European Radiology (2019) 29:4177–4187
https://doi.org/10.1007/s00330-018-5986-x

MAGNETIC RESONANCE

Pretreatment prediction of immunoscore in hepatocellular cancer: a radiomics-based clinical model based on Gd-EOB-DTPA-enhanced MRI imaging

Shuling Chen[1] · Shiting Feng[2] · Jingwei Wei[3,4,5] · Fei Liu[3,4,5] · Bin Li[6] · Xin Li[7] · Yang Hou[8] · Dongsheng Gu[3,4,5] · Mimi Tang[9] · Han Xiao[9] · Yingmei Jia[2] · Sui Peng[6,9] · Jie Tian[3,4,5] · Ming Kuang[1,10]

Received: 26 August 2018 / Revised: 22 November 2018 / Accepted: 18 December 2018 / Published online: 21 January 2019

Abstract

Objectives Immunoscore evaluates the density of CD3+ and CD8+ T cells in both the tumor core and invasive margin. Pretreatment prediction of immunoscore in hepatocellular cancer (HCC) is important for precision immunotherapy. We aimed to develop a radiomics model based on gadolinium-ethoxybenzyl-diethylenetriamine (Gd-EOB-DTPA)-enhanced MRI for pretreatment prediction of immunoscore (0–2 vs. 3–4) in HCC.

Materials and methods The study included 207 (training cohort: $n = 150$; validation cohort: $n = 57$) HCC patients with hepatectomy who underwent preoperative Gd-EOB-DTPA-enhanced MRI. The volumes of interest enclosing hepatic lesions including intratumoral and peritumoral regions were manually delineated in the hepatobiliary phase of MRI images, from which 1044 quantitative features were extracted and analyzed. Extremely randomized tree method was used to select radiomics features for building radiomics model. Predicting performance in immunoscore was compared among three models: (1) using only intratumoral radiomics features (intratumoral radiomics model); (2) using combined intratumoral and peritumoral radiomics features (combined radiomics model); (3) using clinical data and selected combined radiomics features (combined radiomics-based clinical model).

Results The combined radiomics model showed a better predicting performance in immunoscore than intratumoral radiomics model (AUC, 0.904 (95% CI 0.855–0.953) vs. 0.823 (95% CI 0.747–0.899)). The combined radiomics-based clinical model showed an improvement over the combined radiomics model in predicting immunoscore (AUC, 0·926 (95% CI 0·884–0·967) vs.

Shuling Chen and Shiting Feng contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00330-018-5986-x) contains supplementary material, which is available to authorized users.

✉ Jie Tian
tian@ieee.org

✉ Ming Kuang
kuangminda@hotmail.com

[1] Department of Medical Ultrasonics, Institute of Diagnostic and Interventional Ultrasound, The First Affiliated Hospital of Sun Yat-sen University, 58 Zhong Shan Road 2, Guangzhou 510080, China

[2] Department of Radiology, The First Affiliated Hospital of Sun Yat-sen University, 58 Zhong Shan Road 2, Guangzhou 510080, China

[3] Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

[4] Beijing Key Laboratory of Molecular Imaging, Beijing 100190, China

[5] University of Chinese Academy of Sciences, Beijing 100049, China

[6] Clinical Trial Unit, The First Affiliated Hospital of Sun Yat-sen University, 58 Zhong Shan Road 2, Guangzhou 510080, China

[7] GE HealthCare China, Shanghai 200000, China

[8] Department of Mathematics, Jinan University, Guangzhou 510632, China

[9] Department of Gastroenterology and Hepatology, The First Affiliated Hospital of Sun Yat-sen University, 58 Zhong Shan Road 2, Guangzhou 510080, China

[10] Department of Liver Surgery, The First Affiliated Hospital of Sun Yat-sen University, 58 Zhong Shan Road 2, Guangzhou 510080, China

0·904 (95% CI 0·855–0·953)), although differences were not statistically significant. Results were confirmed in validation cohort and calibration curves showed good agreement.

Conclusion The MRI-based combined radiomics nomogram is effective in predicting immunoscore in HCC and may help making treatment decisions.

Key Points

- *Radiomics obtained from Gd-EOB-DTPA-enhanced MRI help predicting immunoscore in hepatocellular carcinoma.*
- *Combined intratumoral and peritumoral radiomics are superior to intratumoral radiomics only in predicting immunoscore.*
- *We developed a combined clinical and radiomicsnomogram to predict immunoscore in hepatocellular carcinoma.*

Keywords Carcinoma · Hepatocellular · Gadolinium ethoxybenzyl DTPA · Magnetic resonance imaging · Immunotherapy

Abbreviations

AFP	Alpha-fetoprotein
AST	Aspartate transaminase
CT	Center of the tumor
DAB	Diaminobenzidine
DCA	Decision curve analysis
Gd-EOB-DTPA	Gadolinium-ethoxybenzyl-diethylenetriamine
GGT	γ-Glutamyl transpeptadase
GLCM	Gray level co-occurrence matrix
GLRCM	Gray level run-length matrix
HBP	Hepatobiliary phase
HCC	Hepatocellular carcinoma
ICB	Immune checkpoint blockade
ICC	Intra-class correlation coefficient
IM	Invasive margin
NPV	Negative predictive value
PD-1	Programmed death receptor 1
PD-L1	Programmed death-ligand 1
PPV	Positive predictive value
TIL	Tumor infiltrating lymphocytes
TME	Tumor microenvironment
VOI	Volumes of interest

Introduction

Despite improved surveillance and development in treatment strategies, the prognosis of hepatocellular carcinoma (HCC) remains poor due to high recurrence rates and detection at advanced stages [1, 2]. Immunotherapy using immune checkpoint blockade (ICB) is a recent advance for the treatment of HCC [2, 3]. However, on top of high cost, the efficacy of immunotherapy varies greatly across individuals with an objective response rate of 20% [2, 3]. Thus, identification of potential responders to immunotherapy is critical. Previous studies showed that response and survival were closely related to the density of tumor infiltrating lymphocytes (TILs) [4–6]. A study revealed that PD-1 was upregulated on TILs and antibodies against PD-ligand 1 (PD-L1) restore functions of TILs in HCC, which suggested that the density of TILs might be associated with the efficacy of ICB [7]. Analyses of the type, functional orientation, density, and spatial location of TILs have been developed into immunoscore in several cancers [8–13]. For HCC, high immunoscore was significantly associated with a low rate of recurrence and a prolonged recurrence-free survival [14–17]. HCC immunoscrore could be crucial for predicting prognosis and selecting relevant candidates for immunotherapy. However, the immunoscore is mainly determined on postoperative histology, prompting the need for non-invasive pretreatment tools for prediction of immune infiltrates.

The association between imaging features and TIL density has been explored [18–20]. However, these features were either extracted by visual imaging or quantitative kinetic analysis. Radiomics can extract far more features than visual imaging [21–23]. Several studies dealing with glioma, breast, skin, and liver cancers showed that many image features extracted by radiomics, notvisually observed, were closely related to specific microscopic features at the molecular level and could characterize the tumor and its tumor microenvironment (TME) [24–28]. These imaging features, including texture features, filter transformed features, and wavelet features, may be used to predict tumor gene expression, pathological classification, molecular subtyping, or treatment response [23–28]. Hepatobiliary-specific MRI contrast agent, gadolinium-ethoxybenzyl-diethylenetriamine (Gd-EOB-DTPA) improve the delineation of tumor boundaries and could facilitate radiomics study [29, 30]. Most previous radiomics studies focused on the intratumoral region and disregarded boundaries [23, 31–34]. Recently, a breast cancer study suggested that the combined use of intratumoral and peritumoral radiomics was more efficient in predicting treatment efficacy [25].

Thus, we compared an intratumoral only radiomics model, an intratumoral and peritumoral combined radiomics model and a radiomics-based clinical model using Gd-EOB-DTPA-enhanced MRI for pretreatment evaluation of immunoscore in HCC patients.

Materials and methods

Patients

A total of 207 patients who underwent surgery between Nov 2011 and Sep 2017 were consecutively included in this study. The inclusion criteria were as follows: (1) resectable solitary HCC lesion or multiple HCC lesions within one liver lobe; (2) underwent hepatectomy; (3) pathological confirmation of HCC; (4) Child-Pugh A or B; (5) received Gd-EOB-DTPA-enhanced MRI of the liver within 1 month before surgery. Exclusion criteria included (1) received other anti-tumor therapies before surgery; (2) with history of any other concurrent malignancies; (3) incomplete clinical data; or (4) insufficient pathological samples. All enrolled patients were allocated to the training ($n = 150$) and validation cohorts ($n = 57$) in a 7:3 ratio according to the time sequence. The Institutional Ethic Review Board has approved our study and the informed consent was waived. The workflow of the study is shown in Fig. 1.

Immunohistochemistry

Slides from paraffin blocks prepared with surgical tissue specimen were incubated (24 h at 4 °C) with monoclonal antibodies against CD3 (2GV6, Ventana) and CD8 (C8/44, Dako). Envision system and DAB-chromogen (Dako) were applied and counterstained with methylene blue. After reviewing the center of the tumor (CT) and invasive margin (IM), three representative areas of CT and IM were respectively captured under high power magnification (× 200) using ZEISS Axio Scan Z1 Slide Scanner (University of Sydney, Bosch Institute) (Fig. 1). The captured images were analyzed using ImagePro Plus software (Media Cybernetics) to quantify the positive cells of CD3 and CD8 (Fig. 2). The density was recorded as the number of positive cells per unit tissue surface unit in square millimeters [9]. The mean density of the three areas was used for statistical analysis.

Immunoscore calculation

The medians of cell density were taken as the cutoff values to stratify patients into groups based on the degree of tumor infiltration [14, 15]. The cutoff thresholds for CD3 and CD8 cell densities were 189, 92 cells/mm^2 in the CT and 474, 284 cells/mm^2 in the IM, respectively. For each patient, a binary score (0 as low, 1 as high) was given for each immune cell type (CD3+ and CD8+) in each tumor region (CT and IM) according to the developed thresholds. An immunoscore for each patient was derived from the summation of four binary scores (range, 0–4) [8, 35]. Immunoscore ≥ 3 were defined as high immunoscore, and immunoscores ≤ 2 were defined as low immunoscore.

MR imaging acquisition

The MRI equipment and settings used in the present study are detailed in Supplemental Materials and Supplemental Table 1.

Volume of interest identification and segmentation

The volumes of interest (VOIs) were delineated enclosing the 3D volume area of the liver lesions outlined in the hepatobiliary phase (HBP) of MRI images by three

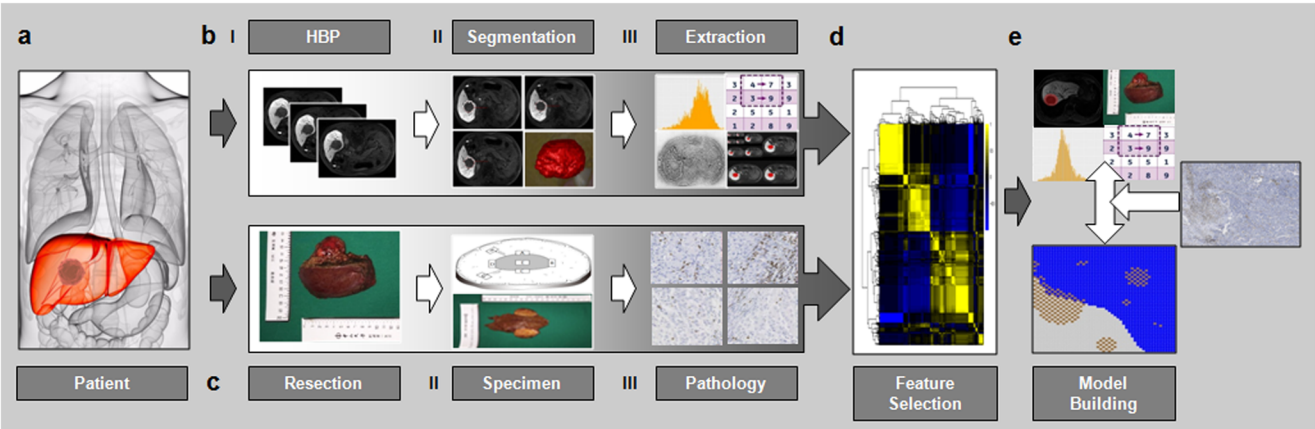

Fig. 1 Flow chart of the study. **a** Enrolling HCC patients who met the inclusion and exclusion criteria; **b** extracting radiomics features: (I) obtaining Gd-EOB-DTPA-enhanced MRI images in the hepatobiliary phase; (II) segmentation: the volume of interest was delineated by experienced radiologists and three-dimensional images were formed; (III) extracting 1044 quantitative features by software; **c** histopathological examination: (I) obtaining gross specimens of tumor tissue; (II) pathologic specimens; (III) immunohistochemistry; **d** data cleaning and feature dimension reduction; **e** establishing the model for predicting immunoscore by machine learning method: features through dimension reduction were applied to establish the model by machine learning method to predict Immunoscore (a diagrammatic sketch of pathological images)

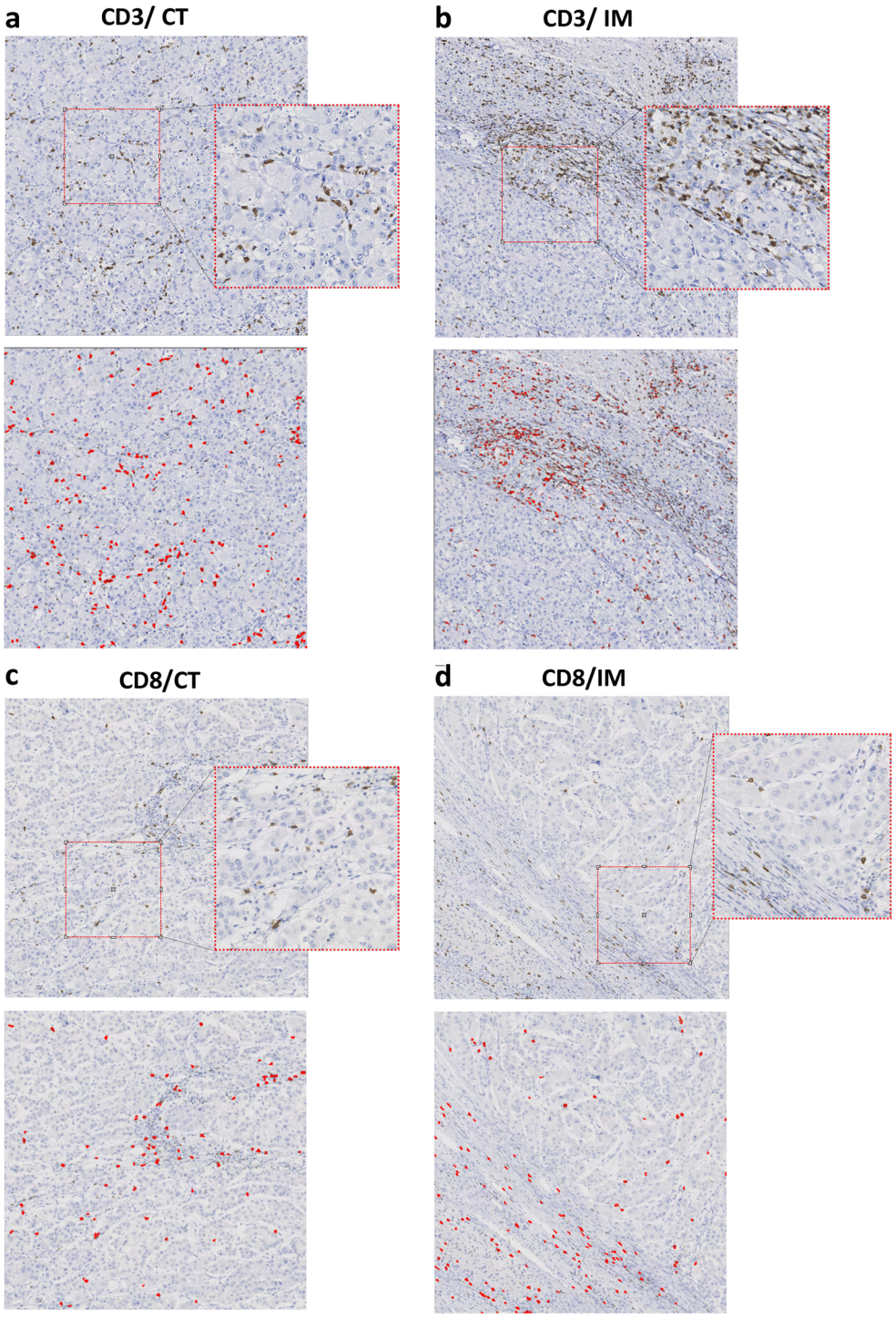

Fig. 2 **a**, **b** Representative examples of CD3 immunostaining in tumor center (CT) and invasive margin (IM) of HCC tissue specimens and digital images of CD3+ T cells analyzed using the image software (ImagePro Plus software). **c**, **d** Representative examples of CD8 immunostaining in CT and IM of HCC tissue specimens and digital images of CD8+ T cells analyzed using the image software. Immunostained cells were brown and tumor cells were blue in color. For each marker, an enlargement of a spot and an enlargement of stained cells are illustrated. In the digitial analyzed images, stained cells were represented in red

independent radiologists with 5 years, 7 years, and 8 years of experience respectively in liver imaging (Fig. 3). The intratumoral region was defined as the area within radiologist-annotated tumor boundaries. Then the annotated intratumoral region was dilated at a radius of 1 cm by topology algorithm, generating the combined intratumoral and peritumoral region (dilated distance was analyzed as dilated pixel counts multiplied by pixel size). Then the reproducibility of the three radiologists to delineate VOI was assessed.

Radiomics feature extraction

The MR image features of all patients were extracted and analyzed by the A.K. software version 2.0.0 (house-made

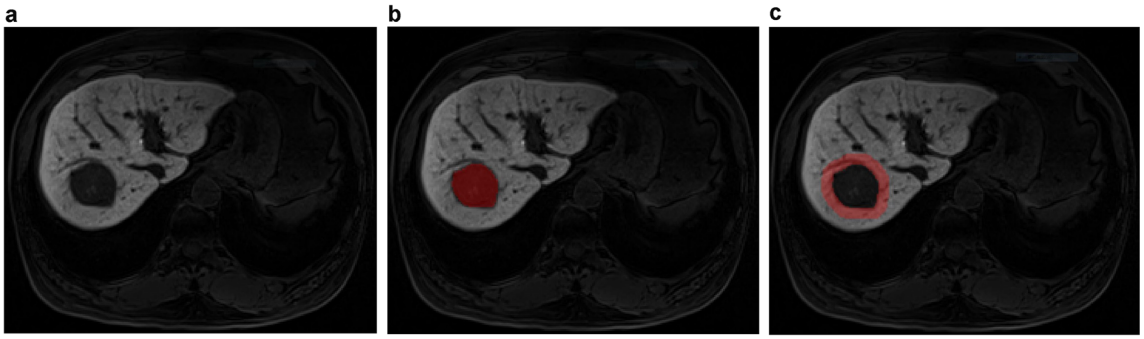

Fig. 3 Representative MR images. **a** MR images showing tumor lesion in the hepatobiliary phase. **b** MR images showing the delineation of intratumoral region. **c** MR images showing the delineation of peritumoral region

software; Analysis-Kit, GE healthcare). A total of 1044 imaging features were extracted, including six kinds of features (Supplemental Table 2): 42 histogram parameters, 10 texture parameters, 9 form factor parameters, 432 gray level co-occurrence matrix (GLCM), 540 gray level run-length matrix (GLRLM), and 11 gray level size zone matrix (GLSZM).

Statistical analysis

The continuous variables were described as mean ± standard deviation or median and quartile, and the categorical variables were described as frequency and percentage. Independent sample *t* test or Kruskal-Wallis (KW) non-parametric rank sum test was used to compare the clinical characteristics between training and validation cohorts for the continuous variables with the former test for variables with normal distribution and the latter one for those with abnormal distribution, while chi-square test or Fisher exact test for categorical variables. Skewness and kurtosis test was used to test the normality of the continuous variables. The inter-observer and intra-observer reproducibility of feature extraction was assessed by the intra-class correlation coefficient (ICC). ICC ≥ 0.8 indicated high consistency, 0.5–0.79 middle, and < 0.5 low [36].

Radiomics feature selection and radiomics model building The extremely randomized tree (Extra-Trees) method was used for building predictive models for immunoscore using features extracted not only from intratumoral region (intratumoral radiomics model) but also from combined intratumoral and peritumoral regions (combined radiomics model) (Fig. 4) [37, 38]. The detailed Extra-Trees method was described in Supplemental Materials. The diagnostic ability of the model was assessed with the area under the characteristics operating curves (AUC), which measures the model's capability to classify immunoscore into 0 (immunoscore ≤ 2) or 1 (immunoscore ≥ 3). The accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated as well.

Development and validation of radiomics-based clinical model In addition to the above two models, we also tried to integrate the radiomics features with clinical data to provide a combined radiomics-based clinical model in immunoscore. The radiomics features and clinical variables were selected into the logistic regression model based on the backward selection with *p* values less than 0.1 in the training cohort. A model was formulated based on the results of the multivariate logistic regression model, and then tested in the validation cohort. A calibration plot was generated to compare the model-predicted versus actual observed probability of high immunoscore.

Comparison of radiomics model and radiomics-based clinical model The comparisons of predicting performance in immunoscore between the combined radiomics model and intratumoral radiomics model, and between the combined radiomics model and combined radiomics-based clinical model were performed using the AUCs and the decision curve analysis (DCA). DCA was conducted to determine the clinical usefulness of these models by quantifying the net benefits at different threshold probabilities. Bonferroni methodology was used to adjust *p* values for multiple comparisons of the above three models.

Association of radiomics features with clinical data A heatmap analysis was performed to present associations between radiomics features and clinical data. Spearman correlation analysis was used to evaluate the associations between radiomics feature and continuous clinical variables, and KW non-parametric rank sum test was used between radiomics feature (continuous variable) and categorical clinical variables.

The Extra-Trees model is built in Python (version 2.7.14), utilizing ExtraTreesClassifier from Scikit-learn. Other statistical analyses were performed by R software version 3.2.3 (Bell Laboratories, Murray Hill, NJ; https://cran.r-project.org/bin/windows/base/old/3.2.3). A two-sided *p* value was considered statistically significant if less than 0.05.

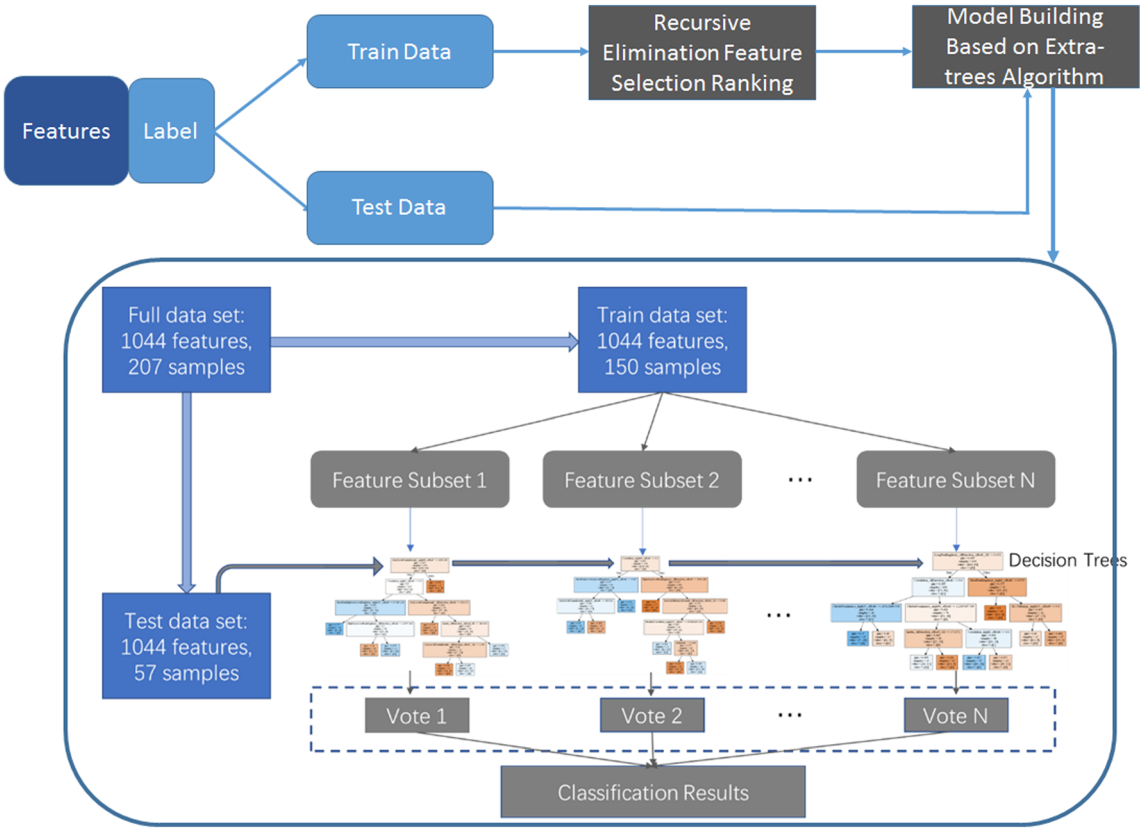

Fig. 4 Flowchart of feature selection and model building

Results

Baseline characteristics

Baseline characteristics were not significantly different between the training and validation cohorts (Table 1). In the training and validation cohorts, there were 32 (21.3%) and 13 (22.8%) patients with high immunoscore, respectively. In the training cohort, platelet count was significantly higher in the group of low immunoscore than that in the group of high immunoscore ($p = 0.041$). In the validation cohort, the proportion of alpha-fetoprotein (AFP) > 200 μg/L was significantly higher in the group of high immunoscore than that in the group of low immunoscore ($p = 0.019$). Other results were seen in Supplemental Table 3.

Inter-observer and intra-observer reproducibility of feature extraction

The inter-observer ICC of the three radiologists was ≥ 0.8, 0.5–0.79, and < 0.5 for 82%, 10%, and 8% of the features, respectively. And the intra-observer ICC of the same radiologist was ≥ 0.8, 0.5–0.79, and < 0.5 for 85%, 14%, and 1% of the features, respectively.

Radiomics feature extraction and selection

We used unsupervised clustering analysis to divide 1044 features in the training cohort into six groups as follows: (I) histogram parameters, (II) texture parameters, (III) form factor parameters, (IV) GLCM, (V) GLRLM, and (VI) GLSZM. The heatmap was used to show the correlation coefficient matrix among features (Supplemental Fig. 1). To reduce dependency and redundancy, we used recursive feature elimination to reduce dimensions of these 1044 features and 863 features were selected. Then, extremely randomized trees were applied and 70 features were selected to establish the radiomics model. These 70 features and their importance were shown in Supplemental Table 4.

Development and validation of radiomics model

The AUC, accuracy, sensitivity, and specificity of our intratumoral radiomics model in predicting immunoscore were 0.823 (95% confidence interval (CI), 0.747–0.899), 0.707, 81.3%, and 67.8% in the training cohort and 0.640 (95% CI, 0.477–0.803), 0.597, 53.9%, and 61.4% in the validation cohort (Table 2). The AUC, accuracy, sensitivity, and specificity of our combined radiomics model were 0.904 (95% CI 0.855–0.953), 0.787,

Table 1 Baseline clinical and pathological characteristics of the training and validation cohorts

	Total ($n = 207$)	Training ($n = 150$)	Validation ($n = 57$)	p values
Age, years, mean ± SD	54.5 ± 11.5	55.0 ± 11.3	53.2 ± 12.0	0.307
Gender				0.901
Male	188 (90.8)	136 (90.7)	52 (91.2)	
Female	19 (9.2)	14 (9.3)	5 (8.8)	
HBsAg				0.274
Negative	35 (16.9)	28 (18.7)	7 (12.3)	
Positive	172 (83.1)	122 (81.3)	50 (87.7)	
Child-Pugh class				0.097
A	200 (96.6)	143 (95.3)	57 (100.0)	
B	7 (3.4)	7 (4.7)	0 (0.0)	
BCLC				0.651
0	16 (7.7)	10 (6.7)	6 (10.5)	
A	108 (52.2)	77 (51.3)	31 (54.4)	
B	33 (15.9)	24 (16.0)	9 (15.8)	
C	50 (24.2)	39 (26.0)	11 (19.3)	
ALT, U/L	36.0 (22.0, 51.0)	37.0 (22.0, 52.0)	34.0 (22.0, 47.0)	0.306
AST, U/L	35.0 (26.0, 58.0)	35.5 (25.0, 56.0)	33.0 (26.0, 59.0)	0.829
GGT, U/L	68.0 (35.0, 135.0)	79.0 (40.0, 142.0)	58.0 (34.0, 107.0)	0.102
Platelet count, 10^9/L	161.0 (121.0, 219.0)	160.0 (118.0, 210.0)	176.0 (128.0, 219.0)	0.709
AFP, μg/L	46.7 (6.1, 856.3)	51.4 (6.8, 789.4)	35.6 (5.4, 856.3)	0.508
AFP group				0.534
≤ 200 μg/L	131 (63.3)	93 (62.0)	38 (66.7)	
> 200 μg/L	76 (36.7)	57 (38.0)	19 (33.3)	
Tumor size	4.5 (2.8, 6.9)	4.6 (2.9, 6.6)	4.3 (2.7, 8.5)	0.674
Tumor size group				0.967
< 5 cm	123 (59.4)	89 (59.3)	34 (59.6)	
≥ 5 cm	84 (40.6)	61 (40.7)	23 (40.4)	
Tumor number	1.0 (1.0, 2.0)	1.0 (1.0, 2.0)	1.0 (1.0, 2.0)	0.221
Tumor number group				0.403
1	136 (65.7)	96 (64.0)	40 (70.2)	
≥ 2	71 (34.3)	54 (36.0)	17 (29.8)	
Macrovascular invasion				0.497
Absent	192 (92.8)	54 (94.7)	138 (92.0)	
Present	15 (7.2)	3 (5.3)	12 (8.0)	
MVI				0.118
Absent	109 (52.7)	25 (43.9)	84 (56.0)	
Present	98 (47.3)	32 (56.1)	66 (44.0)	
Edmondson grade				0.616
1	8 (3.9)	3 (5.3)	5 (3.3)	
2	185 (89.4)	49 (86.0)	136 (90.7)	
3	14 (6.8)	5 (8.8)	9 (6.0)	
Capsule				0.211
Absent	6 (2.9)	3 (5.3)	3 (2.0)	
Present	201 (97.1)	54 (94.7)	147 (98.0)	
Capsule invasion				0.780
Absent	36 (17.4)	9 (15.8)	27 (18.0)	
Present	165 (79.7)	45 (78.9)	120 (80.0)	
Satellite lesion				0.057
Absent	135 (65.2)	43 (75.4)	92 (61.3)	
Present	72 (34.8)	14 (24.6)	58 (38.7)	

Continuous variables are presented as median (inter-quartile range, IQR) unless noted otherwise. Categorical variables are presented as n (%). Independent sample t test was used for continuous variables with normal distribution and Kruskal-Wallis (KW) non-parametric rank sum test was used for continuous variables with abnormal distribution. Skewness and kurtosis test was used to test the normality of continuous variables. Chi-square test or Fisher exact test was used for the comparisons of categorical variables

SD standard deviation, *HBsAg* hepatitis B surface antigen, *BCLC* Barcelona Clinic Liver Cancer, *ALT* alanine aminotransferase, *AST* aspartate aminotransferase, *GGT* gamma-glutamyltransferase, *AFP* alpha-fetoprotein, *MVI* microvascular invasion

93.8%, and 74.6% in the training cohort and 0.899 (95% CI 0.804–0.993), 0.772, 92.3%, and 72.7% in the validation cohort (Table 2, Supplemental Fig. 2a, 2b), which suggested that the combined radiomics model yield an improved performance in immunoscore than intratumoral radiomics model alone ($p = 0.036$ and $p = 0.005$; adjusted $p = 0.072$ and $p = 0.010$ in the training and validation cohorts respectively).

Development and validation of radiomics-based clinical model

By integrating the combined radiomics features with all the significant clinical variables (AFP, GGT, AST), we derived a combined radiomics-based clinical model and found that it yielded an AUC of 0.926 (95% CI 0.884–0.967) in the training cohort (Table 2, Supplemental Fig. 2c). This model presented to be more accurate in predicting immunoscore than the combined radiomics model alone, although the difference was not statistically significant ($p = 0.128$, adjusted $p = 0.256$). In the validation cohort, the combined radiomics-based clinical model (AUC, 0.934 (95% CI 0.865–1.000)) also showed an improved predicting performance in immunoscore over the combined radiomics model alone (Supplemental Fig. 2d), despite the non-significant statistical significance ($p = 0.112$, adjusted $p = 0.224$). The calibration curves showed good agreement between predicted and actual events (Supplemental Fig. 3).

The decision curves analysis showed that the combined radiomics-based clinical model had the highest overall net benefit compared with the combined radiomics model and intratumoral radiomics model at different threshold probabilities between 25% and 74% (Fig. 5). The combined radiomics-based clinical model achieved similar overall net benefit to the combined radiomics model alone at a range of thresholds between 0 and 25%.

Association of radiomics features with clinical data

Heatmap results showed that HaralickCorrelation_angle0_offset9, HaralickCorrelation_angle45_offset2 and HaralickCorrelation_AllDirection_offset7 were associated with most of the listed clinical variables (all $p < 0.05$) except satellite lesions and Edmondson grade. Other detailed results were in Supplemental Fig. 4.

Discussion

In this study, we established and validated a model for pretreatment prediction of immunoscore by combining clinical factors and radiomics features extracted from the intratumoral and peritumoral regions of Gd-EOB-DTPA-enhanced MRI. Results showed that the combined radiomics approach yielded an incremental accuracy in predicting immunoscore over the intratumoral only radiomics model (AUC 0.899 vs. 0.640) while the combined radiomics-based clinical model yielded the highest performance with an AUC of 0.934 despite of no significant difference with the combined radiomics model.

To the best of our knowledge, this is the first study establishing a MRI radiomics model for immunoscore prediction of HCC so far. The main reasons for high performance of our combined radiomics model in both the training and validation cohorts might be as followed. First, many studies have supported the radiomics hypothesis that proteogenomic and phenotypic information of the tumor can be inferred from radiological images [27, 39]. The radiomics analysis of non-visible imaging traits is able to quantify the tumor and its microenvironment in detail [40]. Secondly, radiomics requires accurate discrimination of lesion boundaries, and the use of Gd-EOB-DTPA-enhanced MRI images in our study makes stable segmentation possible [29, 30]. Thirdly, because we delineated the intratumoral and peritumoral regions in all the image slices, a three-dimensional (3D) tumor volume was extracted, making this 3D analysis of radiomics more

Table 2 Comparison of the predictive performance of the three models in immunoscore

Models	AUC (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	PPV (%)	NPV (%)
Intratumoral radiomics model						
Training ($n = 150$)	82.28 (74.70, 89.86)	81.25 (26/32)	67.80 (80/118)	70.67 (106/150)	40.63 (26/64)	93.02 (80/86)
Validation ($n = 57$)	63.99 (47.66, 80.31)	53.85 (7/13)	61.36 (27/44)	59.65 (34/57)	29.17 (7/24)	81.82 (27/33)
Total ($n = 207$)	77.13 (69.87, 84.39)	73.33 (33/45)	66.05 (107/162)	67.63 (140/207)	37.50 (33/88)	89.92 (107/119)
Combined radiomics model						
Training ($n = 150$)	90.41 (85.52, 95.30)	93.75 (30/32)	74.58 (88/118)	78.67 (118/150)	50.00 (30/60)	97.78 (88/90)
Validation ($n = 57$)	89.86 (80.40, 99.32)	92.31 (12/13)	72.73 (32/44)	77.19 (44/57)	50.00 (12/24)	96.97 (32/33)
Total ($n = 207$)	90.16 (85.77, 94.54)	93.33 (42/45)	74.07 (120/162)	78.26 (162/207)	50.00 (42/84)	97.56 (120/123)
Radiomics nomogram						
Training ($n = 150$)	92.58 (88.43, 96.73)	90.63 (29/32)	84.75 (100/118)	86.00 (129/150)	61.70 (29/47)	97.09 (100/103)
Validation ($n = 57$)	93.36 (86.51, 100.00)	84.62 (11/13)	84.09 (37/44)	84.21 (48/57)	61.11 (11/18)	94.87 (37/39)
Total ($n = 207$)	92.73 (89.20, 96.26)	88.89 (40/45)	84.57 (137/162)	85.51 (177/207)	61.51 (40/65)	96.48 (137/142)

AUC the area under the operating characteristic curve, *PPV* positive predictive value, *NPV* negative predictive value

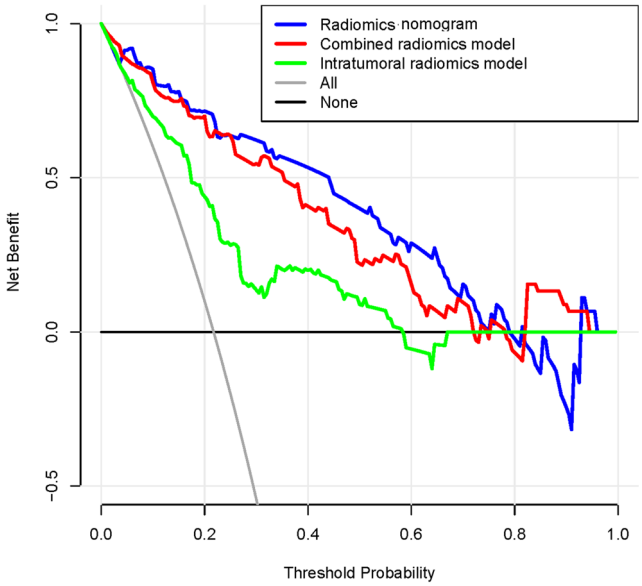

Fig. 5 Decision curve analysis of the radiomics models and radiomics nomogram. The *y*-axis measures the net benefit. The net benefit is determined by calculating the difference between the expected benefit and the expected harm associated with each proposed model [Net benefit = true positive rate–(false positive rate × weighting factor), weighting factor = threshold probability/(1-threshold probability)]. The highest curve (representing the radiomics-based clinical model) at a range threshold probability of 25–74% is the optimal decision-making strategy to maximize the net benefit compared with other models

stable and representative than a 2D analysis [27]. Fourth, peritumoral radiomics improved the model performance by adding information on the tumor boundaries.

In addition, by integrating clinical variables with radiomics features, the derived combined radiomics-based clinical model had a slightly higher, though non-significantly different, predicting performance in immunoscore when compared to the combined radiomics model. This indicated that radiomics features were the stronger component of this derived model while the clinical data had limited value for improving the model.

Indeed, this study is also the first attempt to apply a novel combined intratumoral and peritumoral radiomics approach for predicting TIL density in both the CT and IM. TILs in the peritumoral region are the essential part of the evaluation of immunoscore. That was why we dilated the annotated intratumoral region at a radius of 1 cm to extract both the intratumoral and peritumoral radiomics features. The superiority of the combined radiomics approach over intratumoral radiomics model in this study confirmed that peritumoral radiomics might provide unique and valuable features, suggestive of peritumoral TIL infiltration. This is also consistent with the results observed in breast cancer, suggesting that a combined radiomics model may be better in characterizing immune infiltrates by offering a more comprehensive radiomics profile [25].

Previous studies have found that radiomics features were closely related to tumor microscopic structure and biological behavior [41–44]. Our study discovered 20 radiomics quantitative features most relevant for immunoscore of HCC, not previously reported. Among them, 19 were texture features including histogram feature, GLCM, and GLRLM, and 1 was morphological feature. Among texture features, GLCM and GLRLM reflect signal mixing degree of the lesions by means of relative relationship between distribution and site of the gray level, and thus important markers of intra-tumor homogeneity. The values were higher in tumors with high immunoscore, which indicated that more cells or tissues with similar texture features were aggregated in one direction and the intratumoral heterogeneity was relatively weak. In our study, we observed that TILs were generally cluster-distributed either in the intratumoral or peritumoral region. Clustered TILs manifested as aggregations of similar gray level in radiomics. Highly immune infiltrated tumors were more homogenous, explaining the high value of GLCM. Besides, "surfacevolumeratio," the only morphological feature representing the ratio of the lesion surface area to the volume, was the unique radiomics feature identified in our study, because this feature could reflect three-dimensional radiomics feature. Unlike previous studies in which VOI was only delineated in a single image slice, our study delineated VOIs in all the layers to reconstruct the tumors' 3D morphology.

There are several limitations to this study. First, the sample size is still limited compared with the relatively large number of variables. As such, extremely randomized tree method was used because it uses the whole training sample rather than a bootstrap replica to build a tree in order to minimize bias, and within each tree, it includes a random subset of features and splits nodes by choosing cut-points at random as well. Second, our validation cohort was from the same center as the training cohort, which restricted us to assess the generalizability of our findings to other centers. Third, only TILs were considered in this study while many other components of TME were not.

In conclusion, our study established and validated a Gd-EOB-DTPA-enhanced MRI-based combined radiomics nomogram, for predicting immunoscore in HCC patients. It may be useful in pretreatment individual prediction of immunoscore to guide accurate prognosis prediction and precision immunotherapy for HCC patients.

Funding This study has received funding by grants from the Guangzhou Science and Technology Program key projects (No. 201803010057) and the National Natural Science Foundation of China (No. 81771908, 81571750). This work was supported by Ministry of Science and Technology of China under Grant No. 2017YFA0205200, National Natural Science Foundation of China under Grant No. 81227901, 81527805, Chinese Academy of Sciences under Grant No. GJJSTD20170004 and QYZDJ-SSW-JSC005, Beijing Municipal Science & Technology Commission under Grant No. Z161100002616022, Z171100000117023, the Key International Cooperation Projects of the Chinese Academy of Sciences under Grant

No. 173211KYSB20160053. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Ming Kuang.

Conflict of interest The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Statistics and biometry Two of the authors (Fei Liu, Bin Li) have significant statistical expertise.

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

Ethical approval Institutional Review Board approval was obtained.

Methodology
- retrospective
- diagnostic or prognostic study
- performed at one institution

References

1. Tabrizian P, Jibara G, Shrager B, Schwartz M, Roayaie S (2015) Recurrence of hepatocellular cancer after resection: patterns, treatments, and prognosis. Ann Surg 261:947–955
2. El-Khoueiry AB, Sangro B, Yau T et al (2017) Nivolumab in patients with advanced hepatocellular carcinoma (CheckMate 040): an open-label, non-comparative, phase 1/2 dose escalation and expansion trial. Lancet 389:2492–2502
3. Sangro B, Gomez-Martin C, de la Mata M et al (2013) A clinical trial of CTLA-4 blockade with tremelimumab in patients with hepatocellular carcinoma and chronic hepatitis C. J Hepatol 59:81–88
4. Palucka AK, Coussens LM (2016) The basis of oncoimmunology. Cell 164:1233–1247
5. Kim YJ (2015) Subverting the adaptive immune resistance mechanism to improve clinical responses to immune checkpoint blockade therapy. Oncoimmunology 3:e954868
6. Taube JM (2014) Unleashing the immune system: PD-1 and PD-Ls in the pre-treatment tumor microenvironment and correlation with response to PD-1/PD-L1 blockade. Oncoimmunology 11:e963413
7. Zhou G, Sprengers D, Boor PPC et al (2017) Antibodies against immune checkpoint molecules restore functions of tumor-infiltrating T cells in hepatocellular carcinomas. Gastroenterology 153:1107–1119
8. Galon J, Mlecnik B, Bindea G et al (2014) Towards the introduction of the 'immunoscore' in the classification of malignant tumours. J Pathol 232:199–209
9. Galon J, Costes A, Sanchez-Cabo F et al (2006) Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. Science 313:1960–1964
10. Mahmoud SM, Paish EC, Powe DG et al (2011) Tumor-infiltrating CD8+ lymphocytes predict clinical outcome in breast cancer. J Clin Oncol 29:1949–1955
11. Brunner SM, Rubner C, Kesselring R et al (2015) Tumor-infiltrating, interleukin-33–producing effector-memory CD8+ T cells in resected hepatocellular carcinoma prolong patient survival. Hepatology 61:1957–1967
12. Jiang Y, Zhang Q, Hu Y et al (2018) ImmunoScore signature: a prognostic and predictive tool in gastric cancer. Ann Surg 267:504–513
13. Donnem T, Hald SM, Paulsen EE et al (2015) Stromal CD8+ T-cell density-a promising supplement to TNM staging in non-small cell lung cancer. Clin Cancer Res 21:2635–2643
14. Gabrielson A, Wu Y, Wang H et al (2016) Intratumoral CD3 and CD8 T-cell densities associated with relapse free survival in HCC. Cancer Immunol Res 4:419–430
15. Sun C, Xu J, Song J et al (2015) The predictive value of centre tumour CD8+ T cells in patients with hepatocellular carcinoma: comparison with immunoscore. Oncotarget 6:35602–35615
16. Yao Q, Bao X, Xue R et al (2017) Prognostic value of immunoscore to identify mortality outcomes in adults with HBV-related primary hepatocellular carcinoma. Medicine (Baltimore) 96(17):e67356
17. Garnelo M, Tan A, Her Z et al (2017) Interaction between tumor-infiltrating B cells and T cells controls the progression of hepatocellular carcinoma. Gut 66:342–351
18. Shields AF, Jacobs P, Sznol M et al (2018) Immune modulation therapy and imaging: workshop report. J Nucl Med 59:410–417
19. Ku YJ, Kim HH, Cha JH et al (2016) Correlation between MRI and the level of tumor-infiltrating lymphocytes in patients with triple negative breast cancer. AJR Am J Roentgenol 207:1146–1151
20. Ku YJ, Kim HH, Cha JH et al (2018) Predicting the level of tumor-infiltrating lymphocytes in patients with triple negative breast cancer: usefulness of breast MRI computer-aided detection and diagnosis. J Magn Reson Imaging 47:760–766
21. Kumar V, Gu Y, Basu S et al (2012) Radiomics: the process and the challenges. Magn Reson Imaging 30:1234–1248
22. Sun R, Limkin EJ, Dercle L et al (2017) Computational medical imaging (radiomics) and potential for immuno-oncology. Cancer Radiother 21:648–654
23. Savadjiev P, Chong J, Dohan A et al (2018) Demystification of AI-driven medical image interpretation: past, present and future. Eur Radiol. https://doi.org/10.1007/s00330-018-5674-x
24. Li Y, Liu X, Xu K et al (2018) MRI features can predict EGFR expression in lower grade gliomas: a voxel-based radiomic analysis. Eur Radiol 28:356–362
25. Esteva A, Kuprel B, Novoa RA et al (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 542:115–118
26. Braman NM, Etesami M, Prasanna P et al (2017) Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. Breast Cancer Res 19:57
27. Rutman AM, Kuo MD (2009) Radiogenomics: creating a link between molecular diagnostics and diagnostic imaging. Eur J Radiol 70:232–241
28. Wu M, Tan H, Gao F et al (2018) Predicting the grade of hepatocellular carcinoma based on non-contrast-enhanced MRI radiomics signature. Eur Radiol. https://doi.org/10.1007/s00330-018-5787-2
29. Hamm B, Staks T, Mühler A et al (1995) Phase I clinical evaluation of Gd-EOB-DTPA as a hepatobiliary MR contrast agent: safety, pharmacokinetics, and MR imaging. Radiology 195:785–792
30. Choi JW, Lee JM, Kim SJ et al (2013) Hepatocellular carcinoma: imaging patterns on gadoxetic acid-enhanced MR images and their value as an imaging biomarker. Radiology 267:776–786

31. Zhang B, Tian J, Dong D et al (2017) Radiomics features of multiparametric MRI as novel prognostic factors in advanced nasopharyngeal carcinoma. Clin Cancer Res 23:4259–4269
32. Bakr S, Echegaray S, Shah R et al (2017) Noninvasive radiomics signature based on quantitative analysis of computed tomography images as a surrogate for microvascular invasion in hepatocellular carcinoma: a pilot study. J Med Imaging (Bellingham) 4:041303
33. Huang YQ, Liang CH, He L et al (2016) Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. J Clin Oncol 34:2157–2164
34. Li H, Zhu Y, Burnside ES et al (2016) MR imaging radiomics signatures for predicting the risk of breast cancer recurrence as given by research versions of MammaPrint, Oncotype DX, and PAM50 gene assays. Radiology 281:382–391
35. Galon J, Pagès F, Marincola FM et al (2012) Cancer classification using the immunoscore: a worldwide task force. J Transl Med 10:205
36. Leijenaar RT, Carvalho S, Velazquez ER et al (2013) Stability of FDG-PET radiomics features: an integrated analysis of test-retest and inter-observer variability. Acta Oncol 52:1391–1137
37. Geurts P, Ernst D, Wehenkel L (2006) Extremely randomized trees. Mach Learn 63:3–42
38. Marée R, Geurts P, Wehenkel L (2007) Random subwindows and extremely randomized trees for image classification in cell biology. BMC Cell Biol 8(Suppl 1):S2
39. Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: images are more than pictures, they are data. Radiology 278:563–577
40. Bourgier C, Colinge J, Aillères N et al (2015) Radiomics: definition and clinical development. Cancer Radiother 19:532–537
41. Grossmann P, Stringfield O, El-Hachem N et al (2017) Defining the biological basis of radiomic phenotypes in lung cancer. Elife 6. https://doi.org/10.7554/eLife.23421
42. Fox MJ, Gibbs P, Pickles MD (2016) Minkowski functionals: an MRI texture analysis tool for determination of the aggressiveness of breast cancer. J Magn Reson Imaging 43:903–910
43. Ganeshan B, Goh V, Mandeville HC, Ng QS, Hoskin PJ, Miles KA (2013) Non-small cell lung cancer: histopathologic correlates for texture parameters at CT. Radiology 266:326–336
44. Segal E, Sirlin CB, Ooi C et al (2007) Decoding global gene expression programs in liver cancer by noninvasive imaging. Nat Biotechnol 25:675–680