Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

Research article

Deep learning of pretreatment multiphase CT images for predicting response to lenvatinib and immune checkpoint inhibitors in unresectable hepatocellular carcinoma

Nan-Qing Liao [a,b], Zhu-Jian Deng [b], Wei Wei [c], Jia-Hui Lu [d], Min-Jun Li [b], Liang Ma [b], Qing-Feng Chen [d,*], Jian-Hong Zhong [b,**,1]

[a] School of Medical, Guangxi University, Nanning, China
[b] Hepatobiliary Surgery Department, Guangxi Medical University Cancer Hospital, Nanning, China
[c] Radiology Department, Guangxi Medical University Cancer Hospital, Nanning, China
[d] School of Computer, Electronics and Information, Guangxi University, Nanning, China

ARTICLE INFO

ABSTRACT

Objectives: Combination therapy of lenvatinib and immune checkpoint inhibitors (CLICI) has emerged as a promising approach for managing unresectable hepatocellular carcinoma (HCC). However, the response to such treatment is observed in only a subset of patients, underscoring the pressing need for reliable methods to identify potential responders.

Materials & methods: This was a retrospective analysis involving 120 patients with unresectable HCC. They were divided into training (n = 72) and validation (n = 48) cohorts. We developed an interpretable deep learning model using multiphase computed tomography (CT) images to predict whether patients will respond or not to CLICI treatment, based on the Response Evaluation Criteria in Solid Tumors, version 1.1 (RECIST v1.1). We evaluated the models' performance and analyzed the impact of each CT phase. Critical regions influencing predictions were identified and visualized through heatmaps.

Results: The multiphase model outperformed the best biphase and uniphase models, achieving an area under the curve (AUC) of 0.802 (95% CI = 0.780–0.824). The portal phase images were found to significantly enhance the model's predictive accuracy. Heatmaps identified six critical features influencing treatment response, offering valuable insights to clinicians. Additionally, we have made this model accessible via a web server at http://uhcc net.com/ for ease of use.

Conclusions: The integration of multiphase CT images with deep learning-generated heatmaps for predicting treatment response provides a robust and practical tool for guiding CLICI therapy in patients with unresectable HCC.

1. Introduction

Combined treatment with lenvatinib and immune checkpoint inhibitors (CLICI) improve the survival in unresectable hepatocellular carcinoma (HCC). [1,2] Despite these encouraging outcomes, the benefit of CLICI therapy extends to only a subset of patients, underscoring the need for early identification of likely responders. While imaging modalities such as CT, MRI, and PET are crucial in assessing responses to CLICI by measuring tumor size, morphology, and metabolic activity, the intricacies of tumor biology and treatment effects, like pseudo-progression in immunotherapy, complicate response assessment. Current tumor response criteria, mainly RECIST and WHO (World Health Organization Criteria for Tumor Response), focus on tumor size and may not adequately capture the effects of treatments, nor account for HCC heterogeneity and imaging complexity, leading to challenges in accurate prediction of therapeutic outcomes. [3] Thus, developing and validating a novel method for early prediction of CLICI treatment responses is imperative for identifying patients who stand to gain the most,

offering significant clinical value.

The application of analytical methods for predicting cancer outcomes using clinical imaging has significantly increased in interest, especially with the incorporation of artificial intelligence technologies. Convolutional Neural Networks (CNNs), a branch of artificial intelligence technology that excels at identifying visual patterns in images, have demonstrated their effectiveness through successful deployments in multiple oncology fields. [4–6] Studies utilizing CNNs to analyze CT data have underscored the efficacy of deep learning in facilitating cancer diagnosis, [7] subtype classification, [8] and metastasis. [9] More recent investigations have expanded the scope of deep learning's applicability, demonstrating its capacity to predict therapeutic outcomes across a range of treatments, including chemotherapy and chemoradiotherapy, [10,11] targeted therapy, [12] and immunotherapy. [13,14] Despite these advancements, the specific application of deep learning for predicting responses to CLICI in patients with HCC has yet to be explored, representing a significant gap in the current research landscape.

Several studies have reported the use of CT imaging and deep learning technologies for clinical applications in HCC, such as diagnosis and grading. [15,16] owever, these studies have not utilized multiphase CT images. Multiphase CT imaging provides a more comprehensive analysis by capturing liver cancer from various perspectives, thereby enabling the extraction of intricate features that can highlight subtle variations. Such detailed analysis holds the potential to predict responses to CLICI therapy in patients with unresectable HCC. Furthermore, while several strategies [17–20] have been explored to enhance the predictive capabilities of deep learning models, the challenge of interpretability remains a significant barrier to their broader clinical adoption. An interpretable model is more meaningful to the user than just solely providing model prediction as this allows clinicians to better comprehend and evaluate the predictive process of the model.

In this study, we designed a deep-learning model through multi-input convolutional network architecture aimed at predicting the response to CLICI therapy using pre-treatment multiphase CT images. Furthermore, we explored the contribution of each CT phase to the prediction accuracy through the attentional mechanism. Additionally, we generated characteristic heat maps correlated with the model's predictions. These heat maps enable clinicians to visually identify and concentrate on relevant areas of interest, thereby enhancing their understanding of the model's predictive process and its potential clinical utility.

2. Materials and methods

2.1. Patients

This retrospective study was approved by the ethical committee of Guangxi Medical University Cancer Hospital (LW2023054). Patient consent was waived due to the retrospective nature of the study.

This study retrospectively enrolled consecutive patients who received lenvatinib plus immune checkpoint inhibitors (ICIs) treatment for patients with unresectable HCC between January 2019 and July 2022 at Guangxi Medical University Cancer Hospital, Nanning, China. The inclusion criteria were as follows: (1) patients with unresectable HCC who were not suitable or refused to receive hepatic resection, ablation, radiotherapy, or transarterial therapy (patients who have previously received the above treatments meet the inclusion criteria), (2) received lenvatinib plus ICIs therapy, (3) patients who had at least one measurable lesion as defined by revised RECIST v1.1, [21] (4) patients with Child–Pugh class A or B liver function, an Eastern Cooperative Oncology Group (ECOG) performance status of 0 or 1, and adequate organ function at the time of receiving lenvatinib plus ICIs therapy, (5) patients who had a dynamic enhanced CT imaging study performed within one month prior to initiation of lenvatinib plus ICIs therapy and at least one month after initiation of lenvatinib plus ICIs to evaluate the treatment response, and (6) patients with an observation period of at

least two months. Patients with incomplete medical information, lost to follow-up, a treatment duration of lenvatinib less than one month, or without an enhanced CT scan at baseline were excluded (Fig. 1).

Treatment responses were categorized into four groups as per RECIST v1.1: complete response (CR), partial response (PR), stable disease (SD), and progressive disease (PD). Leveraging HCC specialists' insights, specific criteria were established to categorize patients using the ORR. Accordingly, patients exhibiting CR or PR were deemed responders, whereas those with SD or PD are considered non-responders. In cases involving multiple lesions, the largest intrahepatic tumor is designated as the "target lesion," while the remaining tumors are classified as "non-target lesions." In the evaluation process, CR and PR are primarily assessed by examining the "target lesion." However, the determination of PD involves a comprehensive evaluation of both intrahepatic and extrahepatic lesions. PD is confirmed if any new lesion emerges, or if the "target lesion" exhibits an increase of 20% or more, regardless of the lesion's location.

2.2. Therapeutic regimens

Patients with unresectable HCC were administered oral lenvatinib (Levima®, Eisai, Tokyo, Japan) at a dose of 8 mg/day for patients < 60 kg or 12 mg/day for those ≥ 60 kg plus ICIs therapy. In this retrospective study, ICIs included tislelizumab (BGB-A317, BeiGene), camrelizumab (SHR-1210, Jiangsu HengRui Medicine Co., Ltd.), sintilimab (IBI308, Innovent Biologics [Suzhou] Co. Ltd.), and toripalimab (triprizumab, JS001, Shanghai Junshi Biosciences Co., Ltd.), with dose and duration according to the manufacturer's guidelines. All patients with positive hepatitis B antigen or hepatitis B virus DNA received tenofovir or entecavir antiviral therapy. [22,23].

2.3. CT methods

The CT acquisition parameters are presented as follows: the acquisition encompassed three distinct phases for preprocessing liver CT scans—namely, the plain scan phase, the arterial phase, and the portal phase. The arterial phase was initiated approximately 90 s following the plain scan, while the portal phase commenced 60 s subsequent to the arterial phase. The administration of the contrast agent, iodixanol (300 mgI/mL), occurred immediately after the plain scan, with a dosage of 1.5 mL/kg. This was injected into the antecubital vein using a high-pressure injector, maintaining a flow rate of 3.5 mL/s. CT scans were performed for all study participants within one month prior to the initiation of CLICI treatment, utilizing Siemens Somaton Sensation 64-slice MSCT and GE Discovery 750 HD CT scanners (Scanning parameters: tube voltage 120 kV, tube current 250 mAs, layer thickness 5 mm, layer spacing 5 mm, pitch 0.8, matrix 512 ×512). CT images were acquired while patients maintained an inspiration breath-hold following the contrast injection. All scans adhered to the facilities' established CT chest protocols and utilized standard image reconstruction techniques. Only the axial-phase images were used for subsequent analysis.

2.4. Preprocessing of CT images

In the development of our model, three phases of CT image slices that encompassed the entire liver were identified as regions of interest (ROI). All CT image slices were size of 224 × 224. Considering the range of values across all CT image slices extended from − 1024 to 2000, a comprehensive normalization process was imperative. The normalization was conducted using the following steps: a common window width of 400 and window bit of 0 were selected. Then, truncation normalization was adopted, where the window width minus the window position (minimum and maximum) of the ROI was used to truncate the gray range, and the resulting values were normalized to [0,1].
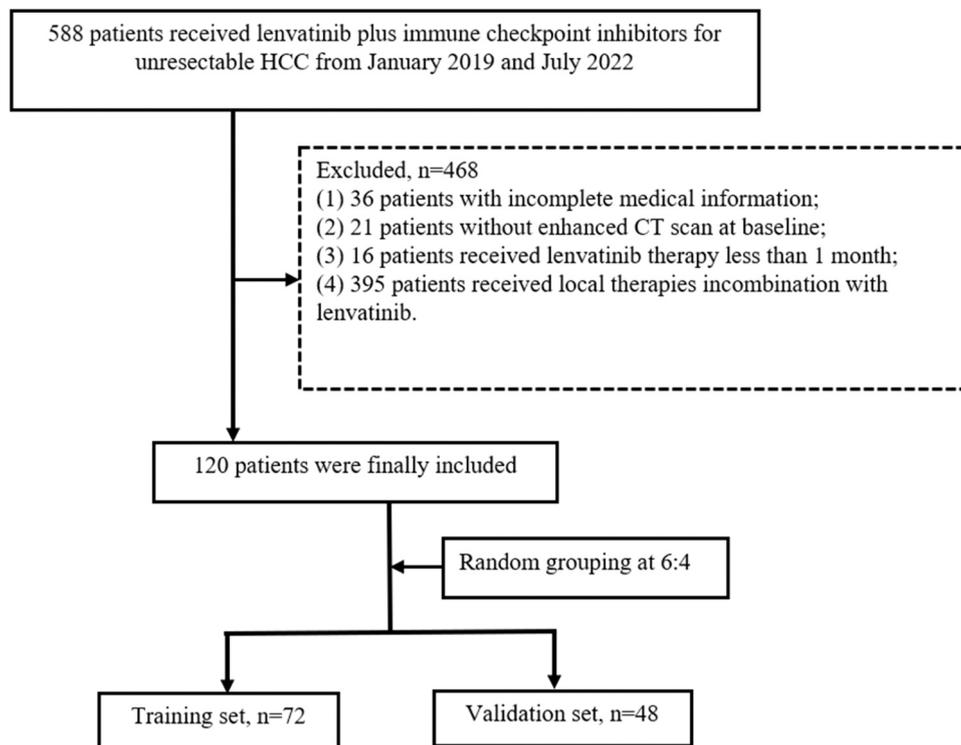
Fig. 1. Flowchart of the selected study patients.

2.5. Model development

Fig. 2a illustrates the detailed architecture of a multi-pathway deep learning network, engineered for the analysis of multiphase CT images. Initially, the network individually processes each imaging sequence. The ResNet-18 model, pre-trained and fine-tuned via transfer learning from the ImageNet dataset, forms the basis. It is then enhanced with several convolutional layers equipped with BatchNorm and ReLU activations. At each pathway's end, 1×1 two-dimensional average pooling layers compile spatial features. These features are integrated across pathways to create a comprehensive multiphase feature vector. To emphasize crucial features, the combined output feeds into a self-attention layer, facilitating dynamic interaction among inputs to prioritize key features within each phase. The network predicts the CLICI response using two fully connected layers and a softmax function, aiming to minimize softmax cross-entropy loss. Dropout at a probability of 0.5 is applied across the fully connected layers to mitigate overfitting.

Our models were developed using the PyCharm (version 2022.1.3) deep-learning framework on a computer equipped with dual Intel Xeon CPUs (Gold 6230, 2.1 GHz) and Tesla T4 GPUs for training and validation. We utilized an ADAM optimizer with a batch size of 20 and an initial learning rate of 0.0001, which was halved every 50 epochs. Training was capped at 13,000 iterations, employing early stopping based on loss and accuracy improvement to halt the process. To enhance model robustness, we augmented the training data with Gaussian noise, blur, brightness and contrast adjustments, and image rotations ranging from $-20°$ to $20°$, all chosen randomly within specific ranges. Considering the unique phases of computed CT imaging, specifically the plain scan phase, arterial phase, and portal vein phase, along with their integrated combinations, we developed models across uniphase, biphase, and multiphase frameworks. These deep-learning models were trained to autonomously analyze CT images and determine the corresponding response.

2.6. Predicting patient in distinct therapeutic response groups through model

For each patient, ROI slices were analyzed across multiphase consecutive slices, extracting features from each slice. Each slice was treated as an individual sample and fed into the model to predict a label. We employed a decision-making approach known as the B-voting strategy to categorize subjects into appropriate groups, based on a majority vote among the labels of CT slices per subject. Consequently, a patient's subgroup was determined by the predominant label across their ROI slices. For example, if the ROI slices of a patient received labels 0, 1, 1, 1, 0 (where 1 indicates a response and 0 indicates no response), the B-voting strategy would classify the patient as 1, suggesting a predicted positive response to CLICI. This workflow is succinctly outlined in Fig. 2b.

2.7. Interpretability of the deep learning models

To enhance clinicians' insights into the model's rationale for its response predictions, we employed a gradient-weighted class activation mapping (Grad-CAM) approach [24] as a method to produce a heatmap from the final convolutional layer of a test image. This method enables the generation of a heatmap that aligns in size with the convolutional feature map, thereby facilitating the visualization of the model's decision-making process. We adjusted the grid size to correspond with the original dimensions of the CT image and then carefully superimposed the relevant image patches, providing a detailed and intuitive depiction of the areas that influence the model's predictions.

2.8. Statistical analysis

The efficacy of the models' classification performance was appraised through the computation of AUC values, with comparisons drawn utilizing Delong's test accompanied by binomial exact confidence intervals. To elucidate the impact of each phase image on the models, key metrics including accuracy, precision (positive predictive value), recall
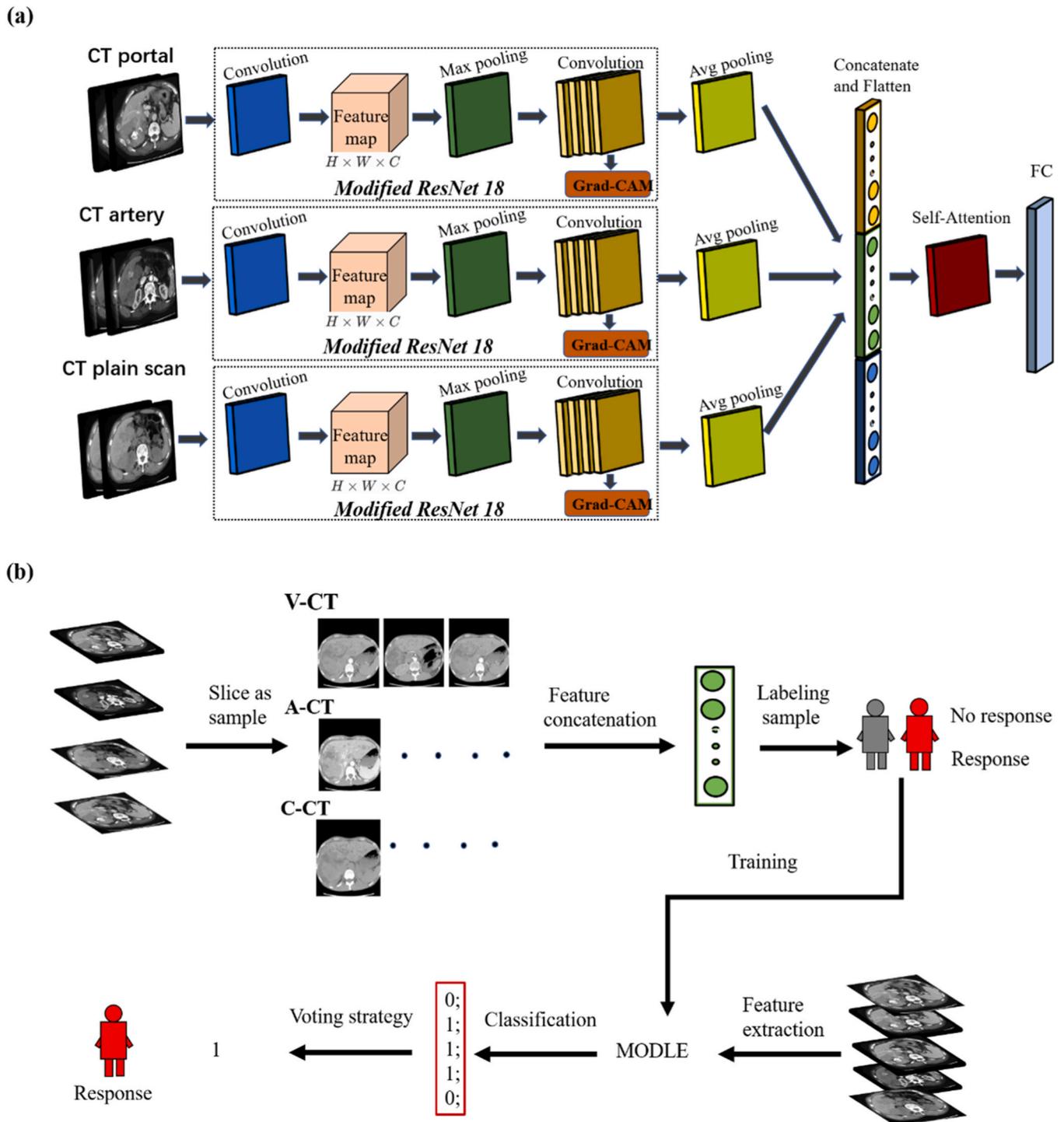
Fig. 2. (a)The schematic deep-learning architecture of multi-input convolutional network is shown. (b) The flowchart displays the process of prediction, where features were extracted for each CT slice.

(sensitivity), and the F1-score (the harmonic mean of recall and precision) were determined for each model. For the estimation of confidence intervals pertaining to sensitivity and specificity, data underwent 1000 bootstrap resampling procedures. A threshold of $P < 0.05$ was established to denote statistical significance. All statistical analyses were conducted using standard statistical software (IBM SPSS, version 26.0).

2.9. Model comparison

Given the effectiveness of pre-trained models trained on large datasets in various medical image classification tasks, we used popular CNN models such as MobileNetV3, VGG19, ResNet34, and ResNet50 to compare their performance. These baseline models, pre-trained with ImageNet dataset parameters through transfer learning, were meticulously fine-tuned on a multiphase CT image dataset, specifically designed for three-input classification tasks. Parameter optimization for these baseline models was performed using grid search, and they were trained using cross-entropy loss function and Adam optimizer. To prevent overfitting, the training employs early stopping, which halts the process based on the lack of improvement in loss and accuracy observed

on the validation set.

3. Results

3.1. Patients' characteristics

Patients within the development dataset were randomly allocated to either the training or validation set, as depicted in (Fig. 1). Deep learning models aimed at predicting the response to CLICI in patients with unresectable HCC were crafted using a dataset comprising 10,815 CT images from 120 patients who underwent live CT scans. The demographic and clinical characteristics of these patients are detailed in Table 1, revealing no significant disparities between the training and validation cohorts. The ORR (responders), based on RECIST 1.1, were observed at 27.8% in the training set and 27.1% in the validation set, respectively.

3.2. Performance of deep learning models

The multi-pathway deep CNN was meticulously designed, incorporating a convolution operator that enabled the network to extract informative features within each pathway (Fig. 2). Using the Grad-CAM technique, features from the terminal convolutional layer of each pathway were harnessed to create a heatmap for each phase image. To identify the most effective base CNN architecture, we evaluated the performance of ResNet-18 in comparison with other prevalent CNN models such as VGG19, [25] ResNet-50 [26] and Mobilenetv3, [27] using only uniphase data (the portal phase). As shown in Fig. 3a, ResNet-18 achieved the best performance with less complexity.

We introduced a self-attention-based multiway model to fine-tune the weighting of live image features. To rigorously test its effectiveness, we carried out ablation studies comparing models with and without the self-attention mechanism. This involved evaluating both our multiphase model and the *CV* model (our highest-performing biphase model trained on datasets including plain scan and portal phase images) against each other in terms of AUC on the validation set. As depicted in Fig. 3b, incorporating self-attention significantly enhanced model performance, yielding an AUC greater than 0.70—noticeably superior to the 0.60–0.70 AUC of models lacking this feature. On average, models with self-attention outperformed their counterparts by a margin of 9.2%

Table 1
Patient baseline demographic and clinical characteristics.

Variables	Training set, n = 72 (%)	Validation set, n = 48 (%)	P value
Age, median, yrs	48 (26-72)	47 (37-63)	0.824
Female	11 (15.2)	10 (20.8)	0.547
Hepatitis B surface antigen, positive	62 (86.1)	38 (79.2)	0.273
α-fetoprotein, ≥ 200 ng/mL	42 (58.3)	30 (62.5)	0.710
Platelet, ≥ 100 × 10⁹/L	57 (79.2)	40 (83.3)	0.780
Child-Pugh grade			0.838
A	52 (72.2)	34 (70.8)	
B	20 (27.8)	14 (29.2)	
Tumor size			0.780
< 5 cm	30 (41.6)	18(37.5)	
≥ 5 cm	42 (58.4)	30 (62.5)	
Tumor number			0.386
Single	23 (31.9)	20 (41.7)	
Multiple	49 (68.1)	28 (58.3)	
Macrovascular invasion, present	50 (69.4)	34 (70.8)	0.843
Extrahepatic metastases, present	32 (44.4)	18 (37.5)	0.646
Tumor responses			0.572
CR	1 (1.4)	0 (0)	
PR	19 (26.4)	13 (27.1)	
SD	29 (40.3)	20(41.7)	
PD	23 (31.9)	15 (31.2)	
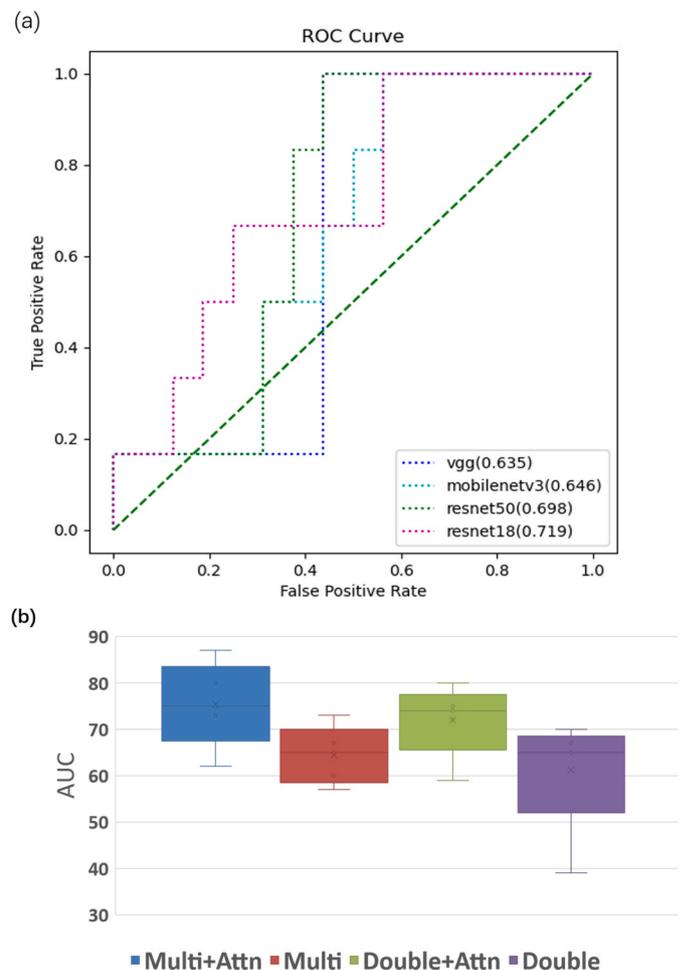
Fig. 3. (a)A total of four deep convolutional neural networks were further compared, including VGG19, Mobilenetv3, ResNet50, ResNet18. (b)AUC distribution for self- attention and no self- attention in multiphase model and the biphase model. The horizontal line represents the median AUC, and the boxes represent the first and third quartiles.

in AUC.

Table 2 and Fig. 4 illustrate the performance outcomes. Our multiphase model, incorporating plain scan (*C*), arterial (*A*), and portal (*V*) phases, demonstrated the highest prediction accuracy for CLICI response in the validation set, achieving an AUC of 0.802. To assess comparative predictive performance, we also developed several biphase models with varying phase combinations. The ROC curves displayed in Fig. 4a for these models reveal that the CV-based model attained an AUC of 0.760, the AV model reached 0.740, and the AC model, showing the lowest efficacy among the biphase models, registered an AUC of 0.719.

3.3. Phase importance

As depicted in Fig. 4, we assessed the impact of individual phases on model performance. The *V* phase emerged as the most effective, achieving superior metrics (AUC = 0.719, ACC = 0.671, Precision = 0.823, Recall = 0.705) compared to the *A* and *C* models (AUCs = 0.708 and 0.688, ACCs = 0.637 and 0.640, Precisions = 0.786 and 0.743, Recalls = 0.697 and 0.696, respectively). Among dual-phase models, combinations including the *V* phase (*VA* and *VC*) demonstrated enhanced performance over the *AC* model, which lacks the *V* phase. Specifically, *VA* excelled in all evaluated metrics (AUC = 0.740, Precision = 0.779, Recall = 0.818, F1 = 0.786), while *VC* surpassed *AC* in three metrics (AUC = 0.760, Recall = 0.762, F1 = 0.775). Removing the *V* phase significantly impacted the models, with the AUC of *CV* dropping

Table 2

The performance of model in the training set and validation set.

Model	Training set (n = 72)				Validation set (n = 48)			
	AUC (95%CI)	Accuracy	Sensitivity	Specificity	AUC (95%CI)	Accuracy	Sensitivity	Specificity
V	0.923 (0.864–0.985)	0.929	0.880	0.929	0.719 (0.620–0.818)	0.671	0.706	0.794
C	0.874 (0.830-0.918)	0.894	0.867	0.894	0.688 (0.633-0.743)	0.640	0.692	0.777
A	0.908 (0.843-0.973)	0.910	0.890	0.910	0.708 (0.626-0.790)	0.637	0.698	0.650
AV	0.937 (0.890-0.984)	0.926	0.871	0.926	0.740 (0.675-0.805)	0.682	0.810	0.816
AC	0.931 (0.901-0.961)	0.930	0.866	0.930	0.719 (0.684-0.754)	0.689	0.715	0.742
CV	0.919 (0.887-0.951)	0.928	0.901	0.928	0.760 (0.716-0.804)	0.664	0.760	0.780
Multi	0.956 (0.931-0.981)	0.941	0.899	0.941	0.802 (0.753-0.851)	0.725	0.829	0.803

(Model V: portal phase, Model C: plain scan phase, Model A: arterial phase, Model AV: arterial phase + portal phase, Model AC: arterial phase +plain scan phase, Model CV: plain scan phase +portal phase, Model Multi: portal phase+ arterial phase +plain scan phase)
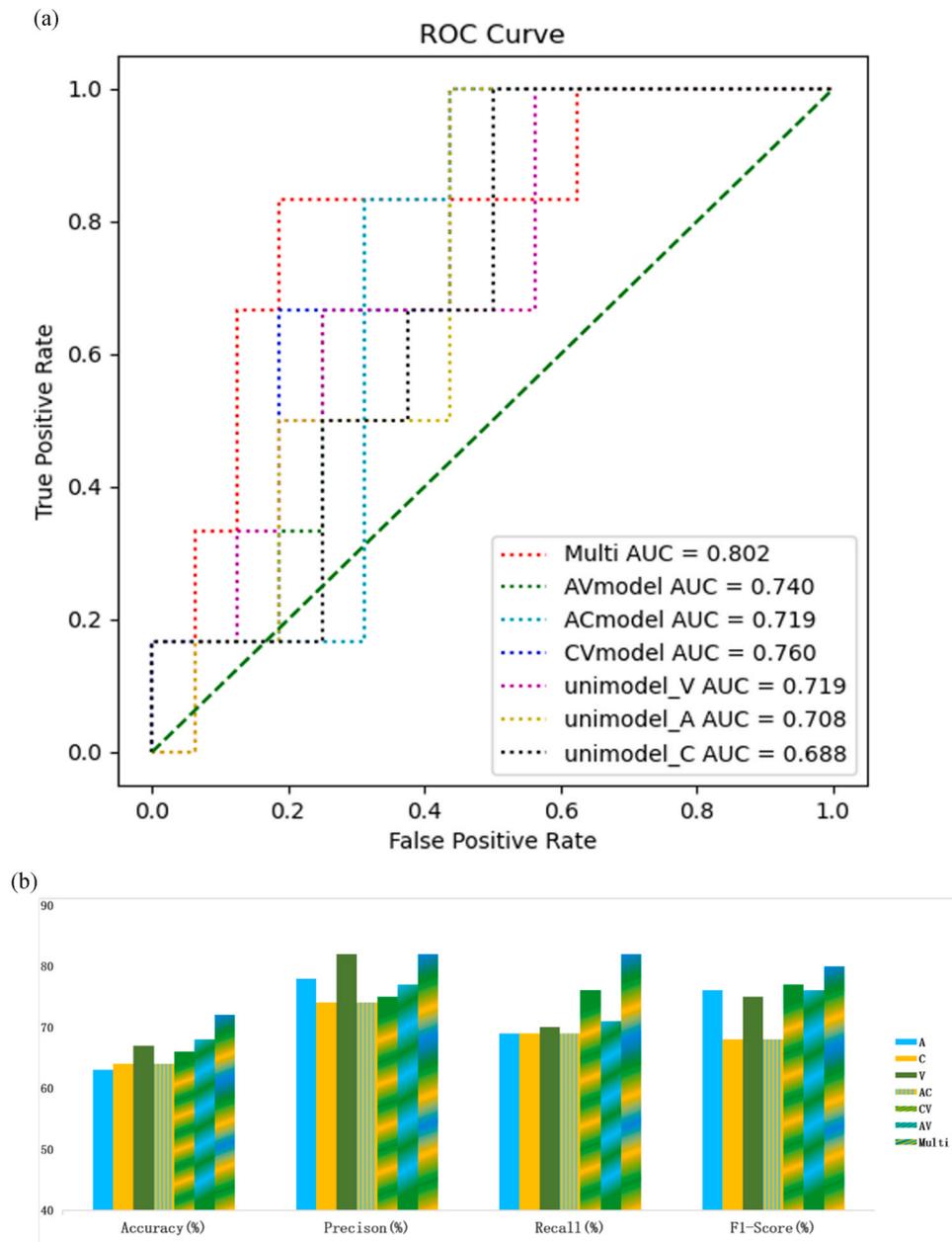
Fig. 4. (a)The ROC curve of the three models is presented for the validation set. (b)Shows all 4 evaluation metrics (accuracy, precision recall, and F1-score) for the possible combinations of phases.

from 0.760 to 0.688 and AV from 0.802 to 0.719.

3.4. Heatmaps for understanding deep-learning models results- predicting

We introduced heatmaps into our analysis to spotlight regions crucial for prediction accuracy, as depicted in Fig. 5a. These visual guides aid in understanding the ROI across different CT imaging phases, highlighting their potential clinical importance (Fig. 5b-m). In the arterial phase images, four key areas emerged as significant predictors for CLICI response in unresectable HCC: necrosis, tumor heterogeneity, intrahepatic multifocal tumors, and areas adjacent to transarterial chemoembolization (TACE) iodide. For plain scan phase images, the heatmaps showcased vasculature, tumor, and peritumoral regions, with a specific focus away from intrahepatic multifocal tumors and areas adjacent to TACE iodide. When analyzing portal phase images, the model similarly identified tumor and peritumoral regions, tumor heterogeneity, intrahepatic multifocal tumors, and regions near TACE iodide as crucial for predicting CLICI response.

3.5. Comparison with other baseline models

During the model's initial development phase, to find the most effective CNN architecture, we evaluated ResNet-18 against other popular CNN models using only the uniphase (portal phase). To delve deeper into our model's performance, we compared the top-performing multiphase model (ResNet18) with other baseline models (Mobilenetv3, VGG19, ResNet34, and ResNet50) using three phases of CT images on the validation set. The metrics of ResNet18 on the validation set were 0.802(0.753–0.851) (AUC (95%CI)), 0.725(*Accuracy*), 0.829(*Sensitivity*), 0.803(*Specificity*). The multiphase model (ResNet18) surpassed all other baseline models in performance metrics, demonstrating its superior robustness. (Table 3).

3.6. Development of a webserver platform for models

We've developed a user-friendly web platform for predicting therapeutic responses in CT images, available at http://uhccnet.com/. Built with Vue3 for the frontend and Django for the backend, this server supports our trained models, enabling the creation of sophisticated analytical applications. It's tailored for the analysis of CT images to forecast responses to CLICI treatment in unresectable HCC. Users can upload DICOM-formatted CT images across three phases. With a simple click on the 'Submit' button, the system activates the models to process the data. It then provides predictions of response or no response by integrating and analyzing input from multiphase CT images (Fig. 6).

4. Discussion

This study leveraged CT images from three distinct phases—plain scan, arterial, and portal—to enhance liver feature extraction and develop a robust prediction model. Our approach encompassed uniphase, biphase, and multiphase models, tailored for potential clinical applicability. Our prediction model effectively integrates these three CT phases to anticipate the CLICI response in unresectable HCC before treatment, showing exceptional performance on validation set. The multiphase model, in particular, significantly outperformed the uniphase and biphase models, underscoring the clinical utility of multiphase CT images in predicting unresectable HCC's response to CLICI treatment. Furthermore, the findings suggest that the *V* phase is a crucial catalyst for predicting CLICI response. We posit that this empirical advantage arises from the *V* phase's ability to convey more extensive information about HCC and its environment relative to the *C* and *A* phases, which enhances the model capacity to represent and interpret data effectively. Heatmaps, providing explainability, revealed six key features crucial for predicting CLICI response in unresectable HCC, including necrosis, vasculature, tumor heterogeneity, tumor and

peritumoral regions, intrahepatic multifocal tumors, and tumor regions surrounding TACE iodide. These results underscore deep learning's ability to capture the complex spatial variability of tumors and the impact of the tumor microenvironment on immune sensitization. By providing such detailed insights, deep learning enhances its clinical relevance, helping clinicians identify key areas of interest and thereby facilitating more informed clinical decisions.

The CLICI has shown increasing efficacy in treating unresectable HCC. [1,2,28] Nevertheless, the necessity for a reliable, non-invasive method to forecast treatment effectiveness remains crucial for selecting appropriate candidates. While MRI is esteemed for its soft tissue resolution in assessing HCC, its limitations—cost, speed, availability, and patient contraindications—render CT scans a more accessible, faster alternative. [29] Multiphase CT images, capturing diverse liver changes, allow for a comprehensive feature extraction, underlining the potential of deep-learning models based on these images for clinical use.

Many previous studies have employed prior knowledge as attention to guide feature extraction. [30] And when it comes to live CT images, it is unreasonable to extract information from each phase of CT images equally, as this inevitably reduces the attention directed towards the diseased region. Our study employed a multi-pathway deep learning approach, incorporating a self-attention mechanism to refine feature discrimination across multiphase models without the need for manual segmentation. Demonstrating effectiveness, with AUC surpassing 0.802, our methodology marks a significant advancement in forecasting responses to CLICI treatment, similar to a prior study predicting treatment response in gastric cancer using deep learning features. [31] This also aligns with findings in some studies where deep learning prediction models outperform handcrafted signature models in terms of identification capabilities. [32] On the other hand, we conducted a performance comparison between the multiphase model and other baseline models to demonstrate the superior reliability and robustness of our model.

Our evaluation explored each CT phase's importance in predicting therapeutic responses, aiming to clarify the model's decision-making process. Notably, the *V* phase is the key to significantly enhance the model performances. While dual-phase applications generally improved results over single-phase models, the standalone *V* phase model's performance matched that of combining *C* and *A* phases. The *V* phase model's precision and recall not only exceeded those of single-phase models but also outperformed *CV* and *AC* biphase configurations.

Radiologists believe that, the *V* phase offers a more precise observation of HCC when compared to the *A* and *C* phases. This superior accuracy stems from the tendency of *C* phase images to be compromised by fatty liver, which can obscure lesion visibility. Although most HCC lesions are significantly enhanced during the *A* phase, it may fail to highlight a small subset of lesions as effectively. Additionally, the presence of hepatic cirrhosis nodules and uneven blood perfusion can affect the observation of lesions in *A* phase. Unlike *A* and *C* phases, *V* phase, on the other hand, shows HCC lesions with a relatively lower density compared to normal liver parenchyma, reducing the impact of fatty liver and uneven blood perfusion. This phase also allows for the identification of more lesions relative to both the *A* phase and *C* phase. [33] Currently, the *V* phase is valuable in the histological grading of HCC [34] and the recognition of HCC differentiation levels. [35] Studies have revealed that *V* phase can reveal evidence of liver disease in most patients [36] and assess the effectiveness of TACE treatment. [37] Moreover, compared to the *A* phase, the *V* phase can more accurately predict the survival rates of patients undergoing surgical resection for HCC [38] and those with advanced HCC receiving sorafenib treatment. [39] The above findings indicate that *V* phase has a superior advantage in reflecting the imaging information of HCC patients, thereby making a critical contribution to the development of models.

The inherent opacity of deep learning models presents a challenge in clinical trust. [40] Research indicates that models offering interpretability not only boost clinicians' confidence in diagnosis but also provide valuable educational feedback, benefiting those less specialized, such as
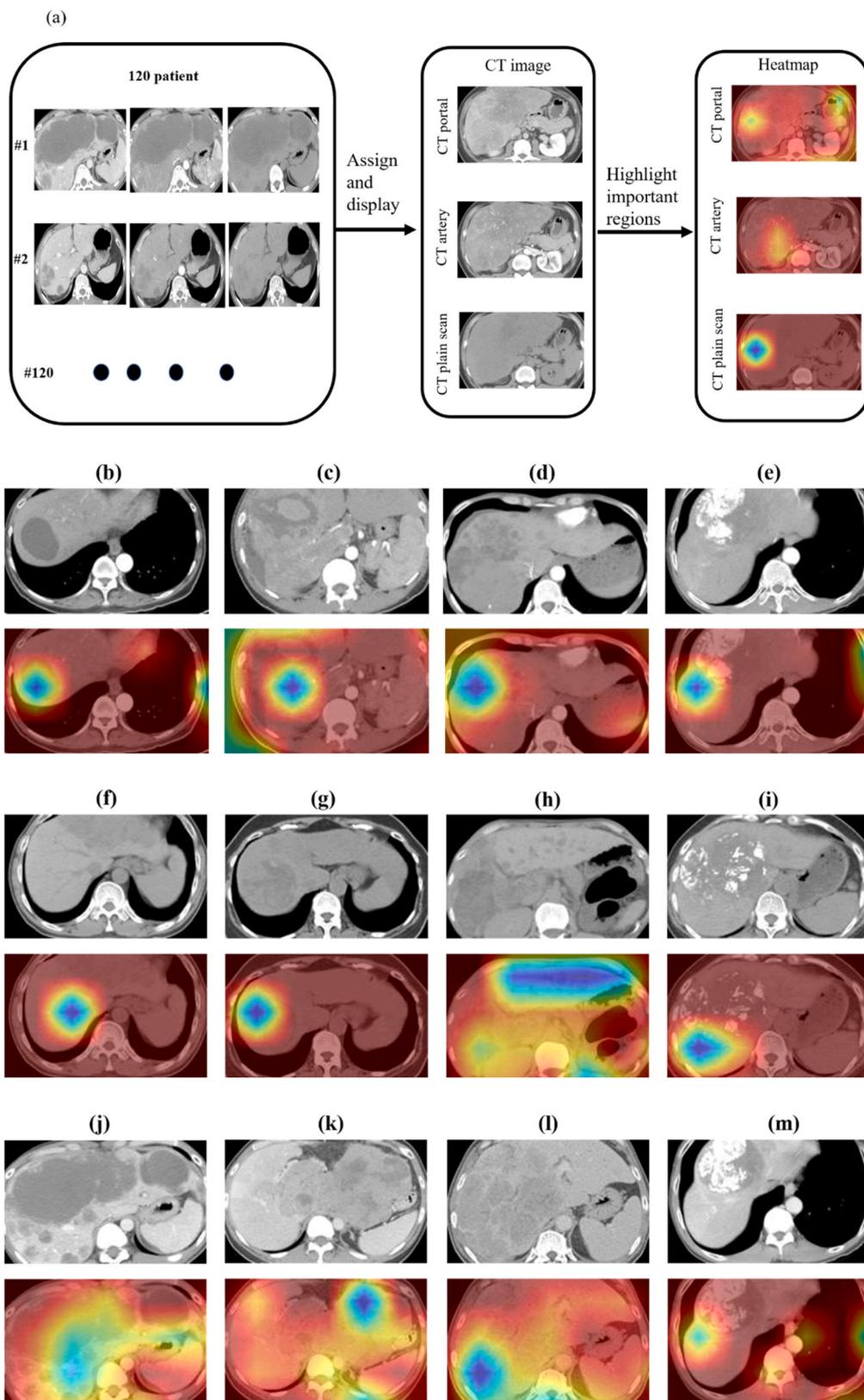
Fig. 5. (a) The workflow of the heatmap generation process is shown. For each patient, information was extracted from three CT images, including plain scan phase, arterial phase and portal phase, and corresponding heatmaps were generated. (b)-(m) Examples of regions with potential clinical value for predicting the response of unresectable HCC to CLICI treatment in the validation set are shown. In each set of images, the 1st row shows raw image, the 2nd row shows corresponding heatmap. (b)-(e)the arterial phase images, (f)-(i) the plain scan phase images. (j)-(m)the portal phase images. The highlighted six regions include necrosis (b), vasculature (f), tumor heterogeneity (c,k), tumor and peritumoral regions (g,j), intrahepatic multifocal tumors (d,h,l), and tumor regions surrounding TACE iodide (e,i,m).

Table 3

Performance comparison of baseline models using multiphase CT images on validation set.

Model	AUC (95%CI)	Accuracy	Sensitivity	Specificity
VGG19	0.732 (0.710–0.754)	0.692	0.763	0.738
Mobilenetv3	0.764 (0.728-0.800)	0.701	0.787	0.746
ResNet34	0.789 (0.742-0.836)	0.718	0.809	0.795
ResNet-50	0.771 (0.730-0.812)	0.722	0.811	0.784
ResNet-18(ours)	0.802 (0.753-0.851)	0.725	0.829	0.803

general radiologists and non-radiologist clinicians. [41] To enhance our model's potential clinical applicability, we've incorporated explainability features like heatmaps, making the decision-making process more transparent to human experts. Heatmaps give clinicians a visual representation of the model's predictive insights, aiding in clinical decision-making. These heatmaps highlight six key features—necrosis, vasculature, tumor heterogeneity, tumor and peritumoral regions, intrahepatic multifocal tumors, and tumor regions surrounding TACE iodide—essential for predicting CLICI response in unresectable HCC.

Studies have revealed that necrosis is a constituent of complex intratumoral components that may stimulate the immune system. [42] Tumor heterogeneity could contribute to the survival of HCC and facilitate the treatment of CLICI. [43,44] The vasculature can be partially explained by the fact that lenvatinib is known to be anti-angiogenic, primarily interacting with VEGF. [45] Consequently, it may be more effective in tumors with ample blood supply and high degrees of enhancement. Neutrophil infiltration in the peritumoral region of HCC is widely distributed around the cancer, rather than within the cancer itself. Peritumoral infiltration from the cancer center can potentially impact the aggregation and immune function of immune cells, thus influencing the prognosis. [46] Furthermore, the peritumoral region is associated with anti-PD-1 therapy for patients with HCC. [47] Intrahepatic multifocal tumors, characterized by genetic diversity with multiple clonal origins, exhibits strong metastatic potential and multiple metastatic patterns, especially linear and parallel diffusion models, leading to a poor prognosis. [48,49] The spatial relationship between TACE iodide deposits and the tumor could be a critical determinant, as prior TACE iodide therapy has been shown to elicit immune responses within and around the tumor in HCC. Research further suggests that the presence of tumor-associated antigen responses tends to be more pronounced in patients who have undergone pre-treatment with TACE. [50] Therefore, it is believed that there are positive antigens that play a role in patient selection for unresectable HCC in CLICI therapy.

This critical insight underscores the significant role these factors play in selecting patients for CLICI therapy in unresectable HCC. The heatmaps serve as crucial tools for identifying key patient characteristics, offering insights that help clinicians identify areas of concern in the predictive model's process. These features deepen understanding of the model's outcomes and aid in interpreting highlighted regions' significance. Future work will aim to enlarge dataset sizes for a more comprehensive evaluation of interpretable AI systems' benefits. Despite the ongoing development of interpretability in deep learning, our methodology shows great promise in improving clinical decision-making.

We have developed a web server with the hope that it can serve as a preliminary exploratory analysis tool for researchers with data akin to ours, allowing for the testing and comparison of their data or model predictions. Paramount to us is the protection of patient privacy in the operation of this web server. To this end, we have enacted stringent security protocols to encrypt and safeguard patient information, in strict compliance with both international and regional data protection statutes, including the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA). In terms of regulation, our web server's development and its open utilization align with the pertinent regulatory standards for health information technology, ensuring, through ongoing dialogue with regulatory agencies, that our platform adheres to all relevant medical norms and directives.
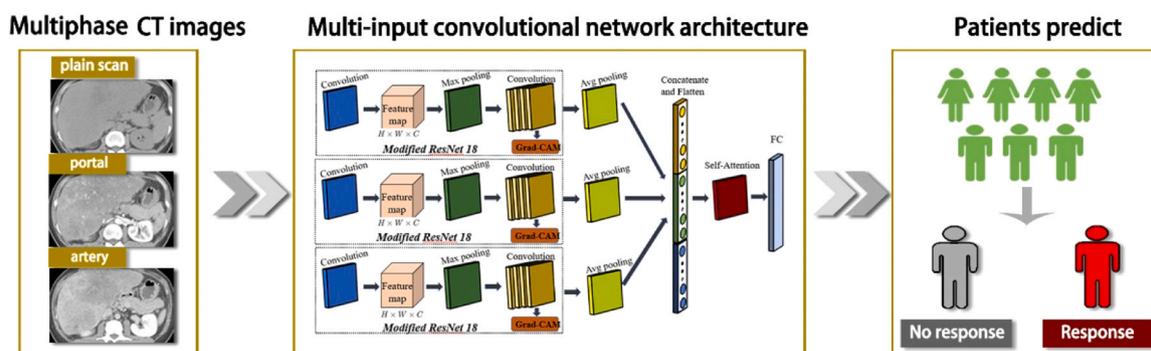
Fig. 6. The content of webserver.

Our study is subject to several limitations: (1) The current accuracy and sensitivity levels of our model require enhancement to fulfill clinical standards. We are committed to refining our model accordingly. (2) The patient data were sourced exclusively from a single hospital, restricting the breadth of our findings' applicability. To bolster the model's generalizability, future validations should incorporate data from multiple healthcare institutions. (3) Although we employed data augmentation techniques to mitigate the effects of data scarcity and overfitting, these methods cannot perfectly emulate real-world physical phenomena and equipment settings.

5. Conclusion

In this study, these findings represent a significant advance in predicting the response to chemoimmunotherapy in unresectable HCC using a deep-learning-based AI model. The integration of CT images and a user-friendly heatmap in the workflow has the potential to guide clinical practice in a prospective setting. In addition, we have developed a web server to make it more convenient to analyze CT images and predict therapeutic responses to CLICI in unresectable HCC.

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the manuscript. The corresponding authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Funding

CRediT authorship contribution statement

Jian-Hong Zhong: Writing – review & editing, Conceptualization, Formal analysis, Funding acquisition. **Nan-Qing Liao:** Conceptualization, Writing – original draft, Writing – review & editing. **Zhu-Jian Deng:** Data curation, Investigation. **Wei Wei:** Data curation, Formal analysis. **Jia-Hui Lu:** Investigation. **Min-Jun Li:** Investigation. **Liang Ma:** Resources. **Qingfeng Chen:** Conceptualization, Funding acquisition, Project administration, Writing – review & editing.

Declaration of Competing Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Acknowledgments

We thank the authors of the cited references for their excellent work.

Author contributions

J.-H.Z and Q.-F.C conceived the study; all authors participated in the acquisition of the data and analyzed the data; N.-Q.L and J.-H.Z drafted and revised the manuscript; all authors read the manuscript and approved the final version to be published. J-HZ and Q.-F.C had full access to all the data in the study and serves as guarantor, taking full responsibility for the integrity of the data and the accuracy of the data analysis.

References

[1] Finn RS, Ikeda M, Zhu AX, et al. Phase Ib study of lenvatinib plus pembrolizumab in patients with unresectable hepatocellular carcinoma. J Clin Oncol Sep 10 2020; 38(26):2960–70.

[2] Chen K, Wei W, Liu L, et al. Lenvatinib with or without immune checkpoint inhibitors for patients with unresectable hepatocellular carcinoma in real-world clinical practice. Cancer Immunol Immunother 2022;71(5):1063–74.

[3] Tsujita Y, Sofue K, Ueshima E, et al. Evaluation and prediction of treatment response for hepatocellular carcinoma. Magn Reson Med Sci 2023;22(2):209–20.

[4] Bera K, Braman N, Gupta A, Velcheti V, Madabhushi A. Predicting cancer outcomes with radiomics and artificial intelligence in radiology. Nat Rev Clin Oncol 2022;19 (2):132–46.

[5] Cao Y, Zhang G, Bao H, et al. Development of a dual-energy spectral CT based nomogram for the preoperative discrimination of mutated and wild-type KRAS in patients with colorectal cancer. Clin Imaging 2021;69:205–12.

[6] Zhang W, Fang M, Dong D, et al. Development and validation of a CT-based radiomic nomogram for preoperative prediction of early recurrence in advanced gastric cancer. Radio Oncol 2020;145:13–20.

[7] Wong LM, King AD, Ai QYH, et al. Convolutional neural network for discriminating nasopharyngeal carcinoma and benign hyperplasia on MRI. Eur Radio 2021;31(6): 3856–63.

[8] Zhuo EH, Zhang WJ, Li HJ, et al. Radiomics on multi-modalities MR sequences can subtype patients with non-metastatic nasopharyngeal carcinoma (NPC) into distinct survival subgroups. Eur Radio 2019;29(10):5590–9.

[9] Dong D, Fang MJ, Tang L, et al. Deep learning radiomic nomogram can predict the number of lymph node metastasis in locally advanced gastric cancer: an international multicenter study. Ann Oncol 2020;31(7):912–20.

[10] Wei H, Yang F, Liu Z, et al. Application of computed tomography-based radiomics signature analysis in the prediction of the response of small cell lung cancer patients to first-line chemotherapy. Exp Ther Med 2019;17(5):3621–9.

[11] Xu Y, Hosny A, Zeleznik R, et al. Deep learning predicts lung cancer treatment response from serial medical imaging. Clin Cancer Res Jun 1 2019;25(11): 3266–75.

[12] Malviya R, Verma S. Targeted therapy using deep learning tools: state of art approach. Curr Drug Targets 2022;23(12):1133–5.

[13] Tunali I, Gray JE, Qi J, et al. Novel clinical and radiomic predictors of rapid disease progression phenotypes among lung cancer patients treated with immunotherapy: an early report. Lung Cancer 2019;129:75–9.

[14] Loi S, Drubay D, Adams S, et al. Tumor-infiltrating lymphocytes and prognosis: a pooled individual patient analysis of early-stage triple-negative breast cancers. J Clin Oncol 2019;37(7):559–69.

[15] Wang L, Wu M, Li R, Xu X, Zhu C, Feng X. MVI-mind: a novel deep-learning strategy using computed tomography (CT)-based radiomics for end-to-end high efficiency prediction of microvascular invasion in hepatocellular carcinoma. Cancers 2022:14 (12).

[16] Mu W, Jiang L, Shi Y, et al. Non-invasive measurement of PD-L1 status and prediction of immunotherapy response using deep learning of PET/CT images. J Immunother Cancer 2021;9(6).

[17] Hu Y, Xie C, Yang H, et al. Computed tomography-based deep-learning prediction of neoadjuvant chemoradiotherapy treatment response in esophageal squamous cell carcinoma. Radio Oncol 2021;154:6–13.

[18] Jiang M, Li CL, Luo XM, et al. Ultrasound-based deep learning radiomics in the assessment of pathological complete response to neoadjuvant chemotherapy in locally advanced breast cancer. Eur J Cancer 2021;147:95–105.

[19] Choi JH, Kim HA, Kim W, et al. Early prediction of neoadjuvant chemotherapy response for advanced breast cancer using PET/MRI image deep learning. Sci Rep 2020;10(1):21149.

[20] Flecken T, Schmidt N, Hild S, et al. Immunodominance and functional alterations of tumor-associated antigen-specific CD8+ T-cell responses in hepatocellular carcinoma. Hepatology 2014;59(4):1415–26.

[21] Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 2009;45(2): 228–47.

[22] Su J.Y., Wu Y.L., Wu X.L., et al. Letter: Entecavir versus tenofovir on serum lipoprotein levels of hepatitis B virus-related hepatocellular carcinoma after curative hepatectomy. *Aliment Pharmacol Ther.* 2023.

[23] Yuan BH, Li RH, Huo RR, Li MJ, Papatheodoridis G, Zhong JH. Lower risk of hepatocellular carcinoma with tenofovir than entecavir treatment in subsets of chronic hepatitis B patients: an updated meta-analysis. J Gastroenterol Hepatol 2022;37(5):782–94.

[24] Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. Ieee I Conf Comp Vis 2017:618–26.

[25] Simonyan K., Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. 2014:arXiv:1409.1556. https://ui.adsabs.harvard.edu/abs/ 2014arXiv1409.1556S. Accessed September 01, 2014.

[26] He KM, Zhang XY, Ren SQ, Sun J. Deep residual learning for image recognition. Proc Cvpr Ieee 2016:770–8.

[27] Howard A., Sandler M., Chu G., et al. Searching for MobileNetV3. 2019 Ieee/Cvf International Conference on Computer Vision (Iccv 2019). 2019:1314–1324.

[28] Sheng R, Zeng M, Jin K, Zhang Y, Wu D, Sun H. MRI-based nomogram predicts the risk of progression of unresectable hepatocellular carcinoma after combined lenvatinib and anti-PD-1 antibody therapy. Acad Radio 2022;29(6):819–29.

[29] Peng J, Huang J, Huang G, Zhang J. Predicting the initial treatment response to transarterial chemoembolization in intermediate-stage hepatocellular carcinoma by the integration of radiomics and deep learning. Front Oncol 2021;11:730282.

[30] Baltrusaitis T, Ahuja C, Morency LP. Multimodal. Mach Learn: A Surv Taxon *Ieee T Pattern Anal* 2019;41(2):423–43.

[31] Cui Y, Zhang J, Li Z, et al. A CT-based deep learning radiomics nomogram for predicting the response to neoadjuvant chemotherapy in patients with locally advanced gastric cancer: A multicenter cohort study. EClinicalMedicine 2022;46: 101348.

[32] Jiang Y, Jin C, Yu H, et al. Development and validation of a deep learning CT signature to predict survival and chemotherapy benefit in gastric cancer: a multicenter, retrospective study. Ann Surg 2021;274(6):e1153–61.

[33] Oliver 3rd JH, Baron RL, Federle MP. Rockette HE, Jr. Detecting hepatocellular carcinoma: value of unenhanced or arterial phase CT imaging or both used in conjunction with conventional portal venous phase contrast-enhanced CT imaging. AJR Am J Roentgenol 1996;167(1):71–7.

[34] Nishie A, Yoshimitsu K, Okamoto D, et al. CT prediction of histological grade of hypervascular hepatocellular carcinoma: utility of the portal phase. Jpn J Radio 2013;31(2):89–98.

[35] Tomino T, Itoh S, Okamoto D, et al. Impact of portal-phase signal intensity of dynamic gadoxetic acid-enhanced magnetic resonance imaging in hepatocellular carcinoma. J Hepatobiliary Pancreat Sci 2023;30(9):1089–97.

[36] Hollett MD, Jeffrey Jr RB, Nino-Murcia M, Jorgensen MJ, Harris DP. Dual-phase helical CT of the liver: value of arterial phase scans in the detection of small (< or = 1.5 cm) malignant hepatic neoplasms. AJR Am J Roentgenol 1995;164(4): 879–84.

[37] Lam A., Fernando D., Sirlin C.C., et al. Value of the portal venous phase in evaluation of treated hepatocellular carcinoma following transcatheter arterial chemoembolisation. Clin Radiol. Nov 2017;72(11):994 e999–994 e916.

[38] Brenet Defour L, Mule S, Tenenhaus A, et al. Hepatocellular carcinoma: CT texture analysis as a predictor of survival after surgical resection. Eur Radio 2019;29(3): 1231–9.

[39] Mule S, Thiefin G, Costentin C, et al. Advanced hepatocellular carcinoma: pretreatment contrast-enhanced CT texture parameters as predictive biomarkers of survival in patients treated with sorafenib. Radiology 2018;288(2):445–55.

[40] Castelvecchi D. Can we open the black box of AI? Nature 2016;538(7623):20–3.

[41] Lee H, Yune S, Mansouri M, et al. An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets. Nat Biomed Eng 2019;3(3):173–82.

[42] Chen J, Wu Z, Xia C, et al. Noninvasive prediction of HCC with progenitor phenotype based on gadoxetic acid-enhanced MRI. Eur Radio 2020;30(2):1232–42.

[43] Jeng KS, Chang CF, Jeng WJ, Sheen IS, Jeng CJ. Heterogeneity of hepatocellular carcinoma contributes to cancer progression. Crit Rev Oncol Hematol 2015;94(3): 337–47.

[44] Friemel J, Rechsteiner M, Frick L, et al. Intratumor heterogeneity in hepatocellular carcinoma. Clin Cancer Res 2015;21(8):1951–61.

[45] Kawamura Y, Kobayashi M, Shindoh J, et al. Pretreatment heterogeneous enhancement pattern of hepatocellular carcinoma may be a useful new predictor of early response to lenvatinib and overall prognosis. Liver Cancer 2020;9(3):275–92.

[46] Zhang T, Lei X, Jia W, et al. Peritumor tertiary lymphoid structures are associated with infiltrating neutrophils and inferior prognosis in hepatocellular carcinoma. Cancer Med 2023;12(3):3068–78.

[47] Yuan G, Song Y, Li Q, et al. Development and validation of a contrast-enhanced CT-based radiomics nomogram for prediction of therapeutic efficacy of anti-PD-1 antibodies in advanced HCC patients. Front Immunol 2020;11:613946.

[48] Xie DY, Fan HK, Ren ZG, Fan J, Gao Q. Identifying clonal origin of multifocal hepatocellular carcinoma and its clinical implications. Clin Transl Gastroenterol 2019;10(2):e00006.

[49] Dong LQ, Peng LH, Ma LJ, et al. Heterogeneous immunogenomic features and distinct escape mechanisms in multifocal hepatocellular carcinoma. J Hepatol 2020;72(5):896–908.

[50] Pinato DJ, Murray SM, Forner A, et al. Trans-arterial chemoembolization as a loco-regional inducer of immunogenic cell death in hepatocellular carcinoma: implications for immunotherapy. J Immunother Cancer 2021;9:9.