation and the test procedures were  $1.4 \times 10^{-4}$  and  $5.8 \times 10^{-3}$  (Delong's test), respectively. Besides, the  $p$  values between two AUCs obtained between the BN- and the WD-based models in the validation and the test procedures were  $3.1 \times 10^{-3}$  and 0.29 (Delong's test), respectively. The BN-based model showed a higher RI of 1.51 compared with the OI-based model (RI: 1.33) or the WD-based model (RI: 1.29).

In the test procedure using the BN-based model, the AUCs for KUH and TCIA were 0.76 (CI: 0.60–0.90) and 0.61 (CI: 0.41–0.79), respectively. Fig 6 shows distributions of the signatures for the *EGFR* mutants and the wildtypes among the datasets from three countries. Euclidean distances of medians of the BN-based signature, which was normalized using z-score, between UMMC and KUH for the mutants and the wildtypes were 0.54 and 1.13, respectively. On the other hand, the distances between UMMC and TCIA for the mutants and the wildtypes were 1.63 and 2.17, respectively.

Table 5 summarizes the identification performance of the present and previous studies that performed the validation and the test of the models [8,9]. Although our dataset consisted of

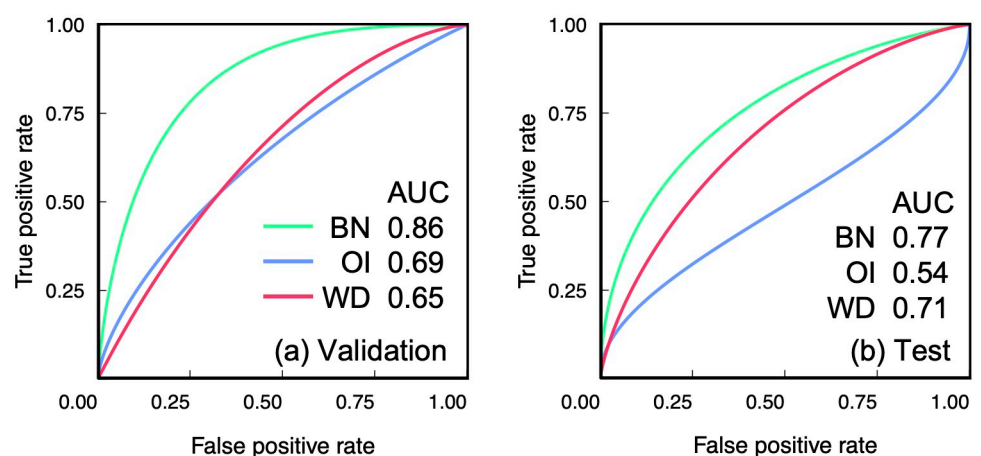
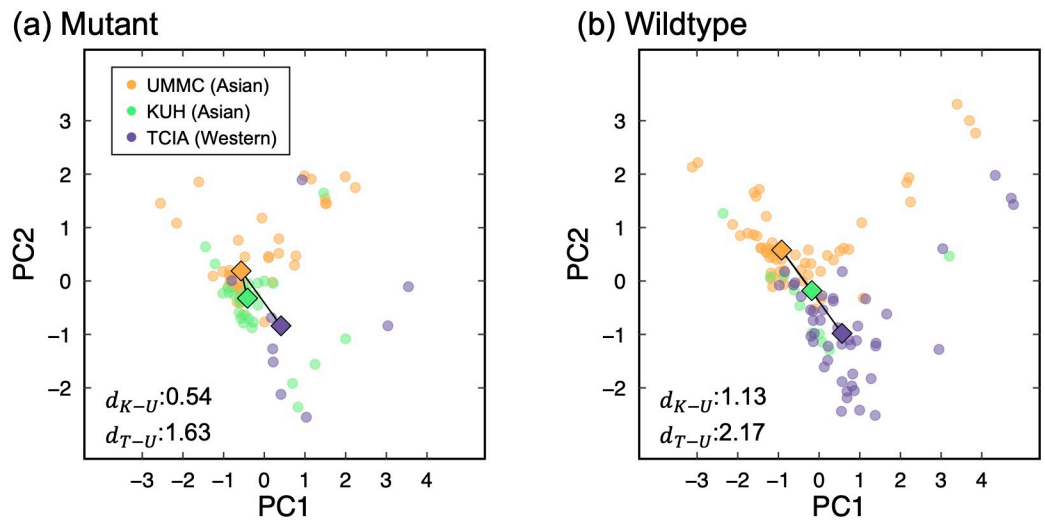


Fig 5. Receiver operating characteristic (ROC) curves for identification of epidermal growth factor receptor mutants using Betti number (BN)-, original image (OI)- and wavelet decomposition (WD)-based models with area under the ROC curves (AUC) in (a) the validation and (b) the test procedures.

<https://doi.org/10.1371/journal.pone.0244354.g005>



**Fig 6. Distributions of signatures for (a) epidermal growth factor receptor mutants and (b) wildtypes among datasets from three countries.** Euclidean distances between medians of the signatures from University of Malaya Medical Centre (UMMC) and The Cancer Imaging Archive (TCIA) ( $d_{T-U}$ ) were larger than that from UMMC and Kyushu University Hospital (KUH) ( $d_{K-U}$ ). PC1 and PC2 represent first and second principal components, respectively, calculated among constituent features in the signature. The medians were represented by square-shaped plots.

<https://doi.org/10.1371/journal.pone.0244354.g006>

**Table 5. Comparison of the results of areas under the receiver operating characteristic curves (AUCs) and robustness indices obtained from the present study and previous studies.**

Author (year)	Number of cases	Feature type	Datasets	Validation and test datasets	Origin of datasets (country)	AUC		Robustness index
						Validation	Test	
Present study	194	Betti number	Training 99 Testing 95	Independent datasets from three institutions	Malaysia, Japan, America	0.86	0.77	1.57
		Original image (OI)				0.69	0.54	1.33
		Wavelet decomposition (WD)				0.65	0.71	1.29
Wang (2019) ref. [8]	844	Deep learning	Training 603 Testing 241	Independent datasets from two institutions	China	0.85	0.81	1.60
Yang (2019) ref. [9]	467	shape + OI +WD	Training 306 Testing 161	Random split datasets from one institution	China	0.83	0.79	1.56

<https://doi.org/10.1371/journal.pone.0244354.t005>

patients from three different countries, the BN-based approach showed similar RIs compared with previous studies that used domestic databases.

### 4. Discussion

Almost 50% of Asian lung cancer patients harbor tumors with TKI-sensitizing *EGFR* mutation positive NSCLC [36,37], whereas prevalence of the *EGFR* mutations in Western lung cancer patients was about 15% [38]. Prevalence rate of the *EGFR* mutations in our datasets was similar to these general proportions (Asian [UMMC and KUH]: 50.00% [*EGFR* mutants:wildtypes = 50:100], Western [TCIA]: 16.67% [*EGFR* mutants:wildtypes = 9:45]). The small number of the *EGFR* mutations in the Western patients could be attributable to the lower AUC of 0.61 in the test for the TCIA dataset as compared with the AUC of 0.76 for the KUH dataset. Aside from that, we also found that Euclidean distances of medians of the BN-based signature between UMMC and TCIA were larger than that from UMMC and KUH (Fig 6). The image

features between Asian and Western could potentially differ from each other. These differences in the image features also lessened the AUCs in the test for the TCIA dataset.

The BN-based model showed stable identification performance in both the validation and test datasets, although the CT images were acquired using various scanning parameters (Tables 2 and 3). This result suggested that the BN-based signature was robust to differences in the scanning parameters. The  $b_0$ ,  $b_1$  and  $b_1/b_0$  maps evaluated the number of the connected components composed of high-intensity pixels, the number of holes caused by low-intensity pixels and the density of low-intensity holes, respectively. These calculation procedures might be suitable for preserving the characteristics of the *EGFR* mutants and dismissing the effects of differences in the quality of the CT images. As a result, the proposed BN-based model could perform as well as the model developed by Wang et al. [8] that used four times more cases as in the present study.

The presence of ground-glass opacity (GGO) has been recognized as one of representative characteristics associated with presence of the *EGFR* mutations [19,20]. The GGO is defined as a hazy area of increased attenuation of the lung with preservation of bronchial and vascular margins which cause low-intensity holes on the CT images [20]. In addition, intra-tumor heterogeneity on CE CT images has also been reported to have the association with the *EGFR* mutations in NSCLC [19]. Those findings were only explored in a qualitative manner. Our results had similar tendency to those of the qualitative analysis [19]. In the present study, we developed the robust model to quantify their heterogeneity using the topologically invariant BNs.

The RI obtained in the previous radiomics study was higher than that obtained using our conventional approaches [9]. Aside from the experiments using the OI- and the WD-based features, we also assessed the feasibility of shape-based features extracted using two- and three-dimensional calculations provided by pyradiomics package in Python. However, there were no representative features which were significantly associated the *EGFR* mutation status. In the previous study conducted using conventional radiomic analysis (Table 5), dataset was composed of larger number of cases from a single institution. Therefore, there might be smaller variability in the image qualities.

A BN- and clinical factor (BC)-based SVM model was also constructed using the signature of the BN and clinical factors of sex and smoking status because they were significantly associated with the *EGFR* mutation status in the training dataset (S1 Table). The optimal SVM parameters for the BN-based model was used for the BC-based model. The identification performance of the BC-based model was similar to that of the BN-based model. In the BC-based model, the AUCs for the validation and the test were 0.81 (CI: 0.72–0.89) and 0.77 (CI: 0.66–0.87), respectively.

Worldwide application of the model for the identification of the *EGFR* mutants requires the investigation of the feasibilities of the model with respect to multiple databases across the world. However, past studies mentioned above [8,9] have not conducted evaluation of the model using patient data from different countries. That fact motivated us to conduct the present study with an international database consisted of patients with NSCLC from three different countries.

The *EGFR* mutations promote cellular proliferation, differentiation, and migration of NSCLC [39]. *EGFR*-TKIs such as gefitinib, erlotinib, afatinib, and osimertinib for patients with advanced *EGFR* mutant NSCLC are the standards of care for first-line treatment. These agents confer significantly longer median progression-free survival compared to standard platinum-based doublet chemotherapy [40]. Therefore, the *EGFR* mutations in patients with NSCLC should be accurately identified for the selection of optimal treatments in precision medicine [41]. Currently, the *EGFR* mutations among patients with NSCLC can be identified

using different platforms such as direct sequencing, real-time polymerase chain reaction, or immunohistochemistry on tissue specimens obtained via image-guided invasive needle biopsies, bronchoscopy biopsies, or surgical resection [42]. However, these invasive procedures are associated with discomfort and potentially serious complications such as pneumothorax, bleeding, airway trauma, infection, and rarely death [43–45]. In addition, invasive procedures may not be feasible in some patients due to physical unfitness, co-morbidities, and reluctance [46]. Therefore, non-invasive approaches to assessing the *EGFR* mutation status are preferable and may provide more patients with the opportunity to undergo targeted treatment with the *EGFR*-TKIs. The present study could facilitate a non-invasive and reliable approach for the detection of the *EGFR* mutations.

The present study had four limitations. First, the number of cases was small although the proposed model showed potential to be robust and feasible in the identification of *EGFR* mutants. Past studies used at least two times as much number of cases as this study [8,9]. If we add more patient data in the analysis, the model's performance could be further improved and confident. Second, three-dimensional computations of the BNs have not been performed. The application of three-dimensional BN features may lead more accurate identification of the *EGFR* mutants by reflecting volumetric information of heterogeneous lung tumors. Third, we did not assess the impact of the time between contrast injection and the CT scan. Yang et al. reported that some conventional radiomic features showed variability depending on the time at which the CT scan was obtained after contrast injection [47]. Since the BN-based features were computed from the BN maps obtained through thresholding to the CE CT images, the timing of the CT scans after injection may affect the quantitative values of the features. Finally, we only focused on identifying the *EGFR* mutations from the others (wildtypes). It would be necessary to have other mutation groups such as *KRAS* mutations and *ALK* fusions.

## Conclusions

The proposed model based on the topologically invariant BN outperformed the conventional identification models. The results of the present study suggested the robustness of the BN-based approach against variations in scanner/scanning parameters of three different countries. Therefore, the BN-based approach showed potential in the non-invasive identification of the *EGFR* mutations and assist physicians to tailor more effective treatment strategies for NSCLC patients.

## Supporting information

**S1 Table. Distributions and significant differences in demographic/clinical characteristics between patients with sensitizing epidermal growth factor receptor (*EGFR*) mutants and wildtypes in a dataset obtained from University of Malaya Medical Centre.**  
(DOCX)

**S2 Table. Distributions and significant differences in demographic/clinical characteristics between patients with sensitizing epidermal growth factor receptor (*EGFR*) mutants and wildtypes in a dataset obtained from Kyushu University Hospital.**  
(DOCX)

**S3 Table. Distributions and significant differences in demographic/clinical characteristics between patients with sensitizing epidermal growth factor receptor (*EGFR*) mutants and wildtypes in a dataset obtained from The Cancer Imaging Archive.**  
(DOCX)

**S4 Table. Case numbers obtained from The Cancer Imaging Archive for constructing a test dataset.**

(DOCX)

**S5 Table. Case numbers selected from The Cancer Imaging Archive for constructing a multi segmentation dataset.**

(DOCX)

**S6 Table. Radiomic features with the feature types.**

(DOCX)

## Acknowledgments

The authors are grateful to all the members of the Arimura Laboratory (<http://web.shs.kyushu-u.ac.jp/~arimura>), whose comments and suggestions made enormous contributions to this study. We would like to thank Editage ([www.editage.jp](http://www.editage.jp)) for their English language editing service.

## Author Contributions

**Conceptualization:** Kenta Ninomiya, Hidetaka Arimura.

**Data curation:** Kenta Ninomiya, Hidetaka Arimura, Wai Yee Chan, Kentaro Tanaka, Nadia Fareeda Muhammad Gowdh, Nur Adura Yaakup, Chong-Kin Liam, Chee-Shee Chai, Kwan Hoong Ng.

**Formal analysis:** Kenta Ninomiya.

**Funding acquisition:** Hidetaka Arimura, Shinichi Mizuno.

**Investigation:** Kenta Ninomiya, Hidetaka Arimura.

**Methodology:** Kenta Ninomiya, Hidetaka Arimura.

**Project administration:** Hidetaka Arimura, Wai Yee Chan.

**Resources:** Kenta Ninomiya, Hidetaka Arimura, Wai Yee Chan, Kentaro Tanaka, Nadia Fareeda Muhammad Gowdh, Nur Adura Yaakup, Chong-Kin Liam, Chee-Shee Chai, Kwan Hoong Ng.

**Software:** Kenta Ninomiya.

**Supervision:** Hidetaka Arimura, Kwan Hoong Ng.

**Validation:** Kenta Ninomiya, Hidetaka Arimura, Kentaro Tanaka.

**Visualization:** Kenta Ninomiya, Hidetaka Arimura.

**Writing – original draft:** Kenta Ninomiya, Hidetaka Arimura, Wai Yee Chan, Kentaro Tanaka.

**Writing – review & editing:** Kenta Ninomiya, Hidetaka Arimura, Wai Yee Chan.

## References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 2018; 68: 394–424. <https://doi.org/10.3322/caac.21492> PMID: 30207593

2. Molina JR, Yang P, Cassivi SD, Schild SE, Adjei AA. Non-small cell lung cancer: Epidemiology, risk factors, treatment, and survivorship. *Mayo Clin Proc.* 2008; 83: 584–594. <https://doi.org/10.4065/83.5.584> PMID: 18452692
3. Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WEE, et al. The IASLC lung cancer staging project: Proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM Classification for lung cancer. *J Thorac Oncol.* 2016; 11: 39–51. <https://doi.org/10.1016/j.jtho.2015.09.009> PMID: 26762738
4. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Cavalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun.* 2014. <https://doi.org/10.1038/ncomms5006> PMID: 24892406
5. Ninomiya K, Arimura H. Homological radiomics analysis for prognostic prediction in lung cancer patients. *Phys Medica.* 2020; 69: 90–100. <https://doi.org/10.1016/j.ejmp.2019.11.026> PMID: 31855844
6. Soufi M, Arimura H, Nagami N. Identification of optimal mother wavelets in survival prediction of lung cancer patients using wavelet decomposition-based radiomic features. *Med Phys.* 2018; 45: 5116–5128. <https://doi.org/10.1002/mp.13202> PMID: 30230556
7. Arimura H, Soufi M, Kamezawa H, Ninomiya K, Yamada M. Radiomics with artificial intelligence for precision medicine in radiation therapy. *J Radiat Res.* 2018; 1–8.
8. Wang S, Shi J, Ye Z, Dong D, Yu D, Zhou M, et al. Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning. *Eur Respir J.* 2019; 53. <https://doi.org/10.1183/13993003.00986-2018> PMID: 30635290
9. Yang X, Dong X, Wang J, Li W, Gu Z, Gao D, et al. Computed tomography-based radiomics signature: A potential indicator of epidermal growth factor receptor mutation in pulmonary adenocarcinoma appearing as a subsolid nodule. *Oncologist.* 2019; 24: e1156–e1164. <https://doi.org/10.1634/theoncologist.2018-0706> PMID: 30936378
10. Tu W, Sun G, Fan L, Wang Y, Xia Y, Guan Y, et al. Radiomics signature: A potential and incremental predictor for EGFR mutation status in NSCLC patients, comparison with CT morphology. *Lung Cancer.* 2019; 132: 28–35. <https://doi.org/10.1016/j.lungcan.2019.03.025> PMID: 31097090
11. Mackin D, Fave X, Zhang L, Fried D, Yang J, Brian T, et al. Measuring CT scanner variability of radiomics features. *Invest Radiol.* 2015; 50: 757–765. <https://doi.org/10.1097/RLI.000000000000180> PMID: 26115366
12. Li Y, Lu L, Xiao M, Dercle L, Huang Y, Zhang Z, et al. CT Slice Thickness and Convolution Kernel Affect Performance of a Radiomic Model for Predicting EGFR Status in Non-Small Cell Lung Cancer: A Preliminary Study. *Sci Rep.* 2018; 8: 1–10.
13. Nakane K, Takiyama A, Mori S, Matsuura N. Homology-based method for detecting regions of interest in colonic digital images. *Diagn Pathol.* 2015; 10: 1–5.
14. Lawson P, Sholl AB, Brown JQ, Fasy BT, Wenk C. Persistent Homology for the Quantitative Evaluation of Architectural Features in Prostate Cancer Histology. *Sci Rep.* 2019; 9: 1–15.
15. Oyama A, Hiraoka Y, Obayashi I, Saikawa Y, Furui S, Shiraishi K, et al. Hepatic tumor classification using texture and topology analysis of non-contrast-enhanced three-dimensional T1-weighted MR images with a radiomics approach. *Sci Rep.* 2019; 9: 8764. <https://doi.org/10.1038/s41598-019-45283-z> PMID: 31217445
16. Nishio M, Kubo T, Togashi K. Estimation of lung cancer risk using homology-based emphysema quantification in patients with lung nodules. *PLoS One.* 2019; 14. <https://doi.org/10.1371/journal.pone.0210720> PMID: 30668605
17. Kadoya N, Tanaka S, Kajikawa T, Tanabe S, Abe K, Nakajima Y, et al. Homology-based radiomic features for prediction of the prognosis of lung cancer based on CT-based radiomics. *Med Phys.* 2020. <https://doi.org/10.1002/mp.14104> PMID: 32096876
18. Kaczynski T, Mischaikow K, Mrozek M. *Computational Homology.* 2004.
19. Liu Y, Kim J, Qu F, Liu S, Wang H, Balagurunathan Y, et al. CT features associated with epidermal growth factor receptor mutation status in patients with lung adenocarcinoma. *Radiology.* 2016; 280: 271–280. <https://doi.org/10.1148/radiol.2016151455> PMID: 26937803
20. Yang Y, Yang Y, Zhou X, Song X, Liu M, He W, et al. EGFR L858R mutation is associated with lung adenocarcinoma patients with dominant ground-glass opacity. *Lung Cancer.* 2015; 87: 272–277. <https://doi.org/10.1016/j.lungcan.2014.12.016> PMID: 25582278
21. Bakr S, Gevaert O, Echegaray S, Ayers K, Zhou M, Shafiq M, et al. Data descriptor: A radiogenomic dataset of non-small cell lung cancer. *Sci Data.* 2018; 5: 1–9.
22. Bakr S, Gevaert O, Echegaray S, Ayers K, Zhou M, Shafiq M, et al. Data for NSCLC Radiogenomics Collection. The Cancer Imaging Archive. 2017. <http://doi.org/10.7937/K9/TCIA.2017.7hs46erv>.
23. ITK-SNAP. <http://www.itksnap.org>.

24. 3D slicer. <https://www.slicer.org>.
25. Herman GT, Zheng J, Bucholtz CA. Shape-based Interpolation. *IEEE Comput Graph Appl*. 1992; 12: 69–79. <https://doi.org/10.1109/38.135915>
26. Shen C, Liu Z, Guan M, Jiangdian S, Yucheng L. 2D and 3D CT Radiomics Features Prognostic Performance Comparison in Non-Small Cell Lung Cancer. *Transl Oncol*. 2017; 10: 886–894. <https://doi.org/10.1016/j.tranon.2017.08.007> PMID: 28930698
27. Meng L, Dong D, Chen X, Fang M, Wang R, Li J, et al. 2D and 3D CT Radiomic Features Performance Comparison in Characterization of Gastric Cancer: A Multi-center Study. *IEEE J Biomed Heal Informatics*. 2020; 2194: 1–1. <https://doi.org/10.1109/JBHI.2020.3002805> PMID: 32750940
28. Kalpathy-Cramer J, Zhao B, Goldgof D, Gu Y, Wang X, Yang H, et al. A Comparison of Lung Nodule Segmentation Algorithms: Methods and Results from a Multi-institutional Study. *J Digit Imaging*. 2016; 29: 476–487. <https://doi.org/10.1007/s10278-016-9859-z> PMID: 26847203
29. Kalpathy-Cramer J, Mamomov A, Zhao B, Lu L, Cherezov D, Napel S, et al. Radiomics of Lung Nodules: A Multi-Institutional Study of Robustness and Agreement of Quantitative Imaging Features. *Tomography*. 2016; 2: 430–437. <https://doi.org/10.18383/j.tom.2016.00235> PMID: 28149958
30. Kalpathy-Cramer J, Napel S, Goldgof D, Zhao B. TCIA multisegmentation. 2017.
31. Chalkidou A, O'Doherty MJ, Marsden PK. False discovery rates in PET and CT studies with texture features: A systematic review. *PLoS One*. 2015; 10. <https://doi.org/10.1371/journal.pone.0124165> PMID: 25938522
32. Vallières M, Freeman CR, Skamene SR, El Naqa I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Phys Med Biol*. 2015; 60: 5471–5496. <https://doi.org/10.1088/0031-9155/60/14/5471> PMID: 26119045
33. e 1071. <https://www.rdocumentation.org/packages/e1071/versions/1.7-3>.
34. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 2016; 15: 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012> PMID: 27330520
35. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics*. 1988. <https://doi.org/10.2307/2531595> PMID: 3203132
36. Han B, Tjulandin S, Hagiwara K, Normanno N, Wulandari L, Laktionov K, et al. EGFR mutation prevalence in Asia-Pacific and Russian patients with advanced NSCLC of adenocarcinoma and non-adenocarcinoma histology: The IGNITE study. *Lung Cancer*. 2017; 113: 37–44. <https://doi.org/10.1016/j.lungcan.2017.08.021> PMID: 29110846
37. Hirsch FR, Bunn PA. EGFR testing in lung cancer is ready for prime time. *Lancet Oncol*. 2009; 10: 432–433. [https://doi.org/10.1016/S1470-2045\(09\)70110-X](https://doi.org/10.1016/S1470-2045(09)70110-X) PMID: 19410185
38. Kohno T, Nakaoku T, Tsuta K, Tsuchihara K, Matsumoto S, Yoh K, et al. Beyond ALK-RET, ROS1 and other oncogene fusions in lung cancer. *Transl Lung Cancer Res*. 2015; 4: 156–164. <https://doi.org/10.3978/j.issn.2218-6751.2014.11.11> PMID: 25870798
39. Wieduwilt MJ, Moasser MM. The epidermal growth factor receptor family: Biology driving targeted therapeutics. *Cell Mol Life Sci*. 2008; 65: 1566–1584. <https://doi.org/10.1007/s00018-008-7440-8> PMID: 18259690
40. Planchard D, Popat S, Kerr K, Novello S, Smit EF, Faivre-Finn C, et al. Corrigendum: Metastatic non-small cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2019; 30: 863–870. <https://doi.org/10.1093/annonc/mdy474> PMID: 31987360
41. Kerr KM, Bubendorf L, Edelman MJ, Marchetti A, Mok T, Novello S, et al. Second ESMO consensus conference on lung cancer: Pathology and molecular biomarkers for non-small-cell lung cancer. *Ann Oncol*. 2014; 25: 1681–1690. <https://doi.org/10.1093/annonc/mdu145> PMID: 24718890
42. Angulo B, Conde E, Suárez-Gauthier A, Plaza C, Martínez R, Redondo P, et al. A comparison of EGFR mutation testing methods in lung carcinoma: Direct sequencing, real-time PCR and immunohistochemistry. *PLoS One*. 2012; 7. <https://doi.org/10.1371/journal.pone.0043842> PMID: 22952784
43. Wu CC, Maher MM, Shepard JAO. Complications of CT-guided percutaneous needle biopsy of the chest: Prevention and management. *Am J Roentgenol*. 2011; 196: 678–682. <https://doi.org/10.2214/AJR.10.4659> PMID: 21606253
44. Stahl D, Richard K, Papadimos T. Complications of bronchoscopy: A concise synopsis. *Int J Crit Illn Inj Sci*. 2015; 5: 189–195. <https://doi.org/10.4103/2229-5151.164995> PMID: 26557489
45. Pei G, Zhou S, Han Y, Liu Z, Xu S. Risk factors for postoperative complications after lung resection for non-small cell lung cancer in elderly patients at a single institution in China. *J Thorac Dis*. 2014; 6: 1230–1238. <https://doi.org/10.3978/j.issn.2072-1439.2014.07.23> PMID: 25276365



46. Manhire A, Charig M, Clelland C, Gleeson F, Miller R, Moss H, et al. Guidelines for radiologically guided lung biopsy. *Thorax*. 2003; 58: 920–936. <https://doi.org/10.1136/thorax.58.11.920> PMID: 14586042
47. Yang J, Zhang L, Fave XJ, Fried DV, Stingo FC, Ng CS, et al. Uncertainty analysis of quantitative imaging features extracted from contrast-enhanced CT in lung tumors. *Comput Med Imaging Graph*. 2016; 48: 1–8. <https://doi.org/10.1016/j.compmedimag.2015.12.001> PMID: 26745258