



Radiomics and machine learning based on preoperative MRI for predicting extrahepatic metastasis in hepatocellular carcinoma patients treated with transarterial chemoembolization

Gang Peng, Xiaojing Cao, Xiaoyu Huang, Xiang Zhou*

Department of Interventional Therapy, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

ARTICLE INFO

Keywords:

Radiomics
Machine learning
Carcinoma
Hepatocellular
Chemoembolization
Therapeutic
Metastasis
Magnetic resonance imaging

ABSTRACT

Purpose: To develop and validate a radiomics machine learning (Rad-ML) model based on preoperative MRI to predict extrahepatic metastasis (EHM) in hepatocellular carcinoma (HCC) patients receiving transarterial chemoembolization (TACE) treatment.

Methods: A total of 355 HCC patients who received multiple TACE procedures were split at random into a training set and a test set at a 7:3 ratio. Radiomic features were calculated from tumor and peritumor in arterial phase and portal venous phase, and were identified using intraclass correlation coefficient, maximal relevance and minimum redundancy, and least absolute shrinkage and selection operator techniques. Cox regression analysis was employed to determine the clinical model. The best-performing algorithm among eight machine learning methods was used to construct the Rad-ML model. A nomogram combining clinical and Rad-ML parameters was used to develop a combined model. Model performance was evaluated using C-index, decision curve analysis, calibration plot, and survival analysis.

Results: In clinical model, elevated neutrophil to lymphocyte ratio and alpha-fetoprotein were associated with faster EHM. The XGBoost-based Rad-ML model demonstrated the best predictive performance for EHM. When compared to the clinical model, both the Rad-ML model and the combination model performed better (C-indexes of 0.61, 0.85, and 0.86 in the training set, and 0.62, 0.82, and 0.83 in the test set, respectively). However, the combined model's and the Rad-ML model's prediction performance did not differ significantly. The most influential feature was peritumoral waveletHLL_firstorder_Minimum in AP, which exhibited an inverse relationship with EHM risk.

Conclusions: Our study suggests that the preoperative MRI-based Rad-ML model is a valuable tool to predict EHM in HCC patients treated with TACE.

1. Introduction

Hepatocellular carcinoma (HCC), accounting for approximately 80% of all liver cancer cases globally. It is the third highest cause of cancer-related mortality worldwide, with over 800,000 deaths annually from HCC [1,2]. Transarterial chemoembolization (TACE) is a minimally invasive procedure that delivers chemotherapy directly to the liver tumor through the hepatic artery while blocking blood flow to the tumor, thereby depriving it of oxygen and nutrients necessary for growth. TACE is commonly used as a palliative therapeutic regimen for

HCC patients with intermediate-stage who are not candidates for curative treatments such as surgery, liver transplantation, radiofrequency or microwave ablation, while systemic therapy is the more recommended treatment for patients with advanced hepatocellular carcinoma [3–5]. However, despite improvements in treatment options, the prognosis remains poor due to high rates of recurrence and extrahepatic metastasis (EHM) [6,7]. Accurate prediction of EHM before treatment is crucial for determining the optimal treatment strategy and improving patient outcomes.

Radiomics is an emerging technique that aims to segment and extract

* Correspondence to: Department of Interventional Therapy, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, No.17, Panjiayuan Nanli, Chaoyang District, Beijing 100021, China.

E-mail address: zhou.xiang@yeah.net (X. Zhou).

<https://doi.org/10.1016/j.ejro.2024.100551>

Received 16 November 2023; Received in revised form 14 January 2024; Accepted 15 January 2024

2352-0477/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

valuable information from routine medical images beyond what can be visually detected [8]. By analyzing thousands of quantitative image features including shape, intensity distribution and texture, radiomics provides a comprehensive characterization of tumors that can assist in diagnosing, monitoring and predicting treatment outcomes [9,10]. Machine learning algorithms process these features to generate predictive models that contribute to the diagnosis, treatment, and prognosis assessment of HCC [11,12]. Prior investigations have demonstrated the practical value of radiomic features extracted from MR or CT images for predicting HCC recurrence and survival [7,13–15]. However, there has been little research on its application in predicting EHM in HCC patients treated with TACE.

The purpose of this study is to develop and validate radiomics coupled with machine learning model based on MRI for predicting EHM in patients with HCC receiving TACE treatment. The findings of this study could lead to more precise risk stratification and personalized treatment planning for HCC patients undergoing TACE.

2. Materials and methods

2.1. Study population

The National Cancer Center/Cancer Hospital's Institutional Review Board provided authorization for this observational research, which was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki. Because of the retrospective nature of this study, informed consent was exempted. The study population included 2303 consecutive patients newly diagnosed with HCC who received their initial TACE treatment between January 2013 and December 2020, as provided by the clinical data warehouse. Inclusion criteria included: (1) HCC diagnosis confirmed by histology or radiography consistent with the American Society for the Study of Liver Disorders guidelines [16]; (2) Eastern Cooperative Oncology Group performance status ≤ 2 ; and (3) Child-Pugh classification A or B. Patients were excluded if they received their initial TACE treatment at another hospital, underwent other treatments prior to TACE progression, had significant comorbidities involving other serious diseases or cancers, experienced severe adverse events related to TACE, did not undergo dynamic enhanced MRI within one month before the initial TACE treatment, were lost to follow-up or received only a single TACE treatment, or were classified as BCLC stage C. A total of 355 patients were finally included and randomized into training and test cohort in a ratio of 7:3 (Supplementary Fig. 1).

2.2. Conventional TACE (cTACE) procedure

A 5-Fr arterial catheter (Radiofocus; Terumo) was punctured into the femoral artery using the Seldinger technique under the guidance of digital subtraction angiography (Allura Xper FD20, Philips). A 5-Fr hepatic catheter (Radiofocus; Terumo) was then inserted into the hepatic artery through the abdominal aorta. The feeding arteries were incrementally injected with an emulsion composed of a chemotherapeutic agent (anthracycline or platinum, 10–50 mg/m²) combined with 5–20 ML of iodized oil (Lipiodol; Guerbet), followed by 350–560 μ m gelatin sponge particles (Gelfoam; Alicon). The endpoint of TACE treatment was achieved when tumor staining diminished and lipiodol filling appeared in the minute peritumoral portal vein branches. The dosages of chemotherapy agents and iodized oil were determined by the tumor's characteristics and the patient's liver function. Additional TACE procedures were performed as needed upon the detection of intrahepatic recurrence or metastasis. Treatment was discontinued in cases of TACE refractoriness [17], severe complications, or withdrawal of consent.

2.3. MRI techniques

The magnetic resonance imaging (MRI) protocol encompassed fat-

saturated T1- and T2-weighted imaging, and diffusion-weighted imaging. A dynamic contrast enhanced MRI sequence was implemented utilizing gadopentetate dimeglumine (Magnevist; Bayer Healthcare) at a rate of 2.0 ML/s and a dosage of 0.2 ML/kg of body weight. Following the administration of the contrast agent, arterial phase (AP), portal venous phase (PVP), and delayed phase (DP) scans were conducted at 35 s, 60 s, and 180 s, respectively.

2.4. MRI segmentation and feature extraction

The workflow of this study is shown in Fig. 1. Two experienced radiologists (10-year and 11-year of experience) independently performed MRI interpretations and semiautomatic segmentations at AP and PVP utilizing 3D Slicer software (version 5.1.0; <https://www.slicer.org/>). They were blinded to clinical baseline data and EHM results. Intratumoral and peritumoral regions of interests (ROIs) were segmented separately. The peritumoral ROI encompassed a 5-mm area with an inward shrinkage and outward expansion of the tumor margin by 2.5 mm, respectively. The resampled voxel size was 1 * 1 * 1 mm; LoG kernel sizes ranged from 1 to 5; voxel intensities utilized a bin-width of 25; and wavelet features were employed. Each three-dimensional ROI consisted of 1316 radiomic features, including 18 first-order features, 14 shape features, 75 textural features, 465 LoG features, and 744 wavelet features. A total of 5264 radiomic features were extracted from the four groups of ROIs (tumoral features in AP, peritumoral features in AP, tumoral features in PVP, and peritumoral features in PVP) using the Pyradiomics package version 3.10 (<https://pyradiomics.readthedocs.io/>).

2.5. Feature selection

Firstly, all features were normalized so that their mean was zero and their variance was one. Secondly, features with an intraclass correlation coefficient for segmentation by two radiologists greater than 0.8 were included. Thirdly, a feature selection algorithm based on the maximal relevance and minimum redundancy and least absolute shrinkage and selection operator (LASSO) regression (Supplementary Fig. 2) was employed using ten-fold cross-validation to choose the optimal features.

2.6. Model construction

We utilized eight machine learning algorithms, including logistic regression (LR), Naive Bayes (NB), support vector machine (SVM), K-Nearest Neighbor (KNN), random forest (RF), Light Gradient Boosting Machine (LightGBM), Adaptive Boosting (AdaBoost), and eXtreme Gradient Boosting (XGBoost), to compare the area under the curve in training and test sets and select the optimal algorithm as the radiomics machine learning (Rad-ML) model based on the remaining features after feature selection. Uni- and multivariable backward stepwise Cox regression analyses were conducted to identify independent risk factors for EHM, which were designated as clinical model parameters. A combined model was constructed using a nomogram that integrated the clinical model and the radiomic score based on the best Rad-ML model.

2.7. Model evaluation and explanation

The time-dependent receiver operating characteristic (ROC) curve and the area under the curve (AUC) were employed to assess the performance at different time points. The concordance index (C-index) was utilized to evaluate the overall discrimination. A calibration curve was plotted to estimate the consistency between the observed and predicted probabilities. Decision curve analysis and survival curves were used to determine the clinical effectiveness. SHapley Additive exPlanations (SHAP) method was utilized to elucidate the Rad-ML model's output. The SHAP plot provides an insight into the importance and direction of the feature's contribution to the result.

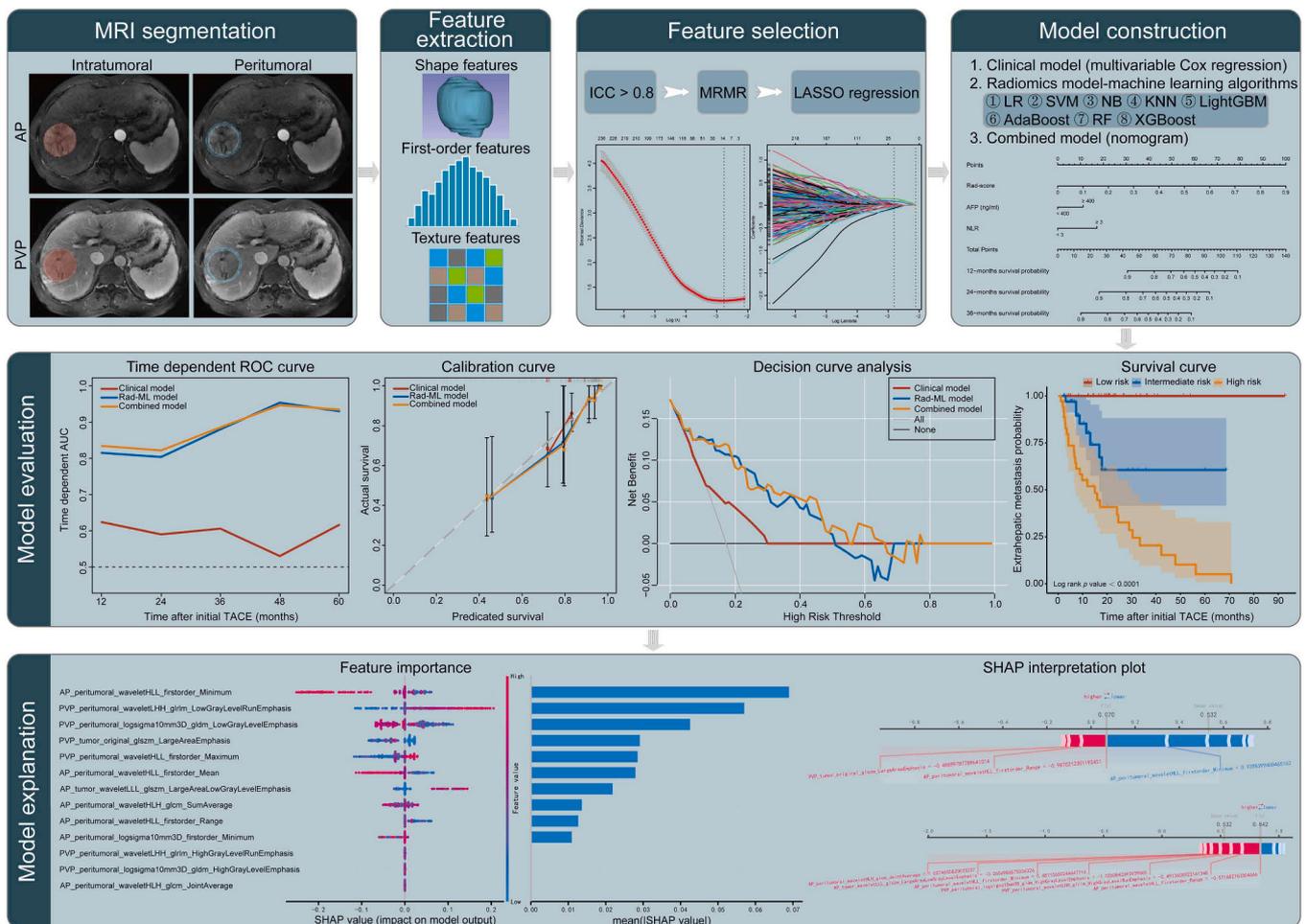


Fig. 1. The workflow in this study. **Abbreviation:** AP, arterial phase; PVP, portal venous phase; ICC, intraclass correlation coefficient; MRMR, maximum relevance minimum redundancy; LASSO, least absolute shrinkage and selection operator; LR, logistic regression; SVM, support vector machine; NB, Naive Bayes; KNN, K-Nearest Neighbor; LightGBM, Light Gradient Boosting Machine; RF, random forest; AdaBoost, Adaptive Boosting; XGBoost, eXtreme Gradient Boosting; ROC, receiver operator characteristic; Rad-ML, radiomics machine learning; TACE, transarterial chemoembolization; SHAP, SHapley Additive exPlanations.

2.8. Outcomes and follow-up

Extrahepatic metastases are defined as tumor lesions that were biopsy proved or newly discovered on imaging. The outcome of this study was the time to EHM. The time from the initial TACE treatment to the occurrence of EHM was used to determine the time of EHM. Patients who did not develop EHM at the last follow-up were censored. Dynamic enhanced MRI scans, liver function, and serum alpha fetoprotein (AFP) tumor markers were assessed every 1–2 months following each TACE session. The study had a median follow-up time of 33 months (95% CI: 26, 40). The last follow-up date was June 30, 2023.

2.9. Statistical analyses

Statistical analyses were performed by R software version 4.1.1 (<http://www.r-project.org/>). A probability value < 0.05 was considered significant for hypothesis tests. Chi-square tests were used to detect statistical differences between the training and test cohort. The inter-observer reliability between the two radiologists was determined by the intraclass correlation coefficients. The DeLong tests were used to compare the C-index of different models. The grouping of survival curves was based on the tripartite division of prediction probabilities generated by the best Rad-ML model, and the differences between groups were compared by use of log-rank test.

3. Results

3.1. Baseline characteristics of participants

The median age was 61 years (interquartile range [IQR], 53–67 years), with 306 (86.20%) males and 49 (13.80%) females. The majority of patients had BCLC stage B, Child-Pugh grade A, HBV infection, serum AFP < 400 ng/mL, albumin-bilirubin grade 1, total bilirubin < 20 umol/L, gamma glutamyl transpeptidase \geq 50 U/L, alanine aminotransferase < 40 U/L, aspartate aminotransferase < 40 U/L, albumin \geq 35 g/L, platelets \geq 100×10^9 /L, prothrombin time < 13 s, international normalized ratio \geq 1.0, neutrophil to lymphocyte ratio (NLR) < 3, and platelet to lymphocyte ratio (PLR) < 125. There were no significant differences in all baseline characteristics between the training and test sets. (Table 1).

3.2. Time to extrahepatic metastasis

During follow-up, 118 (33.24%) of the 355 patients experienced EHM, with a median EHM time of 31 months (95% CI: 23–39). Additionally, 83 patients (33.47%) in the training cohort and 35 patients (32.71%) in the test cohort experienced EHM during the follow-up period. The median EHM time was 31 months (95% CI: 24–38) in the training cohort and 31 months (95% CI: 21–41) in the test cohort ($p = 0.91$).

Table 1
Baseline participant characteristics in the training cohort and test cohort.

Variables	Total (n = 355)	Training cohort (n = 248)	Test cohort (n = 107)	p value
Age < 60 / ≥ 60 (years)	157 (44.23%)/ 198 (55.77%)	110 (44.35%)/ 138 (55.65%)	47 (43.93%)/ 60 (56.07%)	0.940
Gender (Female/ Male)	49 (13.80%)/ 306 (86.20%)	35 (14.11%)/ 213 (85.89%)	14 (13.08%)/ 93 (86.92%)	0.797
HBV infection (No/Yes)	94 (26.48%)/ 261 (73.52%)	68 (27.42%)/ 180 (72.58%)	26 (24.30%)/ 81 (75.70%)	0.541
BCLC stage (A/ B)	73 (20.56%)/ 282 (79.44%)	57 (22.98%)/ 191 (77.02%)	16 (14.95%)/ 91 (85.05%)	0.086
Child-Pugh grade (A/B)	323 (90.99%)/ 32 (9.01%)	222 (89.52%)/ 26 (10.48%)	101 (94.39%)/ 6 (5.61%)	0.141
Serum AFP < 400 / ≥ 400 (ng/mL)	249 (70.14%)/ 106 (29.86%)	171 (68.95%)/ 77 (31.05%)	78 (72.90%)/ 29 (27.10%)	0.456
ALBI grade (1/2 and 3) ^a	227 (63.94%)/ 128 (36.06%)	155 (62.50%)/ 93 (37.50%)	72 (67.29%)/ 35 (32.71%)	0.388
TBIL < 20 / ≥ 20 (umol/L)	256 (72.11%)/ 99 (27.89%)	179 (72.18%)/ 69 (27.82%)	77 (71.96%)/ 30 (28.04%)	0.967
AST < 40 / ≥ 40 (U/L)	187 (52.68%)/ 168 (47.32%)	130 (52.42%)/ 118 (47.58%)	57 (53.27%)/ 50 (46.73%)	0.883
ALT < 40 / ≥ 40 (U/L)	223 (62.82%)/ 132 (37.18%)	153 (61.69%)/ 95 (38.31%)	70 (65.42%)/ 37 (34.58%)	0.505
GGT < 50 / ≥ 50 (U/L)	100 (28.17%)/ 255 (71.83%)	72 (29.03%)/ 176 (70.97%)	28 (26.17%)/ 79 (73.83%)	0.582
ALB < 35 / ≥ 35 (g/L)	30 (8.45%)/ 325 (91.55%)	24 (9.68%)/ 224 (90.32%)	6 (5.61%)/ 101 (94.39%)	0.206
PLT < 100 / ≥ 100 (×10 ⁹ /L)	84 (23.66%)/ 271 (76.34%)	57 (22.98%)/ 191 (77.02%)	27 (25.23%)/ 80 (74.77%)	0.647
PT < 13 / ≥ 13 (s)	275 (77.46%)/ 80 (22.54%)	191 (77.02%)/ 57 (22.98%)	84 (78.50%)/ 23 (21.50%)	0.758
INR < 1.0 / ≥ 1.0	73 (20.56%)/ 282 (79.44%)	46 (18.55%)/ 202 (81.45%)	27 (25.23%)/ 80 (74.77%)	0.153
NLR < 3 / ≥ 3	277 (78.03%)/ 78 (21.97%)	190 (76.61%)/ 58 (23.39%)	87 (81.31%)/ 20 (18.69%)	0.327
PLR < 125 / ≥ 125	264 (74.37%)/ 91 (25.63%)	184 (74.19%)/ 64 (25.81%)	80 (74.77%)/ 27 (25.23%)	0.910

Note: Values are presented as number (percentage). ^aALBI score = $0.66 \times \log_{10}$ [total bilirubin (μmol/L)] - $0.0852 \times$ [albumin (g/L)]; ALBI: grade 1 ≤ -2.60; -2.60 < grade 2 ≤ -1.39; and grade 3 > -1.39.

Abbreviation: HBV, hepatitis B virus; BCLC, Barcelona Clinic Liver Cancer; AFP, alpha fetoprotein; ALBI, albumin-bilirubin; TBIL, total bilirubin; AST, aspartate aminotransferase; ALT, alanine aminotransferase; GGT, gamma glutamyl transpeptidase; ALB, albumin; PLT, platelet; PT, prothrombin time; INR, international normalized ratio; NLR, neutrophil to lymphocyte ratio; PLR, platelet to lymphocyte ratio.

3.3. Univariable and multivariable Cox regression analyses

Univariable Cox regression analysis showed that serum AFP ≥ 400 ng/mL, NLR ≥ 3, and PLR ≥ 125 were associated with faster EHM (hazard ratio [HR] and 95% CI: 1.60 [1.11–2.32], 1.84 [1.21–2.80], 1.70 [1.14–2.55], respectively; Table 2). In backward stepwise multivariable Cox regression analysis, serum AFP ≥ 400 ng/mL and NLR ≥ 3 were also identified as independent risk factors for EHM (HR and 95% CI: 1.54 [1.06–2.23], 1.78 [1.17–2.72], respectively, Table 2). Therefore, AFP and NLR were used as clinical model parameters for predicting EHM.

3.4. Performance of Rad-ML model

In the training cohort, the results of the 10-fold cross-validation showed that the XGBoost algorithm achieved the highest AUC of 0.869 (95% CI, 0.824–0.914), with an accuracy of 0.828, and F1 score 0.733 (Fig. 2A, Table 3). In the test cohort, the XGBoost algorithm also showed the highest AUC of 0.762 (95% CI, 0.664–0.859), with an accuracy of 0.734, and F1 score 0.624 (Fig. 2B, Table 3). Therefore, XGBoost algorithm was used as the Rad-ML model for predicting EHM.

Table 2
Preoperative clinical risk factors for extrahepatic metastasis in hepatocellular carcinoma patients receiving transarterial chemoembolization treatment.

Variables	Univariable analysis		Multivariable analysis	
	HR (95% CI)	p value	HR (95% CI)	p value
Age (≥60 years)	0.79 (0.55-1.13)	0.191		
Gender (Male)	0.97 (0.56-1.67)	0.903		
Etiology (HBV)	0.87 (0.58-1.31)	0.508		
BCLC stage (B)	1.19 (0.76-1.87)	0.448		
Child-Pugh grade (B)	1.12 (0.60-2.09)	0.713		
Serum AFP (≥ 400 ng/ML)	1.60 (1.11-2.32)	0.013	1.54 (1.06-2.23)	0.023
ALBI grade (2 and 3)	0.77 (0.51-1.16)	0.210	0.70 (0.46-1.06)	0.091
TBIL (≥ 20 umol/L)	0.92 (0.61-1.39)	0.698		
AST (≥ 40 U/L)	1.33 (0.92-1.91)	0.126	1.36 (0.93-1.99)	0.110
ALT (≥ 40 U/L)	1.27 (0.87-1.84)	0.214		
GGT (≥ 50 U/L)	1.27 (0.85-1.91)	0.248		
ALB (≥ 35 g/L)	1.55 (0.68-3.52)	0.297		
PLT (≥ 100 ×10 ⁹ /L)	1.33 (0.85-2.07)	0.207		
PT ≥ 13 (s)	1.02 (0.67-1.56)	0.933		
INR (≥ 1.0)	1.35 (0.83-2.17)	0.224		
NLR (≥ 3)	1.84 (1.21-2.80)	0.004	1.78 (1.17-2.72)	0.007
PLR (≥ 125)	1.70 (1.14-2.55)	0.010		

Abbreviation: HBV, hepatitis B virus; BCLC, Barcelona Clinic Liver Cancer; AFP, alpha fetoprotein; ALBI, albumin-bilirubin; TBIL, total bilirubin; AST, aspartate aminotransferase; ALT, alanine aminotransferase; GGT, gamma glutamyl transpeptidase; ALB, albumin; PLT, platelets; PT, prothrombin time; INR, international normalized ratio; NLR, neutrophil to lymphocyte ratio; PLR, platelet to lymphocyte ratio.

3.5. Establishment of combined model and evaluation of all models

A combined model, which incorporated clinical parameters and the Rad-score, was constructed to predict 12-month, 24-month, and 36-month EHM based on a nomogram (Fig. 3). The C-index values for the clinical model, Rad-model, and combined model were 0.61, 0.85, and 0.86 in the training set, respectively, and 0.62, 0.82, and 0.83 in the test set, respectively (Table 4). In both the training and test sets, the C-index values for the Rad-ML model and the combined model were superior to the clinical model ($p < 0.01$), but there was no significant difference of predictive performance between the Rad-ML model and the combined model ($p = 0.731$ in the training set and $p = 0.143$ in the test set, Table 4). The Rad-ML model and the combined model demonstrated comparable and superior performance compared to the clinical model in time-dependent ROC curves, decision curve analyses and calibration curves for predicting 1-year, 2-year, and 3-year EHM in training and test sets (Figs. 4, 5, and 6). Survival curves indicated satisfactory performance in both the training and test groups based on the trisection of the predicted probability values derived from the Rad-ML model (all log rank $p < 0.001$, Fig. 7A and B).

3.6. Rad-ML model explanation

Fig. 8A displays the 13 remaining features of the Rad-ML model

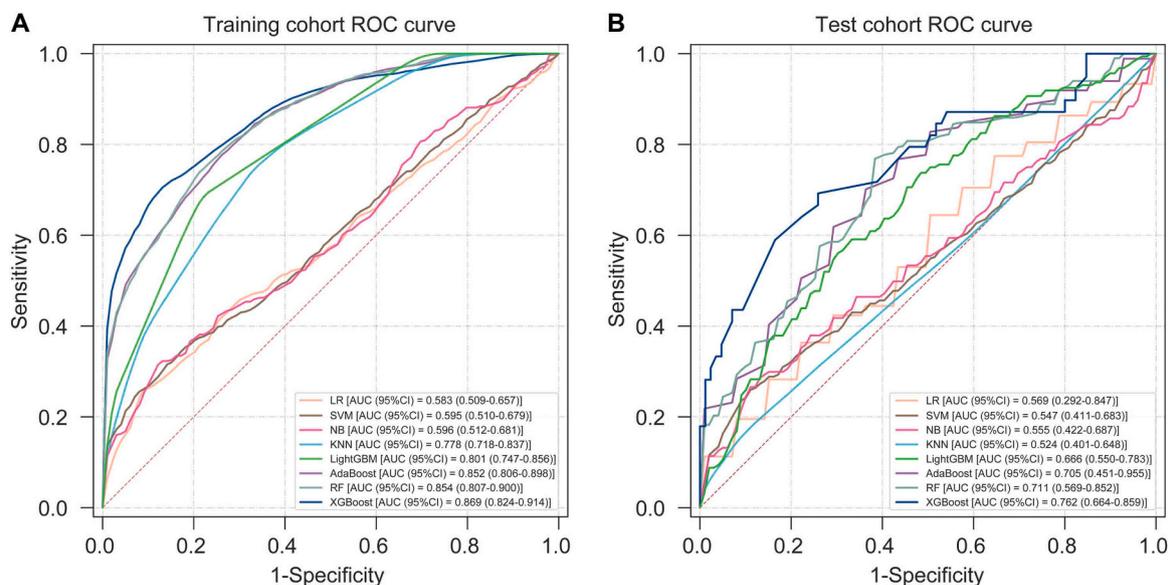


Fig. 2. The receiver operator characteristic curves of eight machine learning algorithms for predicting EHM in HCC patients receiving TACE treatment in the training cohort (A) and test cohort (B). **Abbreviation:** EHM, extrahepatic metastasis; HCC, hepatocellular carcinoma; TACE, transarterial chemoembolization; ROC, receiver operator characteristic; AUC, area under the curve; LR, logistic regression; SVM, support vector machine; NB, Naive Bayes; KNN, K-Nearest Neighbor; LightGBM, Light Gradient Boosting Machine; RF, random forest; AdaBoost, Adaptive Boosting; XGBoost, eXtreme Gradient Boosting.

Table 3

Performance of eight machine learning algorithms for predicting extrahepatic metastasis in hepatocellular carcinoma patients receiving transarterial chemoembolization treatment.

Models	AUC (95% CI)	Accuracy	Sensitivity	Specificity	PPV	NPV	F1 score
Training cohort							
LR	0.583 (0.509-0.657)	0.677	0.355	0.834	0.523	0.730	0.406
SVM	0.595 (0.510-0.679)	0.679	0.384	0.826	0.546	0.735	0.426
NB	0.596 (0.512-0.681)	0.699	0.323	0.882	0.551	0.732	0.407
KNN	0.778 (0.718-0.837)	0.747	0.775	0.651	0.682	0.761	0.724
LightGBM	0.801 (0.747-0.856)	0.744	0.707	0.755	0.821	0.741	0.749
AdaBoost	0.852 (0.806-0.898)	0.765	0.773	0.762	0.619	0.865	0.683
RF	0.854 (0.807-0.900)	0.762	0.805	0.741	0.604	0.882	0.687
XGBoost	0.869 (0.824-0.914)	0.828	0.717	0.882	0.758	0.861	0.733
Test cohort							
LR	0.569 (0.292-0.847)	0.614	0.733	0.586	0.346	0.696	0.389
SVM	0.547 (0.411-0.683)	0.638	0.399	0.804	0.475	0.714	0.381
NB	0.555 (0.422-0.687)	0.675	0.479	0.717	0.491	0.712	0.457
KNN	0.524 (0.401-0.648)	0.632	0.333	0.760	0.425	0.680	0.324
LightGBM	0.666 (0.550-0.783)	0.677	0.640	0.675	0.549	0.715	0.570
AdaBoost	0.705 (0.451-0.955)	0.650	0.722	0.738	0.476	0.768	0.566
RF	0.711 (0.569-0.852)	0.654	0.778	0.648	0.475	0.774	0.588
XGBoost	0.762 (0.664-0.859)	0.734	0.692	0.741	0.568	0.825	0.624

Abbreviation: PPV, positive predictive value; NPV, negative predictive value; LR, logistic regression; SVM, support vector machine; NB, Naive Bayes; KNN, K-Nearest Neighbor; LightGBM, Light Gradient Boosting Machine; AdaBoost, Adaptive Boosting; RF, random forest; XGBoost, eXtreme Gradient Boosting.

based on the XGBoost algorithm following feature selection, comprising 7 AP features (1 tumor-related and 6 peritumoral-related) and 6 PVP features (1 tumor-related and 5 peritumoral-related). The most influential feature was the peritumoral waveletHLL_firstorder_Minimum in AP, yet it exhibited an inverse relationship with the risk of EHM (Fig. 8A and B). Fig. 8C and D illustrate the interpretability of the Rad-ML model for an EHM and a non-EHM patient, respectively. For the first patient, the most critical feature of AP_peritumoral_waveletHLL_firstorder_Minimum was low (0.481), resulting in an elevated risk of EHM (0.842). For the second patient, the most critical feature was high (0.940), indicating a reduced risk of EHM (0.070).

4. Discussion

In this research, we applied radiomics and eight machine learning algorithms based on preoperative MRI to predict EHM in HCC patients

treated with TACE and confirmed that the XGBoost algorithm has the best predictive performance. Then, our results demonstrated that the combined model, consisting of clinical parameters and radiomics features, and Rad-ML model outperformed the clinical model in predicting EHM. However, the combined model did not improve performance compared to the Rad-ML model.

In clinical model, our study found that preoperative serum AFP ≥ 400 ng/mL and NLR ≥ 3 were independent risk factors for EHM in HCC patients receiving TACE treatments. Numerous clinical models have previously documented that elevated serum AFP levels serve as a substantial indicator for predicting EHM in patients with HCC [18–25], a finding that is in line with our research. Nonetheless, discrepancies persist, and some clinical models have not corroborated the predictive value of AFP for EHM [26–30]. This discrepancy may be attributed to variations in study sample sizes, disparate therapeutic regimens, and the heterogeneous nature of the research populations. Furthermore,

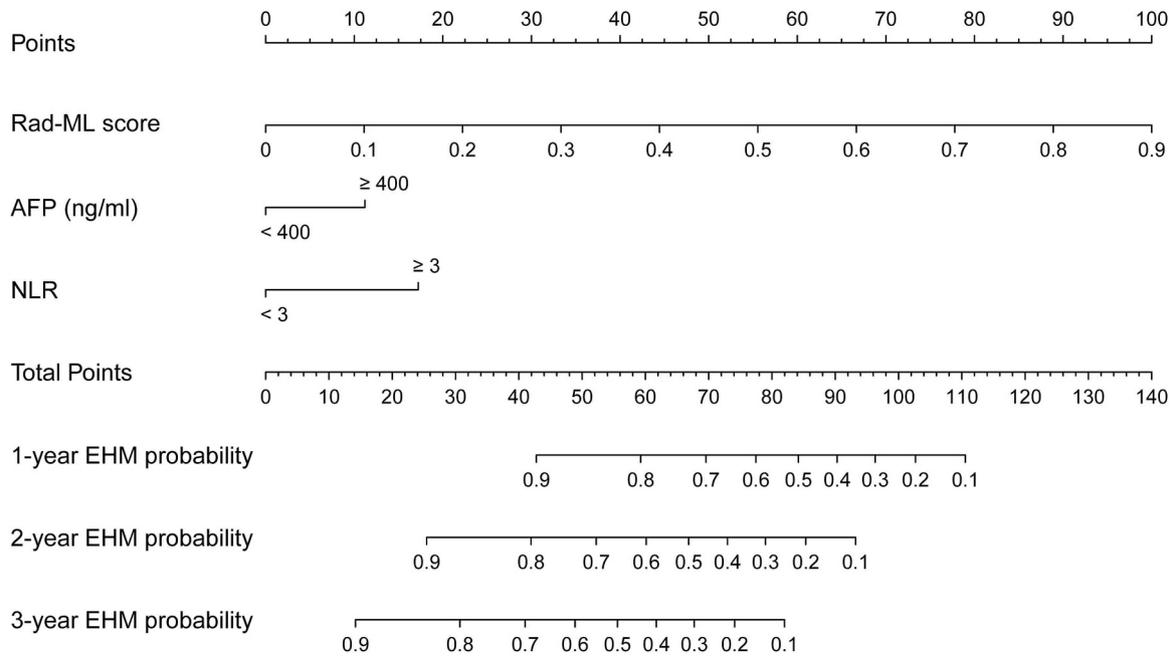


Fig. 3. Nomogram of the combined model for predicting the 1-year, 2-year, and 3-year EHM rates in HCC patients receiving TACE treatment. **Abbreviation:** EHM, extrahepatic metastasis; HCC, hepatocellular carcinoma; TACE, transarterial chemoembolization; Rad-ML, radiomics machine learning; AFP, alpha fetoprotein; NLR, neutrophil to lymphocyte ratio.

Table 4

Performance of different models for predicting extrahepatic metastasis in hepatocellular carcinoma patients receiving transarterial chemoembolization treatment.

Models	C-index (95% CI)	1-year AUC (95% CI)	2-year AUC (95% CI)	3-year AUC (95% CI)	4-year AUC (95% CI)	5-year AUC (95% CI)	p value
Training cohort							
Clinical model	0.61 (0.53-0.68)	0.66 (55.33-75.65)	0.60 (50.96-69.94)	0.62 (51.16-71.80)	0.61 (47.03-73.92)	0.52 (33.35-70.30)	< 0.001 ^a
Rad-ML model	0.85 (0.80-0.90)	0.88 (82.59-93.98)	0.86 (79.42-93.32)	0.87 (77.37-95.62)	0.88 (75.45-101.21)	0.96 (90.74-100.66)	< 0.001 ^b
Combined model	0.86 (0.82-0.90)	0.90 (85.13-95.09)	0.88 (81.50-94.01)	0.89 (80.84-96.31)	0.90 (78.27-101.05)	0.95 (88.69-100.68)	0.731 ^c
Test cohort							
Clinical model	0.62 (0.51-0.73)	0.62 (46.62-78.18)	0.59 (43.42-74.61)	0.61 (45.06-76.12)	0.53 (31.77-74.28)	0.62 (50.22-73.03)	< 0.001 ^a
Rad-ML model	0.82 (0.72-0.91)	0.82 (70.51-92.50)	0.80 (67.19-93.53)	0.88 (75.60-100.41)	0.95 (87.97-102.73)	0.93 (81.18-104.78)	< 0.001 ^b
Combined model	0.83 (0.75-0.92)	0.83 (73.12-93.75)	0.82 (69.97-94.37)	0.89 (77.38-99.74)	0.95 (86.90-102.39)	0.93 (81.38-105.52)	0.143 ^c

Note: ^aClinical model vs Combined model; ^bClinical model vs Rad-ML model; ^cRad-ML model vs Combined model.

Abbreviation: C-index, Concordance index; AUC, area under the curve; Rad-ML, radiomics machine learning.

preoperative NLR as an independent risk factor for EHM has not been previously reported in clinical models [20–29]. Nonetheless, some meta-analyses and reviews have consistently demonstrated that elevated NLR pre-treatment serves as a reliable prognostic indicator of recurrence and survival in HCC patients, potentially due to the crucial role of NLR as an inflammatory marker in reducing immunity, secreting and modulating multiple chemokines and cytokines [31–34]. The C-index of our clinical model, however, was only 0.61 in the training set and 0.62 in the test set, which reflected a suboptimal performance. This finding was consistent with some prior investigations. A recent study, for instance, developed a prediction model for EHM in HCC, identifying neutrophils, prothrombin time, tumor number, and size as independent risk factors, yielding a C-index of 0.672 in the training group and 0.694 in the validation group [26]. Another study determined that a nomogram constructed by tumor number, tumor size, platelet count, serum AFP, and macrovascular invasion was associated with high risk of EHM, with a C-index of 0.733 and 0.739 in the training and validation sets, respectively [19]. These studies corroborate our conclusion that clinical models alone are insufficient for predicting EHM performance, thus motivating us to develop a predictive model incorporating radiomics and machine learning algorithms.

We compared the predictive effects of eight machine learning algorithms and found that the XGBoost algorithm achieved the highest AUC in both the training and test sets. This suggests that the XGBoost

algorithm is well-suited for predicting EHM in HCC patients treated with TACE. XGBoost algorithm employs a computationally efficient approach to gradient boosting, utilizing decision trees added iteratively to rectify classification errors, while utilizing a boosted decision tree structure that bins similar features for reduced computational complexity, resulting in improved performance [35]. A recent study reviewed existing methods for cancer detection using machine learning, focusing on the four most common cancers around the world: breast, prostate, lung, and colorectal cancer. The results showed that the XGBoost algorithm performed the highest overall accuracy of 73% and 75% using two different transforms [36]. Another comparative study of machine learning algorithms found that XGBoost was superior to SVM (AUC of 0.896 and 0.850, respectively) in computer-aided diagnosis of pulmonary nodule classification [37]. However, comparative studies of machine learning algorithms in HCC patients receiving TACE are limited, and our study expands the application of Rad-ML in these patient populations.

The combined model and Rad-ML model demonstrated better performance than the clinical model in predicting EHM. This finding underscores the value of radiomic features to improve predictive performance. A study that applied deep learning to predict EHM and macrovascular invasion in HCC patients produced similar results to ours [38]. It was found that the AUC of clinical model, deep learning model and combined model in the validation set were 0.780, 0.836 and 0.862,

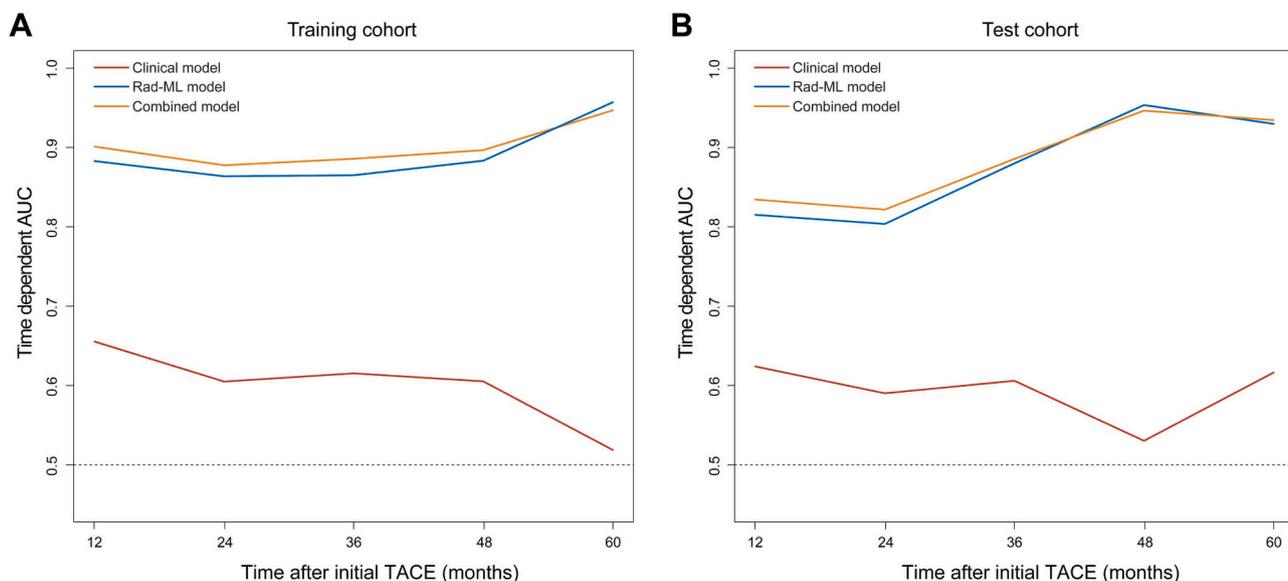


Fig. 4. Time dependent area under the curves of clinical model, Rad-ML model, and combined model for predicting the 1-year, 2-year, 3-year, 4-year, and 5-year EHM in HCC patients receiving TACE treatment in the training cohort (A) and test cohort (B). **Abbreviation:** Rad-ML, radiomics machine learning; EHM, extra-hepatic metastasis; HCC, hepatocellular carcinoma; TACE, transarterial chemoembolization; AUC, area under the curve.

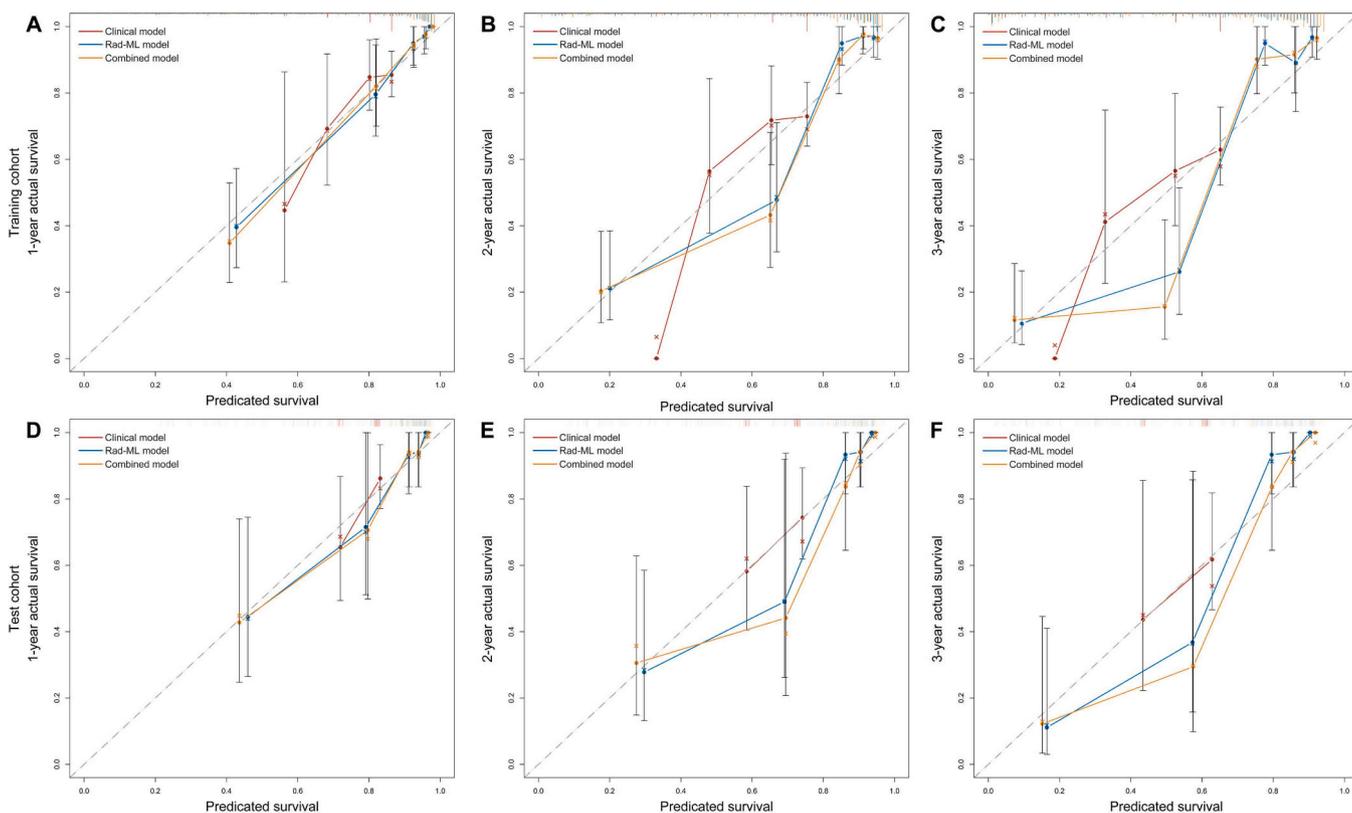


Fig. 5. Calibration and performance of clinical model, Rad-ML model, and combined model for predicting the 1-year, 2-year, and 3-year EHM in HCC patients receiving TACE treatment in the training cohort (A, B, and C) and test cohort (D, E, and F). **Abbreviation:** Rad-ML, radiomics machine learning; EHM, extrahepatic metastasis; HCC, hepatocellular carcinoma; TACE, transarterial chemoembolization.

respectively. Deep learning model and combined model predicted better than clinical model, but there was no significant difference between them. Given that the prediction of the clinical model was unsatisfactory, as mentioned before, its inclusion did not improve the predictive effectiveness of the combined model.

We investigated the features contributing to the Rad-ML model's

prediction of EHM and explained the black box of radiomics using SHAP method. Firstly, we applied the previously reported radiomic segmentation method on the recurrence of HCC in authoritative literature, that is, the tumoral and peritumoral features in AP and PVP were segmented respectively [39], which contains more information than previous studies that only covered single phase and only tumor segmentation [38,

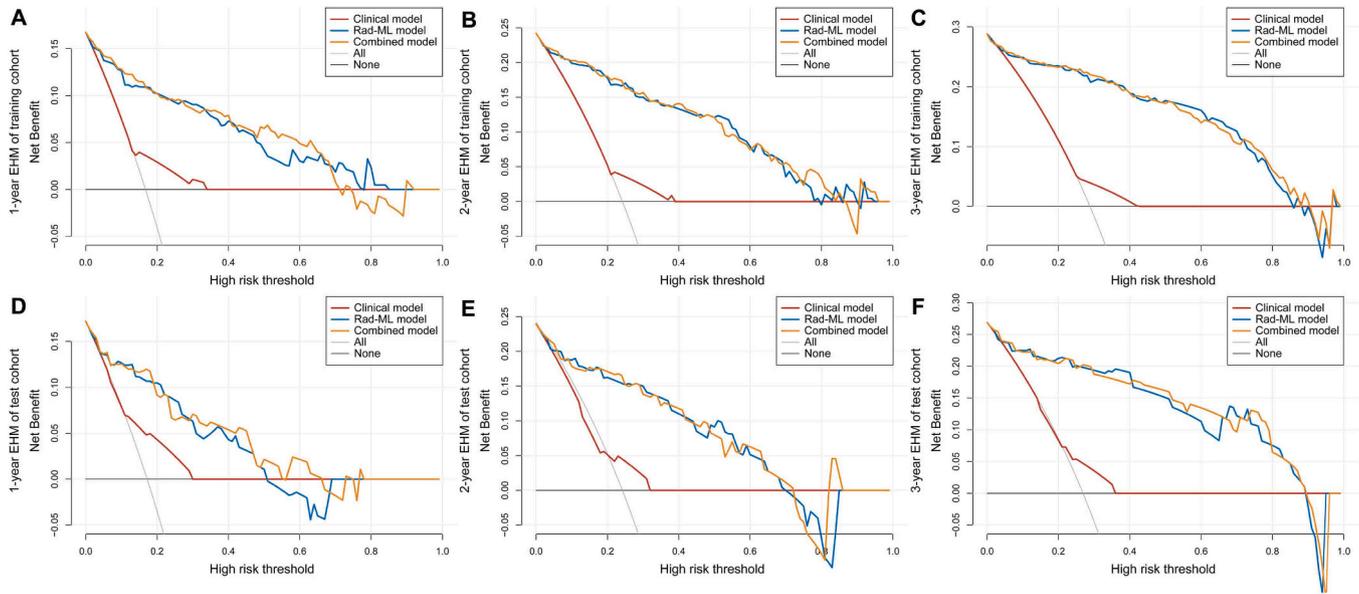


Fig. 6. Decision curve analysis and performance of clinical model, Rad-ML model, and combined model for predicting the 1-year, 2-year, and 3-year EHM in HCC patients receiving TACE treatment in the training cohort (A, B, and C) and test cohort (D, E, and F). **Abbreviation:** Rad-ML, radiomics machine learning; EHM, extrahepatic metastasis; HCC, hepatocellular carcinoma; TACE, transarterial chemoembolization.

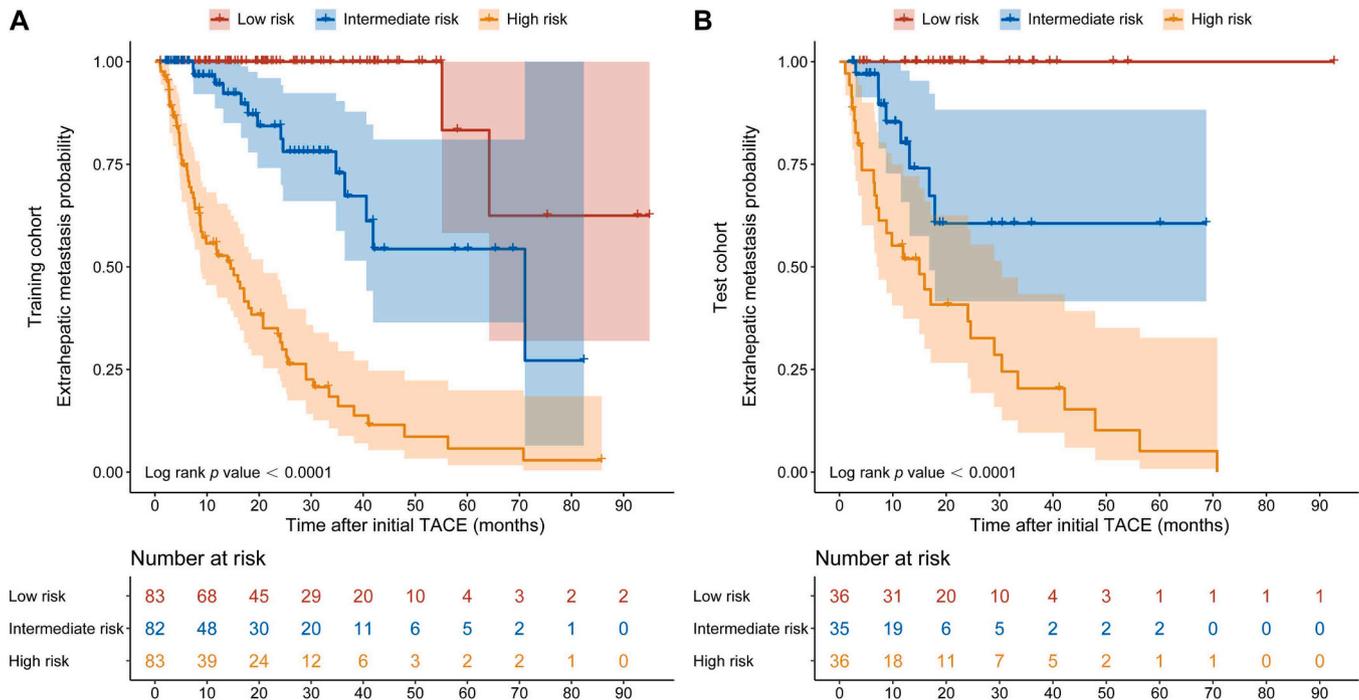


Fig. 7. Kaplan-Meier analysis and performance of the Rad-ML model in predicting EHM in HCC patients receiving TACE treatment in the training cohort (A) and test cohort (B). **Abbreviation:** Rad-ML, radiomics machine learning; EHM, extrahepatic metastasis; HCC, hepatocellular carcinoma; TACE, transarterial chemoembolization.

40]. Secondly, after feature selection, there were 13 remaining features, including 11 peritumoral features and 2 tumor features, and the most important feature was the peritumoral waveletHLL_firstorder_Minimum in the AP. This result suggests that increased peritumoral heterogeneity plays an important role of EHM in HCC patients treated with TACE. A recent study using radiomic analysis based on dynamic enhanced CT for predicting EHM in HCC patients found that the Rad-ML model achieved a high AUC of 0.944 in the test set, while the result in the external validation cohort was not satisfactory with an AUC of 0.744 [40], which may be related to the fact that the study only performed tumor

segmentation in the AP. Up to now, the known signs of non-smooth tumor margin, incomplete tumor capsule, presence of satellite nodules, high peritumoral enhancement in AP, and rim hyperenhancement in AP are correlated with adverse prognosis in HCC [41,42], which could explain the important role of peritumoral radiomic features in predicting EHM. However, the mechanism of peritumoral features in EHM of HCC patients receiving TACE treatments still needs to be further studied.

Based on the Rad-ML score, we classified three types of EHM risk stratification: low risk, intermediate risk, and high risk. Given that the rate of EHM in low-risk patients was 0% at 4 years, we recommend an

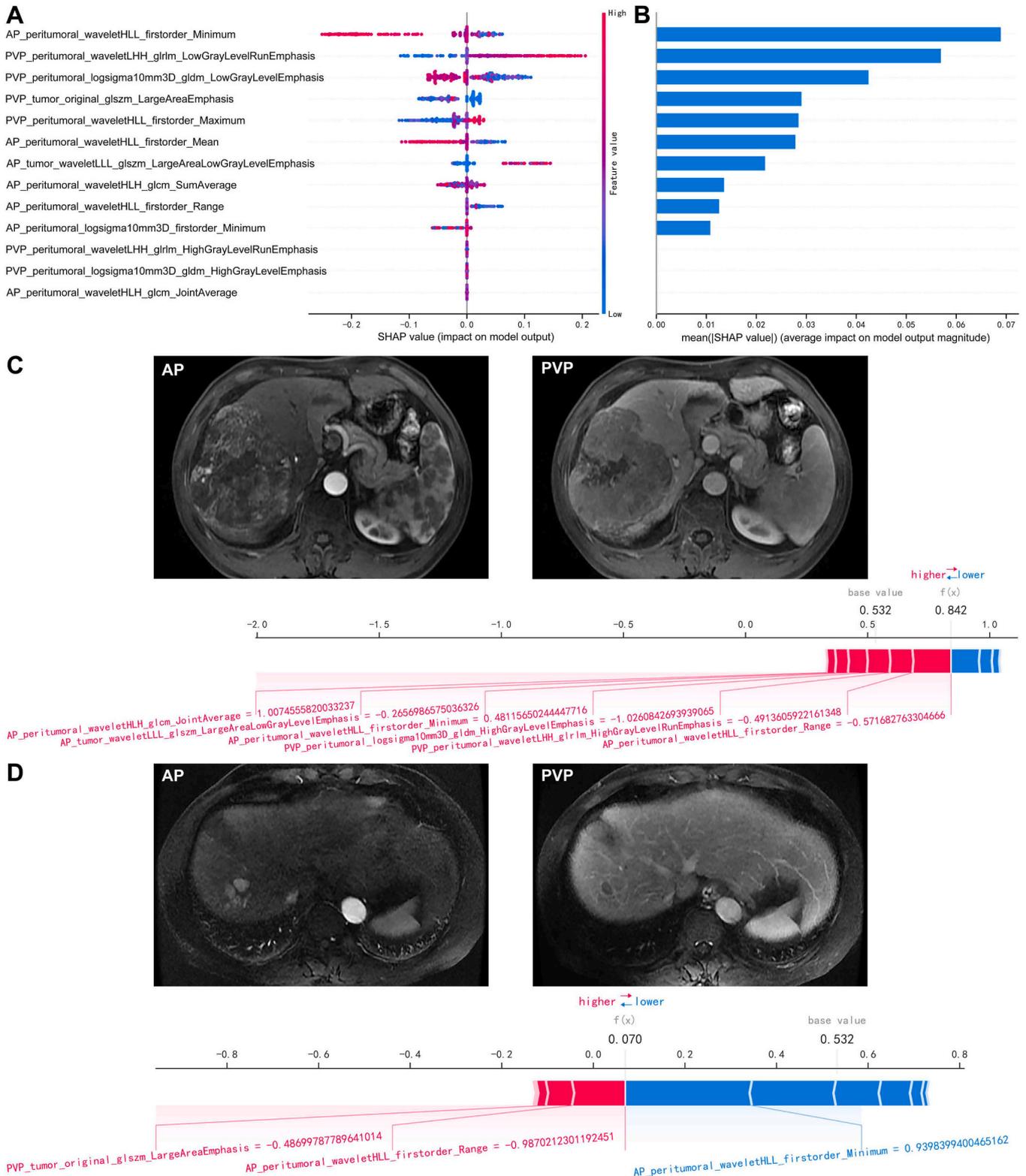


Fig. 8. The SHAP explanatory diagram shows the distribution of features' importance of the Rad-ML model in predicting EHM in HCC patients receiving TACE treatment (A and B). Inference process of the Rad-ML model with an EHM (C) and non-EHM patient (D). **Abbreviation:** SHAP, SHapley Additive explanations; Rad-ML, radiomics machine learning; EHM, extrahepatic metastasis; HCC, hepatocellular carcinoma; TACE, transarterial chemoembolization; AP, arterial phase; PVP, portal venous phase.

appropriate extension of follow-up for reasons of cost-effectiveness of surveillance. For intermediate-risk patients, we recommend following existing follow-up strategy. For high-risk patients, shorter follow-up and early combination of systemic therapy may be advisable.

Our study has several limitations. Firstly, as a retrospective, single-center study, it may contain selection bias, and the findings require validation in larger, multicenter prospective studies. Secondly, as our study predominantly included patients with HBV-related HCC, the

results may not be applicable to patients with other underlying liver diseases. Thirdly, radiological features in the hepatobiliary phase were not evaluated as only a small proportion of patients underwent pre-therapy gadoteric acid-enhanced MRI. Although the above limitations restrict the scope of application of the findings, they do not affect the accuracy of the conclusions, and further prospective multicenter studies are needed to determine whether our findings can be generalized to a wider population.

In conclusion, our study demonstrates the good value of the Rad-ML model based on preoperative MRI for predicting EHM in HCC patients treated with TACE. In addition, the interpretability of the Rad-ML model illustrates the importance of peritumoral features in the AP, which may help clinicians better understand the factors influencing EHM risk and guide personalized treatment strategies for HCC patients. However, further validation of our findings in larger, independent patient cohorts are necessary to confirm the generalizability and clinical utility of our model.

Funding

This work is supported by the National Natural Science Foundation of China (62271509) and Fujian Province Science and Technology Innovation Joint Fund project (2019Y9205).

Ethical statement

The National Cancer Center/Cancer Hospital's Institutional Review Board gave its authorization for this observational research, which was carried out in compliance with the Declaration of Helsinki's ethical principles. The requirement for informed consent was waived as a result of the retrospective study design.

CRedit authorship contribution statement

Zhou Xiang: Supervision. **Huang Xiaoyu:** Software. **Cao Xiaojing:** Software. **Peng Gang:** Writing – original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.ejro.2024.100551](https://doi.org/10.1016/j.ejro.2024.100551).

References

- [1] H. Sung, J. Ferlay, R.L. Siegel, et al., Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries, *CA Cancer J. Clin.* 71 (2021) 209–249.
- [2] A. Vogel, T. Meyer, G. Sapisochin, R. Salem, A. Saborowski, Hepatocellular carcinoma, *Lancet* 400 (2022) 1345–1362.
- [3] M. Reig, A. Forner, J. Rimola, et al., BCLC strategy for prognosis prediction and treatment recommendation: The 2022 update, *J. Hepatol.* 76 (2022) 681–693.
- [4] T. Couri, A. Pillai, Goals and targets for personalized therapy for HCC, *Hepatol. Int.* 13 (2019) 125–137.
- [5] R. Sacco, L. Faggioni, I. Bargellini, et al., Assessment of response to sorafenib in advanced hepatocellular carcinoma using perfusion computed tomography: results of a pilot study, *Dig. Liver Dis.* 45 (2013) 776–781.
- [6] European Association For The Study Of The L, 2012. European Organisation For R, Treatment Of C. EASL-EORTC clinical practice guidelines: management of hepatocellular carcinoma. *J Hepatol* 2012; 56:908–943.
- [7] Jin J., Jiang Y., Zhao Y.L., Huang P.T., 2023. Radiomics-based Machine Learning to Predict the Recurrence of Hepatocellular Carcinoma: A Systematic Review and Meta-analysis. *Acad Radiol* 2023.
- [8] 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 (2017) 749–762.
- [9] Y.P. Zhang, X.Y. Zhang, Y.T. Cheng, et al., Artificial intelligence-driven radiomics study in cancer: the role of feature engineering and modeling, *Mil. Med. Res.* 10 (2023) 22.
- [10] K. Bera, N. Braman, A. Gupta, V. Velcheti, A. Madabhushi, Predicting cancer outcomes with radiomics and artificial intelligence in radiology, *Nat. Rev. Clin. Oncol.* 19 (2022) 132–146.
- [11] S. Feng, J. Wang, L. Wang, et al., Current status and analysis of machine learning in hepatocellular carcinoma, *J. Clin. Transl. Hepatol.* 11 (2023) 1184–1191.
- [12] Z.M. Zou, D.H. Chang, H. Liu, Y.D. Xiao, Current updates in machine learning in the prediction of therapeutic outcome of hepatocellular carcinoma: what should we know? *Insights Imaging* 12 (2021) 31.
- [13] Xia T., Zhao B., Li B., et al., 2023. MRI-Based Radiomics and Deep Learning in Biological Characteristics and Prognosis of Hepatocellular Carcinoma: Opportunities and Challenges. *J Magn Reson Imaging* 2023.
- [14] H. Tian, Y. Xie, Z. Wang, Radiomics for preoperative prediction of early recurrence in hepatocellular carcinoma: a meta-analysis, *Front Oncol.* 13 (2023) 1114983.
- [15] Q. Lai, G. Spoleitini, G. Mennini, et al., Prognostic role of artificial intelligence among patients with hepatocellular cancer: a systematic review, *World J. Gastroenterol.* 26 (2020) 6679–6688.
- [16] J.A. Marrero, L.M. Kulik, C.B. Sirlin, et al., Diagnosis, staging, and management of hepatocellular carcinoma: 2018 practice guidance by the American Association for the Study of Liver Diseases, *Hepatology* 68 (2018) 723–750.
- [17] M. Kudo, Y. Kawamura, K. Hasegawa, et al., Management of hepatocellular carcinoma in Japan: JSH consensus statements and recommendations 2021 update, *Liver Cancer* 10 (2021) 181–223.
- [18] Y. Chen, J. Zeng, P. Guo, J. Zeng, J. Liu, Prognostic significance of platelet-to-lymphocyte ratio (PLR) in extrahepatic metastasis of hepatocellular carcinoma after curative resection, *Cancer Manag Res* 13 (2021) 1395–1405.
- [19] C.H. Lee, C.J. Chang, Y.J. Lin, et al., Nomogram predicting extrahepatic metastasis of hepatocellular carcinoma based on commonly available clinical data, *JGH Open* 3 (2019) 38–45.
- [20] K.B. Chandra, A. Singhal, Predictors of macrovascular invasion and extrahepatic metastasis in treatment naive hepatocellular carcinoma: when is [(18)F] FDG PET/CT relevant? *Nucl. Med Mol. Imaging* 55 (2021) 293–301.
- [21] J.H. Yoon, Y.J. Goo, C.J. Lim, et al., Features of extrahepatic metastasis after radiofrequency ablation for hepatocellular carcinoma, *World J. Gastroenterol.* 26 (2020) 4833–4845.
- [22] T. Yokoo, A.D. Patel, N. Lev-Cohain, A.G. Singal, A.C. Yopp, I. Pedrosa, Extrahepatic metastasis risk of hepatocellular carcinoma based on alpha-fetoprotein and tumor staging parameters at cross-sectional imaging, *Cancer Manag Res* 9 (2017) 503–511.
- [23] S.W. Lee, H.L. Lee, N.I. Han, et al., Early treatment response to transcatheter arterial chemoembolization is associated with time to the development of extrahepatic metastasis and overall survival in intermediate-stage hepatocellular carcinoma, *Cancer Chemother. Pharm.* 79 (2017) 81–88.
- [24] C.H. Lee, Y.J. Lin, C.C. Lin, et al., Pretreatment platelet count early predicts extrahepatic metastasis of human hepatoma, *Liver Int.* 35 (2015) 2327–2336.
- [25] M. Kanda, R. Tateishi, H. Yoshida, et al., Extrahepatic metastasis of hepatocellular carcinoma: incidence and risk factors, *Liver Int* 28 (2008) 1256–1263.
- [26] L. Zhou, L. Ren, W. Yu, et al., Construction and validation of a prediction model of extrahepatic metastasis for hepatocellular carcinoma based on common clinically available data, *Front. Oncol.* 12 (2022) 961194.
- [27] K. Morio, T. Kawaoka, H. Aikata, et al., Preoperative PET-CT is useful for predicting recurrent extrahepatic metastasis of hepatocellular carcinoma after resection, *Eur. J. Radio.* 124 (2020) 108828.
- [28] Y. Morimoto, K. Nouse, N. Wada, et al., Involvement of platelets in extrahepatic metastasis of hepatocellular carcinoma, *Hepatol. Res.* 44 (2014) E353–E359.
- [29] L. Jun, Y. Zhenlin, G. Renyan, et al., Independent factors and predictive score for extrahepatic metastasis of hepatocellular carcinoma following curative hepatectomy, *Oncologist* 17 (2012) 963–969.
- [30] W. Elmoghazy, K. Ahmed, A. Vijay, et al., Hepatocellular carcinoma in a rapidly growing community: epidemiology, clinico-pathology and predictors of extrahepatic metastasis, *Arab J. Gastroenterol.* 20 (2019) 38–43.
- [31] M. Mouchli, S. Reddy, M. Gerrard, L. Boardman, M. Rubio, Usefulness of neutrophil-to-lymphocyte ratio (NLR) as a prognostic predictor after treatment of hepatocellular carcinoma. Review article, *Ann. Hepatol.* 22 (2021) 100249.
- [32] K. Arvanitakis, I. Mitroulis, G. Germanidis, Tumor-associated neutrophils in hepatocellular carcinoma pathogenesis, prognosis, and therapy, *Cancers (Basel)* 13 (2021).
- [33] M. Najjar, S. Agrawal, J.C. Emond, K.J. Halazun, Pretreatment neutrophil-lymphocyte ratio: useful prognostic biomarker in hepatocellular carcinoma, *J. Hepatocell. Carcinoma* 5 (2018) 17–28.
- [34] W.K. Xiao, D. Chen, S.Q. Li, S.J. Fu, B.G. Peng, L.J. Liang, Prognostic significance of neutrophil-lymphocyte ratio in hepatocellular carcinoma: a meta-analysis, *BMC Cancer* 14 (2014) 117.
- [35] S.M. Borstelmann, Machine learning principles for radiology investigators, *Acad. Radio.* 27 (2020) 13–25.
- [36] M. Mokoatle, V. Marivate, D. Mapiye, R. Bornman, V.M. Hayes, A review and comparative study of cancer detection using machine learning: SBERT and SimCSE application, *BMC Bioinforma.* 24 (2023) 112.
- [37] M. Nishio, M. Nishizawa, O. Sugiyama, et al., Computer-aided diagnosis of lung nodule using gradient tree boosting and Bayesian optimization, *PLoS One* 13 (2018) e0195875.
- [38] S. Fu, M. Pan, J. Zhang, et al., Deep learning-based prediction of future extrahepatic metastasis and macrovascular invasion in hepatocellular carcinoma, *J. Hepatocell. Carcinoma* 8 (2021) 1065–1076.

- [39] G.W. Ji, F.P. Zhu, Q. Xu, et al., Radiomic features at contrast-enhanced CT predict recurrence in early stage hepatocellular carcinoma: a multi-institutional study, *Radiology* 294 (2020) 568–579.
- [40] L.W.C. Chan, S.C.C. Wong, W.C.S. Cho, et al., Primary tumor radiomic model for identifying extrahepatic metastasis of hepatocellular carcinoma based on contrast enhanced computed tomography, *Diagnostic* 13 (2022).
- [41] S.H. Hwang, H. Rhee, Radiologic features of hepatocellular carcinoma related to prognosis, *J. Liver Cancer* 23 (2023) 143–156.
- [42] C. An, M.J. Kim, Imaging features related with prognosis of hepatocellular carcinoma, *Abdom. Radio. (NY)* 44 (2019) 509–516.