Development of radiomics model for prognosis prediction in gastric cancer

Jaeseung Shin

Department of Medicine

The Graduate School, Yonsei University

Development of radiomics model for prognosis prediction in gastric cancer

Jaeseung Shin

Department of Medicine

The Graduate School, Yonsei University

Development of radiomics model for prognosis prediction in gastric cancer

Directed by Professor Yongmin Huh

The Doctoral Dissertation
submitted to the Department of Medicine,
the Graduate School of Yonsei University
in partial fulfillment of the requirements for the degree
of Doctor of Medicine

Jaeseung Shin

June 2021

This certifies that the Doctoral Dissertation of Jaeseung Shin is approved.

Thesis Supervisor: Yong-Min Huh

Thesis Committee Member #1: Joon Seok Lim

Thesis Committee Member #2: Hyo Song Kim

Thesis Committee Member #3: Jae-Ho Cheong

Thesis Committee Member #4: Helen Hong

The Graduate School
Yonsei University

June 2021

ACKNOWLEDGEMENTS

I would like to express my gratitude towards my supervisor, Prof. Yong-Min Huh for all the guidance, constant supervision and the encouragement which helped me in completion of this research project.

I am incredibly grateful to my mentor, Prof. Sungwon Kim for sharing his knowledge and valuable advice.

I also wish to thank the committee members, Prof. Joon Seok Lim, Prof. Hyo Song Kim, Prof. Jae-Ho Cheong, and Prof. Helen Hong for sharing their precious time and constructive comments during the course of this dissertation.

Last but not least, I owe my beloved family, in particular my wife Jooah Cheon, a great debt of gratitude for their unconditional support throughout the completion of this dissertation and my life in general. Without their understanding and encouragement, none of this thesis would indeed be possible.

<TABLE OF CONTENTS>

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Development of radiomics model for prognosis prediction in gastric cancer

Jaeseung Shin

Department of Medicine
The Graduate School, Yonsei University

(Directed by Professor Yong-Min Huh)

Purpose : Preoperative therapy has gained wide interest in advanced gastric cancer patients due to its potential advantages of improved disease control. Selection of high risk patients based on preoperative staging is crucial to choose the candidates for neoadjuvant therapy. To evaluate the added value of the radiomic signature for predicting recurrence-free survival (RFS) of locally advanced gastric cancer using preoperative contrast-enhanced abdominal computed tomography (CT), compared with the clinical prediction model.

Methods : Our institutional review board approved this retrospective study and waived the requirement for patient consent. The present study included 349 patients who underwent curative resection for locally advanced gastric cancer in 2010 without neoadjuvant therapies as training cohort. External validation cohort with 61 patients was collected from another hospital. Clinical factors which were available in the preoperative setting including conventional CT staging and endoscopic data were obtained and a total 438 of radiomic features were extracted from the preoperative CT. To predict RFS, the radiomic model was developed using penalized Cox regression with a least absolute shrinkage and selection operator with ten-fold cross-validation. Internal and external validations were

performed using a bootstrapping method (n=1000) for each validation. The incremental values of radiomic features were evaluated by using the integrated area under the receiver operating characteristic curve (iAUC).

Results : With the final 410 patients (58.2±13.0 years-old; 268 female), the radiomic model consisted of seven selected features. The clinical model included two independent factors, CT Borrmann type 4 and extramural nodular infiltration, The merged model was built using both clinical and radiomic features. In both of the internal and the external validation, the integrated area under the receiver operating characteristic curve values of both the radiomic model (0.714 [95% confidence interval(CI) 0.667, 0.759], P<0.001 in internal validation; 0.652 [95% CI 0.628, 0.674], P=0.010 in external validation) and the merged model (0.719 [95% CI 0.674, 0.764], P<0.001; 0.651 [95% CI, 0.630, 0.673], P=0.014) were significantly higher than those of the clinical model (0.616 [95% CI, 0.570, 0.663]; 0.59 4[95% CI 0.544, 0.636]).

Conclusion : The radiomics signature based on preoperative CT images is a possible preoperative imaging biomarker that can improve RFS prediction of the preoperative clinical profile in LAGC. The ability of radiomic signatures to identify high-risk LAGC patients may be helpful in selecting appropriate candidates for neoadjuvant therapy.

Keywords : radiomics, computed tomography, gastric cancer, prognosis prediction

Development of radiomics model for prognosis prediction in gastric cancer

Jaeseung Shin

Department of Medicine
The Graduate School, Yonsei University

(Directed by Professor Yong-Min Huh)

I. INTRODUCTION

Gastric cancer is the fourth most common cancer and third leading cause of cancer-related death worldwide, accounting for an estimated annual 723,100 deaths[1,2]. Complete R0 resection of tumor with subsequent adjuvant chemotherapy has been established as an effective treatment for patients with locally advanced gastric cancer (LAGC)[3]. However, recurrence after complete resection occurs in up to 30%–40% of patients within 5 years[4-6]. Recently, neoadjuvant chemotherapy is widely recommended in international western guidelines for advanced gastric cancer patients because of its potential benefits, including early treatment of micrometastases, delivery of higher dose chemotherapy before surgery, and an improved down-staging change of the primary tumor[3]. Higher R0 resection rate and survival can be achieved with neoadjuvant chemotherapy followed by curative surgery[7,8]. As evidence supporting neoadjuvant chemotherapy accumulates, identification of patients as neoadjuvant candidates becomes important.

Computed tomography (CT) is the modality of choice for preoperative clinical staging of gastric cancer; however, studies have reported limitations

regarding staging accuracy and risk stratification[9]. Due to intrinsic limitations of CT spatial resolution in distinguishing gastric wall layers, tumor staging is suboptimal. Preoperative CT-based node staging is also limited because size-based differentiation of small lymph nodes (LNs) with micrometastasis from large reactive LNs is difficult[10]. Hence, there is a growing need to use biomarkers in conjunction with abdominal CT to predict the prognosis of LAGC.

Radiomics has emerged as a promising tool for discovering new imaging biomarkers by converting digital medical images into high-dimensional quantitative features such as shape, histogram and texture that captures tumor heterogeneity[11-13]. Its potential capacity to capture useful information and to increase the diagnostic and prognostic power has shown in lung, prostate, brain, liver, colorectal cancer[14]. In gastric cancer, several radiomics studies has been performed based to reveal CT texture parameters as noninvasive predictive factor[15-17]. Previous study by Giganti et al.[15] investigated the association between CT texture analysis and overall survival and showed preoperative CT texture features as prognostic factor for risk stratification in gastric cancer. However, those studies were limited by the small sample size and lack of validation. Recently, a large retrospective study[18] demonstrated that the radiomics signature had good performance in predicting prognosis and survival benefit of adjuvant chemotherapy. However, the study included a considerable number of gastric cancer cases with early stage or distant metastasis, which limited the risk stratification of patients with LAGCs who are subject to preoperative chemotherapy.

This study aimed to develop and validate a radiomics-based prognostic model for recurrence-free survival (RFS) using preoperative contrast-enhanced CT in LAGC. Moreover, we assessed the value added by radiomic signatures when integrated with clinical profiles in the preoperative setting and whether the radiomics model can perform risk stratification for tumor recurrence.

II. MATERIALS AND METHODS

This retrospective study was approved by the institutional review board of Severance hospital (Seoul, Korea, Protocol no. 4-2019-0062), with a waiver for informed consent.

1. Patients and data acqusitions

From January 1, 2010 to December 31, 2010., we retrospectively searched a total of 426 consecutive patients with locally advanced gastric cancer (pT2–4) who underwent curative surgery without neoadjuvant therapy at Sinchon Severance hospital. Patients who had double primary cancer (n = 10), endoscopic clipping (n = 6), histology other than adenocarcinoma (n = 1), history of previous endoscopic mucosal resection (n = 3), and less than 6 months follow-up (n = 18) were excluded. Patients with insufficient preoperative CT images with slice thickness more than 5 mm (n = 6) or pixel size larger than 1.0 mm × 1.0 mm (n = 8) were also excluded due to poor CT quality.

For external validation, 83 consecutive patients with the same enrollment criteria were collected from Kangnam Severance hospital. Patients who had double primary cancer (n = 2), less than 6 months follow-up (n = 12), endoscopic clipping (n = 5), and history of previous endoscopic mucosal resection (n = 1) were excluded. After CT image analysis, 25 and two patients in the training and validation cohorts were excluded respectively, due to no identifiable lesion on their CT scans, respectively. The final training and validation cohorts consisted of 349 and 61 patients, respectively (Figure 1).

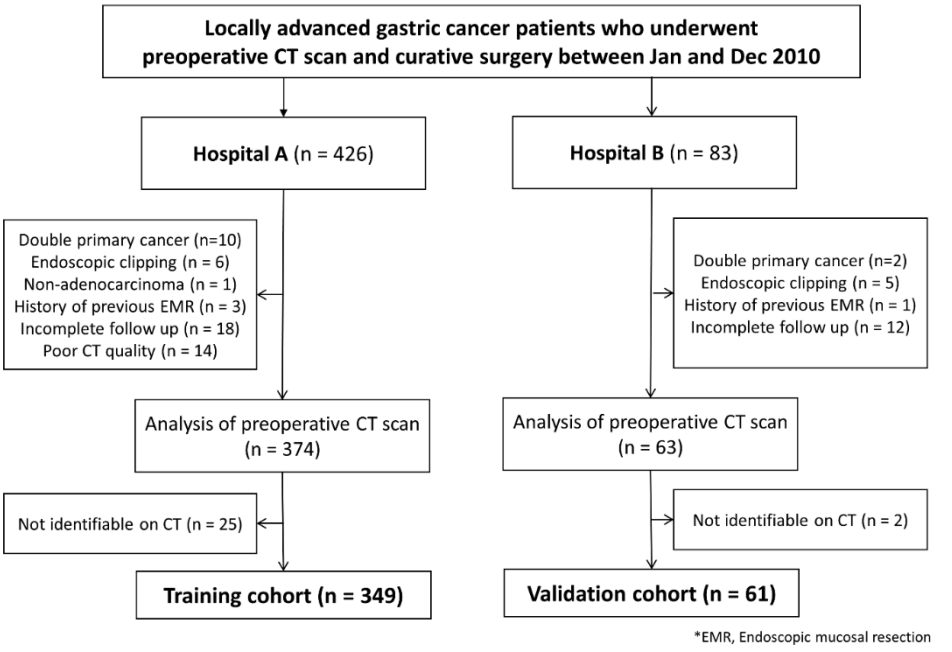


Figure 1. Flowchart for patient selection in training cohort and validation cohort

Clinical, laboratory, endoscopic, and pathological data were retrieved from patients' electronic medical records, including age, gender, serum levels of carcinoembryonic antigen (CEA), carbohydrate antigen (CA) 19-9, tumor location, tumor size, differentiation, Lauren type, and TNM stage. The TNM staging was reclassified according to the eighth edition of the American Joint Committee on Cancer/ Union for International Cancer Control staging system. The following clinical factors were integrated in the preoperative clinical model: Age (≤60 vs. >60 years); Sex (male vs. female); levels of serum carcinoembryonic antigen (< 5 vs. ≥5 U/ml) and carbohydrate antigen 19-9 (< 37 vs. ≥ 37 U/ml); endoscopy result including tumor location (upper vs. middle vs. lower), histological grade from biopsy tissue (well or moderate vs. poorly differentiated), Borrmann type (type 4 vs. others).

After surgical resection, all patients were followed up for 6.5 to 109.2 months (median follow-up: 71.5 months) through December 2018 according to the institutional protocol[19]. Follow-up data (follow-up duration and survival) were collected from hospital records for patients who were lost during follow-up. The follow-up duration was measured from the time of surgery to the last follow-up date, and information regarding the survival status at the last follow-up was collected. The recurrence free survival (RFS) was defined as the time to recurrence at any site (event) or the last follow-up date (censored).

2. CT image acquisition and image analysis

CT scans were performed with a 16- or 64-channel multidetector CT scanner (Somatom Sensation 16 and Sensation 64; Siemens Medical Solutions, Forchhein, Germany; and Lightspeed VCT, GE Healthcare, Milwaukee, WI, USA). Images were acquired from the diaphragm level to the symphysis pubis with detector collimations of 16 × 0.75 mm (Somatom Sensation 16, Simens Medical Solutions), 64 × 0.6 mm (Somatom Sensation 64, Simens Medical Solutions), or 64 × 0.625mm (Lightspeed VCT, GE healthcare). Other scanning parameters were as follows: tube current 160 mAs (Somatom Sensation 16 and Sensation 64, Siemens Medical Solutions) and 100 – 300 mAs of Automated tube current modulation with a noise index of 15 (AutomA; Lightspeed VCT, GE Healthcare); tube voltage 120 kVp; table speed, 24 mm per rotation; and gantry rotation time, 0.5 seconds. The details regarding the acquisition parameters of CT image are presented in Table 1. For gastric distention, either gas distention with two packs of effervescent granules or water distention with 1 L of water was introduced. Scanning was performed during portal phases, as determined with bolus tracking and automated triggering technique after intravenous administration of 120 – 150 mL of nonionic contrast materials (300 mgI/mL) using an automatic injector at a rate of 4 ml/second. The amount of contrast medium per patient was determined

by the total body weight. Axial and coronal images were reconstructed with 3-mm-thick sections and a 3 mm interval with filtered back projection algorithm. From the Picture Archiving and Communication System (Centricity, GE Medical Systems, Milwaukee, WI, USA), portal venous phase CT images were retrieved for qualitative image review and radiomic feature extraction because the tumor tissue was well differentiated from the adjacent normal gastric tissue.

Table 1. Details of CT acquisition Parameters

Parameters	Sensation 16	Sensation 64	Lightspeed VCT
No. of channels	16	64	64
Section collimation*	16 × 0.75	64 × 0.6	64 × 0.625
Slice thickness (mm)	3	3	3
Pitch	1	0.6	0.9840
Tube current (mAs)†	160	160	100–300‡
Rotation time (sec)	0.5	0.5	0.5
Table speed (mm per rotation)	24.0	24.0	24.0
Tube voltage (kV)	120	120	120
Kernel	B30f/B31f		Standard
Matrix	512 × 512	512 × 512	512 × 512

Following CT scanners were used: Sensation 16 and Sensation 64 (Siemens Healthineers); and Lightspeed VCT (GE Healthcare).
*Number of detector rows times section thickness (mm)
†Reference milliampere-seconds
‡AutomA was set between 100 and 300 mA with a noise index of 15.

Preoperative CT images were independently reviewed by two board-certified abdominal radiologists with more than 10 years of subspecialty experience, who arrived at a consensus in cases with discrepancy. The CT imaging characteristics analyzed were tumor depth, lymph node (LN) status, tumor size, and Borrmann type. Tumor depth on CT (CT-Depth) was categorized into two groups, nodular or less than nodular extramural infiltration groups—one of the major discriminating factors for predicting recurrence of AGC in a previous study[20]. LN

involvement on CT (CT-LN) was categorized into two groups, N0 – 1 and N2 – 3, as multidetector CT might be useful for selecting candidates for neoadjuvant therapy with ≥ pN2 disease[21]. LNs were considered metastatic if they had a short-axis diameter > 8 mm. Tumor size (CT-Size) was measured as the longest diameter on the axial or coronal plane. Tumors were classified as CT-Type4 when infiltrative stomach cancer showed no definite ulceration or mass formation on preoperative CT[20].

3. Radiomics feature extraction

A radiologist selected one axial image among the CT images that depicted the largest area of the lesion, under the inspection of an abdominal radiologist. The CT images were resampled by pixel spacing 1.0 mm × 1.0 mm using the BSpline interpolator of Insight Segmentation and Registration Toolkit (ITK) package (https://www.itk.org). A free-form region of interest (ROI) was drawn along the margins of the tumor using semi-automatic methods aided by the CT attenuation threshold, measured using an open-source application, Medical Image Processing, Analysis, and Visualization (MIPAV) (https://mipav.cit.nih.gov). Each selected image and ROI were thoroughly checked by another abdominal radiologist with 16-year subspecialty experience. Disagreements about the ROI were resolved by consensus-based discussion. The radiologists were blinded to the clinical and histopathologic data, except for information on the diagnosis of gastric cancer and the general location of the tumor (upper, middle, lower, or whole) based on findings of the preoperative endoscopy, since we were not evaluating the detection ability.

Pyradiomics (version 2.0.0), the open-source python package, was used to extract radiomics features, including shape-based features, first-order features, and texture features (Appendix 1). Gray value discretization was performed to a fixed bin width of 10 bins. No normalization was performed in PyRadiomics.

First-order statistics describe the distribution of pixel intensities within an CT image such as energy, entropy and kurtosis. Texture features were extracted using Gray Level Co-occurrence Matrix (GLCM) Features, Gray Level Size Zone Matrix (GLSZM) Features, Gray Level Run Length Matrix (GLRLM) Features, Gray Level Dependence Matrix (GLDM) Features.The GLCM and GLRLM features were extracted for each direction separately, after which the average value over all directions was returned as the extracted feature value, with no weighting in calculation. Feature descriptions can be found in the PyRadiomics documentation (PyRadiomics feature definitions: https://pyradiomics.readthedocs.io/en/latest/features.html)

Wavelet transformation decouples textural information by decomposing the original image in low and high frequencies. The original CT image was decomposed into four decompositions (low-high, high-high, high-low, low-low subbands) using two-dimensional coiflet wavelets. For the wavelet-filter, stationary wavelet transform was applied using the “coif1” (coiflet-1) wavelet function. Each image was filtered using either a high band-pass filter or low bandpass filter in x and y directions, yielding 4 different combinations of decompositions. First order features, GLCM features, GLRLM features, GLSZM features, and GLDM features were extracted from each original (unfiltered) as well as filtered images.

Finally, 438 tumor imaging quantifying features were obtained (94 features from original image and 86 × 4 features from the wavelet transformed images).

4. Inter-observer and inter-slice agreement for selected features

To exclude the possibility of interobserver variability affecting the ROI, another radiologist with 5 years of subspecialty experience drew ROIs in 30 randomly selected lesions to analyze interobserver reproducibility. The radiologists were blind to the clinical and histopathological data except for a

general location of the tumor based on the preoperative endoscopic finding, since the present study did not aim to evaluate the detection ability. Interobserver agreement was evaluated by the intraclass correlation coefficient (ICC) with the 95% confidence interval (CI) based on a two-way random effect model.

As only one slice with the largest section of the lesion was selected to draw ROI, inter-slice agreement amoung extracted features was anlaysed with 30 randomly chosen images for three consecutive slices including the largest section in the middle. The inter-slice ICC among the consecutive slices were calculated. The features with both inter-observer and inter-slice ICC greater than 0.75, which were suggested to be categorized into good to excellent reproducibility[22], were included in subsequent analyses.

5. Feature selection and radiomics signature building

The least absolute shrinkage and selection operator method (LASSO) Cox regression model[23] was used to select the most prognostically useful features. Then, a multiple-feature based radiomics signature, namely the radscore, was constructed for predicting survival in the training cohort. The method uses an L1 penalty to shrink some regression coefficients to exactly zero. LASSO has been extended and broadly applied to the Cox proportional hazard regression model for survival analysis with high-dimensional data. We selected λ via 1-SE (standard error) criteria, i.e., the optimal λ is the largest value for which the partial likelihood deviance is within one SE of the smallest value of partial likelihood deviance. Ten-fold cross-validation in the training set was performed to to optimize hyperparameters for model generalizability.

6. Construction of various models

All available clinical factors in the preoperative setting were included in the clinical model building. In the training cohort, the clinical model for predicting RFS was built using the multivariable Cox proportional hazards model with a

backward stepwise approach based on the Akaike information criteria. The following clinical factors are integrated in preoperative clinical model: age, sex, CEA (<5 or ≥5 U/ml), CA 19-9 (<37 or ≥37 U/ml), conventional CT features (CT size, CT-depth [nodular extramural infiltration or not], CT-LN [N0 – 1 or N2 – 3], CT Borrmann type 4), endoscopy result (tumor location [Upper/middle/lower/whole], Histological grade from biopsy tissue [Well or moderate / Poorly diffentiated], Borrmann type 4). The radiomic model was built using radscore in a univariate Cox model. The radomic score was incorperated into the clinical model to build merged clinico-radiomic (merged) model in preoperative setting to evaluate the potential value of the radscore. The performances of the three models were evaluated in the external validation cohort.

7. Radscore-based risk stratification

The potential association of radscore with RFS was assessed in training and validation cohorts using Kaplan-Meier survival analysis. The patients were stratified into high- and low-radscore groups, using a maximally selected log-rank statistic-based threshold[24]. The threshold value determined in the training cohort was applied to the validation cohort. Differences in survival distributions between the two groups were compared using log-rank tests. Subgroup analyses according to CT-Size, CT-LN, CT-Depth, CT-Type4, and adjuvant chemotherapy were performed to determine if there was any survival difference between the high- and low-radscore groups.

8. Assessment of incremental value of radiomics signature in individual RFS estimation

To demonstrate the incremental value of the radiomics signature to preoperative clinical risk factors for individualized assessment of RFS, a nomogram of the merged model was developed in the training cohort. To compare the predicted survival with the actual survival, calibration curves were generated.

To quantify the discrimination performance, the integrated area under the receiver operating characteristic curves (iAUC) were compared between the addition of radiomics and clinical parameters. The iAUC is a weighted mean of the AUC over a follow-up period to measure the model's performance in survival prediction. To quantify the improvement of usefulness added by the radiomics signature, a net reclassification improvement (NRI) calculation was also applied[25]. Finally, a decision curve analysis determined the clinical usefulness of the radiomics nomogram by quantifying the net benefits at different threshold probabilities[26].

9. Statistical analysis

Continuous variables were described using mean ± standard deviation and/or median with interquartile range and compared using independent t-tests. Categorical variables were compared using chi-squared or Fisher's exact tests. RFS was assessed by the Kaplan-Meier method, and differences in survival distributions between groups were compared using log-rank tests. The multivariate Cox proportional hazards model with backward stepwise approach was used to identify independent clinical prognostic factors for RFS. Outcomes were expressed as hazard ratios and 95% CIs. To quantify the discrimination performance, iAUC values and their differences between models were calculated using a bootstrapping method (resampled 1000 times) in training (for internal validation) and validation cohorts (for external validation). 95% CIs for iAUC values and differences were computed by the percentile method39. The iAUC difference was considered statistically significant if the 95% CI of the iAUC difference did not include zero. A P-value of <0.05 was considered statistically significant. Statistical analyses were performed using open source R software (version 3.3.2, https://www.r-project.org/, Supplementary materials). In addition, the radiomics quality score[11] was evaluated to assess the overall quality of the study in a standardized form. Overall workflow of the present study was

presented in Figure 2.

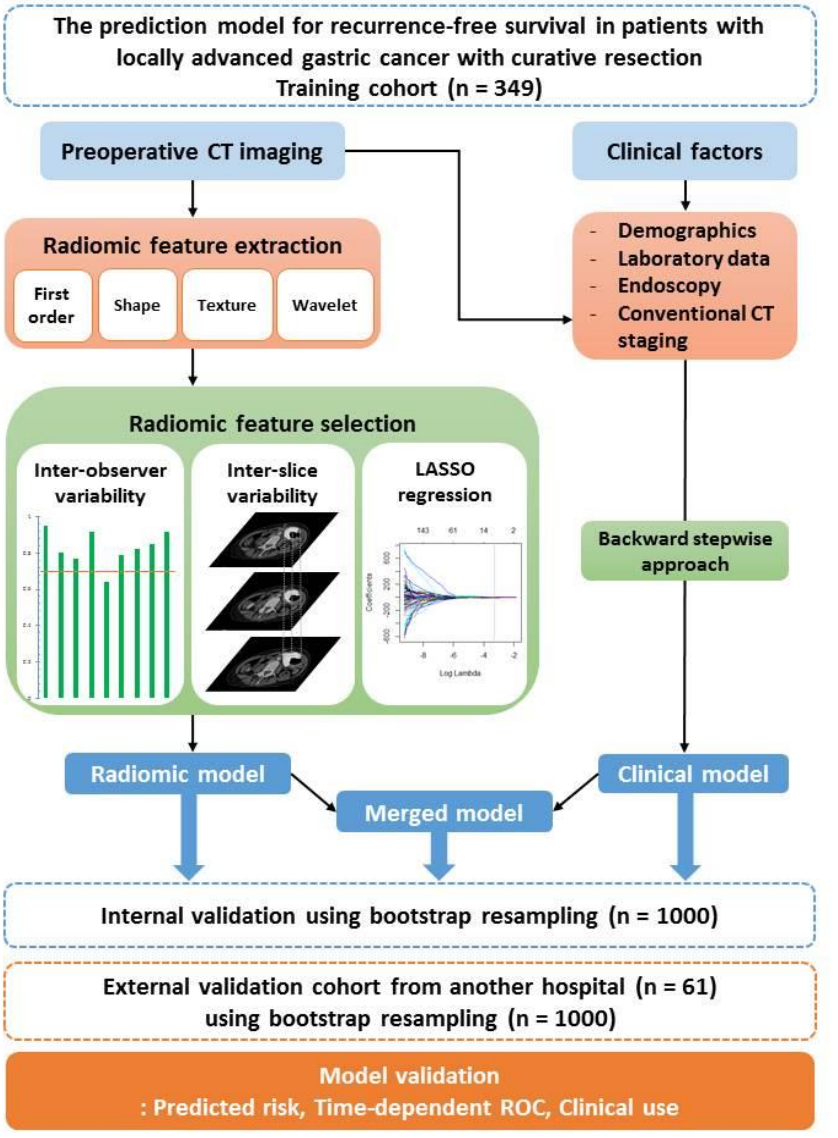


Figure 2. Workflow of the present study

The “glmnet” and “survival” package was used to perform the LASSO Cox regression and survival analysis, respectively. The ”rms” package was used to perform multivariate Cox regression and generate nomograms, and calibration

plots. The “survival ROC” package was used to perform the time-dependent receiver operating characteristic (ROC) curve analysis. The iAUC values were calculated by using the “riskset ROC” package. The calculation of net reclassification improvement was performed with the “survIDINRI” package. Decision curve analysis was performed with the function of “dca.R”.

III. RESULTS

The total radiomics quality score[11] was 15 (adherence rate 15/36, 41.7%) in 16 domains (Appendix 2).

1. Study population Characteristics

A total of 410 patients (mean age, 58.2 ± 13.0 years; 268 men) was included in the final study population, with 349 patients (mean age, 58.3 ± 12.6 years; 232 men) in the training cohort and 61 patients (mean age, 57.3 ± 15.0 years; 40 men) in the validation cohort (Figure 1). There was no significant difference between training and validation cohorts in recurrence, sex, age, carcinoembryonic antigen, carbohydrate antigen 19-9, T stage, N stage, differentiation, Lauren classification, and lymphovascular invasion. Nodular extramural infiltration in CT showed significant difference between the two cohorts (Table 2).

Table 2. Demographic and Clinical Characteristics of Patients in the Training and Validation cohort

	Training (n=349)	Validation (n=61)	p-value
Recurrence (%)	95 (27.2)	21 (33.3)	0.249
Sex (female, %)	117 (33.5)	25 (41.0)	0.259
Age (mean ± SD)	58.3 ± 12.6	57.3 ± 15.0	0.601
CEA (elevated, %)	33 (9.5)	4 (6.6)	0.466
CA 19-9	35 (10.0)	9 (13.8)	0.271
Conventional CT features			
Size (mean ± SD)	47.3 ± 24.9	50.0 ± 19.8	0.347
Tumor depth : Nodular extramural (n, %)	138 (39.5)	12 (19.7)	0.003
cN2 – 3 (n, %)	80 (22.9)	9 (14.8)	0.153
Borrmann type 4 (n, %)	31 (8.9)	8 (13.1)	0.299
Endoscopy data			
Differentiation			0.801
Well/moderate	126 (36.1)	21(34.4)	
Poorly/undifferentiated	223 (63.9)	40 (65.6)	
Location			0.127
Upper	57 (16.3)	14 (23.0)	
Middle	103 (29.5)	10 (16.4)	
Lower	180 (51.6)	34 (55.7)	
Whole	9 (2.6)	3 (4.9)	
Borrmann type 4 (n, %)	11 (3.1)	4 (6.6)	0.191
Surgical pathology data			
T stage			0.607
T2	94 (26.9)	10 (16.4)	
T3	113 (32.4)	21 (34.4)	
T4a	140 (40.1)	30 (49.2)	
T4b	2 (0.6)	0 (0.0)	

N stage			0.785
N0	126 (32.6)	20 (32.7)	
N1	70 (22.0)	10 (16.4)	
N2	70 (20.1)	12 (19.7)	
N3a	55 (16.5)	12 (19.7)	
N3b	28 (8.8)	7 (11.5)	
Differentiation			0.066
Well/moderate	122 (35.0)	14 (23.0)	
Poorly/undifferentiated	227 (65.0)	47 (77.0)	
Lauren classification			0.119
Intestinal	175 (50.1)	24 (39.3)	
Diffuse/mixed	174 (49.8)	37 (60.7)	
LV invasion	154 (44.1)	26 (42.6)	0.827
Borrmann type 4 (n, %)	25 (7.2)	7 (11.5)	0.246

SD, Standard deviation; CEA, Carcinoembryonic antigen; LV invasion, lympho-vascular invasion

In the training cohort, recurrences occurred in 95 of 349 patients (27.2%) and the 1-, 2-, and 5-year cumulative global RFS rates were 92.2%, 85.7%, and 75.1% (95% CI [89.4, 95.1], [82.1, 89.5], [70.6, 80.0]), respectively. In the validation cohort, recurrences occurred in 21 of 61 patients (33.3%) and the 1-, 2-, and 5-year cumulative global RFS rates were 86.0%, 76.5, and 64.1% (95% CI [77.5, 95.5], [66.1, 88.6], [52.2, 78.6]), respectively (Figure 3).

In the final study population of 410 patients, R0 gastrectomy with D2 lymphadenectomy was successfully performed with 140 (34.1%) total gastrectomy and 270 (64.1%) subtotal gastrectomy. TNM stage III patients represented 49.5% (203/410) while TNM stage II and I represented 36.8% (151/410) and 14.4% (59/410), respectively. Overall, 74.1% of the patients received adjuvant chemotherapy (stage III, 94.6%; stage II, 68.2%; stage I, 15.3%;).

2. Interobserver and interslice agreement

The interobserver and interslice intraclass correlation coefficient (ICC) ranges were 0.491–1.000, and 0.360–0.965, respectively (Appendix 1). Therefore, 240 features with ICC>0.75 on both interobserver and interslice reproducibility were used for the further analysis.

3. Feature selection and radiomics signature building

In the LASSO Cox regression model, a value of tuning parameter λ = 0.077 with log (λ) = -2.58 was selected by cross-validation to minimize partial likelihood deviance values among the 240 features. The optimal tuning parameter resulted in seven non-zero coefficients (Table 3, Figure 2). The radscore was calculated by following equation:

Radscore = (-8.661705*original_shape_Sphericity)
+ (-1.974956*original_glcm_Imc1)
+ (1.033112*original_glcm_Imc2)
+ (-0.517325*original_glszm_SmallAreaEmphasis)
+ (1.670901*wavelet.LH_glcm_Idmn)
+ ($5.198918*10^{-5}$ wavelet.LH_gldm_GrayLevelNonUniformity)
+ $-9.272848*10^{-9}$*wavelet.HL_glszm_LargeAreaHighGrayLevelEmphasis)

Table 3. Selected features on the LASSO Cox regression model

Selected Variables	coefficient
original_shape_Sphericity	-8.661705
original_glcm_Imc1	-1.974956
original_glcm_Imc2	1.033112
original_glszm_SmallAreaEmphasis	-0.517325
wavelet.LH_glcm_Idmn	1.670901
wavelet.LH_gldm_GrayLevelNonUniformity	$5.198918*10^{-5}$
wavelet.HL_glszm_LargeAreaHighGrayLevelEmphasis	$-9.272848*10^{-9}$

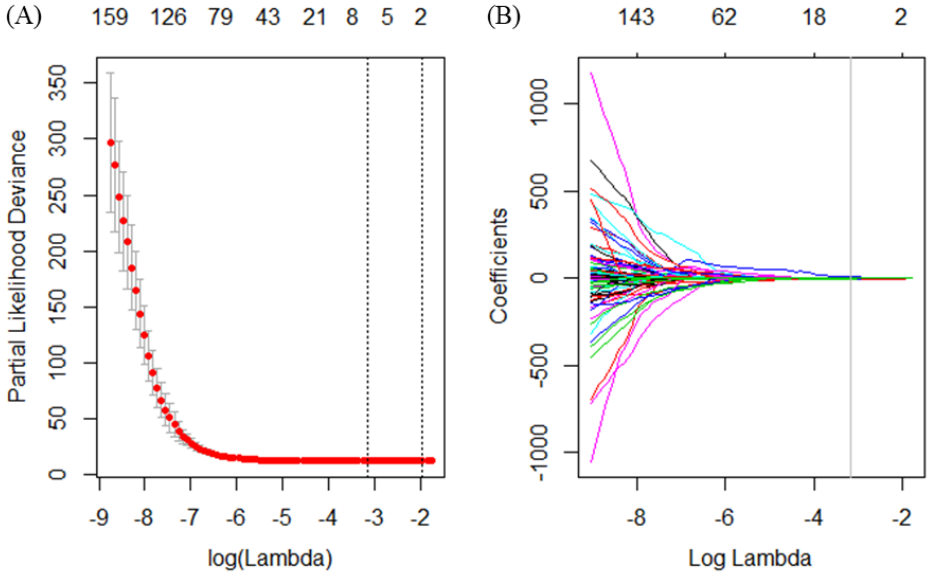


Figure 3. Texture feature selection using the least absolute shrinkage and selection operator (LASSO) Cox regression model. (A) Tuning parameter (λ) selection in the LASSO model used 10-fold cross-validation via minimum criteria. The partial likelihood deviance curve was plotted versus log (λ). Dotted vertical lines were drawn at the optimal values by using the minimum criteria and 1 standard error of the minimum criteria (the 1-SE criteria). A λ value of 0.077, with log (λ) of -2.58 was chosen (1-SE criteria) according to 10-fold cross-validation. (B) LASSO coefficient profiles of the 240 texture features. A coefficient profile plot was produced against the log (λ) sequence. A vertical line was drawn at the value selected using 10-fold cross-validation, where optimal λ resulted in seven nonzero coefficients.

Correlation matrix of the selected radiomics features was shown in Figure 4. Correlation coefficients were ranged -0.92 to 0.6.

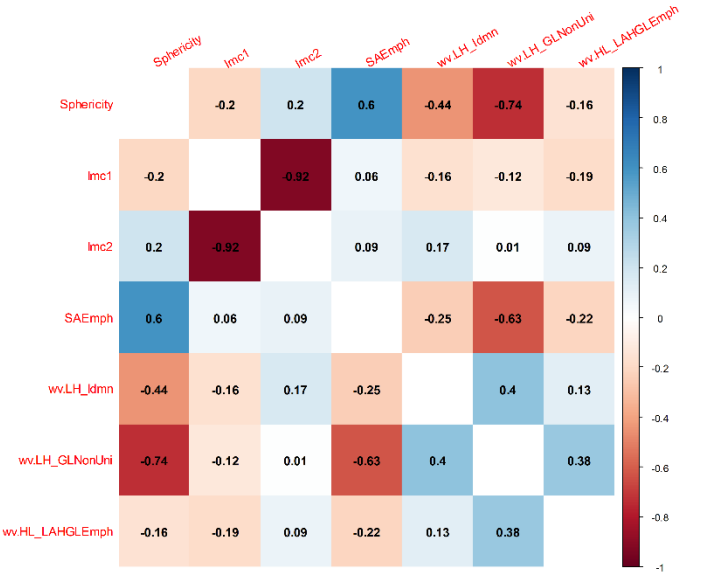


Figure 4. Correlation matrix of the selected radiomics features.
GLNonUni, GrayLevelNonUniformity; HL, Hihg-low pass filter; Idmn, Inverse difference moment normalized; IMC, informational measure of correlation; LAHGLEmph, LargeAreaHighGrayLevelEmphasis; LH, Low-high pass filter; SAEmph, SmallAreaEmphasis; wv, wavelet

Histograms for the radscore in training and validation cohort were shown in Figure 5.

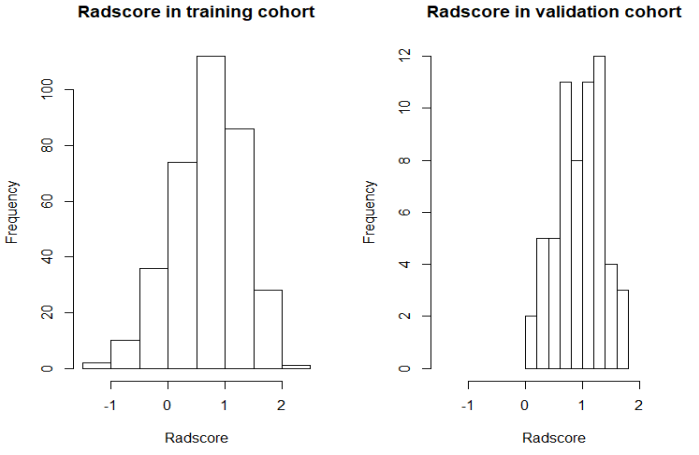


Figure 5. Histograms for the radscore in the training and validation cohort

The defined radscore showed significant difference by tumor recurrence in both training and validation cohort (log-rank test, $p < 0.001$). The optimum cutoff generated by a maximally selected log-rank statistic were 1.116 (Figure 6). Accordingly, patients were classified into a low-radiomics score group (radscore < 1.116) and a high-radiomics score group (radiomics score ≥ 1.116).

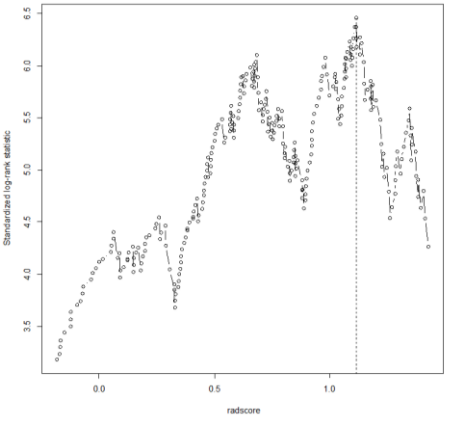


Figure 6. Optimal cutoff selection of radscore (1.116) for high and low risk for tumor recurrence, generated by a maxiamlly selected lor-rak statistic.

4. Construction and validation of prediction models

Performance of the clinical, radiomics, and merged models were evaluated. Among preoperative clinical factors, Tumor depth on CT (CT-Depth) and Tumors classified as Borrmann type 4 on CT (CT-Type 4) were identified as independent factors for predicting RFS using backward stepwise approach (Table 4).

Table 4. Preoperative clinical factors for predicting tumor recurrence-free survival

Clinical feature		**Univariate analysis**		**Multivariate analysis***	
		HR (95% CI)	P value	HR (95% CI)	P value
Age	≤ 60	Reference			
	> 60	1.326 (0.875 – 2.012)	0.184		
Sex	Male	Reference			
	Female	1.189 (0.773 – 1.828)	0.431		
CEA	< 5 U/ml	Reference			
	≥ 5 U/ml	1.557 (0.828 – 2.927)	0.169		
CA 19-9	< 37 U/ml	Reference			
	≥ 37 U/ml	2.143 (1.189 – 3.863)	0.011†		
CT-Size	≤ 4cm	Reference			
	> 4cm	2.469 (1.513-4.030)	<0.001 †		
CT-Depth	Nodular extramural infiltration (-)	Reference		Reference	
	Nodular extramural	2.103 (1.385 – 3.194)	<0.001 †	1.899 (1.237 – 2.915)	0.003†

	infiltration (+)			
CT-LN status	cN0 or cN1	Reference		
	cN2 or cN3	1.696 (1.079 – 2.666)	0.022†	
CT-Borrmann type	Type 1,2, or 3	Reference		Reference
	Type 4	2.646 (1.539 – 4.549)	<0.001 †	2.174 (1.247 – 3.789) 0.006†
Endoscopy-Location	Upper	Reference		
	Middle	2.059 (0.938 – 4.518)	0.072	
	Lower	2.057 (0.973 – 4.349)	0.059	
	Whole	5.155 (1.686 – 15.764)	0.004†	
Endoscopy-Histological grade	Well or moderate.	Reference		
	Poorly differentiated	1.328 (0.841 – 2.097)	0.224	
Endoscopy-Borrmann type	Type 1,2, or 3	Reference		
	Type 4	1.884 (0.764 – 4.645)	0.169	

HR, hazard ratio; CI, confidence interval; CEA, carcinoembryonic antigen; CA, cancer antigen
* The multivariate regression model was built using backward stepwise approach with Akaike information criteria.
† Statistically significant

The radscore prognostic accuracy on time-dependent ROC curves as measured by area under the curves at 1, 2, and 5 years were 0.719, 0.748, and 0.733 in training and 0.795, 0.824, and 0.878, in validation cohorts, respectively (Figure 7).

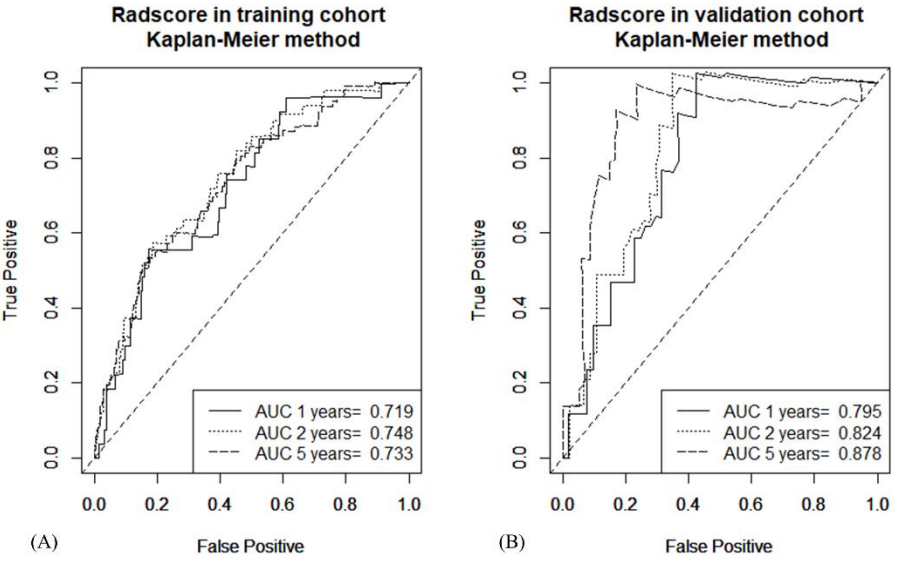


Figure 7. Survival receiver operating characteristic curves at 1, 2, and 5 years with the radscore, (A) in training cohort and (B) in validation cohort

The iAUC values for RFS prediction in the internal validation were: 0.616, 95% CI [0.570, 0.663] in clinical; 0.714, 95% CI [0.667, 0.759] in radiomic; and 0.719, 95% CI [0.674, 0.764] in merged models. In external validation, the iAUC values were: 0.584, 95% CI [0.544, 0.636] in clinical; 0.652, 95% CI [0.628, 0.674] in radiomic; and 0.651, 95% CI [0. 630, 0.673] in merged models, respectively (Table 5).

Table 5. Model performances measured by iAUC for prediction of recurrence-free survival

	Internal validation			External validation		
Model	iAUC (95% CI)	iAUC Difference*	P value	iAUC (95% CI)	iAUC Difference*	P value
Clinical	0.616 (0.570, 0.663)	-	-	0.594 (0.544, 0.636)	-	-
Radiomic	0.714 (0.667, 0.759)	0.098	<0.001†	0.652 (0.628, 0.674)	0.056	0.010†
Clinico-radiomic	0.719 (0.674, 0.764)	0.102	<0.001†	0.651 (0.630, 0.673)	0.057	0.014†

iAUC, the integrated area under the receiver operating characteristic curve; CI, confidence interval
* comparison with clinical model
† Statistically significant

5. Radscore performance evaluation and risk stratification

The patients were classified into low- and high-risk groups based on radscore cutoffs (1.116) selected from the training set using maximally selected log-rank statistics. In both training and validation cohorts, high-risk patients showed significantly lower RFS than low-risk patients. RFS hazard ratios, hazard ratios were 4.209 (95%CI [2.787, 6.357], p<0.001) and 22.061 (95%CI [5.571, 87.36], p<0.001) in training and validation cohorts, respectively (Figure 8).

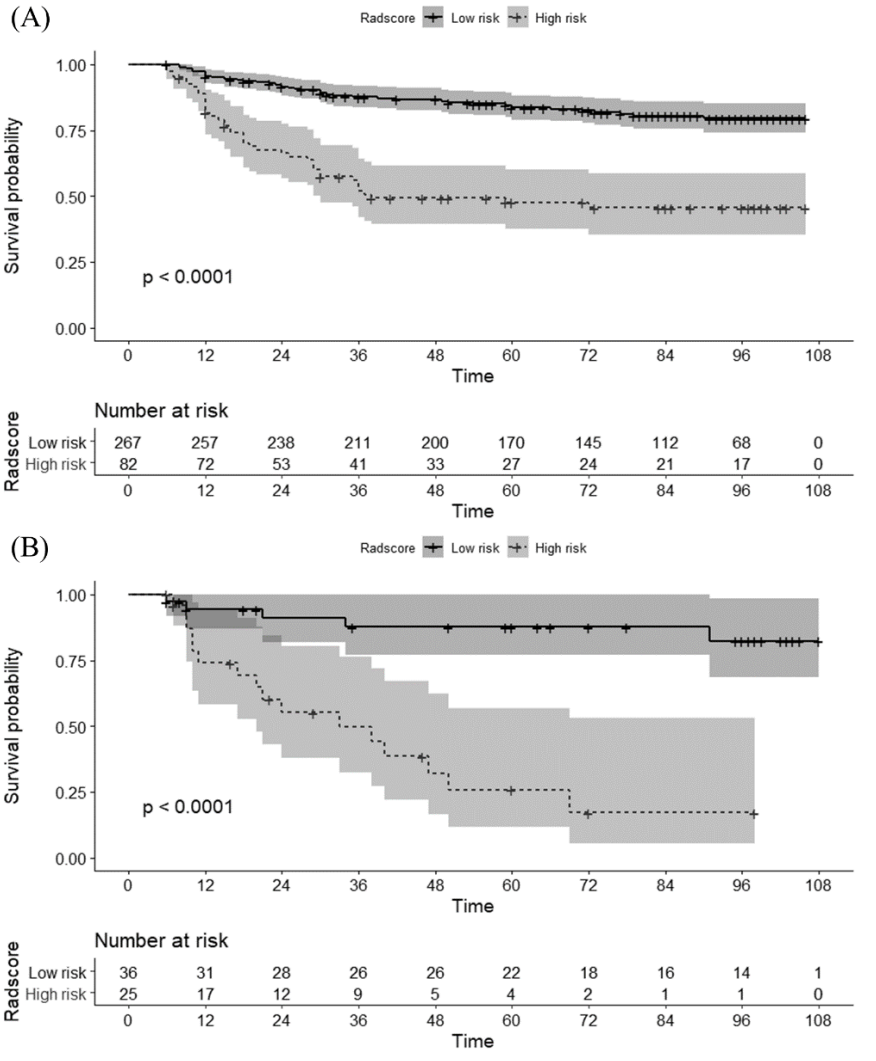


Figure 8. Kaplan-Meier curves and risk tables for recurrence-free survival (RFS) from (A) the training (n = 349) and (B) validation (n = 61) cohorts. Patients were stratified on the basis of the cutoff (radscore = 1.116) to maximize log-rank statistic. The radscore significantly stratified the patients into low- and high-risk groups for RFS in the training cohort ($p < 0.001$) and the validation cohort ($p < 0.001$). Shaded areas represent 95% confidence intervals.

Figure 9 shows estimates of the baseline survival function in training cohort and validation cohort

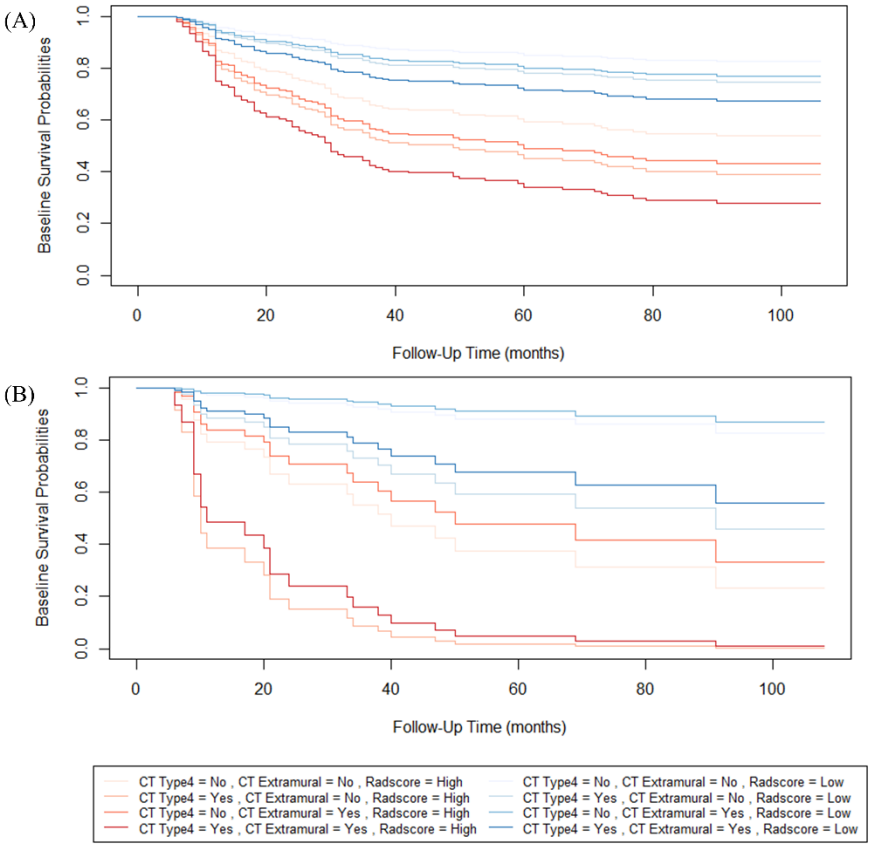


Figure 9. Estimates of the baseline survival function in (A) training cohort and (B) validation cohort.

To assess the ability of radiomics to predict early recurrence, patients who underwent follow-up for more than two years were divided into two groups based on recurrence within two years. There was a significant difference in the radiomic scores between these two groups in both training (0.96 ± 0.50 vs. 0.56 ± 0.59; p<0.001) and validation cohorts (1.20 ± 0.41 vs. 0.76 ± 0.33; p=0.014). When patients were dichotomized according to CT-Size, CT-Depth, and CT-Type4, and

adjuvant chemotherapy, the Kaplan-Meier curves of the high- and low-radscore groups showed a p value <0.05 in the validation group (Figures 10 – Figure 13). However, the Kaplan-Meier curves for RFS of the high- and low-radscore groups were not significantly different in the CT-LN (+) group of the validation cohort (*p*=0.233) (Figure 14).

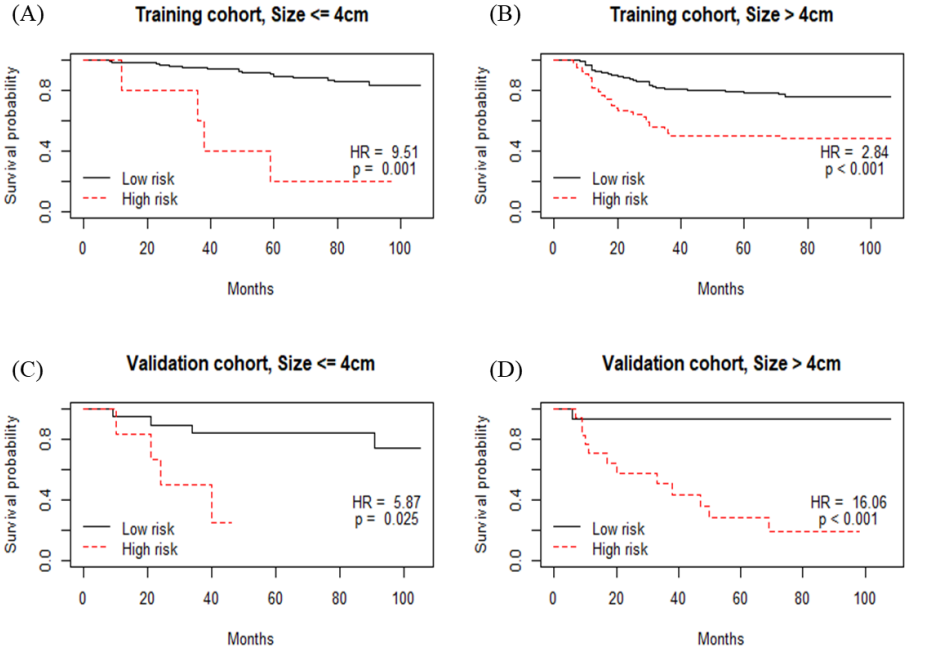


Figure 10. Kaplan-Meier survival analysis of recurrence-free survival according to the radiomics score classifier in CT-Size subgroups. (A) Training cohort, size < 4 cm on CT (n = 136). (B) Training cohort, size ≥ 4 cm on CT (n = 213). (C) Validation cohort, size < 4 cm (n = 27). (D) Validation cohort, size ≥ 4 cm on CT (n = 34).

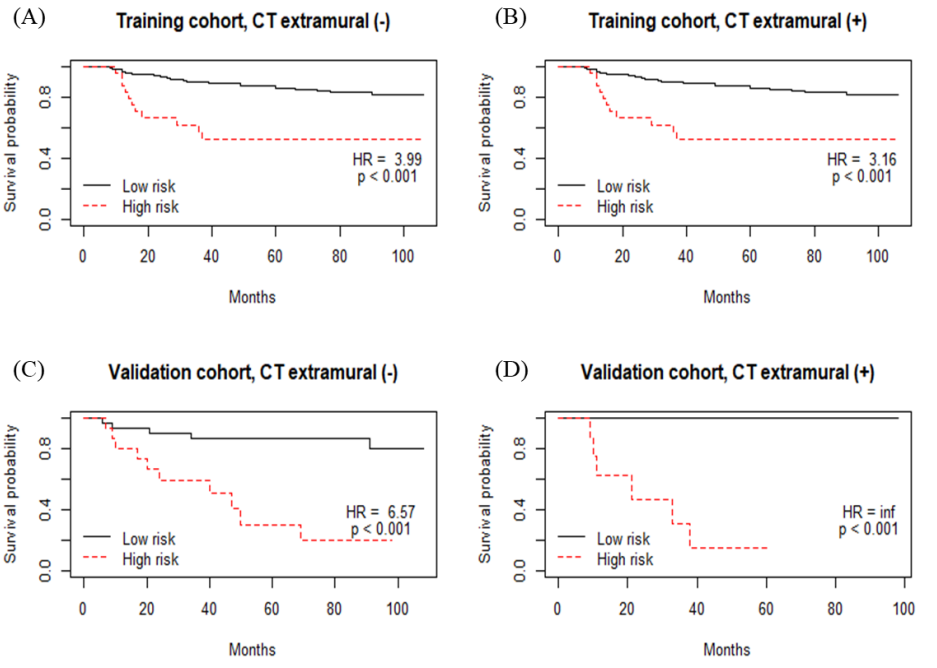


Figure 11. Kaplan-Meier survival analysis of recurrence-free survival according to the radiomics score classifier in CT-depth subgroups. (A) Training cohort, extramural nodular infiltration (-) on computed tomography (CT) (n = 211). (B) Training cohort, extramural nodular infiltration (+) on CT (n = 138). (C) Validation cohort, extramural nodular infiltration (-) (n = 49). (D) Validation cohort, extramural nodular infiltration (+) on CT (n = 12).

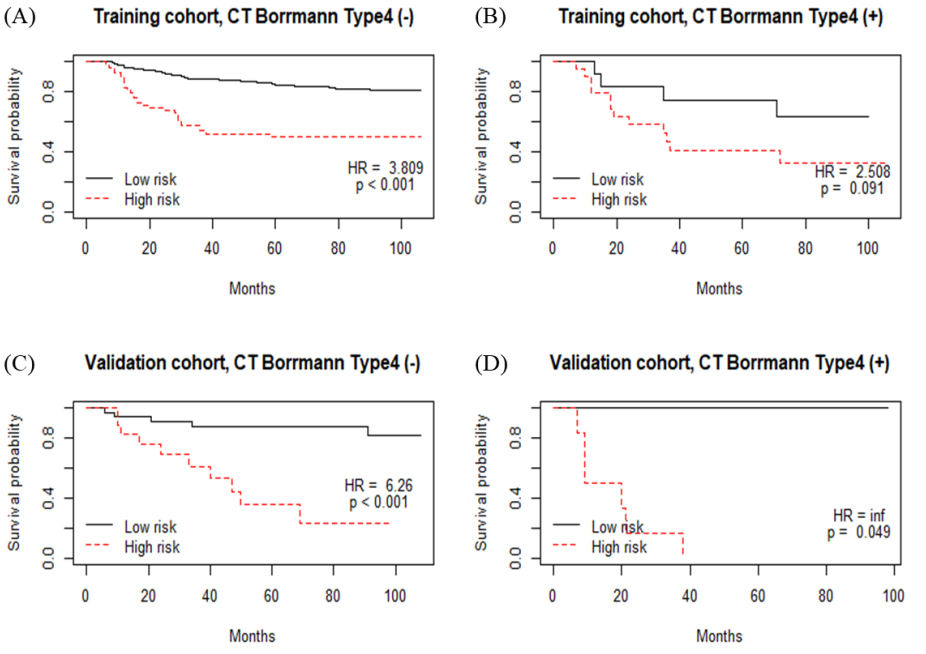


Figure 12. Kaplan-Meier survival analysis of recurrence-free survival according to the radiomics score classifier in CT-type4 subgroups. (A) Training cohort, Borrmann type 4 (-) on computed tomography (CT) (n = 318). (B) Training cohort, Borrmann type 4 (+) on CT (n = 31). (C) Validation cohort, Borrmann type 4 (-) (n = 53). (D) Validation cohort, Borrmann type 4 (+) on CT (n = 8).

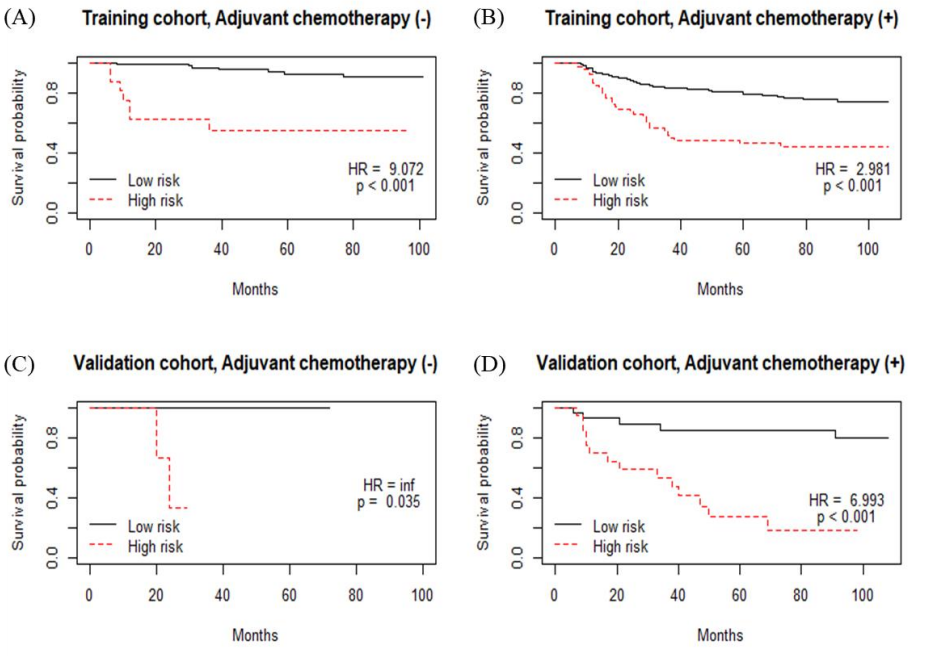


Figure 13. Kaplan-Meier survival analysis of recurrence-free survival according to the radiomics score classifier in adjuvant chemotherapy subgroups. (A) Training cohort, Adjuvant chemotherapy (-) on computed tomography (CT) (n = 110). (B) Training cohort, Adjuvant chemotherapy (+) on CT (n = 239). (C) Validation cohort, Adjuvant chemotherapy (-) (n = 9). (D) Validation cohort, Adjuvant chemotherapy (+) on CT (n = 52).

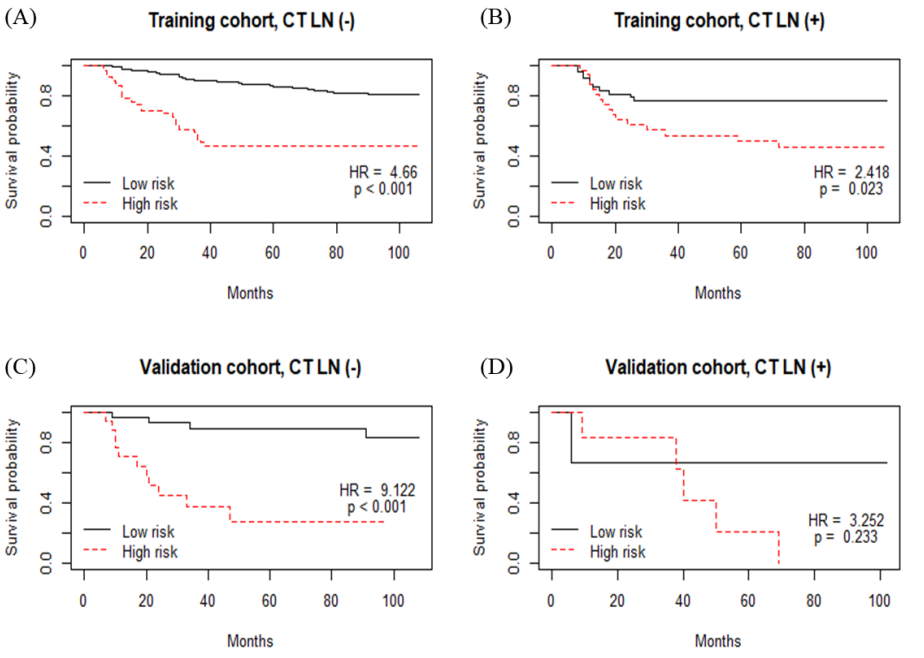


Figure 14. Kaplan-Meier survival analysis of recurrence-free survival according to the radiomics score classifier in CT-LN subgroups. (A) Training cohort, lymph node (LN) stage 0 or 1 on CT (n = 269). (B) Training cohort, LN stage 2 or over on CT (n = 80). (C) Validation cohort, LN stage 0 or 1 on CT (n = 52). (D) Validation cohort, LN stage 2 or over on CT (n = 9).

6. Incremental value of radiomics score in individual RFS estimation

The radiomics nomogram for RFS at 1, 2, and 5 years is presented in Figure 15. The calibration curves of the nomograms at 5 years are shown in Figure 16, with moderate agreement among the estimations with the nomogram of the merged model and actual observations in the training and validation cohorts. Compared to clinical model, a higher overall net benefit of the radiomics nomogram was identified in the decision curve analysis across the majority of the range of reasonable threshold probabilities (Figure 17). NRI (Figure 18) also

showed a better recurrence risk prediction in the merged model compared to the clinical model (NRI [95% CI] 0.350 [0.224, 0.463], $p < 0.001$, in training cohort; 0.350 [0.224, 0.463], $p = 0.02$, in validation cohort)

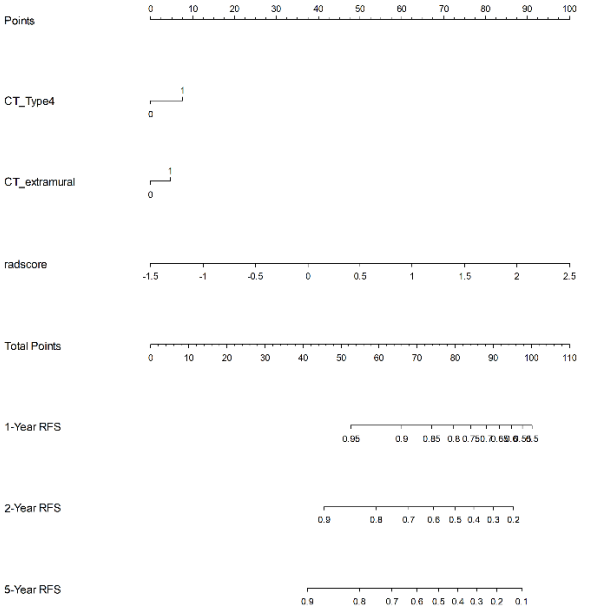


Figure 15. Radiomics nomogram for recurrence free survival (RFS) at 1, 2 and 5 years.

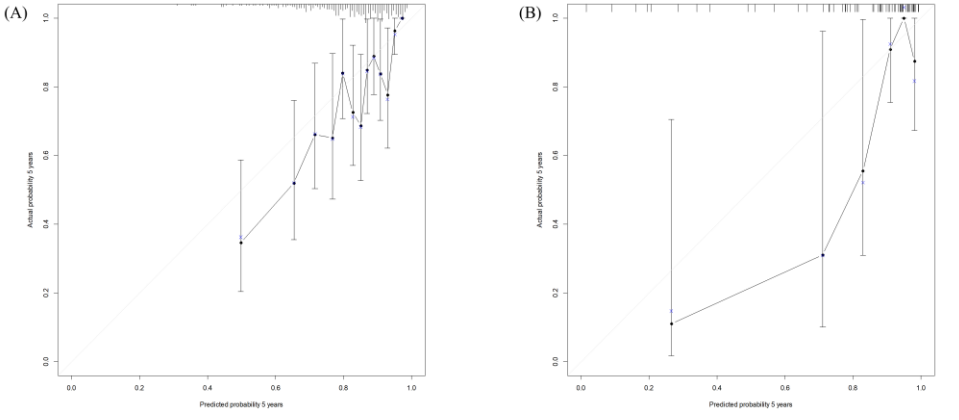


Figure 16. Calibration curve at 5 years after curative surgery of locally advanced gastric cancer (A) in training cohort and (B) in validation cohort.

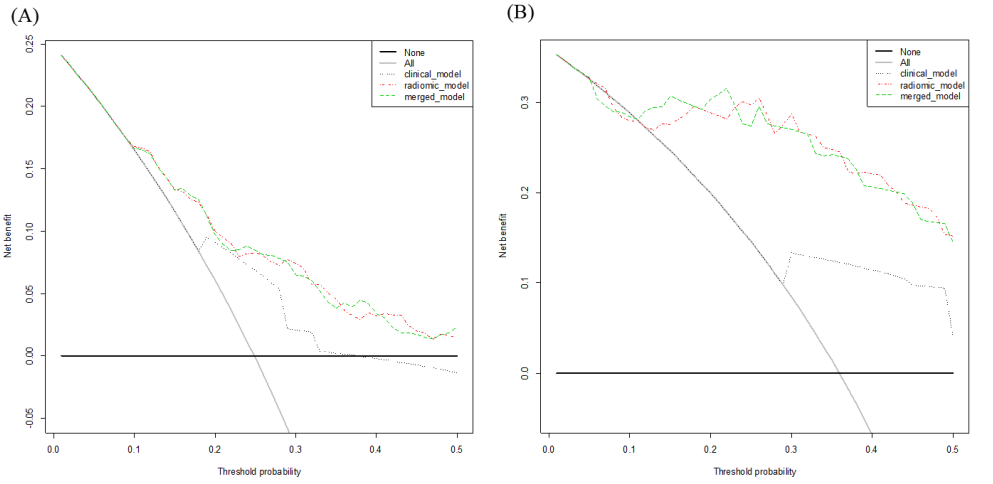


Figure 17. Decision curve analysis for the clinical, radiomics, and merged model (A) in training cohort and (B) in validation cohort. The y-axis measures the net benefit, summing the true positive results and subtracting false positive results weighted by a factor of relative harm from an undetected cancer and unnecessary treatment.

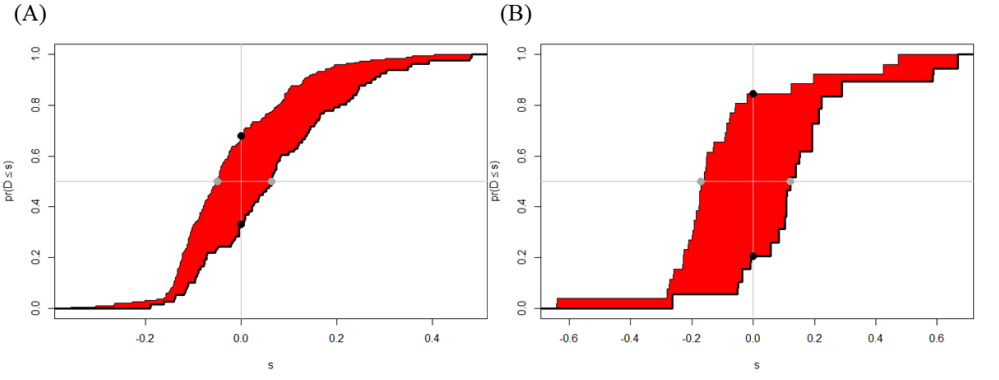


Figure 18. Curve for net reclassification improvement of the merged model compared to the clinical model (A) in training cohort (NRI [95% CI] 0.350 [0.224, 0.463], $p < 0.001$) and (B) in validation cohort (NRI [95% CI] 0.350 [0.224, 0.463], $p = 0.02$).

Examples of patients with high and low risk by the radscore are shown in Figure 19.

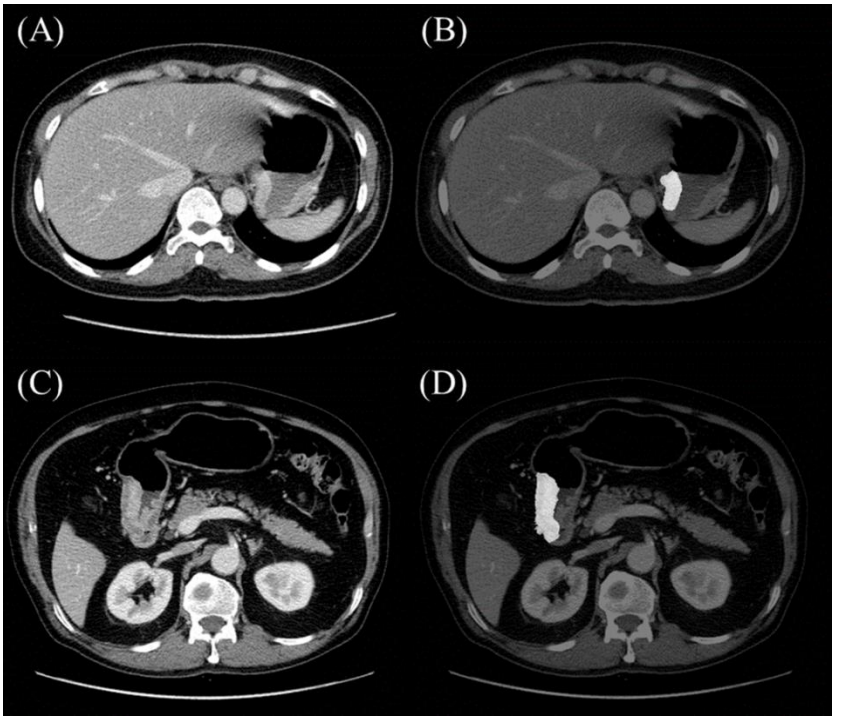


Figure 19. Patients with locally advanced gastric cancer whose recurrence risk was stratified into high and low risk (the radscore cutoff 1.116) (A) CT images on portal venous phase, (B) tumor segmentation in a 52-year-old woman with nodular extramural infiltration on CT whose radscore was 0.98, low risk group. Preoperative carcinoembryonic antigen (CEA) and cancer antigen (CA) 19-9 was within the normal limit. Surgical pathology revealed tumor-node-metastasis (TNM) stage IIb with T4a and N0. There was no tumor recurrence during 96 months after surgery. (C) CT images on portal venous phase, (D) tumor segmentation in a 67-year-old man without nodular extramural infiltration on CT whose radscore was 1.66, high risk group. Preoperative CEA and CA 19-9 was within the normal limit. Surgical pathology revealed TNM stage IIIc with T4a and N3a. Liver metastasis occurred at 12 months after surgery.

IV. DISCUSSION

To predict prognosis for RFS in patients with LAGC using preoperative CT, we identified the radscore consisting of seven radiomic features and verified its value through external validation. In preoperative setting, the radscore was an independent prognostic factor in both training and validation cohorts and showed good RFS predicting performance in LAGC and outperformed the clinical model alone. The merged model showed significantly higher prognostic performance than the clinical model, indicating that the radiomic model added value to the clinical model-based prediction. The results support the clinical application of radiomics in providing additional information for LAGC-treatment decision-making in the preoperative setting, without any additional invasive procedure. Moreover, the high performance of radiomic model on risk-stratification may help in selecting candidates for investigational treatments.

Even though pathologic TNM stage is still the most reliable prognostic factor for long-term outcomes of gastric cancer[27,28], such data can only be obtained after the completion of surgery. Preoperative treatment could alter the pathologic stage, therefore, development of non-invasive biomarkers that provide guidance for adjusting the therapeutic approach is essential for LAGC. Several studies have recently highlighted the prognostic potential of texture analysis or radiomics in patients with gastric cancer[15,18,29]. A large-scale retrospective study demonstrated that radiomics signature has more prognostic value than clinicopathological features[18]. However, their study population included a considerable proportion of patients with early stage gastric cancer or distant metastatic stage. Early gastric cancer is known to have an excellent prognosis without needing chemotherapy and the AGC with distant metastasis is known to require systemic chemotherapy without resection surgery[30,31]. We targeted LAGC since a variety of treatments have been proposed but gray zones persist in treatment determination. Our study revealed that radiomics had higher prognostic performance than the clinical model, suggesting that radiomics could be a

practical imaging biomarker for patients with LAGC in a preoperative setting. However, the merged model did not perform better than the radiomics model. This might be attributed to the possibility that most characteristics of the clinical model (mainly based on conventional imaging characteristics), were already reflected in the radscore.

The usefulness of neoadjuvant chemotherapy in LAGC is still controversial. Large-scale phase III trials in Europe have reported that perioperative chemotherapy has survival benefits over surgical treatment alone[32,33]. However, in majority of cases in these studies, a proper lymphadenectomy was not performed during the surgery. Furthermore, a lack of information about initial tumor staging before treatment could lead to selection bias[10]. In Korea and Japan, D2 lymphadenectomy is generally performed along with gastrectomy in resectable advanced gastric cancer, therefore the usefulness of neoadjuvant chemotherapy has not been concluded yet[30,31]. This controversy could be resolved through risk stratification, i.e., by identifying gastric cancer cases with a high risk of recurrence. In our study, the radscore was not only successfully dichotomized into high- and low-risk groups, but also verified by external validation. Therefore, the radscore could offer guidance for therapeutic strategies depending on recurrence risk, thereby improving the clinical outcome. Particularly, LAGC patients classified as high-risk based on the radscore may be ideal candidates for neoadjuvant treatment, given that its potential benefits outweigh the morbidity risk and higher treatment cost. However, since patients with neoadjuvant treatment were not included in this study, its efficacy could not be assessed using the radscore. Further study is required to evaluate the correlation between the radscore and response to neoadjuvant therapy.

We evaluated the characteristics and applicability of the radscore in various clinical conditions. The radscore showed successful risk stratification in each subgroup dichotomized according to tumor size, tumor depth, or Borrmann types on preoperative CT. This indicates that the radscore could provide a more

sophisticated risk stratification independent of known clinical prognostic factors. The radscore might help predict prognosis of LAGC, regardless of the outcome of current preoperative clinical staging. Interestingly, the radscore could significantly distinguish patients into two risk groups, only in the clinically LN negative subgroup, but not in the clinically LN positive subgroup. However, the number of patients with LN stage 2 or over on CT in the validation cohort was too small (n = 9) to have statistical power. Moreover, clinical N staging by preoperative CT is very limited in LAGC patients[10], even though relatively satisfactory sensitivity and specificity have been reported for ≥ pN2 stage[21], which was used as the cutoff in our study. Although the targets of radiomics were limited to primary tumors, and since metastatic LN was not included in this radiomics analysis, it is needed further study to confirm the prognostic power of the radscore in the LN positive group.

Among the seven features in the radscore, the sphericity, which is selected from the shape features, quantifies the roundness of the shape of the tumor region relative to a circle. Any lesion with low sphericity could be associated with a flat or infiltrative tumor, which has been regarded as Borrmann type 4. Two GLCM-related features in the radscore, informational measure of correlation 1 (IMC1) and IMC2, measure the complexity of the texture patterns[22]. Selected Small-Area-Emphasis of GLSZM features measures the distribution of small size zones, with a greater value indicative of smaller size zones and finer textures[22]. These GLCM and GLSZM features have specific mathematical formula measuring different aspects of textural heterogeneity within the tumor, e.g. tissue necrosis. These GLCM- or GLSZM-based texture features reflecting the interaction between neighboring pixels have shown better quantification of tumor texture and heterogeneity than histogram-based features[34]. In addition, three features from the wavelet decompositions of original images are also included in our radscore. By focusing on different frequency ranges within the

tumor, features from wavelet decompositions might be able to reveal the characteristics of tumors that did not appear in the original image.

This study has several limitations. First, it was a retrospective study with a relatively small sample size; however, the number was similar to those in previous radiomic studies[35,36]. Moreover, external validation with cohort from the spatially separate hospital was performed to overcome this limitation. Future study with a larger sample for both training and validation is required for a robust prediction model. Second, the recurrence rate in the training cohort was 27.2%, imbalanced data. Any approach to rebalance the dataset was not performed to preserve representative of the clinical situation. In addition, in this study, LASSO Cox regression was performed to build the radscore, instead of machine learning technique. Third, the proportion of cases with nodular infiltration on CT was different between the training and validation cohorts, presumably due to different scale and clinical settings of the two spatially separate hospitals. Nevertheless, the radscore showed significant risk stratification in both cohorts. Fourth, since only patients who did not receive neoadjuvant chemotherapy were included, the benefits of neoadjuvant therapy in high versus low radscore groups could not be evaluated. Further study in a large prospective cohort, randomized by neoadjuvant chemotherapy status is needed to integrate this technology into clinical practice. Fifth, clustering for radiomics features to remove redundancy was not performed before model building and highly correlating features (Figure 4), such as IMC1 and IMC2, were included. These features were linear combined in the radscore and the effect of redundancy might be small. Sixth, images from different machines or manufacturers of CT were included in the training cohort and fourteen patients were excluded with poor quality of CT. To minimize variability from different CT scanners, we used a uniform acquisition protocol and resampled the images into the same pixel spacing. Moreover, the validation was performed on the cohort from a different hospital. Standardized protocol for different CT scanners is required for future study and application of radiomics

prediction model in clinical setting. Seventh, feature extraction was performed from a single slice with the largest lesion, similar to the previous study[18]. Although tumor evaluation on a single CT section might not be representative of the entire tumor characteristics[37], previous studies reported that two-dimensional features showed prognostic performance comparable with three-dimensional segmentations in non-small cell lung cancer and rectal cancer[38,39]. However, there is still controversy whether two-dimensional segmentation can replace recommended three-dimensional segmentation which allows comprehensive assessment of whole tumor. Lastly, the tumors were outlined semi-automatically, which may be time-consuming and user-dependent in terms of selecting the slice containing the largest area of the tumor and region of interest placement. To reduce variability in these processes, only features with excellent inter-slice and inter-reader ICCs were included for analysis. In future studies, automated 3D tumor segmentation based on deep learning would allow further automation of the workflow, minimize user bias, and enable larger studies.

V. CONCLUSION

The radiomics signature based on preoperative CT images is a possible preoperative imaging biomarker that can improve RFS prediction of the preoperative clinical profile in LAGC. The ability of radiomic signatures to identify high-risk LAGC patients may be helpful in selecting appropriate candidates for neoadjuvant therapy.

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA Cancer J Clin 2016;66:7-30.
2. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin 2015;65:87-108.
3. Choi AH, Kim J, Chao J. Perioperative chemotherapy for resectable gastric cancer: MAGIC and beyond. World Journal of Gastroenterology : WJG 2015;21:7343-8.
4. Lee J, Lim DH, Kim S, Park SH, Park JO, Park YS, et al. Phase III Trial Comparing Capecitabine Plus Cisplatin Versus Capecitabine Plus Cisplatin With Concurrent Capecitabine Radiotherapy in Completely Resected Gastric Cancer With D2 Lymph Node Dissection: The ARTIST Trial. Journal of Clinical Oncology 2012;30:268-73.
5. Bang YJ, Kim YW, Yang HK, Chung HC, Park YK, Lee KH, et al. Adjuvant capecitabine and oxaliplatin for gastric cancer after D2 gastrectomy (CLASSIC): a phase 3 open-label, randomised controlled trial. Lancet 2012;379:315-21.
6. Aoyama T, Yoshikawa T, Watanabe T, Hayashi T, Ogata T, Cho H, et al. Survival and prognosticators of gastric cancer that recurs after adjuvant chemotherapy with S-1. Gastric Cancer 2011;14:150-4.
7. D'Ugo D, Rausei S, Biondi A, Persiani R. Preoperative treatment and surgery in gastric cancer: friends or foes? Lancet Oncol 2009;10:191-5.
8. Boige V, Pignon J, Saint-Aubert B, Lasser P, Conroy T, Bouché O, et al. Final results of a randomized trial comparing preoperative 5-fluorouracil (F)/cisplatin (P) to surgery alone in adenocarcinoma of stomach and lower esophagus (ASLE): FNLCC ACCORD07-FFCD 9703 trial. 2007;25:4510-.
9. Lim JS, Yun MJ, Kim M-J, Hyung WJ, Park M-S, Choi J-Y, et al. CT and PET in Stomach Cancer: Preoperative Staging and Monitoring of

Response to Therapy. RadioGraphics 2006;26:143-56.

10. Park SR, Kim MJ, Ryu KW, Lee JH, Lee JS, Nam BH, et al. Prognostic value of preoperative clinical staging assessed by computed tomography in resectable gastric cancer patients: a viewpoint in the era of preoperative treatment. Ann Surg 2010;251:428-35.
11. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017;14:749-62.
12. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology 2016;278:563-77.
13. Aerts HJ. The Potential of Radiomic-Based Phenotyping in Precision Medicine: A Review. JAMA Oncol 2016;2:1636-42.
14. Limkin EJ, Sun R, Dercle L, Zacharaki EI, Robert C, Reuze S, et al. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. Ann Oncol 2017;28:1191-206.
15. Giganti F, Antunes S, Salerno A, Ambrosi A, Marra P, Nicoletti R, et al. Gastric cancer: texture analysis from multidetector computed tomography as a potential preoperative prognostic biomarker. Eur Radiol 2017;27:1831-9.
16. Liu S, Liu S, Ji C, Zheng H, Pan X, Zhang Y, et al. Application of CT texture analysis in predicting histopathological characteristics of gastric cancers. Eur Radiol 2017;27:4951-9.
17. Liu S, Zheng H, Zhang Y, Chen L, Guan W, Guan Y, et al. Whole-volume apparent diffusion coefficient-based entropy parameters for assessment of gastric cancer aggressiveness. J Magn Reson Imaging 2018;47:168-75.
18. Jiang Y, Chen C, Xie J, Wang W, Zha X, Lv W, et al. Radiomics signature of computed tomography imaging for prediction of survival and chemotherapeutic benefits in gastric cancer. EBioMedicine 2018;36:171-

82.

19. Chang JS, Lim JS, Noh SH, Hyung WJ, An JY, Lee YC, et al. Patterns of regional recurrence after curative D2 resection for stage III (N3) gastric cancer: implications for postoperative radiotherapy. Radiother Oncol 2012;104:367-73.
20. Park CJ, Seo N, Hyung WJ, Koom WS, Kim HS, Kim MJ, et al. Prognostic significance of preoperative CT findings in patients with advanced gastric cancer who underwent curative gastrectomy. PLoS One 2018;13:e0202207.
21. Ohashi M, Morita S, Fukagawa T, Wada T, Kushima R, Onaya H, et al. Evaluation of 64-Channel Contrast-Enhanced Multi-detector Row Computed Tomography for Preoperative N Staging in cT2-4 Gastric Carcinoma. World J Surg 2016;40:165-71.
22. Zwanenburg A, Vallières M, Abdalah MA, Aerts H, Andrearczyk V, Apte A, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. Radiology 2020;295:328-38.
23. Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2011;73:273-82.
24. Hothorn T, Lausen B. On the exact distribution of maximally selected rank statistics. Computational Statistics & Data Analysis 2003;43:121-37.
25. Pencina MJ, D'Agostino Sr RB, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. 2011;30:11-21.
26. Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. BMC Med Inform Decis Mak 2008;8:53.

27. Ji X, Bu Z-D, Yan Y, Li Z-Y, Wu A-W, Zhang L-H, et al. The 8th edition of the American Joint Committee on Cancer tumor-node-metastasis staging system for gastric cancer is superior to the 7th edition: results from a Chinese mono-institutional study of 1663 patients. 2018;21:643-52.
28. Neves Filho EHC, de Sant'Ana RO, Nunes LVSC, Pires APB, da Cunha MdPSS. Histopathological regression of gastric adenocarcinoma after neoadjuvant therapy: a critical review. 2017;125:79-84.
29. Li Z, Zhang D, Dai Y, Dong J, Wu L, Li Y, et al. Computed tomography-based radiomics for prediction of neoadjuvant chemotherapy outcomes in locally advanced gastric cancer: A pilot study. Chinese journal of cancer research = Chung-kuo yen cheng yen chiu 2018;30:406-14.
30. Guideline Committee of the Korean Gastric Cancer Association DWG, Review P. Korean Practice Guideline for Gastric Cancer 2018: an Evidence-based, Multi-disciplinary Approach. Journal of gastric cancer 2019;19:1-48.
31. Japanese Gastric Cancer A. Japanese gastric cancer treatment guidelines 2014 (ver. 4). Gastric cancer : official journal of the International Gastric Cancer Association and the Japanese Gastric Cancer Association 2017;20:1-19.
32. Cunningham D, Allum WH, Stenning SP, Thompson JN, Van de Velde CJ, Nicolson M, et al. Perioperative chemotherapy versus surgery alone for resectable gastroesophageal cancer. N Engl J Med 2006;355:11-20.
33. Ychou M, Boige V, Pignon JP, Conroy T, Bouche O, Lebreton G, et al. Perioperative chemotherapy compared with surgery alone for resectable gastroesophageal adenocarcinoma: an FNCLCC and FFCD multicenter phase III trial. J Clin Oncol 2011;29:1715-21.
34. Davnall F, Yip CS, Ljungqvist G, Selmi M, Ng F, Sanghera B, et al. Assessment of tumor heterogeneity: an emerging imaging tool for

clinical practice? Insights Imaging 2012;3:573-89.

35. Bae S, Choi YS, Ahn SS, Chang JH, Kang SG, Kim EH, et al. Radiomic MRI Phenotyping of Glioblastoma: Improving Survival Prediction. Radiology 2018;289:797-806.

36. Kim S, Shin J, Kim DY, Choi GH, Kim MJ, Choi JY. Radiomics on Gadoxetic Acid-Enhanced Magnetic Resonance Imaging for Prediction of Postoperative Early and Late Recurrence of Single Hepatocellular Carcinoma. Clin Cancer Res 2019;25:3847-55.

37. Ng F, Kozarski R, Ganeshan B, Goh V. Assessment of tumor heterogeneity by CT texture analysis: can the largest cross-sectional area be used as an alternative to whole tumor analysis? Eur J Radiol 2013;82:342-8.

38. Shen C, Liu Z, Guan M, Song J, Lian Y, Wang S, et al. 2D and 3D CT Radiomics Features Prognostic Performance Comparison in Non-Small Cell Lung Cancer. Translational oncology 2017;10:886-94.

39. Blazic IM, Lilic GB, Gajic MM. Quantitative Assessment of Rectal Cancer Response to Neoadjuvant Combined Chemotherapy and Radiation Therapy: Comparison of Three Methods of Positioning Region of Interest for ADC Measurements at Diffusion-weighted MR Imaging. Radiology 2017;282:418-28.

APPENDICES

Appendix 1. List of radiomic features extracted from the computed tomography images and Inter-slice and Inter-reader intraclass coefficient correlation.

Feature family	Feature class	Inter-slice ICC					Inter-reader ICC				
		Original	Wavelet-filtered				Original	Wavelet-filtered			
			LH	HL	HH	LL		LH	HL	HH	LL
Shape Features (2D) (8 features)	Elongation	0.924	-	-	-	-	0.809	-	-	-	-
	Major axis	0.928	-	-	-	-	0.848	-	-	-	-
	Minor axis	0.984	-	-	-	-	0.922	-	-	-	-
	Maximum 2D diameter (column)	0.934	-	-	-	-	0.900	-	-	-	-
	Maximum 2D diameter (row)	0.982	-	-	-	-	0.832	-	-	-	-
	Maximum 2D diameter (slice)	0.920	-	-	-	-	0.835	-	-	-	-
	Sphericity	0.975	-	-	-	-	0.875	-	-	-	-
	Surface area	0.967	-	-	-	-	0.909	-	-	-	-
First order features (18 features)	Energy	0.971	0.968	0.968	0.967	0.974	0.911	0.910	0.909	0.909	0.912
	Total energy	0.971	0.968	0.968	0.967	0.974	0.911	0.910	0.909	0.909	0.912

		Entropy*	0.691	0.949	0.979	0.995	0.727	0.810	0.828	0.863	0.918	0.793
		Minimum	-0.065	0.924	0.961	0.843	-0.104	0.004	0.747	0.731	0.801	0.010
		Maximum	0.980	0.362	0.668	0.898	0.971	0.843	0.482	0.418	0.746	0.827
		10th percentile	0.358	0.963	0.978	0.994	0.387	0.712	0.773	0.814	0.965	0.731
		90th percentile	0.993	0.891	0.994	0.992	0.995	0.933	0.930	0.891	0.926	0.922
		Mean	0.845	0.944	0.980	0.804	0.842	0.891	0.817	0.814	0.630	0.892
		Median	0.914	0.914	0.965	0.722	0.919	0.907	0.638	0.645	0.393	0.905
		Interquartile range	0.776	0.921	0.977	0.998	0.776	0.756	0.752	0.831	0.963	0.751
		Range	0.050	0.821	0.930	0.910	0.009	0.118	0.709	0.622	0.824	0.131
		Mean absolute deviation	0.536	0.962	0.982	0.998	0.552	0.750	0.800	0.807	0.941	0.759
		Robust mean absolute deviation	0.764	0.930	0.982	0.998	0.769	0.772	0.764	0.836	0.961	0.778
		Root mean squared	0.867	0.942	0.980	0.802	0.882	0.897	0.816	0.815	0.616	0.902
		Skewness	0.174	0.891	0.879	0.849	0.199	0.345	0.734	0.765	0.646	0.367
		Kurtosis†	0.006	0.953	0.941	0.913	0.007	0.020	0.838	0.699	0.741	0.022
		Variance	0.792	0.960	0.983	0.995	0.821	0.793	0.831	0.877	0.920	0.789
		Uniformity‡	0.134	0.968	0.978	0.998	0.143	0.414	0.790	0.715	0.927	0.433
Gray level co-		Autocorrelation	0.005	0.942	0.966	0.855	-0.001	0.034	0.697	0.664	0.721	0.035

occurrence matrix features (GLCM) (22 features)	Joint average	0.055	0.913	0.955	0.820	0.018	0.124	0.730	0.731	0.741	0.123
	Cluster Prominence	0.001	0.983	0.974	0.998	0.001	-0.002	0.700	0.534	0.749	-0.001
	Cluster Shade	0.008	0.978	0.975	0.886	0.008	0.016	0.758	0.635	0.360	0.019
	Cluster Tendency	0.139	0.972	0.977	0.998	0.144	0.455	0.803	0.729	0.906	0.453
	Contrast	0.412	0.967	0.984	0.998	0.383	0.700	0.847	0.751	0.950	0.695
	Correlation	0.753	0.955	0.981	0.980	0.736	0.580	0.832	0.867	0.855	0.569
	Difference average	0.810	0.980	0.989	0.998	0.790	0.890	0.892	0.864	0.959	0.866
	Difference entropy	0.844	0.972	0.980	0.997	0.865	0.905	0.881	0.855	0.940	0.891
	Difference variance	0.211	0.936	0.967	0.998	0.249	0.450	0.759	0.635	0.915	0.548
	Joint energy§	0.864	0.970	0.983	0.995	0.877	0.786	0.829	0.883	0.926	0.747
	Joint entropy	0.785	0.956	0.981	0.996	0.772	0.839	0.864	0.884	0.937	0.750
	Informational measure of correlation 1	0.820	0.964	0.968	0.956	0.857	0.775	0.795	0.820	0.513	0.766
	Informational measure of correlation 2	0.858	0.934	0.972	0.972	0.839	0.836	0.800	0.831	0.609	0.838
	Inverse difference moment	0.954	0.988	0.991	0.998	0.961	0.895	0.900	0.901	0.953	0.869
	Inverse difference moment normalized	0.706	0.878	0.937	0.738	0.807	0.645	0.766	0.737	0.447	0.691

	Inverse difference	0.949	0.987	0.991	0.998	0.953	0.895	0.899	0.901	0.952	0.874
	Inverse difference normalized	0.564	0.887	0.945	0.784	0.683	0.618	0.793	0.746	0.645	0.681
	Inverse variance	0.931	0.980	0.985	0.996	0.956	0.912	0.883	0.814	0.944	0.872
	Maximum probability‖	0.861	0.978	0.992	0.996	0.885	0.666	0.815	0.880	0.936	0.623
	Sum entropy	0.706	0.931	0.972	0.994	0.709	0.758	0.797	0.871	0.910	0.734
	Sum of squares¶	0.155	0.971	0.979	0.998	0.159	0.484	0.796	0.727	0.933	0.487
Gray level run length matrix features (GLRLM) (16 features)	Short run emphasis	0.967	0.985	0.989	0.997	0.968	0.839	0.870	0.900	0.941	0.843
	Long run emphasis	0.978	0.990	0.989	0.995	0.974	0.849	0.865	0.917	0.922	0.842
	Gray level non-uniformity	0.983	0.972	0.968	0.962	0.985	0.943	0.923	0.911	0.903	0.944
	Gray level non-uniformity normalized	0.764	0.956	0.977	0.994	0.804	0.807	0.829	0.865	0.904	0.798
	Run length non-uniformity	0.966	0.975	0.977	0.988	0.960	0.904	0.915	0.914	0.936	0.900
	Run length non-uniformity normalized	0.967	0.986	0.991	0.997	0.969	0.840	0.871	0.895	0.944	0.840
	Run percentage	0.975	0.989	0.991	0.997	0.973	0.853	0.879	0.903	0.938	0.844
	Gray level variance	0.101	0.968	0.976	0.998	0.126	0.323	0.789	0.707	0.914	0.385

	Run variance	0.983	0.990	0.989	0.992	0.977	0.851	0.841	0.915	0.894	0.828
	Run entropy	0.596	0.926	0.959	0.972	0.590	0.710	0.774	0.758	0.760	0.687
	Low gray level run emphasis	-0.053	0.850	0.938	0.525	0.097	0.382	0.437	0.631	0.444	0.421
	High gray level run emphasis	0.000	0.940	0.965	0.868	-0.006	0.027	0.697	0.662	0.733	0.029
	Short run low gray level emphasis	-0.009	0.840	0.928	0.557	0.123	0.396	0.412	0.632	0.415	0.439
	Short run high gray level emphasis	0.003	0.937	0.964	0.913	-0.001	0.032	0.690	0.654	0.779	0.035
	Long run low gray level emphasis	-0.178	0.893	0.958	0.516	-0.011	0.338	0.554	0.694	0.571	0.337
	Long run high gray level emphasis	0.000	0.953	0.967	0.599	-0.017	0.028	0.747	0.692	0.555	0.018
Gray level size zone matrix features (GLSZM) (16 features)	Small area emphasis	0.850	0.961	0.907	0.895	0.930	0.802	0.733	0.719	0.672	0.828
	Large area emphasis	0.973	0.943	0.992	0.940	0.966	0.740	0.698	0.712	0.770	0.803
	Gray level non-uniformity	0.980	0.979	0.980	0.992	0.982	0.939	0.926	0.914	0.942	0.942
	Gray level non-	0.624	0.944	0.965	0.874	0.728	0.789	0.756	0.822	0.686	0.788

uniformity normalized										
Size-zone non-uniformity	0.954	0.989	0.983	0.996	0.931	0.896	0.928	0.929	0.950	0.861
Size-zone non-uniformity normalized	0.868	0.962	0.924	0.906	0.935	0.802	0.725	0.732	0.687	0.822
Zone percentage	0.962	0.987	0.990	0.996	0.967	0.895	0.868	0.878	0.952	0.859
Gray level variance	0.044	0.949	0.967	0.969	0.086	0.137	0.740	0.681	0.810	0.263
Zone variance	0.975	0.940	0.992	0.940	0.965	0.705	0.686	0.704	0.769	0.758
Zone entropy	0.753	0.904	0.928	0.903	0.730	0.732	0.678	0.661	0.763	0.732
Low gray level zone emphasis	-0.026	0.842	0.937	0.593	0.100	0.423	0.462	0.661	0.439	0.438
High gray level zone emphasis	-0.005	0.938	0.961	0.864	-0.009	0.019	0.694	0.659	0.714	0.024
Small area low gray level zone emphasis	0.130	0.797	0.888	0.576	0.172	0.456	0.412	0.651	0.445	0.474
Small area high gray level zone emphasis	0.009	0.931	0.954	0.868	0.009	0.035	0.686	0.642	0.707	0.047
Large area low gray level zone emphasis	-0.011	0.917	0.859	0.493	-0.063	0.510	0.404	0.461	0.635	0.289

	Large area high gray level zone emphasis	0.007	0.971	0.988	0.940	-0.011	0.108	0.681	0.865	0.814	0.060
Gray level dependence matrix features (GLDM) (14 features)	Small dependence emphasis	0.950	0.989	0.987	0.997	0.959	0.890	0.857	0.871	0.954	0.854
	Large dependence emphasis	0.980	0.989	0.990	0.996	0.977	0.821	0.873	0.909	0.932	0.821
	Gray level non-uniformity	0.983	0.972	0.964	0.960	0.986	0.942	0.916	0.910	0.903	0.943
	Dependence non-uniformity	0.968	0.979	0.973	0.970	0.956	0.907	0.919	0.915	0.917	0.895
	Dependence non-uniformity normalized	0.985	0.993	0.990	0.990	0.976	0.847	0.875	0.875	0.870	0.816
	Gray level variance	0.135	0.968	0.978	0.998	0.143	0.415	0.792	0.715	0.928	0.434
	Dependence variance	0.986	0.989	0.980	0.982	0.986	0.702	0.830	0.851	0.806	0.706
	Dependence entropy	0.702	0.884	0.959	0.992	0.704	0.674	0.669	0.791	0.894	0.669
	Low gray level emphasis	-0.055	0.851	0.939	0.513	0.092	0.365	0.429	0.615	0.463	0.407
	High gray level emphasis	0.002	0.941	0.966	0.868	-0.004	0.030	0.698	0.663	0.735	0.031

	Small dependence low gray level emphasis	0.275	0.802	0.806	0.753	0.257	0.492	0.365	0.647	0.462	0.512
	Small dependence high gray level emphasis	0.046	0.938	0.959	0.979	0.037	0.083	0.695	0.616	0.880	0.091
	Large dependence low gray level emphasis	-0.231	0.902	0.961	0.534	-0.122	0.307	0.589	0.675	0.562	0.148
	Large dependence high gray level emphasis	0.008	0.959	0.971	0.584	-0.017	0.036	0.759	0.705	0.523	0.020

ICC, Intraclass correlation coefficient; LL, Low-Low pass filter; LH, Low-High pass filter; HH, High-High pass filter; LL, Low-Low pass filter

* Defined by image biomarker standardization initiative (IBSI) as Intensity Histogram Entropy

† The IBSI feature definition implements excess kurtosis, where kurtosis is corrected by -3, yielding 0 for normal distributions. The PyRadiomics kurtosis is not corrected, yielding a value 3 higher than the IBSI kurtosis.

‡ Defined by IBSI as Intensity Histogram Uniformity

§ Defined by IBSI as Angular Second Moment.

‖ Defined by IBSI as Joint maximum

¶ Defined by IBSI as Joint Variance

Appendix 2. Radiomics quality score

RQS criteria			Points	
1	Image protocol quality	Well-documented image protocols (for example, contrast, slice thickness, energy, etc.) and/or usage of public image protocols allow reproducibility/replicability	+ 1 (if protocols are well-documented) + 1 (if public protocol is used)	+1
2	Multiple segmentations	Segmentation by different physicians/algorithms/software, perturbing segmentations by (random) noise, segmentation at different breathing cycles. Analyse feature robustness to segmentation variabilities	+ 1	+1
3	Phantom study on all scanners	Detect inter-scanner differences and vendor-dependent features. Analyse feature robustness to these sources of variability	+ 1	0
4	Imaging at multiple time points	Collect images of individuals at additional time points. Analyse feature robustness to temporal variabilities (for example, organ movement, organ expansion/ shrinkage)	+ 1	0
5	Feature reduction or adjustment for multiple testing	Decreases the risk of overfitting. Overfitting is inevitable if the number of features exceeds the number of samples. Consider feature robustness when selecting features	- 3 (if neither measure is implemented) + 3 (if either measure is implemented)	+3
6	Multivariable analysis with non-radiomics features	(for example, EGFR mutation) - is expected to provide a more holistic model. Permits correlating/inferencing between radiomics and non radiomics features	+1	+1
7	Detect and	Demonstration of phenotypic differences (possibly	+1	0

	discuss biological correlates	associated with underlying gene–protein expression patterns) deepens understanding of radiomics and biology		
8	Cut-off analyses	Determine risk groups by either the median, a previously published cut-off or report a continuous risk variable. Reduces the risk of reporting overly optimistic results	+1	+1
9	Discrimination statistics	Report discrimination statistics (for example, C-statistic, ROC curve, AUC) and their statistical significance (for example, p-values, confidence intervals). One can also apply resampling method (for example, bootstrapping, cross-validation)	+ 1 (if a discrimination statistic and its statistical significance are reported) + 1 (if a resampling method technique is also applied)	+1
10	Calibration statistics	Report calibration statistics (for example, Calibration-in-the-large/slope, calibration plots) and their statistical significance (for example, *P*-values, confidence intervals). One can also apply resampling method (for example, bootstrapping, cross-validation)	+ 1 (if a calibration statistic and its statistical significance are reported) + 1 (if a resampling method technique is also applied)	+1
11	Prospective study registered in a trial database	Provides the highest level of evidence supporting the clinical validity and usefulness of the radiomics biomarker	+ 7 (for prospective validation of a radiomics signature in an appropriate trial)	0
12	Validation	The validation is performed without retraining and without adaptation of the cut-off value, provides crucial information with regards to credible clinical performance	- 5 (if validation is missing) + 2 (if validation is based on a dataset from the same institute) + 3 (if validation is based on a dataset from another institute) + 4 (if validation is based on two datasets from two distinct institutes)	+3

			+ 4 (if the study validates a previously published signature) + 5 (if validation is based on three or more datasets from distinct institutes)	
13	Comparison to 'gold standard'	Assess the extent to which the model agrees with/is superior to the current 'gold standard' method (for example, TNM-staging for survival prediction). This comparison shows the added value of radiomics	+2	+2
14	Potential clinical utility	Report on the current and potential application of the model in a clinical setting (for example, decision curve analysis).	+2	0
15	Cost-effectiveness analysis	Report on the cost-effectiveness of the clinical application (for example, QALYs generated)	+1	0
16	Open science and data	Make code and data publicly available. Open science facilitates knowledge transfer and reproducibility of the study	+ 1 (if scans are open source) + 1 (if region of interest segmentations are open source) + 1 (if code is open source) + 1 (if radiomics features are calculated on a set of representative ROIs and the calculated features and representative ROIs are open source)	+1
			Total points (36 = 100%)	15 (41.7%)

Appendix 3. Checklist for reporting on radiomics studies by image biomarker standardization initiative guideline (Reference: Zwanenburg A, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. Radiology 2020;295:328-38).

Topic	Modality	Item	Description	Page
Patient				
Region of interest		1	Describe the region of interest that is being imaged.	9
Patient preparation		2a	Describe specific instructions given to patients prior to image acquisition, e.g. fasting prior to imaging.	7
		2b	Describe administration of drugs to the patient prior to image acquisition, e.g. muscle relaxants.	7
		2c	Describe the use of specific equipment for patient comfort during scanning, e.g. ear plugs.	N/A
Radioactive tracer	PET, SPECT	3a	Describe which radioactive tracer was administered to the patient, e.g. 18F-FDG.	N/A
	PET, SPECT	3b	Describe the administration method.	N/A
	PET, SPECT	3c	Describe the injected activity of the radioactive tracer at administration.	N/A
	PET, SPECT	3d	Describe the uptake time prior to image acquisition.	N/A
	PET, SPECT	3e	Describe how competing substance levels were controlled.	N/A
Contrast agent		4a	Describe which contrast agent was administered to the patient.	7
		4b	Describe the administration method.	7
		4c	Describe the injected quantity of contrast agent.	7
		4d	Describe the uptake time prior to image acquisition.	7

		4e	Describe how competing substance levels were controlled.	N/A
Comorbidities		5	Describe if the patients have comorbidities that affect imaging.	N/A
Acquisition				
Acquisition protocol		6	Describe whether a standard imaging protocol was used, and where its description may be found.	7, 8
Scanner type		7	Describe the scanner type(s) and vendor(s) used in the study.	7, 8
Imaging modality		8	Clearly state the imaging modality that was used in the study, e.g. CT, MRI.	7, 8
Static/dynamic scans		9a	State if the scans were static or dynamic.	7
	Dynamic scans	9b	Describe the acquisition time per time frame.	N/A
	Dynamic scans	9c	Describe any temporal modelling technique that was used.	N/A
Scanner calibration		10	Describe how and when the scanner was calibrated.	N/A
Patient instructions		11	Describe specific instructions given to the patient during acquisition, e.g. breath holding.	N/A
Anatomical motion correction		12	Describe the method used to minimise the effect of anatomical motion.	N/A
Scan duration		13	Describe the duration of the complete scan or the time per bed position.	7
Tube voltage	CT	14	Describe the peak kilo voltage output of the X-ray source.	7, 8
Tube current	CT	15	Describe the tube current in mA.	7, 8
Time-of-flight	PET	16	State if scanner time-of-flight capabilities are used during acquisition.	N/A
RF coil	MRI	17	Describe what kind RF coil used for acquisition, incl. vendor.	N/A

Scanning sequence	MRI	18a	Describe which scanning sequence was acquired.	N/A
	MRI	18b	Describe which sequence variant was acquired.	N/A
	MRI	18c	Describe which scan options apply to the current sequence, e.g. flow compensation, cardiac gating.	N/A
Repetition time	MRI	19	Describe the time in ms between subsequent pulse sequences.	N/A
Echo time	MRI	20	Describe the echo time in ms.	N/A
Echo train length	MRI	21	Describe the number of lines in k-space that are acquired per excitation pulse.	N/A
Inversion time	MRI	22	Describe the time in ms between the middle of the inverting RF pulse to the middle of the excitation pulse.	N/A
Flip angle	MRI	23	Describe the flip angle produced by the RF pulses.	N/A
Acquisition type	MRI	24	Describe the acquisition type of the MRI scan, e.g. 3D.	N/A
k-space traversal	MRI	25	Describe the acquisition trajectory of the k-space.	N/A
Number of averages/ excitations	MRI	26	Describe the number of times each point in k-space is sampled.	N/A
Magnetic field strength	MRI	27	Describe the nominal strength of the MR magnetic field.	N/A
Reconstruction				
In-plane resolution		28	Describe the distance between pixels, or alternatively the field of view and matrix size.	7, 8
Image slice thickness		29	Describe the slice thickness.	7, 8
Image slice spacing		30	Describe the distance between image slices.	7, 8
Convolution kernel	CT	31a	Describe the convolution kernel used to reconstruct the image.	7, 8
	CT	31b	Describe settings pertaining to iterative reconstruction algorithms.	7, 8

Exposure	CT	31c	Describe the exposure (in mAs) in slices containing the region of interest.	7, 8
Reconstruction method	PET	32a	Describe which reconstruction method was used, e.g. 3D OSEM.	N/A
	PET	32b	Describe the number of iterations for iterative reconstruction.	N/A
	PET	32c	Describe the number of subsets for iterative reconstruction.	N/A
Point spread function modelling	PET	33	Describe if and how point-spread function modelling was performed.	N/A
Image corrections	PET	34a	Describe if and how attenuation correction was performed.	N/A
	PET	34b	Describe if and how other forms of correction were performed, e.g. scatter correction, randoms correction, dead time correction etc.	N/A
Reconstruction method	MRI	35a	Describe the reconstruction method used to reconstruct the image from the k-space information.	N/A
	MRI	35b	Describe any artifact suppression methods used during reconstruction to suppress artifacts due to undersampling of k-space.	N/A
Diffusion-weigh ted imaging	DWI-MRI	36	Describe the b-values used for diffusion-weigh ting.	N/A
Image registration				
Registration method		37	Describe the method used to register multi-modality imaging.	N/A
Image processing-data conversion				
SUV normalisation	PET	38	Describe which standardised uptake value (SUV) normalisation method is used.	N/A
ADC computation	DWI-MRI	39	Describe how apparent diffusion coefficient (ADC) values were calculated.	N/A

Other data conversions		40	Describe any other conversions that are performed to generate e.g. perfusion maps.	N/A
Image processing-postacquisition processing				
Anti-aliasing		41	Describe the method used to deal with anti-aliasing when down-sampling during interpolation.	N/A
Noise suppression		42	Describe methods used to suppress image noise.	N/A
Post-reconstruc tion smoothing filter	PET	43	Describe the width of the Gaussian filter (FWHM) to spatially smooth intensities.	N/A
Skull stripping	MRI (brain)	44	Describe method used to perform skull stripping.	N/A
Non-uniformity correction	MRI	45	Describe the method and settings used to perform non-uniformity correction.	N/A
Intensity normalisation		46	Describe the method and settings used to normalise intensity distributions within a patient or patient cohort.	10
Other post-acquisition processing methods		47	Describe any other methods that were used to process the image and are not mentioned separately in this list.	9, 10
Segmentation				
Segmentation method		48a	Describe how regions of interest were segmented, e.g. manually.	9
		48b	Describe the number of experts, their expertise and consensus strategies for manual delineation.	9
		48c	Describe methods and settings used for semi-automatic and fully automatic segmentation.	9
		48d	Describe which image was used to define segmentation in case of multi-modality imaging.	N/A
Conversion to mask		49	Describe the method used to convert polygonal or mesh-based segmentations to a voxel-based mask.	N/A
Image processing-image interpolation				

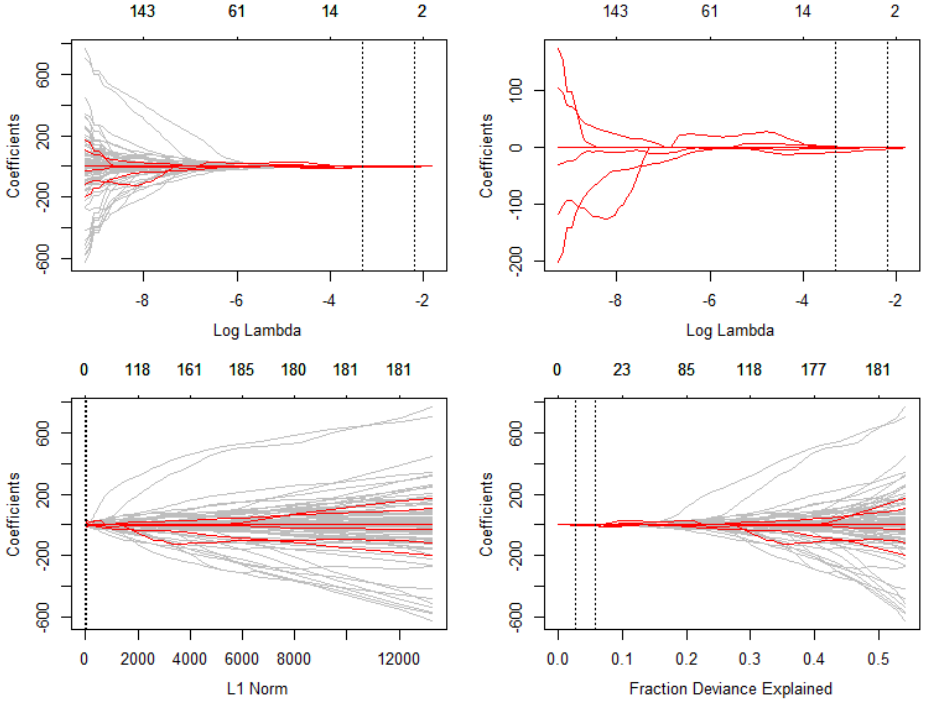
Interpolation method	50a	Describe which interpolation algorithm was used to interpolate the image.	9
	50b	Describe how the position of the interpolation grid was defined, e.g. align by center.	9
	50c	Describe how the dimensions of the interpolation grid were defined, e.g. rounded to nearest integer.	9
	50d	Describe how extrapolation beyond the original image was handled.	N/A
Voxel dimensions	51	Describe the size of the interpolated voxels.	9
Intensity rounding CT	52	Describe how fractional Hounsfield Units are rounded to integer values after interpolation.	N/A
Image processing-ROI interpolation			
Interpolation method	53	Describe which interpolation algorithm was used to interpolate the region of interest mask.	9
Partially masked voxels	54	Describe how partially masked voxels after interpolation are handled.	N/A
Image processing-resegmentation			
Re-segmentation methods	55	Describe which methods and settings are used to re-segment the ROI intensity mask.	N/A
Image processing-discretization			
Discretisation method	56a	Describe the method used to discretise image intensities.	10
	56b	Describe the number of bins (FBN) or the bin size (FBS) used for discretisation.	10
	56c	Describe the lowest intensity in the first bin for FBS discretisation.	10
Image processing-image transformation			

Image filter	57	Describe the methods and settings used to filter images, e.g. Laplacian-of-Gaussian.	10
Radiomics feature computation			
Biomarker set	58	Describe which set of image biomarkers is computed and refer to their definitions or provide these.	App-endix
IBSI compliance	59	State if the software used to extract the set of image biomarkers is compliant with the IBSI benchmarks.	App-endix
Robustness	60	Describe how robustness of the image biomarkers was assessed, e.g. test-retest analysis.	13, 14
Software availability	61	Describe which software and version was used to compute image biomarkers.	13, 14
Radiomics feature computation-texture parameters			
Texture matrix aggregation	62	Define how texture-matrix based biomarkers were computed from underlying texture matrices.	App-endix
Distance weighting	63	Define how CM, RLM, NGTDM and NGLDM weight distances, e.g. no weighting.	App-endix
CM symmetry	64	Define whether symmetric or asymmetric co-occurrence matrices were computed.	App-endix
CM distance	65	Define the (Chebyshev) distance at which co-occurrence of intensities is determined, e.g. 1.	App-endix
SZM linkage distance	66	Define the distance and distance norm for which voxels with the same intensity are considered to belong to the same zone for the purpose of constructing an SZM, e.g. Chebyshev distance of 1.	App-endix
DZM linkage distance	67	Define the distance and distance norm for which voxels with the same intensity are considered to belong to the same zone for the purpose of constructing a DZM, e.g. Chebyshev distance of 1.	App-endix

DZM zone distance norm	68	Define the distance norm for determining the distance of zones to the border of the ROI, e.g. Manhattan distance.	App-endix
NGTDM distance	69	Define the neighbourhood distance and distance norm for the NGTDM, e.g. Chebyshev distance of 1.	N/A
NGLDM distance	70	Define the neighbourhood distance and distance norm for the NGLDM, e.g. Chebyshev distance of 1.	N/A
NGLDM coarseness	71	Define the coarseness parameter for the NGLDM, e.g. 0.	N/A
Machine learning and radiomics analysis			
Diagnostic and prognostic modelling	72	See the TRIPOD guidelines for reporting on diagnostic and prognostic modelling.	12
Comparison with known factors	73	Describe where performance of radiomics models is compared with known (clinical) factors.	13
Multicollinearity	74	Describe where the multicollinearity between image biomarkers in the signature is assessed.	20
Model availability	75	Describe where radiomics models with the necessary pre-processing information may be found.	18
Data availability	76	Describe where imaging data and relevant meta-data used in the study may be found.	N/A

N/A, Not applicable

Appendix 4. Coefficient paths for lasso-penalized Cox proportional hazard regression models. The features with highlighted paths have non-zero coefficients in the model with the optimal λ value as determined by ten-fold cross-validation. The top plots show the coefficient path scaled to reflect log(λ) on the x-axis (top left: full path, top right: zoomed in to only show the selected features). The bottom plots show the coefficient paths relative to the L1-norms of the estimated coefficient vector (left) and to the fraction of the null partial log-likelihood deviance explained (right). The dotted vertical lines indicate the λ values with minimal deviance and with the largest λ value within one standard deviation of the minimal deviance.



(Reference: Sill M, Hielscher T, Becker N, Zucknick M. c060: Extended Inference with Lasso and Elastic-Net Regularized Cox and Generalized Linear Models. 2014 2014;62:22 Journal of Statistical Software).

ABSTRACT (IN KOREAN)

위암 환자 예후 예측을 위한 라디오믹스 모형 개발

<지도교수 허용민>

연세대학교 대학원 의학과

신재승

목적: 진행성 위암 환자에서 수술 전 항암화학요법은 예후 개선에 대한 가능성으로 인해 많은 관심을 불러 일으키고 있다. 수술 전 병기 결정을 기반으로 위험도가 높은 환자를 선별하는 것은 수술 전 항암화학요법 여부를 선택하는 데 중요하다. 임상 예측 모델과 비교하여 수술 전 조영증강 복부 전산화단층촬영을 이용한 라디오믹스 모델이 국소 진행 위암의 무재발 생존 기간을 더 잘 예측할 수 있는지 평가하고자 하였다.

방법: 2010년에 신촌 세브란스 병원에서 수술 전 항암화학요법치료를 받지 않고, 국소 진행성 위암으로 수술적 절제술을 받은 349명의 환자를 모형 훈련군으로 후향적으로 분석하였고, 같은 기준으로 강남 세브란스에서 수술 받은 환자 61명을 검증군으로 사용하였다. 기존의 CT 병기 및 내시경 데이터를 포함한 수술 전 검사에서 얻을 수 있는 임상 적 인자들을 얻었고 수술 전 CT에서 총 438 개의 라디오믹스 변수를 추출하였다. 10배 교차 유효성 검사를 사용한 LASSO 회귀분석을 통해 변수를 선택하고, 라디오믹스 점수를 정의하였다. 내부 및 외

부 검증은 각각 부트스트랩 방법(1000회 실시)를 통해 수행되었고, 라디오믹스 모형의 예측력 향상은 통합 수신자 조작 특성 곡선(integrated receiver operating curve)을 통해 곡선 아래 면적(iAUC, integrated area under the curve)을 계산하여 비교되었다.

결과: 최종 410명 (58.2±13.0 세, 여성 268명)의 환자군이 최종 연구에 포함되었다. 라디오믹스 모형 구성에는 7가지 특성이 선택되었고, 임상 모형은 CT에서의 Borrmann 4형과 결절형 침윤 여부가 포함되었다 병합 모형은 임상 정보와 라디오믹스 점수를 모두 포함하여 구성되었다. 내부 및 외부 검증 모두에서 라디오믹스 (iAUC [95 % 신뢰 구간], 0.714 [0.667, 0.759], P<0.001, 내부검증; 0.652 [0.628, 0.674], P=0.010, 외부검증)과 병합 모형(0.719 [0.674, 0.764], P<0.001, 내부검증; 0.651 [0.630, 0.673], P=0.014, 외부검증)이 각각 임상 모형 (0.616 [0.570, 0.663], 내부검증; 0.594 [0.544, 0.636], 외부검증)과 비교했을 때 통계적으로 유의하게 무재발 생존 기간 예측이 향상되었다.

결론: 조영증강 전산화단층촬영 영상을 기반으로 한 라디오믹스 모형은 수술 전 임상 정보와 통합되어 무재발 생존 기간을 더 잘 예측할 수 있으며 수술 전 영상 생체표지자로서 가능성을 가지고 있다. 따라서 라디오믹스 모형을 통해 재발 고위험군을 분류한다면, 진행성 위암 환자에서 수술 전 항암화학요법 여부를 결정하는데 도움을 줄 수 있다.

핵심되는 말: 라디오믹스, 조영증강 전산화단층촬영, 진행성 위암, 예후 예측