Surgical Oncology 28 (2019) 78–85

Contents lists available at ScienceDirect

Surgical Oncology

journal homepage: www.elsevier.com/locate/suronc

A radiomics-based nomogram for the preoperative prediction of posthepatectomy liver failure in patients with hepatocellular carcinoma

Wei Cai[a,b], Baochun He[b], Min Hu[a], Wenyu Zhang[a,b], Deqiang Xiao[b], Hao Yu[c], Qi Song[d], Nan Xiang[a], Jian Yang[a], Songsheng He[a], Yaohuan Huang[a], Wenjie Huang[a], Fucang Jia[b,*], Chihua Fang[a,**]

[a] Department of Hepatobiliary Surgery, Zhujiang Hospital, Southern Medical University, Guangzhou, China
[b] Research Lab for Medical Imaging and Digital Surgery, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
[c] Department of Radiology, Zhujiang Hospital, Southern Medical University, Guangzhou, China
[d] School of Electronic Engineering and Computer Science, Washington State University, Pullman, WA, USA

ARTICLE INFO

Keywords:
Hepatocellular carcinoma
Liver failure
Radiomics
Nomogram

ABSTRACT

Objectives: To develop and validate a radiomics-based nomogram for the preoperative prediction of post-hepatectomy liver failure (PHLF) in patients with hepatocellular carcinoma (HCC).

Methods: One hundred twelve consecutive HCC patients who underwent hepatectomy were included in the study pool (training cohort: n = 80, validation cohort: n = 32), and another 13 patients were included in a pilot prospective analysis. A total of 713 radiomics features were extracted from portal-phase computed tomography (CT) images. A logistic regression was used to construct a radiomics score (Rad-score). Then a nomogram, including Rad-score and other risk factors, was built with a multivariate logistic regression model. The discrimination, calibration and clinical utility of nomogram were evaluated.

Results: The Rad-score could predict PHLF with an AUC of 0.822 (95% CI, 0.726–0.917) in the training cohort and of 0.762 (95% CI, 0.576–0.948) in the validation cohort; however, the approach could not completely outmatch the existing methods (CP [Child-Pugh], MELD [Model of End Stage Liver Disease], ALBI [albumin-bilirubin]). The individual predictive nomogram that included the Rad-score, MELD and performance status (PS) showed better discrimination with an AUC of 0.864 (95% CI, 0.786–0.942), which was higher than the AUCs of the conventional methods (nomogram vs CP, MELD, and ALBI at $P < 0.001$, $P < 0.005$, and $P < 0.005$, respectively). In the validation cohort, the nomogram discrimination was also superior to those of the other three methods (AUC: 0.896; 95% CI, 0.774–1.000). The calibration curves showed good agreement in both cohorts, and the decision curve analysis of the entire cohort revealed that the nomogram was clinically useful. A pilot prospective analysis showed that the radiomics nomogram could predict PHLF with an AUC of 0.833 (95% CI, 0.591–1.000).

Conclusions: A nomogram based on the Rad-score, MELD, and PS can predict PHLF.

Hepatocellular carcinoma (HCC) is the fourth leading cause of cancer-related death globally, and China alone accounts for approximately 50% of the total number of HCCs. This neoplasm has been associated with a huge health crisis [1,2]. Among the therapeutic strategies used, surgical resection with partial hepatectomy remains the most effective potential treatment option for select HCC patients and is associated with a high complete response rate [3,4]. However, many HCC patients develop chronic liver disease, thereby limiting the feasibility of hepatic resection due to impaired liver function, which results in a high risk of postoperative complications and even death [5]. Posthepatectomy liver failure (PHLF) is a major cause of mortality after liver resection, so operative safety assessments and pre-surgical identifications of patients who are at risk of PHLF are crucial steps in the management of patients with this disease [3,6]. Over the past few decades, many studies have been conducted to address this issue [2–4,6,7]. Conventional blood function tests, such as liver enzyme, bilirubin and albumin tests, and clinical scoring systems (CP [Child-Pugh] and MELD [Model of End Stage Liver Disease]) are widely used

* Corresponding author.
** Corresponding author.
E-mail addresses: fc.jia@siat.ac.cn (F. Jia), fangch_dr@163.com (C. Fang).

https://doi.org/10.1016/j.suronc.2018.11.013
Received 10 June 2018; Received in revised form 15 September 2018; Accepted 12 November 2018

due to their abilities to identify severe hepatic parenchymal disease, but their predictive abilities for the perioperative outcomes of candidates for resection are not satisfactory [4]. A recent study revealed that a new model, named the albumin-bilirubin (ALBI) score [8], could reach a relatively good predictive accuracy for PHLF in HCC patients who underwent liver resection [5].

As an alternative approach, computed tomography (CT) is a major tool that has informed medical science and treatment and that has been routinely used to assess tumoral and anatomical tissue characteristics for HCC management [9]. However, during preoperative surgical planning, CT has typically only been used to calculate the remnant liver volume to help determine whether a patient could safely undergo liver resection to some extent. The large quantity of additional useful information beyond the anatomical data has been underutilized. Due to technological improvements, medical imaging should not only be used for qualitative diagnosis but also evolve into a quantitative science. In recent years, increasing attention has been directed at radiomics that can convert images into mineable data, with high fidelity and high throughput [10]. The hypothesis of radiomics is that massive imaging features can be quantified to enable investigation for treatment monitoring and outcome predictions in cancer sites or as potential imaging biomarkers [11]. Several studies have shown that radiomics has a great potential value in cancer management and may enable advancements in the understanding of tumor biology and better implementation of precision medicine [12]. For most HCC patients with chronic liver disease, the ideal prognosis and treatment algorithms need to account for both tumoral characteristics and liver function [13,14]. To our knowledge, no previous study has simultaneously evaluated the tumor and normal liver parenchyma using a radiomics approach, integrated these data into a radiomics scoring (Rad-score) system or signature and determined whether this new method could enable a superior prediction of PHLF for HCC patients.

Thus, the aim of this study was to develop and validate a radiomics nomogram that incorporated both the radiomics features and clinicopathologic risk factors for individual preoperative prediction of PHLF in HCC patients underwent hepatectomy.

1. Methods

1.1. Patient population

One hundred and twelve consecutive HCC patients who underwent hepatectomy between October 2008 and June 2016 were enrolled, according to the following inclusion criteria: (1) standard contrast-enhanced CT scans performed within 7 days before surgical resection, with the image data stored intact and (2) the utilization of radical hepatectomy and presence of histologically confirmed HCC. The exclusion criteria were as follows: (1) preoperative treatment (transarterial chemoembolization, ablation or portal venous embolization) was performed or (2) other malignant tumors existed simultaneously. The patients were divided into two independent cohorts at a ratio of 3:1 using a random number table. Eighty patients constituted the training cohort and the other thirty-two formed the validation cohort.

Demographic and pretreatment clinical characteristics, including gender, age, alpha-fetoprotein (AFP), total bilirubin (TBIL), albumin, platelet counts (PLA), creatinine, and the international normalized ratio (INR), were collected from medical records. Portal venous-phase CT images (thickness: 1.0 mm) were retrieved from the picture archiving and communication system (PACS). This study was approved by the ethics committee of our hospital. The CT image acquisition, segmentation of region of interests (ROIs) are described in Supplementary Information 1 and 2.

1.2. Radiomics feature extraction

We evaluated a total of 713 CT image features from the tumor and liver parenchyma. Because the tumor was totally removed after radical hepatectomy, we did not consider the preoperative heterogeneous analysis of lesions to be necessary. Finally, we evaluated several targeted features for the liver parenchyma and tumor, and these features were divided into the following four groups: intensity for liver parenchyma (n = 19), hepatic textural features (n = 59), wavelet (n = 624), and tumor shape and size-based features (n = 11). The mathematical description of all features can be found in the online document of software PyRadiomics [15] and in the Supplementary Information 3.

1.3. Definitions and calculation of clinical score values

Major liver resection was defined as the resection of at least three Couinaud liver segments [16]. PHLF and its grade were defined according to the International Study Group of Liver Surgery (ISGLS) definition (described in detail in the Supplementary Information 4) [17]. For each patient, the following composite variables were calculated and recorded. The CP score was obtained using five parameters: the albumin and TBIL serum concentration, prothrombin time, and presence and severity of encephalopathy and ascites [5]. The MELD score was calculated with the following formula: $3.8 \times \ln(\text{bilirubin mg/dL}) + 11.2 \times \ln(\text{INR}) + 9.6 \times \ln(\text{creatinine mg/dL}) + 6.4 \times$ (etiology: 0 if cholestatic or alcoholic, 1 otherwise), and the score was rounded to the nearest integer [18]. The equation for the ALBI score was $0.66 \times \lg(\text{bilirubin } \mu\text{mol/L}) - 0.085 \times (\text{albumin g/L})$ [8]. The body mass index (BMI) was determined as weight in kilograms divided by height in meters squared [19].

1.4. Data analysis

1.4.1. PHLF-related radiomics feature selection and the construction of the Rad-score system

To avoid the “curse of dimensionality” [20], which would cause a large false positive result, we used a logistic regression with least absolute shrinkage and selection operator (LASSO) regularization to select the most informative features from the training cohort. The LASSO logistic regression model was used with penalty parameter tuning that was conducted by 10-fold cross-validation based on minimum criteria. Then, a Rad-score system was constructed with a calculation for each patient via a linear combination of selected features that were weighted by their respective coefficients [21].

1.4.2. Predictive validation of Rad-score and comparison to other existing methods

The potential predictive value of the Rad-score with PHLF was first assessed in the training cohort and then validated in the validation cohort by using the area under the receiver operator characteristic (ROC) curve (AUC). The AUCs of the CP, MELD, and ALBI scores in predicting PHLF were also calculated, and the comparisons between the Rad-score and other methods were performed using DeLong's test in both cohorts.

1.4.3. Development of a radiomics nomogram

The Rad-score and other clinical risk variables [3,22,23], including TBIL, albumin, alkaline phosphatase (ALP), PLA, INR, CP, MELD, ALBI, BMI and the Eastern Cooperative Oncology Group (ECOG) performance status (PS), were tested in a multivariate logistic regression model to predict PHLF in the training cohort. The model selection was performed with a backward stepwise selection process by using the likelihood ratio test with Akaike's information criterion [21]. Then, a comprehensive nomogram was built upon the basis of the multivariate logistic regression model.

1.4.4. Assessment of nomogram performance

In the training cohort, a calibration curve [24] was used to assess

the calibration of the nomogram. Bootstraps with 1000 resamplings were used for this activity, and the Hosmer-Lemeshow test was performed to assess the goodness-of-fit of the nomogram. The AUCs were calculated to quantify the discrimination performance of the nomogram and were compared with the AUCs of the CP, MELD, and ALBI scores.

1.4.5. Validation of the nomogram

In the validation cohort, the internal validation of the nomogram was performed. Each patient in the validation cohort received a Rad-score. Then, the Hosmer-Lemeshow test and calibration were performed; the AUCs were calculated and compared per the aforementioned methods.

1.4.6. Clinical use of the radiomics nomogram

To estimate the clinical usefulness of the radiomics nomogram, decision curve analysis (DCA) [25] was conducted to quantify the net benefits at different threshold probabilities in the combined training and validation cohorts.

1.5. Statistical analysis

Statistical analyses were performed with R software version 3.4.3 (https://www.r-project.org). Continuous variables were compared using the Mann-Whitney *U* test, and categorical variables were compared using the Chi-square test. The "glmnet" package was used to perform LASSO logistic regression analysis. The "pROC" package was used in the calculation and comparisons of AUCs and creation of ROC curves. The "rms" package was used in the construction of nomogram and calibration plots. The "generalhoslem" package was used to execute the Hosmer-Lemeshow (H-L) test. The "rmda" package was used in DCA. A two-sided $P < 0.05$ was considered as indicative of statistical significance.

2. Results

2.1. Patient demographics and clinicopathological characteristics

The characteristics of the study participants are presented in Table 1, and there were no differences between the two cohorts. The median ages of the training and validation cohorts were 48.0 and 47.0 years, respectively. There was a preponderance of males in the training and validation cohorts (83.7% and 81.2%, respectively). There were 29 and 11 patients who experienced PHLF in the training and validation cohorts, respectively.

2.2. Reproducibility of feature extraction, radiomics feature selection and Rad-score calculation

A total of 713 radiomics features of the liver parenchyma and hepatic tumors were extracted from preoperative CT images. The intra-observer ICC ranged from 0.842 to 0.925 for the liver, from 0.860 to 0.934 for the tumor, and the inter-observer ICC ranged from 0.870 to 0.938 for the liver and from 0.884 to 0.945 for the tumor (Table 2), indicating satisfactory feature extraction reproducibility. In the training cohort, 7 potential predictors of PHLF-related features with non-zero coefficients were selected in the LASSO logistic regression model (Fig. 1A and B). Among the 7 features, 6 were derived from the liver parenchyma, and 1 was derived from the tumor. Then, the formula for the Rad-score calculation was generated as follows:

$$\text{Rad-score} = -1.054862 \times 10^{0} + 2.467326 \times 10^{-5} \times \text{LHL\_GLRLM\_Long\_Run\_High\_Gray\_Level\_Emphasis} + 7.205613 \times 10^{-12} \times \text{LHL\_GLSZM\_Large\_Area\_High\_Gray\_Level\_Emphasis} + 1.971394 \times 10^{0} \times \text{HLL\_First-Order\_Features\_Mean} + 3.226196 \times 10^{-1} \times \text{HLL\_GLCM\_Cluster\_Shade} + 1.161079 \times 10^{0} \times \text{HLH\_GLSZM\_Size\_Zone\_Non\_Uniformity\_Normalized} - 8.577448 \times 10^{0} \times \text{HHL\_First-Order\_Features\_Mean} + 1.256704 \times 10^{-6} \times \text{Tumor\_Volume}$$

Table 1
Demographic and clinicopathologic characteristics of the study participants.

Demographic or Characteristic	Training cohort (n = 80)	Validation cohort (n = 32)	*P* value
Age, median(IQR), years	48.0(39.5–58.0)	47.0(38.8–53.8)	0.384
Sex			0.751
Male	67(83.7)	26(81.2)	
Female	13(16.3)	6(18.8)	
BMI, mean ± SD, kg/m^2	22.98 ± 3.01	21.72 ± 3.34	0.055
Etiology of viral hepatitis			0.608
None	8(10.0)	5(15.6)	
HBV	68(85.0)	25(78.1)	
HCV	2(2.5)	0(0.0)	
HBV co-infected with HCV	2(2.5)	2(6.3)	
AFP, median(IQR), μg/L	88.8(7.1–895.6)	64.2(3.9–1345.7)	0.376
ALB, median(IQR), g/L	40.1(37.3–41.7)	39.9(37.3–43.1)	0.946
TBIL, median (IQR), μmol/L	10.7(8.7–15.4)	12.1(9.6–16.7)	0.415
ALT, median(IQR), U/L	34.0(24.3–45.0)	33.5(25.5–64.8)	0.451
GGT, median (IQR), U/L	56.5(32.3–117.3)	57.0(35.3–127.0)	0.875
ALP, median(IQR), U/L	90.5(73.3–133.5)	90.5(73.3–113.3)	0.640
Cr, median(IQR), μmol/L	88.0(73.3–103.8)	81.5(69.8–97.5)	0.467
PLA, median(IQR), $10^9/L$	180.0(130.5–245.5)	173.0(128.3–265.8)	0.936
INR, median(IQR)	1.05(1.01–1.12)	1.03(0.99–1.10)	0.235
Child-Pugh score, median(IQR)	5.0(5.0–5.8)	5.0(5.0–6.0)	0.759
Child-Pugh grade			1.000
A	75(93.7)	30(93.7)	
B	5(6.3)	2(6.3)	
C	0(0)	0(0)	
MELD score, median(IQR)	4.7(2.3–7.0)	4.2(3.6–5.4)	0.964
ALBI score, median (IQR)	−2.71(-2.90～-2.49)	−2.72(-2.91～-2.37)	0.877
ECOG PS			0.805
0	49(61.2)	18(56.2)	
1	20(25.0)	11(34.4)	
2	11(13.8)	3(9.4)	
3-4	0(0)	0(0)	
Cirrhosis			0.299
Yes	73(91.2)	31(96.9)	
No	7(8.8)	1(3.1)	
Size, median(IQR), cm	6.9(4.5–8.9)	5.2(2.8–8.1)	0.104
Extent of resection			0.507
Minor resection	32(40.0)	15(46.9)	
Major resection	48(60.0)	17(53.1)	
Differentiation degree			0.547
poorly	13(16.2)	8(25.0)	
moderately	62(77.5)	21(65.6)	
well	5(6.3)	3(9.4)	
PHLF			0.852
Yes	29(36.2)	11(34.4)	
No	51(63.8)	21(65.6)	

Abbreviations: IQR, interquartile range; BMI, body mass index; SD, standard deviation; HBV/HCV, hepatitis B/C virus; AFP, α-fetoprotein; ALB, albumin; TBIL, total bilirubin; ALT, alanine transaminase; GGT, γ-glutamyl transferase; ALP, alkaline phosphatase; Cr, creatinine; PLA, platelet counts; INR, international normalized ratio; MELD, Model of End Stage Liver Disease; ALBI, albumin-bilirubin; ECOG PS, Eastern Cooperative Oncology Group performance status; PHLF, posthepatectomy liver failure.

Patients who had not experienced PHLF generally had lower Rad-scores than those who suffered from PHLF. A significant difference was evident between the Rad-scores (median [interquartile range]) of these

Table 2
Evaluation of reproducibility of features extraction.

ROI	inter-class correlation coefficient	intra-class correlation coefficient
Liver	0.891(0.842–0.925)	0.910(0.870–0.938)
Tumor	0.904(0.860–0.934)	0.921(0.884–0.945)

two groups in the training cohort (−0.8718[-1.1002～-0.7755] vs −0.5576[-0.6863～-0.4319], respectively, $P < 0.0001$) and in the validation cohort (−0.7033 [-0.9346～-0.5531] vs −0.3606[-0.7086～-0.3125], respectively, $P < 0.025$).

2.3. Discrimination power of the Rad-score, CP score, MELD score, and ALBI score

The predictive accuracy of the Rad-score was quite favorable, with an AUC of 0.822 (95% CI, 0.726–0.917) in the training cohort and 0.762 (95% CI, 0.576–0.948) in the validation cohort. In the training cohort, the AUC of the Rad-score was higher than those of the CP and ALBI scores for predicting PHLF ($P < 0.05$), while no significant difference was evident between the Rad-score and MELD score (P = 0.169). In the training cohort, no statistical differences were evident between the Rad-score and the other scoring systems (Rad-score vs CP, P = 0.373; Rad-score vs ALBI, P = 0.421; Rad-score vs MELD, P = 0.321). The detailed results of the discrimination power are shown in Fig. 1C and D.

2.4. Development, apparent performance and validation of an individualized prediction nomogram

Logistic regression analysis identified the Rad-score, MELD score, and PS as independent predictors of PHLF in patients who underwent hepatectomy (Table 3). Then, a nomogram that incorporated the above independent predictors was formulated based upon the results of a multivariate analysis (Fig. 2A). In the training cohort, the AUC of the nomogram (Fig. 2B) was 0.864 (95% CI, 0.786–0.942), which was higher than those of the conventional methods (nomogram vs CP, MELD, and ALBI; $P < 0.001$, $P < 0.005$, $P < 0.005$, respectively). In the validation cohort, the AUC (Fig. 2C) was 0.896 (95% CI, 0.774–1.000), and significant statistical differences were also observed between the nomogram and other prediction algorithms (nomogram vs CP, MELD, and ALBI; $P < 0.02$, $P < 0.01$, $P < 0.01$, respectively).

Fig. 3A and B shows the calibration curve of the nomogram. The calibration curve and lack of a statistical significance in the H-L test (P = 0.100 in the training cohort and P = 0.386 in the validation

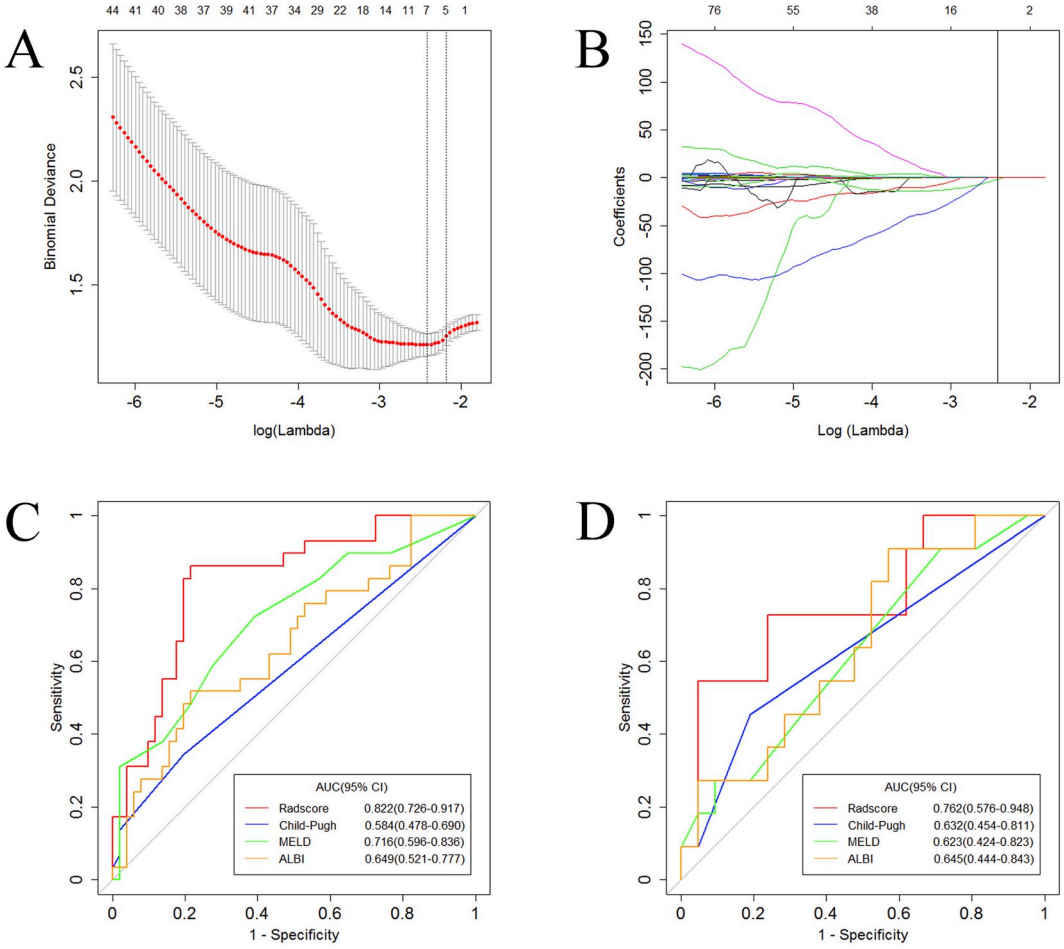

Fig. 1. Radiomics feature selection using logistic regression with LASSO regularization and the discrimination performance of the Rad-score and other clinical scoring systems. (A) 10-fold cross-validation based on minimum criteria was used to select tuning parameter (lambda) in the LASSO model. The binomial deviance was plotted vs log(lambda), and the numbers along the upper x-axis indicate the number of predictors. Two dotted vertical lines were drawn at the optimal values by using the minimum criteria and one standard error of the minimum criteria. An optimal lambda value of 0.0893, with log(lambda) = -2.4158 was chosen through 10-fold cross-validation. (B) LASSO coefficient profiles of the 713 radiomics features. The y-axis indicates the value of the coefficient, and the lower x-axis indicates log (lambda). The upper x-axis represents the number of non-zero coefficients. Each curve represents the trajectory of the change of each independent variable. The vertical line was drawn at the value selected using 10-fold cross-validation, where optimal lambda resulted in seven coefficients. (C) and (D) show the ROC curves for the Rad-score, Child-Pugh score, MELD score and ALBI score in predicting PHLF in the training and validation cohort, respectively.

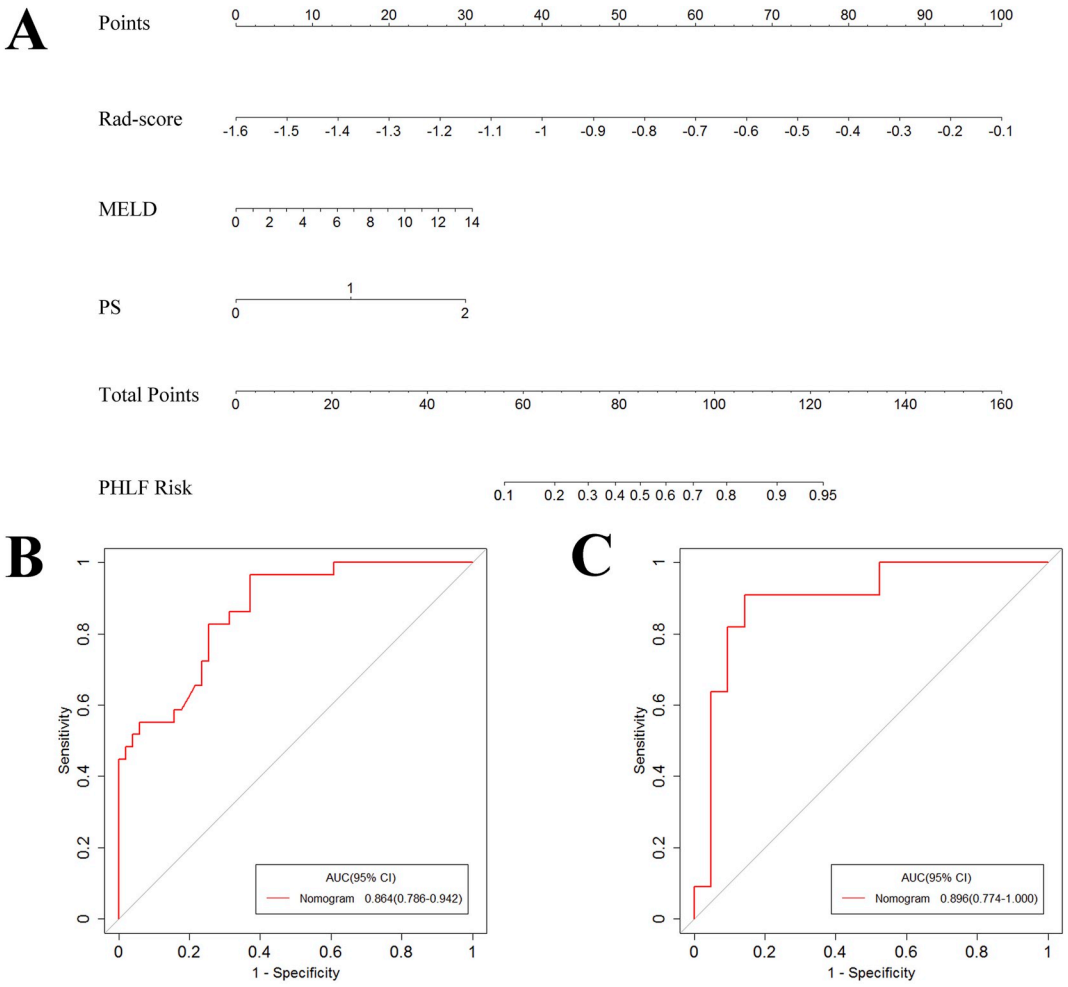

Fig. 2. Radiomics nomogram for the prediction of PHLF and the discrimination performance of this method. (A) The radiomics nomogram was developed in the training cohort and incorporated the Rad-score, the MELD score and PS. The above three predictors (2nd ~ 4th lines) are represented and correspond to a point on the 1st axis. The sum of these three points is located on the total points axis, and a line is drawn downward to the risk of PHLF. Each patient received an individual value. (B) and (C) show the ROC curves for the radiomics nomogram in predicting PHLF in the training and validation cohort, respectively.

cohort) indicated good calibration. These results suggest that the nomogram performed well in both cohorts.

2.5. Clinical utility

Fig. 3C presents the DCA of the nomogram and other methods. The DCA indicated that when the threshold probability for a doctor or patient was within a range from 0.01 to 0.89, this nomogram added more net benefit than the "treat all" or "treat none" scheme. While the threshold probability was with a range from 0.30 to 0.74, 0.19 to 0.61, 0.14 to 0.57 for Child-Pugh, MELD, ALBI, respectively. The clinical utility of the nomogram is the best.

2.6. Prospective pilot analysis

We used the radiomics nomogram to prospective analysis another 13 HCC patients who underwent hepatectomy. There were ten men and three women, and the median age was 54.0 years old. There were four patients experienced PHLF. The characteristics of these patients are presented in Table 4.

The radiomics nomogram could predict PHLF with an AUC of 0.833 (95% CI, 0.591–1.000) (Fig. 4A) and the DCA indicated that when the threshold probability for a doctor or patient was within a range from 0.05 to 0.70 (Fig. 4B), this nomogram added more net benefit than the "treat all" or "treat none" scheme, suggesting the clinical utility of the nomogram is still good. The results of prospective analysis indicated that this nomogram showed a good performance in a prospective setting.

3. Discussion

The results of this study support our hypothesis that a potential PHLF-predictive imaging biomarker, which integrates both liver and tumor features in a radiomics context, can be obtained from the preoperative CT data of HCC patients. The performance of the Rad-score is good but not completely superior to the performances of other existing methods. Therefore, we further developed and validated a diagnostic, radiomics-based nomogram for the prediction of PHLF. This nomogram consisted of three parameters, namely, the Rad-score, MELD score, and PS. The predictive ability was better than the other methods, and this nomogram could enhance the preoperative individualized prediction of PHLF. DCA showed that this simple-to-use nomogram could be applied as a routine clinical tool.

Texture analysis is a form of quantitative image processing that assesses the spatial interrelationships of pixel intensities [26] and is useful in medical practice [27,28]. Texture analysis constitutes a larger class of quantitative imaging analyses known as radiomics, which contains much more valuable information and is considered as a potential bridge between medical imaging and personalized medicine [9,10,29]. Radiomics is a newly established field that adopts advanced

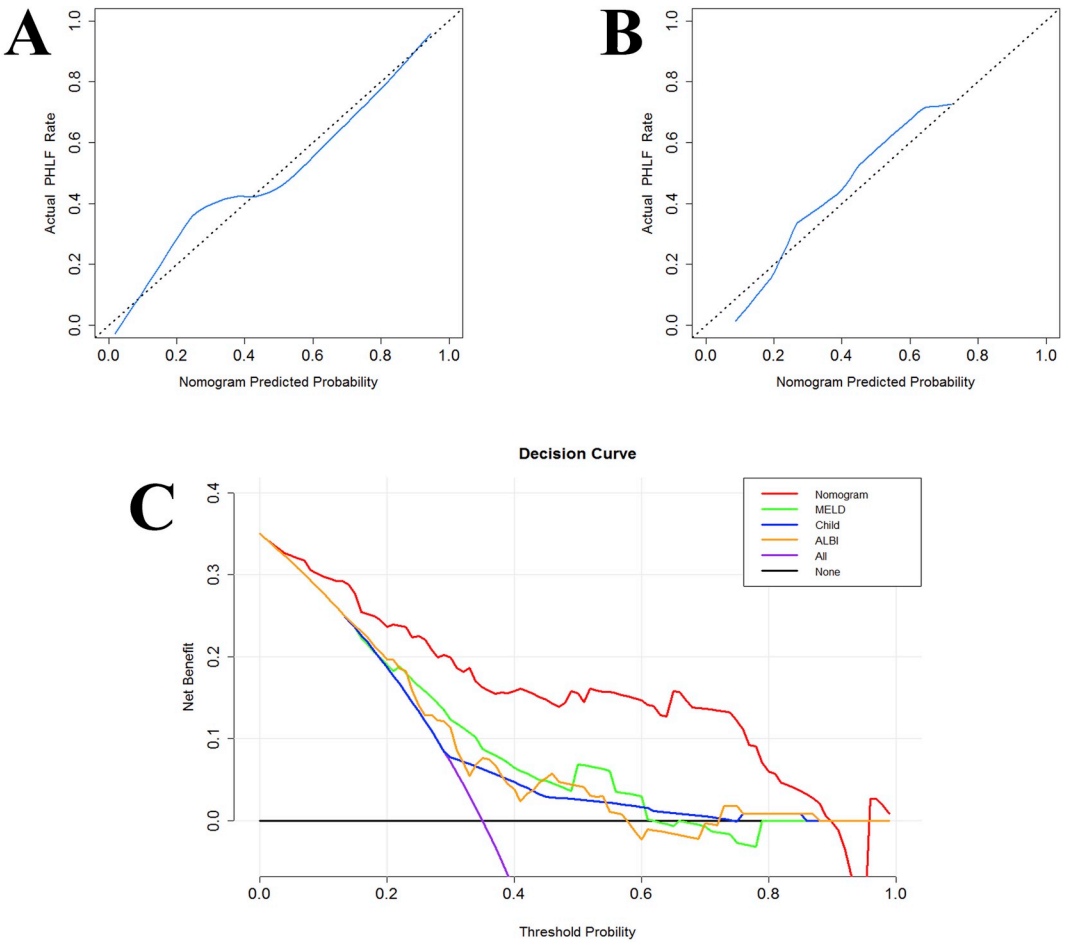

Fig. 3. Calibration curves and decision curve analysis for the radiomics nomogram. (A) The calibration curve of the radiomics nomogram in the training cohort. (B) The calibration curve of the radiomics nomogram in the validation cohort. The x-axis indicates the nomogram predicted probability of PHLF, and the y-axis represents the actual PHLF rate. The dotted line represents a perfect prediction, and the blue solid line represents the predictive performance of this radiomics nomogram. The closer the blue line fit is to the dotted line, the better the prediction of the radiomics nomogram will be. (C) Decision curve analysis of the radiomics nomogram and other methods in the combined training and validation cohorts. The x-axis represents the threshold probability. The y-axis represents the net benefit. The red line represents the radiomics nomogram. The green line represents the MELD. The blue line represents the Child-Pugh. The orange line represents the ALBI. The purple line represents the assumption that all patients have PHLF, and the black line represents the assumption that no patient has PHLF. The threshold probability is where the expected benefit of treatment is equal to the expected benefit of avoiding treatment. The decision curve shows that if the threshold probability is within a range from 0.01 to 0.89, more net benefit is added by using the nomogram for predicting PHLF than by treating either no or all patients. While the threshold probability was with a range from 0.30 to 0.74, 0.19 to 0.61, 0.14 to 0.57 for Child-Pugh, MELD, ALBI, respectively. The clinical utility of the nomogram is the best.

computational methodologies to mine medical imaging data deeply and convert these data into quantitative parameters that can be applied for cancer diagnosis, staging, prognosis, treatment response predictions, disease monitoring and surveillance [12,29]. We have selected CT instead of magnetic resonance imaging (MRI) because the standardization of MRI radiomics is considerably more difficult and because more nuanced sequences and protocols exist between institutions for MRI than for CT [11]. Thus, the findings derived from CT imaging could be generalized sooner.

Radiomics was originally focused on tumor tissues [10,12]. Then,

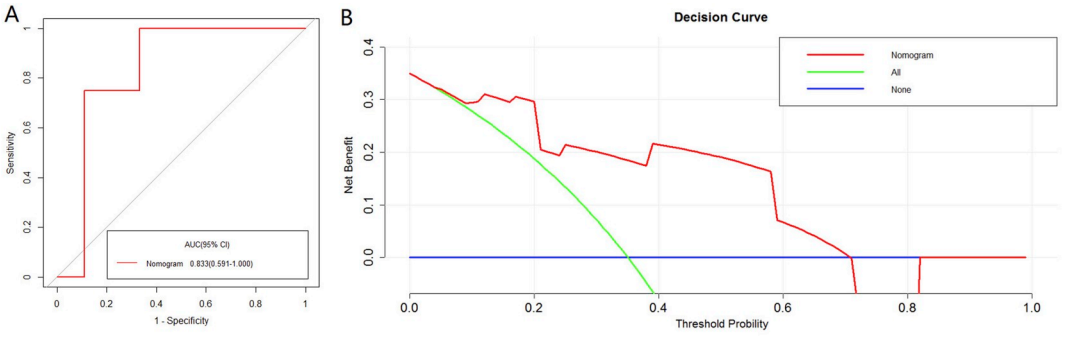

Fig. 4. The AUC (A) and the DCA (B) of the nomogram.

Table 3
Risk factors for PHLF in patients with HCC.

Variable	Univariate logistic regression		Multivariate logistic regression	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value
PLA	0.9944 (0.9881–1.0007)	0.080	NA	NA
ALP	1.0002 (0.9929–1.0076)	0.952	NA	NA
INR (per 0.1 increase)	1.324 (0.777–2.262)	0.305	NA	NA
TB	1.0069 (0.9438–1.0741)	0.836	NA	NA
AL	0.86 (0.76–0.98)	0.022*	NA	NA
Child-Pugh	1.84 (0.96–3.54)	0.068	NA	NA
MELD	1.28 (1.09–1.51)	0.003*	1.32 (1.01–1.74)	0.044*
ALBI	4.27 (1.16–15.7)	0.029*	NA	NA
BMI (< 25 vs ≥ 25)	0.5 (0.16–1.56)	0.232	NA	NA
PS	2.47 (1.28–4.78)	0.007*	4.87 (1.28–18.45)	0.02*
Radscore (per 0.1 increase)	1.652(1.298–2.103)	< 0.001*	1.714 (1.257–2.335)	< 0.001*

Abbreviations: PHLF, posthepatectomy liver failure; HCC, hepatocellular carcinoma; NA, not available. In the multivariate logistic regression analysis, these values were not available because they were eliminated; *$P < 0.05$.

several findings suggested that radiomics could also applied to non-tumor circumstances, such as identifying high-risk atherosclerotic plaque [30], diagnosing and subtyping attention deficit hyperactivity disorder [31], and detecting Meniere's disease [32]. Hepatectomy is a mainstay of treatment for HCC patients, and PHLF is a possible and undesirable outcome, which may lead to hepatectomy-related mortality. In general, the precise prediction of PHLF remains difficult, so we attempted to address this problem with a new approach. To build a Rad-score system, logistic regression with LASSO regularization was used because it is an effective selection strategy for continuous models, especially for those where the number of predictors far exceeds the number of observations [33]. After LASSO logistic regression, seven selected features were combined into a Rad-score as a predictor, according with the statistical standpoint that a combination of multiple prognostic factors could show a better predictive performance than just one factor [34]. One recent study shows that a combined radiomics signature can be used as an independent biomarker to estimate of disease-free survival in patients with early stage non-small cell lung cancer [35]. Similarly, in our study, this radiomics signature demonstrated adequate discrimination with an AUC of 0.822 in the training cohort and 0.762 in the validation cohort. Therefore, we propose that changes in liver parenchymal radiomics features from CT imaging are associated with the degree of underlying hepatic fibrosis and that tumoral properties also influence surgical safety.

Table 4
The demographic and clinicopathologic characteristics of the prospective cohort study.

Demographic or Characteristic	Prospective cohort
Age, median (IQR), years	54.0(48.5–60.5)
Sex	
Male	10 (76.9)
Female	3(23.1)
History of hepatitis	
Yes	12(92.3)
No	1(7.7)
Child-Pugh score, median (IQR)	5.0(5.0–6.0)
A	13(100)
B	0(0)
C	0(0)
MELD score, median (IQR)	4.63(1.19–7.03)
ALBI score, median (IQR)	−2.67(-2.89～-2.48)
Radscore, median (IQR)	−1.407(-1.740～-0.053)
ECOG	
0	10(76.9)
1	3(23.1)
2	0(0)
3-4	0(0)
PHLF	
Yes	4(30.8)
No	9(69.2)

Pre-hepatectomy safety evaluations and PHLF risk prediction have been hot research areas for several decades [36,37]. The CP score is widely used for prognostication and treatment allocation for HCC patients. However, the CP score is limited the five parameters that share the same weighting because class B includes a wide range of patients, whose prognosis may be entirely different [3,13]. The MELD score was initially developed to predict the mortality of patients receiving transjugular intrahepatic portosystemic shunts and then applied to predict PHLF, showing good application prospects [38]. Although the MELD score is promising, the predictive value may be limited to patients with cirrhosis [39]. The ALBI score is an emerging simple and informative clinical predictive model, which only contains two objective laboratory variables, and has been garnered increasing attention in recent years [5,8]. However, some researchers have suggested that grade 2 ALBI comprised patients with prognoses that were different, which may affect the distinguishing ability of the ALBI score [13]. All of the above methods have drawbacks to some extent, but their usefulness cannot be ignored. Therefore, based on a statistical viewpoint that accounting for biomarkers from different perspectives could enhance clinical management [40], we chose several predictive factors from different aspects to establish a multiscale nomogram.

Among the three constituent parts of the nomogram, the PS was an important reference index that determined prognosis in patients with HCC [22,41]. Cancer-related symptoms have been identified as an independent predictive factor of prognosis [6,42], and the PS is a useful assessment tool to assess HCC patients' actual level of function and capability of self-care. So, the clinical manifestations of HCC could be evaluated by PS to a certain degree. The results of the nomogram showed good calibration and discrimination in the training and validation cohorts, and notably, the AUC of this nomogram was higher than those of the other three predictive methods, with the significant differences in both cohorts manifesting the superiority of the nomogram for predicting the PHLF of HCC patients who would undergo hepatectomy.

4. Limitations

This study has several potential limitations. We only performed a single-center study without a large capacity of patients, and lacked external validation from other hospitals. Polycentric large-scale samples are necessary to obtain high-quality evidence for future clinical application.

5. Conclusions

A novel radiomics-based nomogram was developed and validated for predicting PHLF in patients with HCC. The nomogram provides a new approach toward identifying PHLF risk and improving decision support in the treatment of HCC at low cost.

Conflict of interest

None.

Funding

This work was supported in part by the NSFC-Union Program (U1401254 and U1613221), the National Key Research and Development Program (2016YFC0106500/1/2 and SQ2017ZY040217/03), the Guangdong Scientific and Technology Program (2015B020214005 and 2016A030313180), the Shenzhen Key Basic Science Program (JCYJ20170413162213765), and the Shenzhen Key Laboratory Program (ZDSYS201707271637577).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.suronc.2018.11.013.

References

[1] Global Burden of Disease Liver Cancer Collaboration, The burden of primary liver cancer and underlying etiologies from 1990 to 2015 at the global, regional, and national level: results from the global burden of disease study 2015, JAMA Oncol 3 (12) (2017) 1683–1691, https://doi.org/10.1001/jamaoncol.2017.3055.
[2] J.M. Llovet, A. Villanueva, A. Lachenmayer, et al., Advances in targeted therapies for hepatocellular carcinoma in the genomic era, Nat. Rev. Clin. Oncol. 12 (8) (2015) 436, https://doi.org/10.1038/nrclinonc.2015.121.
[3] A. Forner, J.M. Llovet, J. Bruix, Hepatocellular carcinoma, Lancet 379 (9822) (2012) 1245–1255, https://doi.org/10.1016/S0140-6736(11)61347-0.
[4] A.L. Simpson, L.B. Adams, P.J. Allen, et al., Texture analysis of preoperative CT images for prediction of postoperative hepatic insufficiency: a preliminary study, J. Am. Coll. Surg. 220 (3) (2015) 339–346, https://doi.org/10.1016/j.jamcollsurg.2014.11.027.
[5] Y.Y. Wang, J.H. Zhong, Z.Y. Su, et al., Albumin-bilirubin versus Child-Pugh score as a predictor of outcome after liver resection for hepatocellular carcinoma, Br. J. Surg. 103 (6) (2016) 725–734, https://doi.org/10.1002/bjs.10095.
[6] A. Forner, M. Gilabert, J. Bruix, et al., Treatment of intermediate-stage hepatocellular carcinoma, Nat. Rev. Clin. Oncol. 11 (9) (2014) 525–535, https://doi.org/10.1038/nrclinonc.2014.122.
[7] F. Gani, M. Cerullo, N. Amini, et al., Frailty as a risk predictor of morbidity and mortality following liver surgery, J. Gastrointest. Surg. 21 (5) (2017) 822–830, https://doi.org/10.1007/s11605-017-3373-6.
[8] P.J. Johnson, S. Berhane, C. Kagebayashi, et al., Assessment of liver function in patients with hepatocellular carcinoma: a new evidence-based approach-the ALBI grade, J. Clin. Oncol. 33 (6) (2015) 550–558, https://doi.org/10.1200/JCO.2014.57.9151.
[9] H.J. Aerts, E.R. Velazquez, R.T. Leijenaar, et al., Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach, Nat. Commun. 5 (2014) 4006, https://doi.org/10.1038/ncomms5006.
[10] R.J. Gillies, P.E. Kinahan, H. Hricak, Radiomics: images are more than pictures, they are data, Radiology 278 (2) (2016) 563–577, https://doi.org/10.1148/radiol.2015151169.
[11] V. Verma, C.B. Simone 2nd, S. Krishnan, et al., The rise of radiomics and implications for oncologic management, J. Natl. Cancer Inst. 109 (7) (2017), https://doi.org/10.1093/jnci/djx055 djx055.
[12] E.J. Limkin, R. Sun, L. Dercle, et al., Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology, Ann. Oncol. 28 (6) (2017) 1191–1206, https://doi.org/10.1093/annonc/mdx034.
[13] J. Edeline, J.F. Blanc, P. Johnson, et al., A multicentre comparison between Child Pugh and Albumin-Bilirubin scores in patients treated with sorafenib for hepatocellular carcinoma, Liver Int. 36 (12) (2016) 1821–1828, https://doi.org/10.1111/liv.13170.
[14] European Association For The Study Of The Liver, European Organisation for Research and Treatment of Cancer. EASL-EORTC clinical practice guidelines: management of hepatocellular carcinoma, J. Hepatol. 56 (4) (2012) 908–943, https://doi.org/10.1016/j.jhep.2011.12.001.
[15] J.J.M. van Griethuysen, A. Fedorov, C. Parmar, et al., Computational radiomics system to decode the radiographic phenotype, Cancer Res. 77 (21) (2017) e104–e107, https://doi.org/10.1158/0008-5472.CAN-17-0339.
[16] N.N. Rahbari, C. Reissfelder, M. Koch, et al., The predictive value of postoperative clinical risk scores for outcome after hepatic resection: a validation analysis in 807 patients, Ann. Surg. Oncol. 18 (13) (2011) 3640–3649, https://doi.org/10.1245/s10434-011-1829-6.
[17] N.N. Rahbari, O.J. Garden, R. Padbury, et al., Posthepatectomy liver failure: a definition and grading by the international study group of liver surgery (ISGLS), Surgery 149 (5) (2011) 713–724, https://doi.org/10.1016/j.surg.2010.10.001.
[18] P.S. Kamath, R.H. Wiesner, M. Malinchoc, et al., A model to predict survival in patients with end-stage liver disease, Hepatology 33 (2) (2001) 464–470, https://doi.org/10.1053/jhep.2001.22172.
[19] J.K. Saunders, A.S. Rosman, D. Neihaus, T.H. Gouge, M. Melis, Safety of hepatic resections in obese veterans, Arch. Surg. 147 (4) (2012) 331–337, https://doi.org/10.1001/archsurg.2011.1404.
[20] C. Ferte, A.D. Trister, E. Huang, et al., Impact of bioinformatic procedures in the development and translation of high-throughput molecular classifiers in oncology, Clin. Canc. Res. 19 (16) (2013) 4315–4325, https://doi.org/10.1158/1078-0432.CCR-12-3937.
[21] S. Wu, J. Zheng, Y. Li, et al., A radiomics nomogram for the preoperative prediction of lymph node metastasis in bladder cancer, Clin. Canc. Res. 23 (2) (2017) 6905–6911, https://doi.org/10.1158/1078-0432.CCR-17-1510.
[22] F. Farinati, A. Vitale, G. Spolverato, et al., Development and validation of a new prognostic system for patients with hepatocellular carcinoma, PLoS Med. 13 (4) (2016) e1002006, , https://doi.org/10.1371/journal.pmed.1002006.
[23] M.S. Didolkar, J.L. Fitzpatrick, E.G. Elias, et al., Risk factors before hepatectomy, hepatic function after hepatectomy and computed tomographic changes as indicators of mortality from hepatic failure, Surg. Gynecol. Obstet. 169 (1) (1989) 17–26.
[24] A.C. Alba, T. Agoritsas, M. Walsh, et al., Discrimination and calibration of clinical prediction models: users' guides to the medical literature, J. Am. Med. Assoc. 318 (14) (2017) 1377–1384, https://doi.org/10.1001/jama.2017.12126.
[25] A.J. Vickers, E.B. Elkin, Decision curve analysis: a novel method for evaluating prediction models, Med. Decis. Making 26 (6) (2006) 565–574.
[26] R.M. Summers, Texture analysis in radiology: does the emperor have no clothes? Abdominal Radiology (NY) 42 (2) (2017) 342–345, https://doi.org/10.1007/s00261-016-0950-1.
[27] J.E. Koss, F.D. Newman, T.K. Johnson, D.L. Kirch, Abdominal organ segmentation using texture transforms and a Hopfield neural network, IEEE Trans. Med. Imag. 18 (7) (1999) 640–648.
[28] S. Rao, D.M. Lambregts, R.S. Schnerr, et al., CT texture analysis in colorectal liver metastases: a better way than size and volume measurements to assess response to chemotherapy? United European Gastroenterol J 4 (2) (2016) 257–263, https://doi.org/10.1177/2050640615601603.
[29] P. Lambin, R.T.H. Leijenaar, T.M. Deist, et al., Radiomics: the bridge between medical imaging and personalized medicine, Nat. Rev. Clin. Oncol. 14 (12) (2017) 749–762, https://doi.org/10.1038/nrclinonc.2017.141.
[30] M. Kolossváry, J. Karády, B. Szilveszter, et al., Radiomic features are superior to conventional quantitative computed tomographic metrics to identify coronary plaques with Napkin-Ring Sign, Circle: Cardiovasc Imaging 10 (12) (2017) e006843, , https://doi.org/10.1161/CIRCIMAGING.117.006843.
[31] H. Sun, Y. Chen, Q. Huang, et al., Psychoradiologic utility of MR imaging for diagnosis of attention deficit hyperactivity disorder: a radiomics analysis, Radiology (2017) 170226, , https://doi.org/10.1148/radiol.2017170226.
[32] E.L. van den Burg, M. van Hoof, A.A. Postma, et al., An exploratory study to detect Meniere's disease in conventional MRI scans using radiomics, Front. Neurol. 7 (2016) 190, https://doi.org/10.3389/fneur.2016.00190.
[33] T.T. Wu, Y.F. Chen, T. Hastie, E. Sobel, K. Lange, Genome-wide association analysis by lasso penalized logistic regression, Bioinformatics 25 (6) (2009) 714–721, https://doi.org/10.1093/bioinformatics/btp041.
[34] R.D. Riley, J.A. Hayden, E.W. Steyerberg, et al., Prognosis research strategy (PROGRESS) 2: prognostic factor research, PLoS Med. 10 (2) (2013) e1001380, , https://doi.org/10.1371/journal.pmed.1001380.
[35] Y. Huang, Z. Liu, L. He, et al., Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (I or II) non-small cell lung cancer, Radiology 281 (3) (2016) 947–957, https://doi.org/10.1148/radiol.2016152234.
[36] T. Nishio, K. Taura, Y. Koyama, et al., Prediction of posthepatectomy liver failure based on liver stiffness measurement in patients with hepatocellular carcinoma, Surgery 159 (2) (2016) 399–408, https://doi.org/10.1016/j.surg.2015.06.024.
[37] N. Yamanaka, E. Okamoto, K. Kuwata, N. Tanaka, A multiple regression equation for prediction of posthepatectomy liver failure, Ann. Surg. 200 (5) (1984) 658–663.
[38] S.H. Teh, D.M. Nagorney, S.R. Stevens, et al., Risk factors for mortality after surgery in patients with cirrhosis, Gastroenterology 132 (4) (2007) 1261–1269.
[39] A. Annamalai, M.Y. Harada, M. Chen, et al., Predictors of mortality in the critically III cirrhotic patient: is the model for end-stage liver disease enough? J. Am. Coll. Surg. 224 (3) (2017) 276–282, https://doi.org/10.1016/j.jamcollsurg.2016.11.005.
[40] M. Birkhahn, A.P. Mitra, R.J. Cote, Molecular markers for bladder cancer: the road to a multimarker approach, Expert Rev. Anticancer Ther. 7 (12) (2014) 1717–1727.
[41] J.M. Llovet, C. Bru, J. Bruix, Prognosis of hepatocellular carcinoma: the BCLC staging classification, Semin. Liver Dis. 19 (3) (1999) 329–338.
[42] G. Cabibbo, M. Enea, M. Attanasio, J. Bruix, A. Craxì, C. Cammà, A meta-analysis of survival rates of untreated patients in randomized clinical trials of hepatocellular carcinoma, Hepatology 51 (4) (2010) 1274–1283, https://doi.org/10.1002/hep.23485.