ORIGINAL ARTICLE

Prediction of TACE Treatment Response in a Preoperative MRI via Analysis of Integrating Deep Learning and Radiomics Features

Yuchi Tian[1] · Temitope Emmanuel Komolafe[2] · Tao Chen[3] · Bo Zhou[4] · Xiaodong Yang[1,5]

Abstract

Purpose To evaluate the efficiency of an integrated model on MRI scans of hepatocellular carcinoma (HCC) patients for preoperative prediction of transcatheter arterial chemoembolization (TACE) treatment response.

Methods Radiomics and deep learning features were integrated to build a prediction model for preoperative procedures so as to obtain a fast and accurate prediction of TACE treatment response. This is a retrospective study and the data consists of 71 HCC patients who underwent TACE treatment in a single center. These patients were divided into two groups: progressive disease (PD) response (20 patients) and non-progressive disease (N-PD) response (51 patients). fivefold cross-validation was applied to the data set to validate model performance. A receiver operating characteristic (ROC) curve was used to assess the predictive ability of the model. Quantification of its results was performed by calculating the area under the receiver operating characteristic curve (AUC). The accuracy, recall, specificity, precision and f1_score were also calculated for the cutoff value that maximized the AUC value.

Results As assessed by the fivefold cross-validation, the integrated model had the best prediction ability, with a value of AUC 0.947 ± 0.069, an accuracy of 0.893 ± 0.088, f1-score of 0.700 ± 0.245, specificity of 0.700 ± 0.245, precision of 0.700 ± 0.245 and a recall of 0.600 ± 0.279. This was followed by the deep learning-based model with an AUC of 0.867 ± 0.121 and the radiomics-based model with an AUC of 0.848 ± 0.128.

Conclusion The experiment results demonstrate that a feature set that combines radiomics and deep learning features tends to be effective in predicting TACE treatment response as opposed to using only one feature. However, due to the limited amount of data, more data will be needed to verify the effectiveness of this method in the future.

Keywords Radiomics · Deep learning · Hepatocellular carcinoma · Therapeutic chemoembolization

✉ Bo Zhou
zhou.bo@zs-hospital.sh.cn

✉ Xiaodong Yang
xiaodong.yang@sibet.ac.cn

Tao Chen
eetch-en@fudan.edu.cn

[1] Academy of Engineering and Technology, Fudan University, Shanghai 200433, China

[2] School of Biomedical Engineering, ShanghaiTech University, Shanghai 201210, China

[3] Academy of Information Science and Technology, Fudan University, Shanghai 200433, China

[4] Department of Interventional Radiology, Zhongshan Hospital, Shanghai Medical College, Fudan University, Shanghai 200032, China

[5] Department of Medical Imaging, Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou 215163, China

1 Introduction

Hepatocellular carcinoma (HCC) is the third leading cause of cancer-related death worldwide [1]. It is the 2nd and 6th common cancer worldwide in men and women respectively. Every year, there are approximately 905,677 new cases and 830,180 deaths globally [2]. Transcatheter arterial chemoembolization (TACE) is considered a routine and standard treatment for unresectable liver cancer, which has proven effective for the treatment of liver cancer in its intermediate stages [3–7]. However, due to the highly heterogeneous biological behavior of tumor cells, the results of TACE treatment vary between individuals [8]. The objective response rates of progressive disease following TACE range from 15 to 61% [9–12]. Therefore, it is necessary to quickly and accurately select ideal candidates for TACE therapy preoperation, in order to improve treatment efficacy and

overall survival rate. Previous studies have reported that genes and proteins might be biomarkers for TACE treatment response [13, 14]. However, obtaining these biomarkers is quite time-consuming and expensive, which further increases the financial burden on the patients. Radiomics, a high-dimensional quantitative feature analysis approach, can extract high-throughput features from medical imaging and perform quantitative analysis on tumor heterogeneity. Several studies have shown that radiomics features can characterize the tumor and the tumor microenvironment (TME) [15–21]. These were closely related to specific microscopic features at the genetic, protein and molecular levels. The use of radiomic features has been suggested in predicting molecular subtyping, tumor gene expression, pathological classification, treatment response and survival rates [17–25]. Deep learning is a type of machine learning method that can extract a large number of higher-level deep features from deep hidden layers of the convolution neural network (CNN) and has been widely successful in image recognition and classification compared with the handcrafted features, these deep features contain more abstract medical image information and provide more insight for predictive patterns. Incorporating deep learning into the current radiomics model can enrich the judgment factor of the model and improve its prediction performance. Although there are many studies using radiomics or deep learning to predict the response of TACE in HCC patients, there are still relatively few studies combining deep learning and radiomics features for preoperatively predicting TACE treatment response.

In this study, we propose an integrated model for TACE treatment response prediction of HCC patients by integrating deep learning and radiomics features in MRI scans and carry out a quantitative analysis of the features. Our goal is to aid doctors in making decisions regarding pre-treatment based on our model. We only use the deep learning method as a feature extractor for the model and then integrate deep learning features into the radiomics analysis model, which enriches the judgment factor of the model and improves its prediction performance with limited training data. Patients with high levels of positive predicted response outcomes are selected for TACE treatment, while patients with poorly predicted treatment outcomes can have their treatment plan adjusted accordingly while there is still time.

2 Materials and Methods

2.1 Patients

The data in this study consisted of 71 HCC patients who received TACE treatment in our center from February 2016 to August 2020. This is a retrospective study from a single center. The hospital ethical review committee approved our

Table 1 Clinical characteristics of the patients

	Total	PD	N-PD	P value
Age	54.69 ± 12.19	54.42 ± 12.85	53.37 ± 10.58	0.909
Sex				0.001
Male	51	9	42	
Female	20	11	9	
AFP(ng/ml)				0.632
≥ 25	60	16	38	
< 25	11	4	13	
HBV-DNA				0.375
+	40	11	29	
−	31	9	22	
Tumor number				0.607
≤ 3	32	10	22	
> 3	39	10	29	
Tumor size (cm)	9.98 ± 3.7	9.95 ± 3.4	10.1 ± 3.7	0.698
EOCG				0.952
0	44	8	20	
1	27	12	31	
CNLC HCC stage				0.375
Ib	33	11	22	
IIa	38	9	29	

research protocol and waived the requirement for informed consent (B2019-336). Table 1 shows the clinical information of the patients. The patients were followed up once in the first month after the operation and subsequently every 2–3 months. The modified RECIST criteria were used to determine whether there was a treatment response in a 6-month follow-up venous phase MRI scan. The criteria for patient inclusion are as follows: (1) A multifunctional MR test was performed within one week before surgery; (3) The patient has a confirmed clinical diagnosis of HCC; (4) TACE was used as the initial treatment protocol; (5) Complete clinical information on the patient was available. The criteria for exclusion of patients were as follows: (1) Poor image quality (blurry images); (2) The follow-up time was less than six months; (3) There were other malignant tumors; (4) Other surgical or chemotherapy interventions were received before TACE; (5) Expected survival time is less than 12 months. Finally, as shown in Fig. 1, 20 patients with progressive disease (PD) response and 51 patients with non-progressive disease (N-PD) response after TACE treatment were selected.
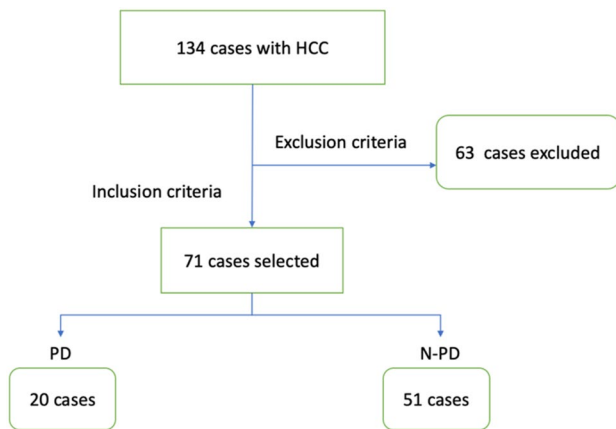
Fig. 1 A flowchart describing how the authors enrolled and excluded patients

Fig. 2 An example of the tumor ROI segmentation on one slice of T2WI

2.2 TACE Procedure

The TACE procedure was performed by six interventional radiologists (IR) with more than 5 and 10 years of clinical experience with TACE respectively. A 5F catheter (RH catheter) was used to selectively perform diagnostic angiography of the celiac trunk and superior mesenteric artery. The process of super-selective catheterization for the tumor-feeding artery was conducted using a 2.7-F microcatheter (Progreat; Terumo). Thereafter, a mixture of 75 mg/m^2 oxaliplatin and 500 mg/m^2 5-fluorouracil was infused. This was mixed with 5–20 mL of iodized oil (Lipiodol Ultrafluido; Guerbet), 30–50 mg/m^2 epirubicin (Pfizer Inc.) and gelatin sponge (Bi-Trumed Biotech Co., Ltd.) for chemical embolization. The embolic material was applied under the guidance of a fluoroscope. The chemical embolization preprocess concluded when the contrast agent stopped being removed from the blood vessel after 10 consecutive heartbeats.

2.3 Image Acquisition and Preprocessing

MRI scans were acquired on a 3.0 T MRI scanner device (Verio; Siemens, Erlangen, Germany). The MRI protocol consisted of a T2-weighted TSE sequence (TR/TE = 3320 ms/83 ms; an acquired resolution of 0.74×0.74 mm^2, slice thickness = 6 mm; matrix = 320×320). The segmentation of the entire tumor volume of interest (VOI) was manually dissected slice-by-slice using the T2-WI by a radiologist with 10 years of experience. Thereafter, each segmentation slice was reviewed and modified by a chief radiologist with over 20 years of experience in HCC MRI analysis. The volume of interests(VOIs) covered the whole tumor. The Medical Imaging Interaction Toolkit (MITK) software was applied to draw the tumor VOI. Figure 2 shows
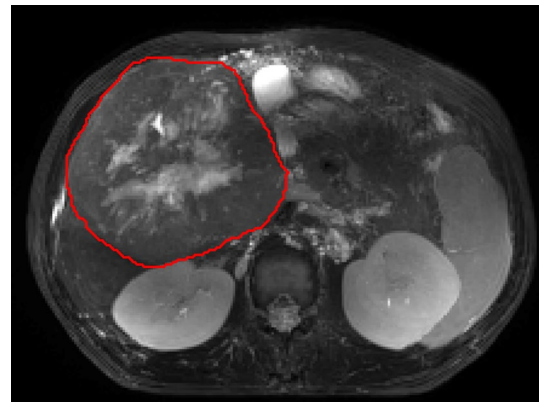
an example of the tumor VOI in a sequence. The dimension of each MRI image was 256×256×40, the intensity value for the T2-WI was normalized by N4BiasFieldCorrection [26] and the intensity range was standardized using histogram matching [27].

2.4 Feature Extraction

2.4.1 Radiomics Features (RsF)

A total of 1595 3D radiological features were extracted from each VOI with Pyradiomics [28]. These radiomic features can be divided into three categories: texture characteristics, intensity characteristics and geometry characteristics. The texture characteristics of VOI can be described by 16 Gylevel size zone matrix (GLSZM) features, 24 Gy-level co-occurrence matrix (GLCM) features, 5 neighboring gray tone difference matrix (NGTDM) features, 16 Gy-level run length matrix (GLRLM) features and 14 Gy-level dependence matrix (GLDM) features. The intensity characteristics within the tumor were reflected by 18 first-order statistical features. The geometry characteristics of the tumor were described using 14 three-dimensional shape features. In addition, eight different image filters were also applied to the original image to yield its corresponding derived image. The filters are gradient, wavelet, square, square root, logarithm, exponential, Laplacian of Gaussian (LoG), and local binary pattern 3D (LBP-3D). The above radiomic features excluding shape features were also extracted from these derived images.

2.4.2 Deep Learning Features (DLF)

A total of 1024 deep learning features for each patient were extracted from a 3D CNN, which consists of two 3D convolution layers and two fully connected layers. The deep

Table 2 3D CNN architecture for extraction of deep learning features

| Layers	Parameter setting
Conv 3D-1	Size $= 5 \times 5 \times 5$; stride $= 1$; zero-padded
Relu-1	Alpha $= 0.2$
Max pool 3D-1	Size $= 4 \times 4 \times 4$; stride $= 4$; zero-padded
Conv 3D-2	Size $= 5 \times 5 \times 5$; stride $= 1$; zero-padded
Relu-2	Alpha $= 0.2$
Max pool 3D-2	Size $= 4 \times 4 \times 4$; stride $= 4$; zero-padded
Fully connected-1	
Flat-1	
Relu-3	Alpha $= 0.2$
Dropout-1	P $= 0.5$
Fully connected-2	
SoftMax	

features were obtained from the outputs of the first fully connected layer after applying the rectified linear unit (ReLU) [29] activation function which maintained values at 0 if the values were < 0. The number 1024 for deep learning features was found to be very effective after a series of trials. The architectures and parameters used are described in Table 2. 3D CNN architecture was designed with Tensorflow [30]. The training batch size was set to 20 and the model was trained using the Adam optimizer [31] with a learning rate of 10^{-3} and 50 epochs.

2.4.3 Radiomics + Deep Learning Features

The output features of the first fully connected layer of the 3D CNN framework were connected with the latter half of the radiomics process so that the extracted 1595 radiomics features and 1024 deep learning features could be merged into an integrative feature space. The total number of features after the fusion was 2619 as shown in Fig. 3. After feature reduction, these were input into the machine learning classifier for joint training.

2.4.4 Feature Selection and Classifier Modeling

We applied normalization to the feature matrix. Each feature vector was subtracted from the average value of the vector and then divided by its length. Due to the high dimensionality of the feature space, we utilized Pearson correlation coefficient (PCC) [32] analysis to identify redundant features. One of the feature pairs would be removed when the absolute value of the PCC was larger than 0.86 as it was considered to be redundant. After that, the dimensionality of the feature space was reduced and each feature became independent of the others. Before building the classifier model, a recursive feature selection approach support vector machine-recursive feature elimination (SVM-RFE) was used to select features. The SVM-RFE method has proven to be very effective in finding worthwhile and significant features for improving classification performance [33, 34]. It selects features based
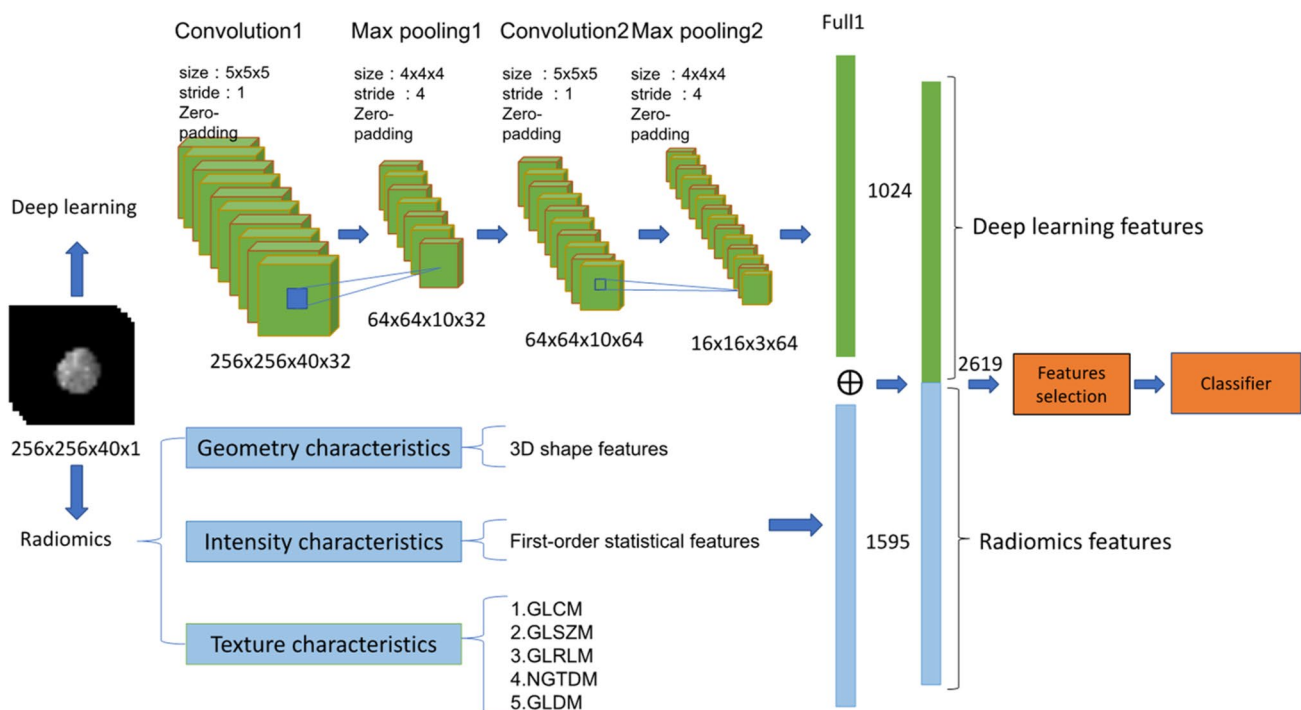
Fig. 3 The predictive models from the integration of radiomics and deep features

on the SVM classifier by recursively considering the smaller size of feature sets. The SVM-RFE algorithm obtained a ranking list of all features by eliminating only one feature that had the least impact on the prediction of the SVM model during each cycle [35, 36]. The first item in the ranking list was the most relevant feature and the last item was the least relevant feature. Finally, the ranking list of the top N features was selected to build the SVM model. Here we used a linear kernel function for building the SVM classifier model [37] with these selected features, which made it easier to interpret the characteristic coefficients of the final model [38]. When the SVM classifier model was built, each selected feature would get a corresponding coefficient. Finally, the predicted RsF + DLF_Score would be calculated by the linear weighted summation of the selected features. Since the number of samples was limited, we also applied an imbalance strategy named SMOTE (Synthetic Minority Oversampling Technique) in the training process [39].

2.5 Statistical Analysis

The receiver operating characteristic (ROC) curve was used to assess the predictive ability of the model and quantification of its results was performed by estimating the area under the receiver operating characteristic curve (AUC). The accuracy, recall, specificity, precision and f1_score were also calculated at the cutoff value that maximized the AUC value. To validate the performance of the model, we used fivefold cross-validation on the data set. In fivefold cross-validation, the data set was randomly divided into five unique subsets S = [s1, s2, s3, s4, s5] to train five independent models. The first model was trained using the subsets [s2, s3, s4, s5] and tested using s1, while the second model was trained using [s1, s3, s4, s5] and tested using s2. This procedure was repeated until all five subsets had been tested. We ensured that there was no patient overlap between the subsets. The mean and standard deviation of the above-mentioned metrics were estimated to evaluate overall model performance. The P value is calculated by univariable association analyses between clinical parameters and TACE treatment response status with the statistical significance set at 0.05. Keras was used to conduct feature selection, feature extraction, classification and statistical analysis. The experiment code was implemented on an Nvidia GeForce GTX 1070 GPU with 8 GB of GDDR5 memory.

3 Results

Table 3 shows the predictive performance of the radiomics-based model, deep learning-based model and integrated model on the data set after fivefold cross-validation. The integrated model had the best predictive ability, with an AUC value of 0.947 ± 0.069, an accuracy of 0.893 ± 0.088, a f1-score of 0.700 ± 0.245, a specificity of 0.700 ± 0.245, a precision of 0.700 ± 0.245 and a recall of 0.600 ± 0.279. This was followed by the deep learning-based model with an AUC of 0.867 ± 0.121 and lastly the radiomics-based model with an AUC of 0.848 ± 0.128. The ROC curves for the three models are shown in Fig. 4. The blue line represents the radiomics feature-based model; the green line corresponds to the deep learning feature-based model; the orange line represents the model with radiomics and deep learning features (the best predictive ability). We also performed a Delong test with the P-value [40] (0.036) between the AUC values of RsF + DLF and the DLF was less than 0.05. However, the difference in P-value between the RsF + DLF and RsF is 0.214 and the difference in P-value between the DLF and RsF is 0.896.

In Table 1, all clinical information is not statistically significant except sex, with the P value being less than 0.05. However, the correlation between sex and TACE efficacy has not been confirmed in medicine, so this factor
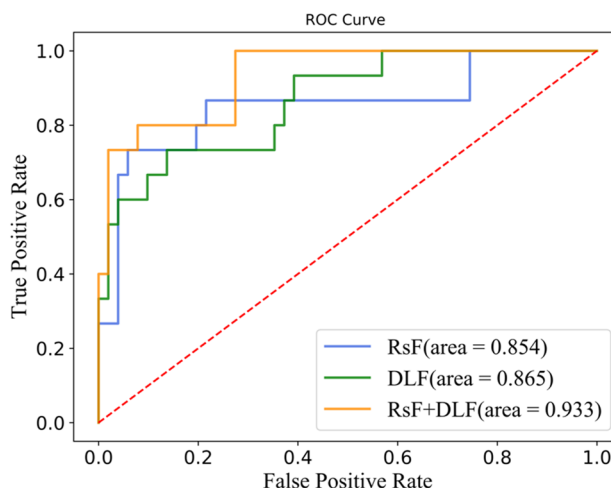

Fig. 4 ROC curves of the prediction models

Table 3 The performance of radiomics-based model, deep learning-based model and integrated model on 5-Fold cross-validation

Model	AUC	95% AUC CI	Accuracy	F1-score	Specificity	Precision	Recall
RsF	0.848 ± 0.128	[0.720 0.987]	0.866 ± 0.032	0.627 ± 0.128	0.960 ± 0.055	0.867 ± 0.183	0.533 ± 0.183
DLF	0.867 ± 0.121	[0.759 0.971]	0.878 ± 0.088	0.673 ± 0.239	0.960 ± 0.055	0.833 ± 0.236	0.600 ± 0.279
RsF + DLF	0.947 ± 0.069	[0.869 0.998]	0.893 ± 0.088	0.700 ± 0.245	0.980 ± 0.045	0.900 ± 0.224	0.600 ± 0.279

is not included in the model analysis. Finally, no clinical characteristics are taken into consideration when building the model. For the integrated model, 864 features remained after PCC screening. The top 30 features were selected by SVM-RFE for model building. The integrated predictive model with the best AUC used 20 selected features and the details are shown in Table 4. By linearly combining the 20 features, the integrated score can be computed as

$$RsF + DLF\_Score = (0.379) \times wavelet - LHL\_glcm\_Idmn$$
$$+ 0.600) \times wavelet - HLH\_firstorder\_Skewness$$
$$+ (-0.356) \times wavelet - HHL\_firstorder\_Mean$$
$$+ (0.385) \times wavelet - HHL\_glcm\_MCC$$
$$+ (0.764) \times wavelet - HHH\_firstorder\_Kurtosis$$
$$+ (0.233) \times wavelet - HHH\_ngtdm\_Complexity$$
$$+ (0.991) \times exponential\_ngtdm\_Busyness$$
$$+ (-0.993) \times lbd - 3D - m2\_glszm\_GrayLevelVariance$$
$$+ (-0.618) \times lbd - 3D - k\_ngtdm\_Contrast$$
$$+ (0.981) \times lbd - 3D - k\_ngtdm\_Busyness$$
$$+ (-0.481) \times deep\_feature\_15$$
$$+ (1.423) \times deep\_feature\_151$$
$$+ (0.999) \times deep\_feature\_231$$
$$+ (-0.550) \times deep\_feature\_253$$
$$+ (0.242) \times deep\_feature\_387$$
$$+ (0.692) \times deep\_feature\_454$$
$$+ (0.662) \times deep\_feature\_508$$
$$+ (0.863) \times deep\_feature\_594$$
$$+ (0.762) \times deep\_feature\_611$$
$$+ (1.019) \times deep\_feature\_757$$

4 Discussion

Previous studies have shown that early response assessment of TACE treatment response is crucial for successful treatment, which may help to modify the treatment plan in time for further effective treatment [41]. Fast and accurate prediction of treatment response before TACE treatment is of great significance for improving overall patient prognosis. Currently, radiomics features have been proven to be inextricably linked to clinical prognosis and tumor genomic patterns [42–46]. Recently, the utilization of radiomics features for preemptively predicting TACE response in HCC patients before surgery has attracted a lot of attention in literature. However, studies that integrate radiomics and deep learning features for preoperative assessment of TACE response are very few.. In our study, we presented an integrated model for quickly and accurately predicting the response of

Table 4 A multivariate analysis of preoperative factors

Features	Coefficient	P-value
wavelet-LHL_glcm_Idmn	0.379	0.208
wavelet-HLH_firstorder_Skewness	0.600	0.007
wavelet-HHL_firstorder_Mean	− 0.356	0.770
wavelet-HHL_glcm_MCC	0.385	0.373
wavelet-HHH_firstorder_Kurtosis	0.764	0.747
wavelet-HHH_ngtdm_Complexity	0.233	0.453
exponential_ngtdm_Busyness	0.991	0.285
lbp-3D-m2_glszm_GrayLevelVariance	− 0.993	0.007
lbp-3D-m2_ngtdm_Contrast	− 0.618	0.036
lbp-3D-k_ngtdm_Busyness	0.981	0.390
deep_feature_15	− 0.481	0.537
deep_feature_151	1.423	0.019
deep_feature_231	0.999	0.227
deep_feature_253	− 0.550	0.010
deep_feature_387	0.242	0.048
deep_feature_454	0.692	0.647
deep_feature_508	0.662	0.876
deep_feature_594	0.863	0.762
deep_feature_611	0.762	0.267
deep_feature_757	1.019	0.047
Rs_DL Score		0.005

Statistically significant level: significant (P value < 0.05); highly significant (P value < 0.01)

unresectable HCC patients before TACE therapy based on analysis using both radiomics and deep learning features.

As assessed using fivefold cross validation, the integrated model achieved the highest AUC of 0.947 ± 0.069, while the deep learning-based and radiomics-based model had AUC values of 0.867 ± 0.121 and 0.848 ± 0.128 respectively. We also performed a Delong test. The P-value between the AUC values of RsF + DLF and DLF was 0.036 which is less than 0.050. However, the P-value between the RsF + DLF and RsF is 0.214 and the P-value between the DLF and RsF is 0.896. In theory, more dimensional feature information should improve model performance. However, the statistically significant difference between the integrated model and radiomics is not very clear. It is only the statistically significant difference between the integrated model and the deep learning model that is obvious. This may be a deviation in results from theory and could be caused by the small amount of data. More data will be collected in the future to verify the proposed theory. The experiment results may not be sufficient to strongly demonstrate that the feature set integrating radiomics and deep learning features is more effective than using one feature type alone in predicting response to TACE treatment. Nonetheless, the integrated model still achieved good performance for predicting TACE treatment response from preoperative MRI scan analyses of HCC patients.

Additionally, we formulated the predictive score equation based on 20 features from the integrated model with the highest AUC, of which 10 were radiomics features and 10 were deep learning features. In our integrated model, both the deep learning and radiomics features contribute to the prediction of TACE treatment efficacy. In deep learning features, deep_feature_151 and deep_feature_253 are strongly correlated with efficacy prediction with P-values of 0.019 and 0.010 respectively, followed by deep_feature_387 (p = 0.048) and deep_feature_757 (p = 0.047). The number in the deep_feature represents the order of neurons in the fully connected layer. For example, deep_feature_151 indicates that this feature is taken from the 151st neuron of the fully connected layer. Among the radiomics features, wavelet-HLH_firstorder_Skewness and lbd-3D-m2_glszm_GrayLevelVariance have an extremely high correlation with efficacy prediction (p = 0.007 and p = 0.007 respectively), followed by lbd-3D-k_ngtdm_Contrast (p = 0.036).

Wavelet filters are mainly used to optimize radiomics features, which can quantify the heterogeneity of tumors at different scales. Previous studies have demonstrated that wavelet features had a strong ability to predict treatment outcomes and could be an important predictor for constructing radiomic features [47–49], which is consistent with the results of our study. First-order radiomics can reflect the distribution of voxel intensities within the tumor ROI region. A significant difference is observed between wavelet-HLH_firstorder_Skewness values of the PD and N-PD cohorts (Fig. 5, left panel) and the median in the PD group is higher than that for the N-PD group. This result demonstrates that voxel intensity information in the tumor area may be related to the TACE treatment response and that with stronger voxel intensity, the therapeutic response may be weaker. The neighborhood gray tone difference matrix (NGTDM) textural features can describe the differences between the gray value of a voxel and the average gray value of its neighboring voxels [50]. The NGTDM features have been applied to tumor heterogeneity analysis [51–53]. In our result, the N-PD group is mostly distributed in the negative value area of the lbd-3D-k_ngtdm_Contrast feature while the N-PD group is mostly distributed in the positive value area. (Fig. 5, middle panel). This may indicate that NGTDM texture features are a good predictor that can provide doctors with more information about TACE treatment response. The Gray Level Size Zone Matrix (GLSZM) can quantify the number of connected voxels that share the same gray level intensity in the tumor area [54]. The GLSZM features are particularly efficient at characterizing texture homogeneity, non-periodicity or speckle-like textures, which could provide better characterizations than granulometry for medical image analysis [55–58]. Figure 5 (right panel) shows that the PD group has a lower median for the lbd-3D m2_glszm_GrayLevelVariance feature than the N-PD group. The variance in
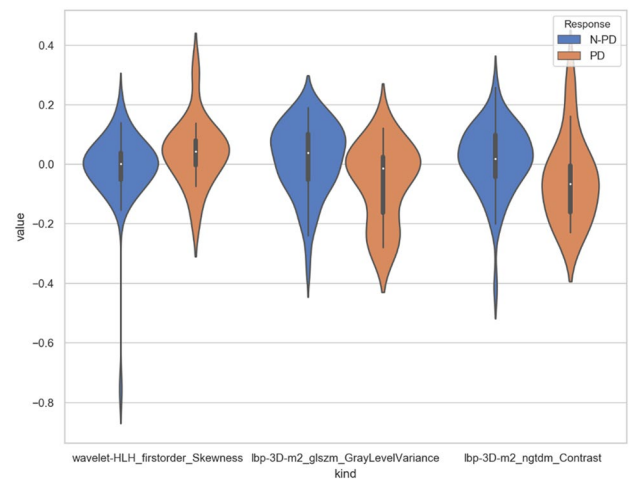

Fig. 5 The violin plot of three radiomics features value distributions between the PD response and N-PD response

gray level intensities for tumor zones is significantly different between the PD and N-PD groups, which indicates that this feature may have an impact on TACE response prediction. The Rs_DL Score is high and statistically significant (0.005), as seen in Table 4. Figure 6 shows the distribution of the Rs_DL Score between the N-PD group and the PD group. The distribution gap is obvious and the case scores in the N-PD group mainly concentrated between 0.0 and −0.5, while the case scores in the PD group are above 0. This indicates that the characteristics of the two groups on the Rs_DL Score are different, which can be used to predict the therapeutic effect of TACE.

However, there are some limitations to this study. First, it is a retrospective study from a single center without additional validation by other hospitals. In addition, the number of HCC cases was limited, especially the proportion of patients with treatment responses. Therefore, the validity of our research results may have been impaired. In future research, more cases and multi-center research are needed. Second, although we have provided the features that contribute to predicting TACE efficacy in the combined model, only the radiomics features can be interpreted, while interpretation of the deep learning features still poses a major problem. The possible correlation between the tumor biological mechanisms and these deep learning features will be investigated from molecular, protein and genetic levels in future work. Also, due to data collection issues, only the T2-WI sequence of MRI scans was available for analysis in this study, which may have omitted information from other kinds of sequences.

In conclusion, we proposed a predictive model to integrate the radiomics and deep learning features for quick and accurate prediction of TACE treatment response for unresectable HCC patients before TACE therapy. The
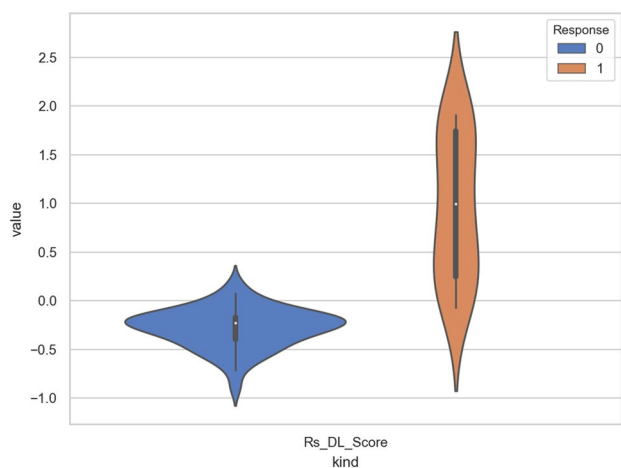
Fig. 6 The violin plot of Rs_DL Score features value distributions between the PD response and N-PD response

experiment results demonstrate that although a feature set that combines radiomics and deep learning features tends to be effective in predicting response to TACE treatment, further validation studies are needed using multi-center data to support this study in the future.

Declaration

Conflict of interest The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

1. Ozakyol, A. (2017). Global epidemiology of hepatocellular carcinoma (HCC epidemiology). *Journal of gastrointestinal cancer, 48*(3), 238–240.
2. Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians, 71*(3), 209–249.
3. Yang, J., Dvornek, N. C.,, Zhang, F., Chapiro, J., Lin, M., Abajian, A., Duncan, J. S. (2019). Hepatocellular carcinoma intra-arterial treatment response prediction for improved therapeutic decision-making. arXiv preprint arXiv:1912.00411.
4. Piasecki, P., Wierzbicki, M., Majewska, A., Kieda, C., & Narloch, J. (2021). Initial experience with novel Embocure Plus microspheres for transarterial chemoembolization (TACE) of liver metastatic colorectal cancer tumours–a clinical and in vitro study. *Polish Journal of Radiology, 86*, 136.
5. Heimbach, J. K., Kulik, L. M., Finn, R. S., Sirlin, C. B., Abecassis, M. M., Roberts, L. R., Zhu, A. X., Hassan Murad, M., & Marrero, J. A. (2018). AASLD guidelines for the treatment of hepatocellular carcinoma. *Hepatology, 67*(1), 358–380.
6. Wang, Q., Xia, D., Bai, W., Wang, E., Sun, J., Huang, M., Mu, W., et al. (2019). Development of a prognostic score for recommended TACE candidates with hepatocellular carcinoma: a multicentre observational study. *Journal of Hepatology, 70*(5), 893–903.
7. Huang, F., Gong, H., & Xiong, Z. (2018). The safety and efficacy of raltitrexed combined with oxaliplatin in TACE treatment for BCLC B/C hepatocellular carcinomas. *Journal of Interventional Radiology, 27*(1), 76–79.
8. Wang, J., Zhong, X., Zhang, Y., Wu, X., Li, S., Jian, P., Ling, Y., et al. (2017). Cezanne predicts progression and adjuvant TACE response in hepatocellular carcinoma. *Cell Death & Disease, 8*(9), 3043–3043.
9. Cheng, A. L., Amarapurkar, D., Chao, Y., Chen, P.-J., Geschwind, J.-F., Goh, K. L., Han, K.-H., et al. (2014). Re-evaluating transarterial chemoembolization for the treatment of Hepatocellular Carcinoma: Consensus recommendations and review by an International Expert Panel. *Liver International, 34*(2), 174–183.
10. Lencioni, R., de Baere, T., Soulen, M. C., Rilling, W. S., & Geschwind, J.-F.H. (2016). Lipiodol transarterial chemoembolization for hepatocellular carcinoma: a systematic review of efficacy and safety data. *Hepatology, 64*(1), 106–116.
11. Guan, Y.-S., He, Q., & Wang, M.-Q. (2012). Transcatheter arterial chemoembolization: history for more than 30 years. *International Scholarly Research Notices, 2012*, 1–8.
12. Fohlen, A., Tasu, J. P., Kobeiter, H., Bartoli, J. M., Pelage, J. P., & Guiu, B. (2018). Transarterial chemoembolization (TACE) in the management of hepatocellular carcinoma: Results of a French national survey on current practices. *Diagnostic and interventional imaging, 99*(9), 527–535.
13. Xu, X., Gao, D., Yuan, X., Liu, L. I., Zhang, X., Liang, X., Chen, S., et al. (2019). β-Catenin expression correlates with prognosis in hepatocellular carcinoma patients treated with transcatheter arterial chemoembolization. *Anticancer Research, 39*(3), 1129–1134.
14. Li, Z., Xiao, Du., Li, X., Zhan, P., Wang, J., & Zhang, H. (2019). Early recurrence detected in hepatocellular carcinoma patients after transcatheter arterial chemoembolization treatment with plasma cell-free DNA. *European Journal of Gastroenterology & Hepatology, 31*(7), 885–892.
15. Moon, S. H., Kim, J., Joung, J.-G., Cha, H., Park, W.-Y., Seok Ahn, J., Ahn, M.-J., et al. (2019). Correlations between metabolic texture features, genetic heterogeneity, and mutation burden in patients with lung cancer. *European Journal of Nuclear Medicine and Molecular Imaging, 46*(2), 446–454.
16. Tomaszewski, M. R., & Gillies, R. J. (2021). The biological meaning of radiomic features. *Radiology, 298*, 202553.
17. Li, Y., Liu, X., Kaibin, Xu., Qian, Z., Wang, K., Fan, X., Li, S., Wang, Y., & Jiang, T. (2018). MRI features can predict EGFR expression in lower grade gliomas: A voxel-based radiomic analysis. *European Radiology, 28*(1), 356–362.
18. Goyen, M. (2014). Radiogenomic imaging-linking diagnostic imaging and molecular diagnostics. *World Journal of Radiology, 6*(8), 519.
19. Wu, M., Tan, H., Gao, F., Hai, J., Ning, P., Chen, J., Zhu, S., Wang, M., Dou, S., & Shi, D. (2019). Predicting the grade of hepatocellular carcinoma based on non-contrast-enhanced MRI radiomics signature. *European Radiology, 29*(6), 2802–2811.
20. Zhang, W., Huang, Z., Zhao, J., He, D., Li, M., Yin, H., Tian, S., Zhang, H., & Song, B. (2021). Development and validation

of magnetic resonance imaging-based radiomics models for preoperative prediction of microsatellite instability in rectal cancer. *Annals of Translational Medicine, 9*(2), 134.

21. Li, X., Jin, J., Yu, Y., Wang, X., Guo, Y., & Sun, H. (2021). PET-CT radiomics by integrating primary tumor and peritumoral areas predicts E-cadherin expression and correlates with pelvic lymph node metastasis in early-stage cervical cancer. *European Radiology, 31*, 1–13.

22. Wang, X.-H., Long, L.-H., Cui, Y., Jia, A. Y., Zhu, X.-G., Wang, H.-Z., Wang, Z., Zhan, C.-M., Wang, Z.-H., & Wang, W.-H. (2020). Mri-Based radiomics model for preoperative prediction of 5-year survival in patients with hepatocellular carcinoma. *British Journal of Cancer, 122*(7), 978–985.

23. Valentinuzzi, D., Vrankar, M., Boc, N., Ahac, V., Zupancic, Z., Unk, M., Skalic, K., et al. (2020). [18F] FDG PET immunotherapy radiomics signature (iRADIOMICS) predicts response of non-small-cell lung cancer patients treated with pembrolizumab. *Radiology and Oncology, 54*(3), 285.

24. Haider, S. P., Burtness, B., Yarbrough, W. G., & Payabvash, S. (2020). Applications of radiomics in precision diagnosis, prognostication and treatment planning of head and neck squamous cell carcinomas. *Cancers of the Head & Neck, 5*, 1–19.

25. Wang, J., Zheng, X., Zhang, J., Xue, H., Wang, L., Jing, R., Chen, S., et al. (2021). An MRI-based radiomics signature as a pretreatment noninvasive predictor of overall survival and chemotherapeutic benefits in lower-grade gliomas. *European Radiology, 31*(4), 1785–1794.

26. Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging, 29*(6), 1310–1320.

27. Nyúl, L. G., Udupa, J. K., & Zhang, X. (2000). New variants of a method of MRI scale standardization. *IEEE Transactions on Medical Imaging, 19*(2), 143–150.

28. Van Griethuysen, J. J. M., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R. G. H., Fillion-Robin, J.-C., Pieper, S., & Aerts, H. J. W. L. (2017). Computational radiomics system to decode the radiographic phenotype. *Cancer Research, 77*(21), e104–e107.

29. Agarap, A. F. (2018). Deep learning using rectified linear units (relu)." arXiv preprint arXiv:1803.08375.

30. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., et al. (2016). Tensorflow: A system for large-scale machine learning. In 12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16), pp. 265–283.

31. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

32. Benesty, J., Chen, J., Huang, Y., & Cohen, I. (2009). Pearson correlation coefficient. In I. Cohen (Ed.), *Noise Reduction in Speech Processing* (pp. 1–4). Berlin: Springer.

33. Fehr, D., Veeraraghavan, H., Wibmer, A., Gondo, T., Matsumoto, K., Alberto Vargas, H., Sala, E., Hricak, H., & Deasy, J. O. (2015). Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images. *Proceedings of the National Academy of Sciences, 112*(46), 6265–6273.

34. Larroza, A., Moratal, D., Paredes-Sánchez, A., Soria-Olivas, E., Chust, M. L., Arribas, L. A., & Arana, E. (2015). Support vector machine classification of brain metastasis and radiation necrosis based on texture analysis in MRI. *Journal of Magnetic Resonance Imaging, 42*(5), 1362–1368.

35. Wei, X., Ai, J., Deng, Y., Guan, X., Johnson, D. R., Ang, C. Y., Zhang, C., & Perkins, E. J. (2014). Identification of biomarkers that distinguish chemical contaminants based on gene expression profiles. *BMC Genomics, 15*(1), 1–17.

36. Ota, K., Oishi, N., Ito, K., Fukuyama, H., Sead-J Study Group. (2014). A comparison of three brain atlases for MCI prediction. *Journal of Neuroscience Methods, 221*, 139–150.

37. Chang, Y.-W., Lin, C.-J. (2008). Feature ranking using linear SVM. In Causation and prediction challenge, pp. 53–64. PMLR.

38. Mladenić, D., Brank, J., Grobelnik, M., Milic-Frayling, N. (2004). Feature selection using linear classifier weights: interaction with classification models. In Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 234–241.

39. Chawla, N. V., et al. (2002). SMOTE: synthetic minority oversampling technique. *Journal of Artificial Intelligence Research, 16*, 321–357.

40. Wang, C.-W., et al. (2018). JOURNAL CLUB: The Warthin tumor score: a simple and reliable method to distinguish Warthin tumors from pleomorphic adenomas and carcinomas. *American Journal of Roentgenology, 210*, 1330–1337.

41. Zhang, Y.-Q., Jiang, L.-J., Wen, J., Liu, D.-M., Gui-Hua Huang, Yu., Wang, W.-Z., & Li, J.-P. (2018). Comparison of α-fetoprotein criteria and modified response evaluation criteria in solid tumors for the prediction of overall survival of patients with hepatocellular carcinoma after transarterial chemoembolization. *Journal of Vascular and Interventional Radiology, 29*(12), 1654–1661.

42. Zinn, P. O., Singh, S. K., Kotrotsou, A., Zandi, F., Thomas, G., Hatami, M., Luedi, M. M., et al. (2016). 139 Clinically applicable and biologically validated MRI radiomic test method predicts glioblastoma genomic landscape and survival. *Neurosurgery, 63*(1), 156–157.

43. Choi, S. W., Cho, H.-H., Koo, H., Rae Cho, K., Nenning, K.-H., Langs, G., Furtner, J., et al. (2020). Multi-habitat radiomics unravels distinct phenotypic subtypes of glioblastoma with clinical and genomic significance. *Cancers, 12*(7), 1707.

44. Capobianco, E., & Dominietto, M. (2020). From medical imaging to radiomics: Role of data science for advancing precision health. *Journal of personalized medicine, 10*(1), 15.

45. Kim, J., Joon Choi, S., Lee, S.-H., Lee, H. Y., & Park, H. (2018). Predicting survival using pretreatment CT for patients with hepatocellular carcinoma treated with transarterial chemoembolization: comparison of models using radiomics. *American Journal of Roentgenology, 211*, 1026–1034.

46. Grossmann, P., Stringfield, O., El-Hachem, N., Bui, M. M., Rios Velazquez, E., Parmar, C., Leijenaar, R. T. H., et al. (2017). Defining the biological basis of radiomic phenotypes in lung cancer. *Elife, 6*, e23421.

47. Jeong, W. K., Jamshidi, N., Felker, E. R., Raman, S. S., & Shinkuo Lu, D. (2019). Radiomics and radiogenomics of primary liver cancers. *Clinical and Molecular Hepatology, 25*(1), 21.

48. Wu, W., Parmar, C., Grossmann, P., Quackenbush, J., Lambin, P., Bussink, J., Mak, R., & Aerts, H. J. W. L. (2016). Exploratory study to identify radiomics classifiers for lung cancer histology. *Frontiers in Oncology, 6*, 71.

49. Mumtaz, W., Xia, L., Azhar Mohd Yasin, M., Azhar Ali, S. S., & Saeed Malik, A. (2017). A wavelet-based technique to predict treatment outcome for major depressive disorder. *PLoS ONE, 12*(2), e0171409.

50. Huang, Y., Liu, Z., He, L., Chen, X., Pan, D., Ma, Z., Liang, C., Tian, J., & Liang, C. (2016). 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), 947–957.

51. Chicklore, S., Goh, V., Siddique, M., Roy, A., Marsden, P. K., & Cook, G. J. R. (2013). Quantifying tumour heterogeneity in 18 F-FDG PET/CT imaging by texture analysis. *European Journal of Nuclear Medicine and Molecular Imaging, 40*(1), 133–140.

52. Cook, G. J. R., Yip, C., Siddique, M., Goh, V., Chicklore, S., Roy, A., Marsden, P., Ahmad, S., & Landau, D. (2013). Are pretreatment 18F-FDG PET tumor textural features in non–small cell lung

cancer associated with response and survival after chemoradiotherapy? *Journal of Nuclear Medicine, 54*(1), 19–26.

53. Mahmood, U., Apte, A., Deasy, J. O., Schmidtlein, C. R., & Shukla-Dave, A. (2017). Investigating the robustness Neighborhood Grey Tone Difference Matrix (NGTDM) and Grey Level Co-Occurrence Matrix (GLCM) radiomic features on clinical Computed Tomography systems using Anthropomorphic Phantoms: Evidence from a multivendor study. *Journal of Computer Assisted Tomography, 41*(6), 995.

54. Ekert, K., Hinterleitner, C., & Horger, M. (2019). Prognosis assessment in metastatic gastrointestinal stromal tumors treated with tyrosine kinase inhibitors based on CT-texture analysis. *European journal of radiology, 116*, 98–105.

55. Thibault, G., Fertil, B., Navarro, C., Pereira, S., Cau, P., Levy, N., Sequeira, J., & Mari, J.-L. (2013). Shape and texture indexes application to cell nuclei classification. *International Journal of Pattern Recognition and Artificial Intelligence, 27*(01), 1357002.

56. Duron, L., Balvay, D., Vande Perre, S., Bouchouicha, A., Savatovsky, J., Sadik, J.-C., Thomassin-Naggara, I., Fournier, L., & Lecler, A. (2019). Gray-level discretization impacts reproducible MRI radiomics texture features. *PLoS ONE, 14*(3), e0213459.

57. Buvat, I., Orlhac, F., & Soussan, M. (2015). Tumor texture analysis in PET: Where do we stand? *Journal of Nuclear Medicine, 56*(11), 1642–1644.

58. Ekert, K., Hinterleitner, C., Baumgartner, K., Fritz, J., & Horger, M. (2020). Extended texture analysis of non-enhanced whole-body MRI image data for response assessment in multiple myeloma patients undergoing systemic therapy. *Cancers, 12*(3), 76.