



Predicting Response to Transarterial Chemoembolization in Hepatocellular Carcinoma Using Machine Learning Models

Niharika Dutta¹ Pankaj Gupta¹

¹Department of Radiodiagnosis and Imaging, Postgraduate Institute of Medical Education and Research, Chandigarh, India

Address for correspondence Pankaj Gupta, MD, Department of Radiodiagnosis and Imaging, Postgraduate Institute of Medical Education and Research, Chandigarh 160012, India (e-mail: Pankajgupta959@gmail.com).

Indian J Radiol Imaging 2026;36:183–191.

Abstract

Background Hepatocellular carcinoma (HCC) is a leading cause of cancer-related deaths, with transarterial chemoembolization (TACE) being a key treatment for intermediate-stage cases. Accurate prediction of TACE response remains challenging, prompting the exploration of machine learning (ML) models.

Aim This study aims to investigate ML models for predicting the response to TACE in HCC patients.

Materials and Methods We utilized the public “WAW-TACE” data set. This data set comprises clinical data and multiphase computed tomography (CT) images with their corresponding masks. We divided this data set randomly into 183 training and validation cases and 50 held-out test cases. Four models were trained: (A) clinical model incorporating demographic and laboratory parameters, (B) radiomic model using PyRadiomics-extracted features, (C) deep neural network (DNN) using multiphase CT images processed with MaskedAttentionViT, and (D) combined clinico-radiological model. Performance was assessed using fivefold cross-validation and testing on a held-out data set to predict a lack of response to TACE.

Results There were 64 (37%) responders and 109 (63%) nonresponders in the training set. There were 13 (26%) responders and 37 (74%) nonresponders in the test set. In the held-out test set, the clinical support vector machine model achieved an accuracy of 70%, sensitivity of 78.9%, specificity of 50%, and area under the curve (AUC) of 0.778 for predicting failure of TACE. The radiomic logistic regression model demonstrated an accuracy of 76.1%, sensitivity of 85.4%, specificity of 18.2%, and AUC of 0.740. The DNN had an accuracy of 63%, sensitivity of 65.7%, specificity of 54.5%, and AUC of 0.601. The combined clinico-radiological model yielded an accuracy of 55.6%, sensitivity of 50%, specificity of 72.7%, and AUC of 0.639.

Conclusion We utilized a multimodal approach to predict response to TACE in HCC patients. Further optimization and multicenter data sets are required to enhance predictive accuracy further.

Keywords

- ▶ computed tomography
- ▶ deep neural network
- ▶ hepatocellular carcinoma
- ▶ machine learning
- ▶ transarterial chemoembolization
- ▶ treatment response

article published online
May 27, 2025

DOI <https://doi.org/10.1055/s-0045-1809372>.
ISSN 0971-3026.

© 2025. Indian Radiological Association. All rights reserved.
This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial-License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)
Thieme Medical and Scientific Publishers Pvt. Ltd., A-12, 2nd Floor, Sector 2, Noida-201301 UP, India

Introduction

Hepatocellular carcinoma (HCC) is the most common primary liver cancer, with an increasing global incidence, particularly in regions with high prevalence of chronic liver disease, such as Asia and sub-Saharan Africa.¹ According to the World Health Organization, liver cancer ranks as the sixth most common cancer and the third leading cause of cancer-related deaths worldwide.² As a result, HCC represents a significant public health challenge. One of the critical issues in managing HCC is the delayed diagnosis, often due to vague or nonspecific complaints in its initial stages. Many patients are diagnosed at an intermediate or advanced stage not amenable to surgical resection or liver transplantation.³ Transarterial chemoembolization (TACE) is a widely used locoregional therapy for treating intermediate and advanced HCC. Despite its efficacy in improving survival rates, TACE is not universally effective, and the response to treatment varies significantly across individuals.^{4,5} Predicting the response to TACE remains a significant challenge, as treatment outcomes depend on a multitude of factors, including tumor characteristics, liver function, and the underlying liver disease.^{6–8} Accurate prediction of the response to TACE could help guide patient selection, optimize treatment planning, and improve patient outcomes. Traditional prediction methods, such as clinical staging systems and imaging evaluation, often fail to provide sufficient accuracy to predict outcomes effectively.^{6–8} This limitation highlights the need for novel approaches incorporating diverse data sources, including clinical, laboratory, and imaging features, to improve predictive accuracy.

Over the past decade, there has been growing interest in using machine learning (ML) techniques to enhance medical imaging and clinical decision-making. ML has shown promise in automating image analysis, identifying subtle patterns that may not be evident to the human eye, and developing predictive models incorporating complex, multidimensional data. In the context of HCC, ML models have been applied to early detection, tumor classification, and treatment response prediction.⁹ Previous studies exploring ML predicting the response to TACE have primarily focused on clinical data, radiomic features, or imaging-based approaches, often with mixed results.^{10–12} While some studies have demonstrated promising outcomes using ML models based on clinical parameters and imaging data, integrating multiple data types and optimizing ML algorithms for better performance remain active areas of research. The current study aims to compare the performance of clinical, radiomics, image-based, and combined ML models for predicting the response to TACE in patients with HCC.

Materials and Methods

The “WAW-TACE: A Hepatocellular Carcinoma Multiphase CT Dataset with Segmentations, Radiomics Features, and Clinical Data” is a publicly available data set designed for research on HCC and its treatment responses.¹³ The data set includes multiphase computed tomography (CT) images of HCC patients who have undergone TACE, along with segmentations

of liver lesions and extraction of radiomic features. In addition to imaging data, the data set provides comprehensive clinical data, including patient demographics, laboratory values, tumor characteristics, and response to TACE. The data set was split into a training set of 183 patients and a held-out test set of 50 patients (► **Table 1**). We used data from the lesion. The response was documented after the first cycle of TACE using the 2017 LR-TR (Liver Imaging Reporting and Data System [LI-RADS] Tumor Response) criteria.¹⁴ We classified tumor response into binary with nonviable tumors assigned to responders and equivocal and viable tumors assigned to non-responders. The ethics committee approval and informed written consent were not applicable to our study as a public data set was used. Below, we discuss the methodology for developing clinical, radiomics, image-based, and combined clinicroadiological models.

Clinical Model

The clinical parameters used for model development included age, gender, lesion number, lesion localization, maximum tumor diameter in the axial plane in the portal venous phase, lesion LI-RADS score, laboratory values (albumin, creatinine, bilirubin, alpha-fetoprotein, international normalized ratio, alanine aminotransferase), Child–Pugh score, Barcelona Clinic Liver Cancer (BCLC) stage, etiology (alcoholic, viral, and other), as well as other scores related to hepatic function (hepatoma arterial embolization prognostic [HAP] score,¹⁵ modified HAP score,¹⁶ Albumin-Bilirubin Transarterial Embolization score,¹⁷ and six-and-twelve score).¹⁸

The training data set was preprocessed as follows: first, the features were standardized to ensure that each feature contributed equally to the model training. Missing values in the clinical features were handled using mean imputation. The training data was then split into an 80% training set and a 20% validation set. Five-model cross-validation was performed. A set of five commonly used ML algorithms was trained and tested for their ability to predict treatment response: random forest (RF), support vector machine (SVM), logistic regression (LR), gradient boosting (GB), and XGBoost (XGB). These models were selected for their ability to handle small and large data sets, interpretability, and general applicability to clinical data. To optimize model hyperparameters, grid search was used with a predefined hyperparameter grid for each model. Once the optimal hyperparameters were identified, the models were trained on the whole training set, and their performance was evaluated on the held-out test set.

Radiomic Models

Radiomic features were extracted from tumor regions using segmentation masks corresponding to the imaging phase in which the tumor was best visualized. Segmentation masks were resampled to match the image size and ensure alignment during feature extraction. The PyRadiomics package (version 3.0.1) was used to extract various features.¹⁹ Features were standardized, and the top 30 features were selected using LASSO (Least Absolute Shrinkage and Selection Operator). RF, SVM, LR, GB, and XGB were trained.

Table 1 Baseline characteristics of the test set (n=50)

ID	Age	Sex	N	Lobe	Dia	LR	Alb	Cr	Bil	AFP	INR	ALT	CPS	D >30	BCIC	Etio	HAP	MHAP	ALBI-TAE	6_12	6_12_score	LR_TR
13	76	0	1	2	42	4	4.2	0.85	0.31	4.44	1.05	24	1	1	0	2	0	0	0	5.2	0	1
19	83	0	2	2	59	5	3.9	1.34	0.55	13	1	56	1	1	1	2	0	1	0	7.9	1	1
26	56	0	1	2	65	5	4.1	0.85	0.63	20	1	63	1	1	0	2	0	0	0	7.5	1	1
31	68	0	2	2	57	5	4.5	0.8	0.98	6.99	1.14	182	1	1	1	2	0	1	0	7.7	1	1
34	56	1	1	1	30	5	3.1	0.91	0.95	8310	1.55	98	1	1	0	2	2	2	2	4	0	1
36	56	0	2	2	99	5	4.1	0.76	1.36	1665	1.15	348	1	1	1	2	3	4	3	11.9	1	1
45	80	1	1	2	36	5	4.5	1.35	1.08	30.58	1.07	27	1	1	0	3	1	1	0	4.6	0	0
50	82	0	1	2	104	5	3.6	0.87	5	73048	1.12	107	2	1	0	1	3	3	3	11.4	1	1
53	46	0	2	1	25	5	3.9	0.91	0.48	638	1.15	189	1	0	1	2	1	2	1	8	1	1
61	64	1	1	2	16	5	4.3	0.56	1.56	1023	1.24	29	1	0	0	2	2	2	1	2.6	0	0
62	79	0	1	2	54	5	4.2	0.99	0.29	4.35	0.92	49	1	1	0	2	0	0	0	6.4	0	0
65	74	0	3	1	97	5	4.6	0.56	0.44	12.9	0.91	84	1	1	1	3	1	2	1	12.7	2	1
66	50	0	1	1	43	5	4.4	0.91	0.69	43.46	1.28	43	1	1	0	2	0	0	0	5.3	0	0
72	59	0	1	1	68	5	3.1	0.72	2.5	22551	1.35	37	2	1	0	1	3	3	2	7.8	1	1
73	62	0	1	1	50	5	3.7	0.88	0.64	2.9	1.03	30	1	1	0	3	0	0	1	6	0	0
74	82	1	1	1	18	5	3.4	0.82	0.29	4.66	1.05	30	1	0	0	3	1	1	1	2.8	0	0
77	69	0	1	2	20	5	3.5	0.72	0.47	2425	1.37	31	2	0	0	2	2	2	2	3	0	1
79	60	0	1	2	105	5	4.1	1.23	0.57	292	1.03	32	1	1	0	1	1	1	2	11.5	1	1
94	61	0	4	2	20	5	3.9	1.04	1.29	4111	1.26	32	1	0	0	1	2	3	2	6.3	0	1
99	65	0	1	2	57	5	3.4	0.61	7	19.43	1.06	40	2	1	0	3	2	2	1	6.7	1	1
111	68	0	1	2	29	5	4.1	0.75	0.46	1.8	1.69	49	1	0	0	3	0	0	0	3.9	0	1
114	69	0	1	2	21	5	4.2	1.04	0.51	4.45	1.03	31	1	0	0	1	0	0	0	3.1	0	1
116	63	0	4	1	35	5	3.7	0.78	1.94	38.7	1.16	285	1	1	1	2	1	2	1	8.3	1	1
118	67	0	2	1	36	5	4	0.92	0.69	4.03	1.23	25	1	1	1	1	0	1	0	5.6	0	1
139	66	0	2	2	27	5	4.2	0.77	0.92	1720	1.01	88	1	0	0	2	1	2	1	4.8	0	1
142	75	0	1	2	91	5	3.5	0.56	0.5	6.7	1.42	28	1	1	0	3	2	2	1	10.1	1	1
144	72	1	2	2	54	5	3.8	0.93	0.83	37.9	1.07	147	1	1	1	2	0	1	1	7.4	1	0
168	61	1	1	2	40	5	3.5	0.75	1.26	10.9	1.34	27	1	1	0	2	2	2	1	5	0	0
179	48	1	3	2	14	5	4.8	0.86	0.28	9.64	0.97	28	1	0	0	2	0	1	0	4.8	0	1

(Continued)

Table 1 (Continued) Baseline characteristics of the test set (n=50)

ID	Age	Sex	N	Lobe	Dia	LR	Alb	Cr	Bil	AFP	INR	ALT	CPS	D >30	BCLC	Etio	HAP	MHAP	ALBI-TAE	6_12	6_12_score	LR_TR
198	85	0	1	1	71	5	4	1	0.5	43	1.15	131	1	1	0	2	1	1	0	8.1	1	1
303	51	0	1	2	23	5	3.2	2.79	5.3	5.16	1.08	32	2	0	0	2	2	2	1	3.3	0	1
319	68	0	1	1	51	5	4.4	0.87	0.46	1.07	1.53	15	1	1	0	2	0	0	0	6.1	0	1
335	62	0	2	2	28	5	4.4	0.99	0.69	7.58	1.3	38	1	0	0	2	0	1	0	4.8	0	1
349	65	0	2	1	47	5	3.8	0.81	1.45	28299	1.39	43	1	1	1	1	2	3	2	6.7	1	1
358	36	0	2	2	20	4	3.9	0.94	0.5	3	1.21	28	1	0	0	2	0	1	0	4	0	0
360	64	0	1	1	58	5	4.5	0.87	1.81	4	1.32	57	1	1	0	1	1	1	0	6.8	1	1
379	51	0	1	1	84	5	2.7	0.57	1.5	5	1.3	270	1	1	0	2	3	3	1	9.4	1	1
439	54	0	1	2	39	5	3.7	1.1	0.8	10	1	23	1	1	0	2	0	0	1	4.9	0	1
442	71	1	1	2	75	5	4.9	0.7	0.4	17	1.11	40	1	1	0	2	1	1	0	8.5	1	1
446	74	0	3	1	97	5	4.8	0.7	0.4	22	1	31	1	1	1	1	1	2	1	12.7	2	1
452	63	0	1	2	59	5	4.3	1	0.7	6700	0.92	90	1	1	0	1	1	1	1	6.9	1	1
453	77	1	2	2	60	5	3.9	1	0.73	3	1	104	1	1	1	2	0	1	1	8	1	0
491	79	1	2	2	44	5	4.4	1	0.6	4	1	18	1	1	1	1	0	1	0	6.4	0	1
493	71	0	1	1	53	5	4.2	1.22	0.5	1811	1	42	1	1	0	1	1	1	1	6.3	0	0
508	45	0	1	1	100	5	3.5	0.85	3.44	3	1.32	81	2	1	0	2	3	3	1	11	1	1
512	74	0	1	2	75	5	4.6	1.31	0.31	166	1	32	1	1	0	1	1	1	0	8.5	1	1
522	65	0	3	2	60	5	4.4	0.93	0.4	6860	1	125	1	1	1	2	1	2	1	9	1	1
525	70	0	1	1	27	5	3	1	1.92	3	1.23	45	1	0	0	1	2	2	1	3.7	0	1
527	83	0	1	2	17	5	4.2	1.32	0.9	1469	1	40	1	0	0	1	1	1	1	2.7	0	0
539	72	1	1	2	43	5	3.8	1.7	0.65	4	1	105	1	1	0	2	0	0	1	5.3	0	1

Note: ID, correspond to PATPRI in the WAW-TACE data set; Sex: 0-male, 1-female; N: number of lesions; Lobe: left-1, right-2; Dia: diameter (mm) of the largest lesion; LR: LI-RADS (Liver Imaging Reporting and Data System); Alb, albumin (g/L); Cr, creatinine (mg/dL); AFP, alpha-fetoprotein (ng/mL); INR, international normalized ratio; ALT, alanine aminotransferase (IU/L); CPS, Child-Pugh score; D >30, lesion diameter greater than 30 mm, 1-yes, 0-no; BCLC, Barcelona Clinic Liver Cancer stage, 0-A, 1-B; Etio, etiology, 1-viral, 2-alcoholic, 3-others; HAP, hepatoma arterial embolization score, albumin < 36 g/dl; AFP > 400 ng/ml; bilirubin > 17 umol/L; MHAP, modified HAP, HAP criteria + Tumor number ≥ 2; ALBI-TAE: albumin bilirubin TAE score, AFP > 200 ng/ml; Up-to-11 criteria; 6_12, six-and-twelve score - largest tumor diameter; Tumor number - continuous; 6_12_score, six-and-twelve score (0/1/2); LR-TR, LI-RADS tumor response, 0-nonviable; 1-equivocal + viable.

Hyperparameter tuning was performed using grid search with fivefold stratified cross-validation.

Image-Based Deep Neural Network

Each image was resized to 224×224 pixels to match the input requirements of the pretrained Vision Transformer (ViT) model (—Fig. 1). The images and masks were transformed into tensors. Data augmentation techniques such as random horizontal and vertical flips were used. A custom neural network model named MaskedAttentionViT was developed, leveraging a pretrained ViT model from the timm library. The original classification head of the ViT model was replaced with a custom head consisting of a dropout layer and a linear classifier. The model takes both the image and the corresponding mask as inputs, applying the mask to the image before feeding it into the ViT model to focus on the regions of interest. The focal loss function and weighted random sampling were employed to address the class imbalance and improve model robustness. Focal loss function adjusts the learning process by focusing more on hard-to-classify samples. Weighted random sampling ensured that each class was equally represented in each batch. The AdamW optimizer was used with a learning rate scheduler to adjust the learning rate based on the validation loss. An early stopping mechanism was implemented to prevent overfitting. Training was terminated if the validation loss did not improve for a specified number of epochs. The training and evaluation were done in a system with Intel (R) Xeon(R) Gold 5218 processor and four Nvidia Tesla V100 32 GB graphics processing unit (GPU).

Combined Model

This model was trained using a hybrid approach integrating clinical data (see above) and imaging data (comprising images with their corresponding masks). We used a custom neural network model named MaskedAttentionViT to extract features from the imaging data. The clinical data was preprocessed by one-hot encoding categorical variables and normalizing quantitative variables. These features were then concatenated with the imaging features extracted by the ViT.

The combined data set was used to train a neural network. The training and evaluation were done in a system with Intel (R) Xeon(R) Gold 5218 processor and four Nvidia Tesla V100 32 GB GPUs.

Statistical Analysis

The performance metrics included accuracy, sensitivity, specificity, and F1 score. Sensitivity and specificity were calculated using the recall values for the positive and negative classes, respectively. Additionally, the receiver operating characteristic (ROC) curve and area under the curve (AUC) were computed to assess the discriminative power of the models. The importance of each model's features was visualized using bar plots and heat maps. To evaluate the model's performance at the patient level, predictions for each patient were averaged across all folds. The final patient-level metrics were calculated using these averaged predictions. The statistical analyses were performed using SciPy 1.1.0 (Austin, Texas, United States).

Results

Baseline Characteristics

The baseline characteristics of all the overall groups are given at <https://pubs.rsna.org/doi/full/10.1148/ryai.240296>.¹³ The median age is 66 (28–86) years. There are 185 (79.4%) males. Most ($n = 149$) patients have single HCC. The baseline characteristics of the test set are given in —Table 1. The median age is 66.5 years (interquartile range 13.75). There are 78% males. Single HCC is present in 64% of the patients.

Responder versus Nonresponders

Overall, there were 77 (33%) responders and 156 (67%) nonresponders. In the training set, there were 64 (37%) responders and 109 (63%) nonresponders. In the test set, there were 13 (26%) responders and 37 (74%) nonresponders.

Clinical Model

The clinical model achieved an average accuracy of 70%, sensitivity of 76.3%, specificity of 50%, and AUC of 0.693.

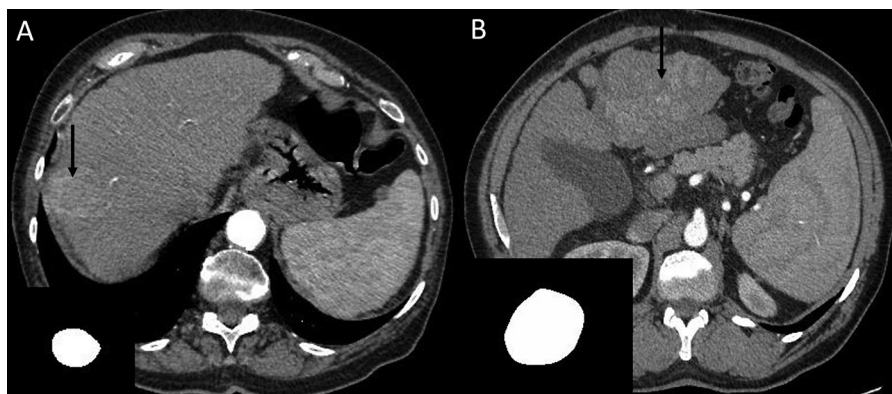


Fig. 1 Representative computed tomography (CT) images from the WAW-TACE data set. (A) Axial arterial phase CT image and the corresponding mask (inset) of a patient who responded to transarterial chemoembolization shows a well-defined arterial phase hyperenhancing mass in segment 8 (arrow). (B) Axial arterial phase CT image and the corresponding mask (inset) of a patient who had a failure to transarterial chemoembolization shows a large ill-defined arterial phase hyperenhancing mass in segment 3 (arrow).

Table 2 Performance of various machine learning models

Model	Accuracy	Sensitivity	Specificity	F1 score	AUC
Clinical					
Random forest	70	78.9	41.6	80	0.692
SVM	72	78.9	50	81	0.778
Logistic regression	58	57.8	58.3	67.6	0.719
XGBoost	68	81.5	25	79.4	0.632
Gradient boosting	66	78.9	25	77.9	0.661
Radiomics					
Random forest	60.8	74.2	18.2	74.2	0.403
SVM	67.3	80.5	0	88.5	0.542
XGBoost	76.1	86.4	0	100	0.496
Gradient boosting	76.1	86.4	0	100	0.488
Logistic regression	76.1	85.7	18.2	94.2	0.742
DNN					
Masked_ViT	63	65.7	54.5	73	0.601
Combined					
Clinical + Masked_ViT	55.5	50	72.2	62.9	0.639

Abbreviation: AUC, area under the curve; DNN, deep neural network; SVM, support vector machines; ViT, Vision Transformer; XGB, extreme gradient boosting.

Among individual algorithms, SVM performed the best, with an accuracy of 72%, sensitivity of 78.9%, specificity of 50%, and AUC of 0.779 (►Table 2).

Radiomic Model

A total of 131 radiomic features were extracted. The radiomic models demonstrated high sensitivity but low specificity. Logistic regression achieved the highest AUC of 0.743, with an accuracy of 76.1%, sensitivity of 85.7%, and specificity of 18.2% (►Table 2).

Image-Based Deep Neural Network

Utilizing the MaskedAttentionViT architecture, the deep learning model achieved a moderate accuracy of 63%, sensitivity of 65.7%, specificity of 54.5%, and an AUC of 0.601 (►Table 2).

Combined Model

The combined model yielded an accuracy of 55.6%, sensitivity of 50%, specificity of 72.7%, and AUC of 0.639 (►Table 2).

►Fig. 2 shows the ROC curves of various models, and ►Fig. 3 shows the feature importance map of the clinical and radiomics models. The top five clinical features contributing to model performance were the six-and-twelve score, tumor diameter, serum albumin, bilirubin, and creatinine. The top five radiomics features were shape and gray level size zone matrix (glszm) features. ►Fig. 4 shows the heat map of a patient where the combined model accurately predicted the response to TACE.

Discussion

Our study explored clinical, radiomics, image-based deep neural network (DNN), and combined models for predicting the failure of the first session of TACE in HCC patients. Our results indicate that different models excelled in distinct performance metrics, highlighting the tradeoffs between sensitivity and specificity across approaches. The clinical model demonstrated reliable predictive capabilities, with SVM emerging as the top performer. Its balanced accuracy and sensitivity suggest that clinical parameters such as tumor characteristics and liver function contribute significantly to predicting TACE failure. Radiomics models, with their high sensitivity, proved effective in identifying non-responders, likely due to the rich quantitative features derived from CT images. Nevertheless, the lack of specificity suggests overfitting to the nonresponder class, limiting their generalizability. The image-based DNNs had moderate accuracy, sensitivity, and specificity. The combined model yielded the best specificity but modest sensitivity. These results suggest the potential of different models in evaluating TACE response, yet highlighting that further research and experimentation with large multicenter data sets is critical to improving the accuracy and generalization further.

Previous studies utilizing CT data to predict response to TACE in HCC have been published. Morshid et al reported that the RF classifier model utilizing the BCLC stage and quantitative CT features performed better (accuracy of 74.2%) than the BCLC stage alone (accuracy of 62.9%). The AUC of the combined model was 0.73.¹⁰ However, this study did not

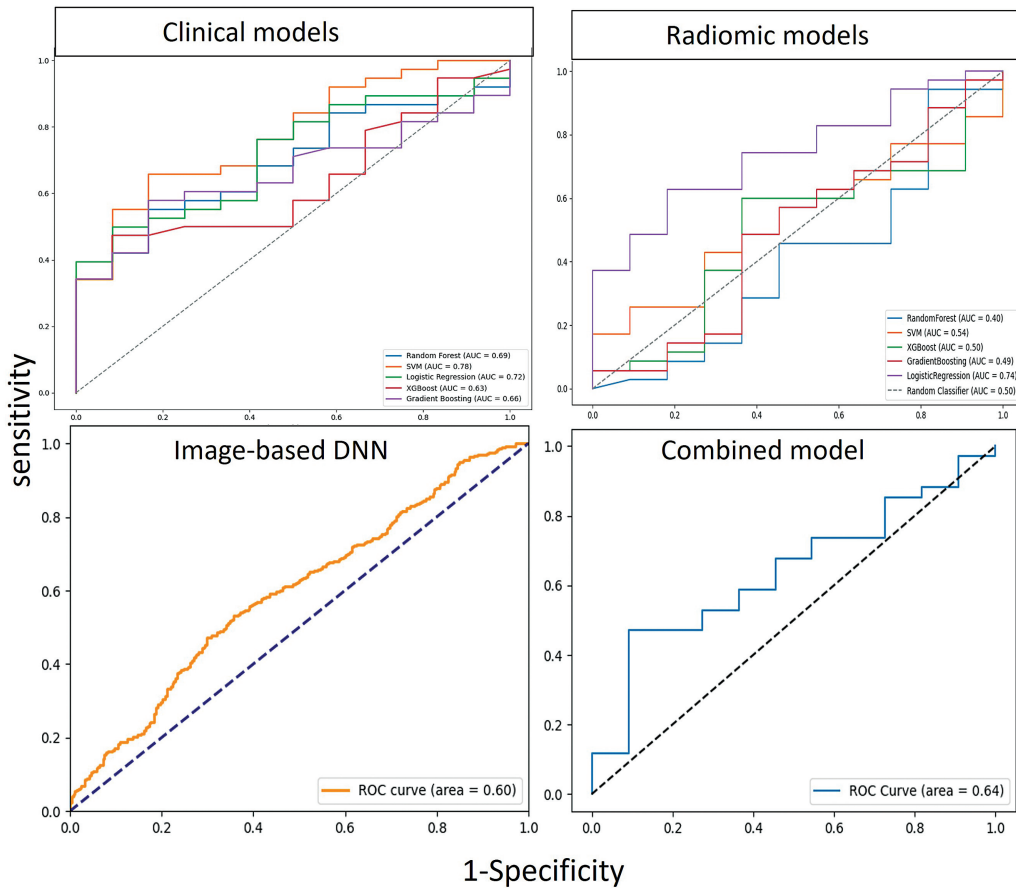


Fig. 2 Receiver operating characteristic curves of different models.

report the detailed metrics and was limited by the majority of the tumors being BCLC C and D compared to the WAW-TACE data set, where all tumors are BCLC A or B. A study by Zhang et al, comprising 110 patients, utilized portal vein tumor thrombosis type, albumin level, and distribution of tumors within the liver, for predictive model building.¹¹ The RF model showed the best performance, with accuracy, sensitivity, specificity, and AUC of 78.4%, 90.4%, 48%, and 0.802, respectively. The authors, however, did not report the performance in a held-out test set. The lower specificity is

similar to our clinical and radiomic models. In another recent study, a combined clinical-CT RF model comprising mean diameter, Eastern Cooperative Oncology Group performance status, cirrhosis, and mean attenuation values of target lesions on multiphase contrast-enhanced CT, arterial, portal venous, and arterial portal venous enhancement ratios had the best performance (sensitivity 75%, specificity 75.4%, and AUC 0.800), for predicting response to TACE.¹² However, to our knowledge, none of the reported studies explored the potential of utilizing multiple clinical parameters (as

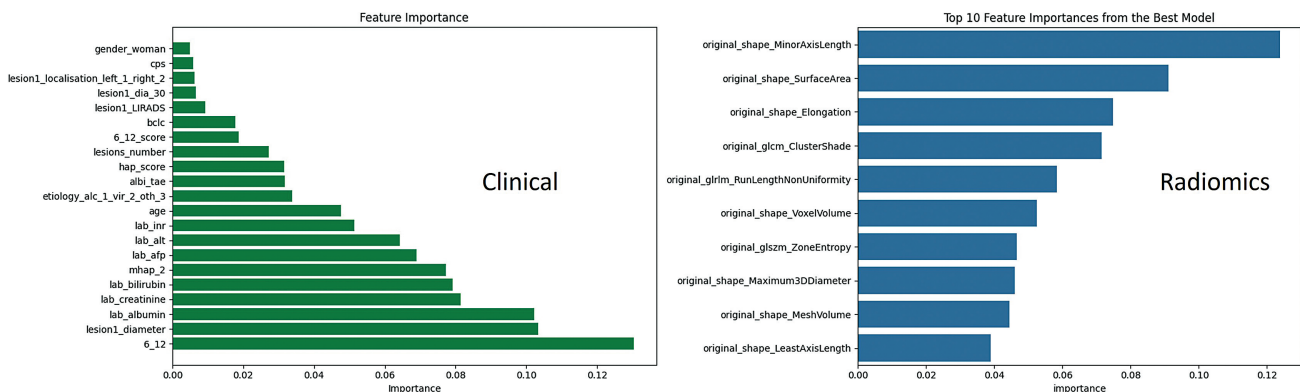


Fig. 3 Feature importance map for the clinical model.

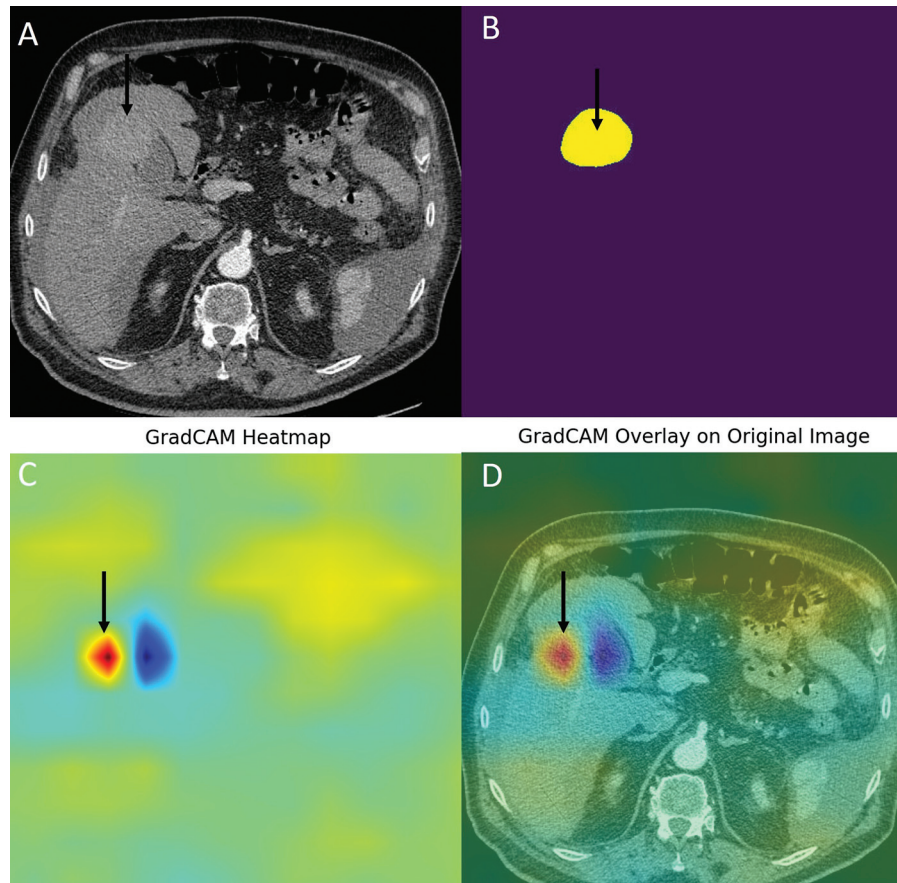


Fig. 4 Gradient class activation map. (A) Axial late arterial phase computed tomography (CT) image shows an arterial phase hyperenhancing lesion (arrow). (B) The segmentation mask is shown (arrow). (C and D) The heat map and overlay images show attention over the tumor (arrows).

reported in the WAW-TACE data set), DNN features, and combining clinical and DNN features. However, our model's moderate performance may reflect data size and heterogeneity limitations. Our approach utilizing the WAW-TACE data set may encourage further research on multicenter data sets that may yield a more realistic performance for response prediction.

There were a few limitations to our study. First, the WAW-TACE data set is a single-center data. Second, the data set is heterogeneous in terms of lesion characteristics and type of CT scanner. Third, the diagnosis of HCC was based on LI-RADS comprising categories LR-4, 5, and M, and histological confirmation was unavailable. Fourth, the treatment response was assessed by a single radiologist based on the LI-RADS treatment response criteria. Finally, as all the contrast phases were not available in all patients, we utilized the images and corresponding masks where the tumor was best visualized, potentially affecting the performance of the image-based models.

In conclusion, we explored a multimodal approach to assess TACE response. However, the models achieved a moderate performance due to the data set limitations. Further research incorporating multicenter large data sets could refine model performance, paving the way for personalized treatment planning in HCC.

Authors' Contributions

N.D.: Methodology, writing - original draft, and writing - review and editing.

P.G.: Conceptualization, methodology, writing - original draft, writing - review and editing, and formal analysis.

Conflict of Interest

None declared.

References

- de Martel C, Georges D, Bray F, Ferlay J, Clifford GM. Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. *Lancet Glob Health* 2020;8(02):e180–e190
- Filho AM, Laversanne M, Ferlay J, et al. The GLOBOCAN 2022 cancer estimates: data sources, methods, and a snapshot of the cancer burden worldwide. *Int J Cancer* 2025;156(07):1336–1346
- Ghouri YA, Mian I, Rowe JH. Review of hepatocellular carcinoma: epidemiology, etiology, and carcinogenesis. *J Carcinog* 2017;16:1
- Sangro B, Salem R. Transarterial chemoembolization and radioembolization. *Semin Liver Dis* 2014;34(04):435–443
- Han K, Kim JH. Transarterial chemoembolization in hepatocellular carcinoma treatment: Barcelona clinic liver cancer staging system. *World J Gastroenterol* 2015;21(36):10327–10335
- Ma W, Jia J, Wang S, et al. The prognostic value of 18F-FDG PET/CT for hepatocellular carcinoma treated with transarterial chemoembolization (TACE). *Theranostics* 2014;4(07):736–744

- 7 Pinato DJ, Sharma R, Allara E, et al. The ALBI grade provides objective hepatic reserve estimation across each BCLC stage of hepatocellular carcinoma. *J Hepatol* 2017;66(02):338–346
- 8 Sieghart W, Huckle F, Pinter M, et al. The ART of decision making: retreatment with transarterial chemoembolization in patients with hepatocellular carcinoma. *Hepatology* 2013;57(06):2261–2273
- 9 Feng S, Wang J, Wang L, et al. Current status and analysis of machine learning in hepatocellular carcinoma. *J Clin Transl Hepatol* 2023;11(05):1184–1191
- 10 Morshid A, Elsayes KM, Khalaf AM, et al. A machine learning model to predict hepatocellular carcinoma response to transcatheter arterial chemoembolization. *Radiol Artif Intell* 2019;1(05):e180021
- 11 Zhang L, Jin Z, Li C, et al. An interpretable machine learning model based on contrast-enhanced CT parameters for predicting treatment response to conventional transarterial chemoembolization in patients with hepatocellular carcinoma. *Radiol Med* 2024;129(03):353–367
- 12 Dong Z, Lin Y, Lin F, et al. Prediction of early treatment response to initial conventional transarterial chemoembolization therapy for hepatocellular carcinoma by machine-learning model based on computed tomography. *J Hepatocell Carcinoma* 2021;8:1473–1484
- 13 Bartnik K, Bartczak T, Krzyziński M, et al. WAW-TACE: a hepatocellular carcinoma multiphase CT dataset with segmentations, radiomics features, and clinical data. *Radiol Artif Intell* 2024;6(06):e240296
- 14 American College of Radiology. Liver Imaging Reporting and Data System version 2018 Manual. 2018. Accessed May 14, 2025 at: <https://www.acr.org/-/media/ACR/Files/Clinical-Resources/LIR-ADS/LI-RADS-2018-Manual-5Dec18.pdf?la=en>
- 15 Kadalayil L, Benini R, Pallan L, et al. A simple prognostic scoring system for patients receiving transarterial embolisation for hepatocellular cancer. *Ann Oncol* 2013;24(10):2565–2570
- 16 Park Y, Kim SU, Kim BK, et al. Addition of tumor multiplicity improves the prognostic performance of the hepatoma arterial-embolization prognostic score. *Liver Int* 2016;36(01):100–107
- 17 Lee IC, Hung YW, Liu CA, et al. A new ALBI-based model to predict survival after transarterial chemoembolization for BCLC stage B hepatocellular carcinoma. *Liver Int* 2019;39(09):1704–1712
- 18 Wang Q, Xia D, Bai W, et al; China HCC-TACE Study Group. Development of a prognostic score for recommended TACE candidates with hepatocellular carcinoma: a multicentre observational study. *J Hepatol* 2019;70(05):893–903
- 19 van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res* 2017;77(21):e104–e107