

RESEARCH

Open Access



Deep learning based on intratumoral heterogeneity predicts histopathologic grade of hepatocellular carcinoma

Shaoming Song^{1†}, Gong Zhang^{2,3†}, Zhiyuan Yao^{2,4†}, Ruiqiu Chen¹, Kai Liu², Tianchen Zhang¹, Guineng Zeng^{2,5}, Zizheng Wang^{6*} and Rong Liu^{1,2*}

Abstract

Objectives The potential of medical imaging to non-invasively assess intratumoral heterogeneity (ITH) is increasingly being recognized. This study aimed to investigate the value of the ITH-based deep learning model for preoperative prediction of histopathologic grade in hepatocellular carcinoma (HCC).

Materials and methods A total of 858 patients from primary cohort and two external cohorts were included. 3.0T or 1.5T axial portal venous phase MRI images were collected. We conducted radiomics feature-driven K-means clustering for automatic partition to reveal ITH. 2.5D and 3D deep learning models based on ResNet architecture were trained to extract deep learning hidden features of each subregion. The selected features were used to train Random Forest classifier, which constructed the feature-fusion model.

Results The extracted voxel-level radiomics features were unsupervised clustered by K-means to generate three subregions. In the 2.5D deep learning, the feature-fusion model based on ITH had superior predictive efficacy than the whole-tumor model (AUC: 0.82 vs. 0.72; $p=0.004$). Even in the validation and external test sets, this model maintained a high AUC of 0.78–0.83, and net reclassification indices indicated that it could improve prediction by 25–28%. Regarding the prognostic value, overall survival (OS) and recurrence-free survival (RFS) could be significantly stratified by the 2.5D feature-fusion model, and multivariable Cox regressions indicated its signature was identified as a risk predictor for OS and RFS ($p < 0.05$).

Conclusion The ITH-based feature-fusion model provided a non-invasive method for classifying tumor differentiation in HCC, which may serve as a promising strategy for stratification management.

Keywords Intratumoral heterogeneity, Deep learning, Histopathologic grade, Magnetic resonance imaging, Hepatocellular carcinoma

[†]Shaoming Song, Gong Zhang and Zhiyuan Yao contributed equally to this work.

*Correspondence:
Zizheng Wang
wzzh301@163.com
Rong Liu
liurong301@126.com

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Introduction

Hepatocellular carcinoma (HCC) ranks as the sixth most common malignant tumor and the fourth leading cause of cancer-related deaths globally [1, 2]. Despite the widespread use of radical resection for HCC over recent decades, the recurrence rate remains high, reaching approximately 70% within five years after surgery [3]. Studies have indicated that poorly differentiated (PD) tumor is a significant factor resulting in the high recurrence rate and poor overall survival in patients with HCC [4, 5]. Early determination of the degree of differentiation of HCC is currently an important basis for clinical diagnosis, staging and selection of therapeutic decisions.

In recent years, radiomics has shown significant potential in clinical research. Radiomics can extract quantitative feature information from lesions in a high throughput, which contributes to the diagnosis and prognosis of HCC. Mao et al. retrospectively included 122 patients and extracted 242 radiomics features based on patients' preoperative arterial enhancement MRI [6]. Using LR algorithm modeling, the MRI-based radiomics features were found to be valuable in predicting the degree of HCC differentiation (AUC: 0.805 and 0.777) [6]. With the rapid development of artificial intelligence, its sub-field deep learning has been widely used in medical image processing and analysis [7–9]. Based on preoperative contrast enhanced CT images, Wei et al. trained a multi-scale and multi-region convolutional neural network [10]. The results showed that the AUC of the deep learning model in the validation set was 0.86, which was better than the conventional radiomics, radiological, and clinical models (AUCs: 0.765, 0.695, and 0.612, respectively) [10]. Both techniques, radiomics and deep learning, have shown promising performance in predicting the degree of differentiation of HCC preoperatively. However, currently available publications treated the entire tumor as a homogeneous entity, overlooking the heterogeneity within the tumor [11].

The intratumoral heterogeneity (ITH) can lead to differences in oxygenation, nutrient abundance, immune activity, and stromal composition across different tumor regions, resulting in spatial heterogeneity that manifests as morphological variations [12, 13]. In basic research, ITH is a widely studied topic because it is crucial for understanding the biological characteristics, response to therapy, and progression of tumors. However, in non-invasive assessment by medical images, the research investigating ITH has been relatively limited [14, 15].

In this study, we proposed a fusion framework that not only considers the spatial heterogeneity of HCC, but also combines radiomics, unsupervised K-means clustering, deep learning, and machine learning to improve the accuracy of predicting histopathologic grade of HCC. Specifically, the entire tumor was divided into subregions

based on the spatial heterogeneity of the tumor using the K-means method and local voxel-level radiomics features. Then, deep learning models were trained to acquire representative information from the subregions and extract key features and build machine learning models from it. This improvement allows us to better capture the complex structures inside the tumor. We also explored the potential value of predictive modeling in the prognostic stratification of HCC patients and the biological basis of model's prediction.

Materials and methods

Patient cohorts

The study was approved by the ethics board of our hospital. Informed consent was waived due to the retrospective nature of this study. We collected three independent patient cohorts. First, a primary cohort comprised 655 patients from the medical center A was included and split into training and validation set. These patients were diagnosed with HCC and underwent surgery between January 2018 and July 2022 (Appendix S1). The inclusion criteria were (1) histopathologically confirmed HCC, (2) MRI scans available within four weeks before surgery, and (3) radical surgical resection. The exclusion criteria are as follows: (1) history of previous liver surgery, (2) anti-tumor therapies before surgery, (3) incomplete medical records, (4) patients with portal vein invasion or metastasis, and (5) patients with poor-quality MRI data.

Second, we collected 65 patients for an external test cohort. The data for this cohort were obtained from the medical center B (40 patients) and the Cancer Imaging Archive database (25 patients). Third, we collected a set of 138 patients with HCC diagnosed at the same medical center A between August 2022 and June 2023 as an internal temporal test set. The inclusion and exclusion criteria were the same as described earlier.

Pathological differentiation analysis

Tumor differentiation data was obtained from standard pathological reports, serving as the benchmark for classification. Tumors were graded into PD, moderately differentiated (MD), and well differentiated (WD) categories. Tumors with MD and WD characteristics were collectively designated as non-poorly differentiated hepatocellular carcinoma (nPD) HCC. In instances where HCC exhibited mixed differentiation patterns, the predominant grade was used to establish the definitive diagnosis.

Image acquisition and processing

The portal venous phase of contrast-enhanced T1WI was obtained on 1.5T or 3.0T axial MRI scanners. The detailed parameters were described in Appendix S2. Image preprocessing included resampling ($1 \times 1 \times 1$ mm), normalization, and N4 bias field correction by Python

SimpleITK package. We utilized 3D U-Net to detect and segment tumors in the liver region automatically. Details of tumor segmentation were provided in Appendix S3. After the segmentation is completed, two experienced radiologists checked the segmentation results of each case one by one. If the segmentation is found to be incorrect or incomplete, they will made corrections. In cases where there was disagreement, a third expert intervened and provided a final decision.

Pattern partition for tumor heterogeneity

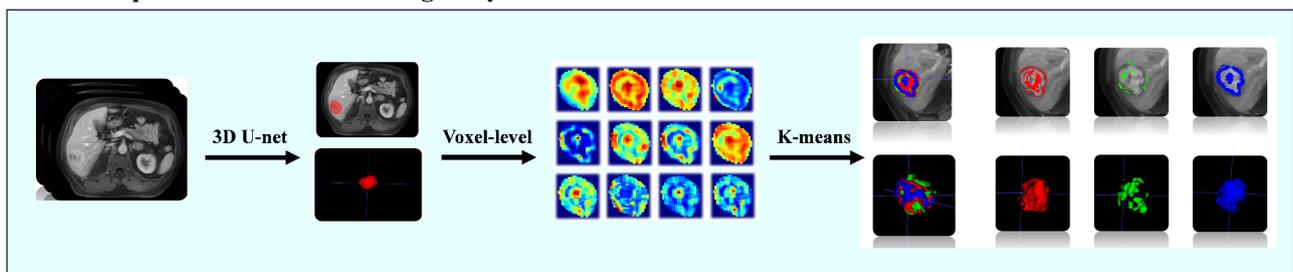
Since it is hardly possible to define the subregions responding to ITH from visual assessment precisely, we employed a data-driven K-means clustering algorithm for automatic and effective partition (Fig. 1) [16]. First, the PyRadiomics package extracted a total of 77 features using a 5×5 matrix centered on each voxel, including first order and texture features (Appendix S4). Second,

we collected each patient’s voxel-level radiomics features to perform unsupervised K-means clustering using the Scikit-learn package. To determine the optimal number of subregions, we conducted K-means with different cluster numbers from 3 to 9, respectively. The Calinski-Harabasz, Silhouette Coefficient, and Davies-Bouldin indices were then calculated to assess the clustering performance [17–19]. After determining the best cluster number, subregions responding to ITH for each patient were automatically partitioned by voxel-based K-means clustering methods, and each exhibits a coherent pattern of imaging attributes.

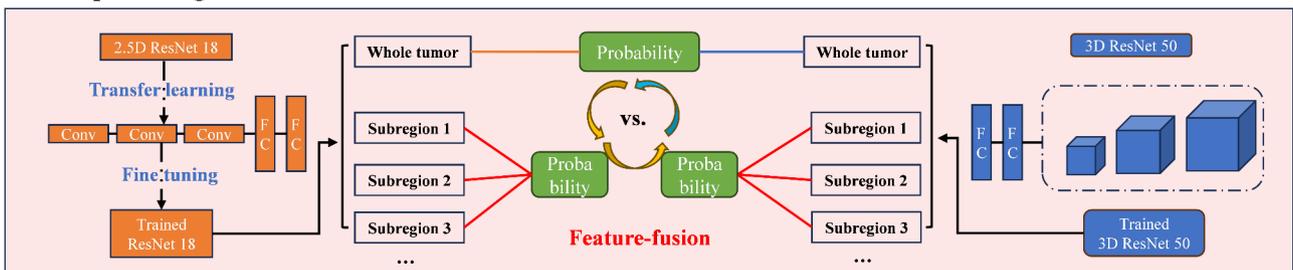
Deep learning procedures and model construction

We cropped the largest tumor cross-sections and its adjacent four slices in Subregion 1. Since the model used multiple 2D images from different layers, it may be in the middle of 2D and 3D, hence the name 2.5D approach

I. Pattern partition for tumor heterogeneity



II. Deep learning model construction



III. Model efficacy and Clinical utility

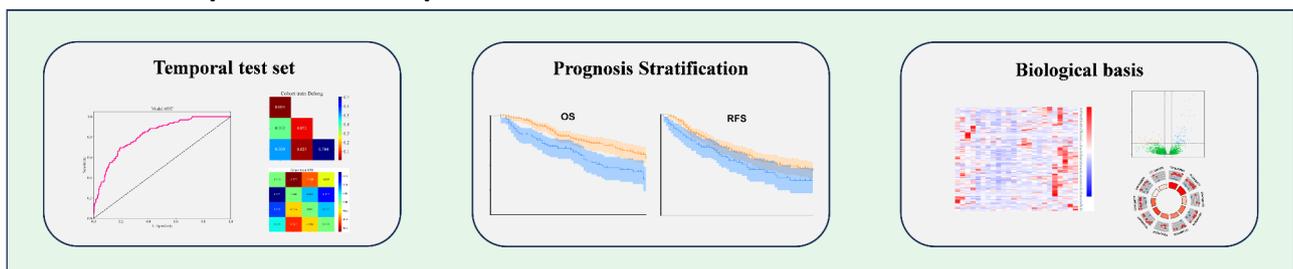


Fig. 1 The flowchart illustrates the study design. Step 1 included image preprocessing, automatic tumor segmentation, extraction of voxel-level radiomics features, and K-means clustering. Step 2 The deep learning model was built for each subregion or whole-tumor, and then the extracted deep learning hidden features from three subregions were combined to train feature-fusion model. Step 3 evaluated the predictive efficacy and clinical utility of the optimal model

[20]. After data augmentation and necessary preprocessing, they were input into five parallel and identical backbone networks with shared parameters (ResNet18). The ResNet18 model, which was pre-trained on the ImageNet dataset, was used for transfer learning supported by OnekeyAI platform (<http://medai.icu/>). Model training was conducted by updating the network weights using a cross-entropy loss function for the prediction task. We employed an initial learning rate of 0.001, training the model for 200 epochs with a batch size of 32. The network parameters were fixed after training was completed and the fixed model was used as a feature extractor [21]. We applied principal component analysis to reduce the dimensionality of the deep learning hidden features, ensuring a balanced representation between features. By reducing the feature dimensions to 200, we aimed to enhance the model's generalization ability and mitigate the risk of overfitting. The same method was used to extract deep learning hidden features from Subregion 2 and Subregion 3. After combination, z-score normalization, Spearman correlation analysis, and LASSO analysis to select the most relevant features, they were subsequently used to train the Random Forest classifier, which constructed the feature-fusion model (Fig. 1) [22].

In 3D deep learning, each subregion within the tumors was cropped into a 3D cube, with a dimension of 96*96*96. Deep learning hidden features were extracted from pre-trained CNN via transfer learning. In the study, the 3D ResNet50 model was employed, which was pre-trained on the Med3D dataset and was supported by OnekeyAI platform. The model was trained with 200 epochs, and the initial learning rate was set to 0.01. Inversion strategies across the X, Y, and Z axes were utilized to augment and diversify the dataset. Similarly, the deep learning hidden features of each subregion were obtained and compressed. These features were then combined and utilized for the Random Forest feature-fusion model using 3D input.

Regarding the construction of the whole-tumor model, we continued the above approach of 2.5D and 3D subregion models. Briefly, for the 2.5D whole-tumor model, we selected the largest cross-section of the entire tumor and its upper and lower layers (five layers in total) as the input dataset. In the 3D whole-tumor model, the entire tumor was segmented into 3D cubes. The training process of the whole-tumor model and the subregion model was similar, with only slight differences in individual parameters. The training and analysis were performed using various software tools, including ITK-SNAP v.3.8.0, and custom code written in Python v.3.7.12. Python packages used in the analysis include Pandas v.1.2.4, NumPy v.1.20.2, PyTorch v.1.8.0, Onekey v.3.1.3, Seaborn v.0.11.1, Matplotlib v.3.4.2, SciPy v.1.7.3, scikit-learn v.1.0.2, and lifelines v.0.27.0.

Performance evaluation and statistical analysis

The predictive performance was assessed and compared quantitatively by receiver operating characteristic curves, the area under the curve (AUC), the Delong test, and the net reclassification index (NRI). Hyperparameter optimization for the model was conducted after analyzing the outcomes of five-fold cross-validation on the primary training set. Patient characteristics were compared using the t-test for continuous variables and χ^2 or Fisher exact test for categorical variables. $P < 0.05$ was considered indicative of a statistically significant difference. Survival analyses were conducted using the Kaplan-Meier curves with log-rank tests, and Cox proportional hazards models were employed to assess potential prognostic variables.

Biological basis exploration

We conducted gene analysis on 24 patients with RNA sequencing data from the Cancer Imaging Archive database to explore the potential biological basis of the model based on ITH (<https://www.cancerimagingarchive.net/collection/tcga-lihc/>) [23]. First, we stratified the 24 patients into two groups based on the threshold of the 2.5D feature-fusion model. Then, we performed differential expression analysis using the edgeR package, considering genes with a log2 fold change greater than 2 and an adjusted p -value less than 0.1 as differentially expressed. Subsequently, Gene Ontology (GO) enrichment analysis was conducted using the clusterProfiler package, mapping differential expression genes (DEGs) to their corresponding Entrez Gene IDs and analyzing for significant enrichment in all three GO domains.

Results

Baseline characteristics

This study included a total of 858 patients. The sample sizes of the training, validation, external test, and internal temporal test sets were 524, 131, 65, and 138, respectively. Of the 858 patients, 620 (72.3%) were under 65 years old and 238 (27.7%) were 65 years or older, including 706 men (82.3%) and 152 women (17.7%). In terms of liver fibrosis, most patients had a diagnosis of liver cirrhosis, primarily due to hepatitis B virus infection. Most HCCs were pathologically confirmed as nPD (690, 80.4%), and PD HCC was identified in 168 patients (19.6%). Compared to patients with PD HCC, a higher proportion of AFP levels less than 20 ng/mL were observed in nPD patients. Regarding other variables, there was no significant difference between the PD and nPD groups in most cases, as detailed in Table 1.

Pattern partition for ITH

The voxel-level radiomics features extracted from tumor segmentation in each patient were subjected to K-means clustering to reflect the ITH (Fig. 2A). Calinski-Harabasz

Table 1 Baseline characteristics of included patients

	Training set			Validation set			External test set			Internal temporal test set		
	nPD HCC (n = 428)	PD HCC (n = 96)	p-value	nPD HCC (n = 108)	PD HCC (n = 23)	p-value	nPD HCC (n = 40)	PD HCC (n = 25)	p-value	nPD HCC (n = 114)	PD HCC (n = 24)	p-value
Age			0.96			0.33			0.82			0.34
<65	316(73.83)	70(72.92)		81(75.00)	20(86.96)		28(70.00)	16(64.00)		71(62.28)	18(75.00)	
≥ 65	112(26.17)	26(27.08)		27(25.00)	3(13.04)		12(30.00)	9(36.00)		43(37.72)	6(25.00)	
Sex			0.31			0.54			0.92			0.14
Female	60(14.02)	18(18.75)		23(21.30)	3(13.04)		13(32.50)	7(28.00)		20(17.54)	8(33.33)	
Male	368(85.98)	78(81.25)		85(78.70)	20(86.96)		27(67.50)	18(72.00)		94(82.46)	16(66.67)	
Cirrhosis			0.04			0.32			0.60			0.42
No	157(36.68)	24(25.00)		43(39.81)	6(26.09)		15(37.50)	7(28.00)		51(44.74)	8(33.33)	
Yes	271(63.32)	72(75.00)		65(60.19)	17(73.91)		25(62.50)	18(72.00)		63(55.26)	16(66.67)	
AFP												
< 20 ng/mL	240(56.07)	22(22.92)	< 0.001	64(59.26)	5(21.74)	0.002	23(57.50)	7(28.00)	0.04	77(67.54)	8(33.33)	0.36
20–400 ng/mL	108(25.23)	30(31.25)	0.28	24(22.22)	10(43.48)	0.06	11(27.50)	12(48.00)	0.16	21(18.42)	7(29.17)	0.02
> 400 ng/mL	80(18.69)	44(45.83)	< 0.001	20(18.52)	8(34.78)	0.15	6(15.00)	6(24.00)	0.56	16(14.04)	9(37.50)	0.24
Tumor size			0.98			0.32			1.00			0.24
< 50 mm	262(61.21)	58(60.42)		60(55.56)	16(69.57)		25(62.50)	16(64.00)		63(55.26)	17(70.83)	
≥ 50 mm	166(38.79)	38(39.58)		48(44.44)	7(30.43)		15(37.50)	9(36.00)		51(44.74)	7(29.17)	
Number of tumors			1.00			0.72			0.11			0.55
Solitary	388(90.65)	87(90.62)		95(87.96)	19(82.61)		34(85.00)	25(100.00)		108(94.74)	24(100.00)	
Multiple	40(9.35)	9(9.38)		13(12.04)	4(17.39)		6(15.00)	0(0)		6(5.26)	0(0)	

PD: poorly differentiated; HCC: hepatocellular carcinoma; nPD: non-poorly differentiated, including moderately differentiated, and well differentiated; AFP: alpha-fetoprotein;

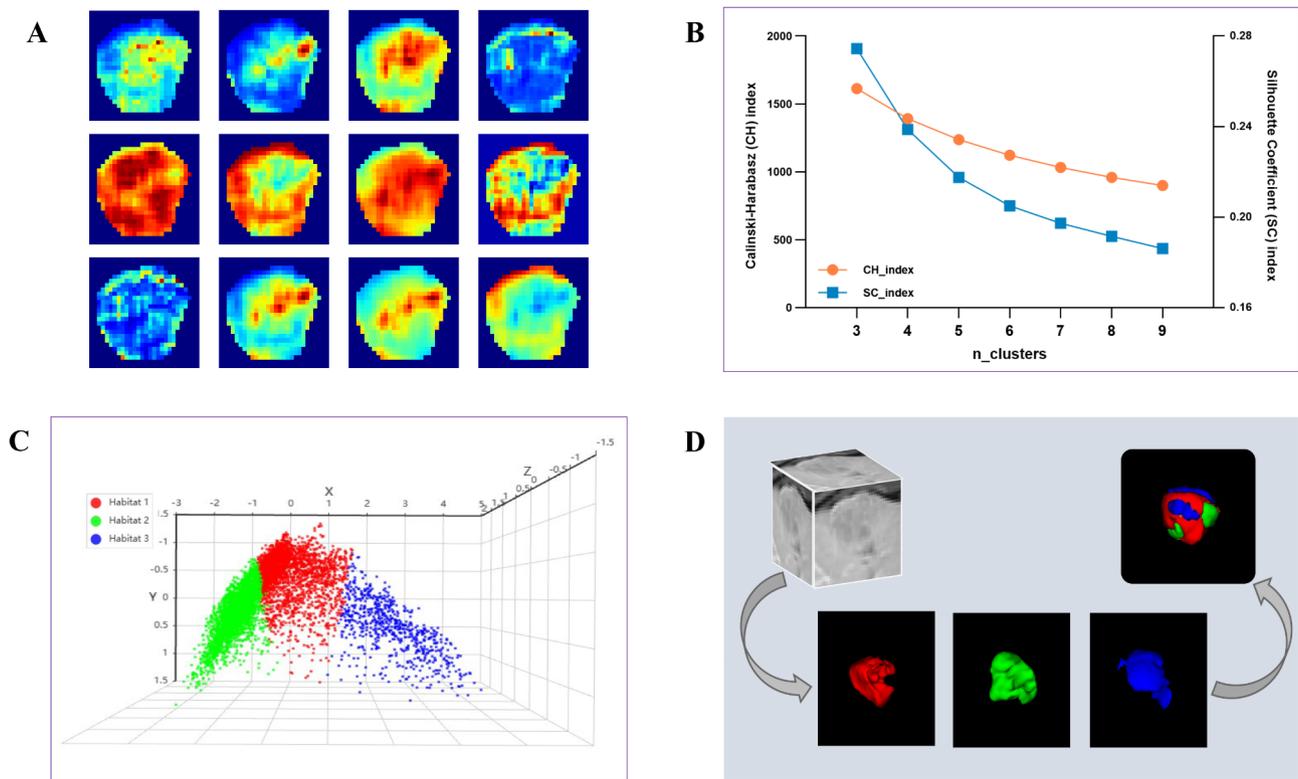


Fig. 2 The diagram shows the procedure of pattern partition for tumor heterogeneity. **(A)** Representative plots for visualization of first-order features and texture features. All local features of each voxel were used for K-means unsupervised clustering; **(B)** The indices were calculated to evaluate the effectiveness of K-means clustering, and the best cluster number was determined accordingly; **(C)** 3D scatterplot for visualization of K-means clustering results for a candidate patient. Each data point represents a voxel and there are 77 local features per voxel. The different colors represent the different clusters (Subregions 1, 2, and 3); **(D)** Through clustering, the tumor was divided into different subregions. These subregions were mapped back to the original tumor volume to reveal the overall intra-tumor heterogeneity

measures the ratio of inter-cluster variance to intra-cluster variance, with higher indices indicating better clustering. We observed that the mean value of this index reached its maximum value when k is set to 3, which means that the clustering results are most favorable at this point (Fig. 2B). The contour coefficient is a metric used to assess the tightness and separation of clusters. In this study, this index similarly supported the result that three-class clustering is optimal.

In the 3D clustering visualization, the different colors represent the different clusters (Subregions 1, 2, and 3) (Fig. 2C). Due to the varying distribution of heterogeneous subregions across different patients, we remapped the subregion labels generated for each patient to ensure similar physical meaning of subregions across all patients. Subsequently, we investigated the predictive potential of applying deep learning methods at the subregional level for HCC tumor differentiation (Fig. 2D).

Model efficiency evaluation

In the training set, compared to the deep learning model of the whole tumor (0.72; 95% CI: 0.66–0.78), the 2.5D feature-fusion model had a significantly higher AUC

(0.82, 95% CI: 0.78–0.86; $p = 0.004$). The model achieved an AUC of 0.825 (95% CI, 0.74–0.92) in the validation set, with an accuracy, sensitivity, and specificity of 0.78, 0.79, and 0.78, respectively. Even in the external test set, the 2.5D feature-fusion model maintained an AUC of 0.78 (95% CI, 0.66–0.90). The predictive performances for the models were presented in Table 2. Regarding the selected features, a total of seven features were included in the 2.5D feature-fusion model after Spearman correlation analysis and LASSO, of which three were from Subregion 1, one from Subregion 2, and three from Subregion 3 (Figure S1).

The 3D feature-fusion model (AUC=0.74, 95%CI, 0.69–0.80) did not show better performance compared to the 3D whole-tumor model (AUC:0.76, 95%CI, 0.70–0.82; $p = 0.70$). This was also the case in the validation (0.69, 95% CI: 0.59–0.79) and external test sets (0.68, 95% CI: 0.55–0.81) (Table 2). By comparison, although not statistically significant by Delong tests, the 2.5D feature-fusion model showed better and more stable performance (Fig. 3). The NRI results also indicated that this model improved predictive accuracy by 25–28% compared to the whole-tumor model. Regarding the selected

Table 2 Predictive performance of different models

Data Set and Model	Whole-tumor	Feature-fusion of subregions from ITH					pvalue*	NRI
	AUC (95%CI)	AUC (95%CI)	ACC	SEN	SPE			
Training set (n = 524)								
2.5D ResNet18	0.716 (0.657, 0.775)	0.821 (0.777, 0.864)	0.784	0.684	0.807	0.004	0.172	
3D ResNet50	0.756 (0.695, 0.817)	0.741 (0.687, 0.795)	0.687	0.674	0.690	0.704	-0.063	
Validation set (n = 131)								
2.5D ResNet18	0.706 (0.580, 0.831)	0.825 (0.735, 0.915)	0.779	0.792	0.776	0.187	0.250	
3D ResNet50	0.732 (0.618, 0.847)	0.690 (0.586, 0.793)	0.641	0.667	0.636	0.593	-0.111	
External test set (n = 65)								
2.5D ResNet18	0.677 (0.538, 0.816)	0.783 (0.662, 0.904)	0.785	0.680	0.850	0.309	0.260	
3D ResNet50	0.705 (0.568, 0.841)	0.679 (0.546, 0.812)	0.662	0.760	0.60	0.787	0.02	
Temporal test set (n = 138)								
2.5D ResNet18	0.693 (0.570, 0.817)	0.808 (0.709, 0.907)	0.819	0.708	0.842	0.195	0.279	
3D ResNet50	0.728 (0.622, 0.834)	0.666 (0.550, 0.782)	0.667	0.583	0.684	0.359	-0.140	

ITH: intratumoral heterogeneity; AUC: the area under the curve; ACC: accuracy; SEN: sensitivity; SPE: specificity; *: Compared with the whole-tumor model; NRI: net reclassification index

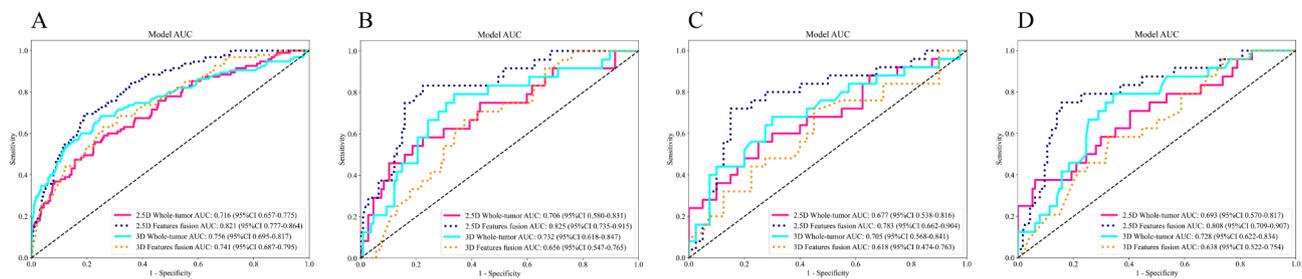


Fig. 3 Performances for classifying tumor differentiation of hepatocellular carcinoma. (A) in the training set; (B) in the validation set; (C) in the external test set; (D) in the temporal test set. These figures demonstrated the 2.5D feature-fusion model had superior predictive efficacy than other three models

features, a total of ten features were included in the 3D feature-fusion model, of which one were from Subregion 1, four from Subregion 2, and five from Subregion 3 (Figure S2).

Grad-CAM from the deep learning model showed differences in attention areas between the 2.5D whole-tumor model and the 2.5D feature-fusion model (Fig. 4). Brighter colors suggested the more significant impact of the features at those locations on the model's decision-making. Overall, the hotspots of each subregion on the Grad-CAM maps were more concentrated and had more explicit boundaries than the 2.5D whole-tumor model.

Clinical utility investigation

In the temporal test set, the 2.5D feature-fusion model based on ITH exhibited satisfactory predictive efficacy, achieving an AUC of 0.81 (95% CI: 0.71–0.91). This model's accuracy, sensitivity, and specificity was 81.9%, 70.8%, and 84.2%, respectively.

We also explored the prognostic value of this selected model. Figure 5AB demonstrated the OS and RFS in the primary cohort ($n = 354$) could be significantly stratified by the signature of this model ($p < 0.05$). In addition, the results of multivariable Cox regressions indicated that the signature from the 2.5D feature-fusion model

was identified as a risk predictor of poor OS (HR: 2.04, 95%CI: 1.42–2.92, $p < 0.001$) and RFS (HR: 1.49, 95%CI: 1.03–2.11, $p = 0.02$) (Table 3).

Biologic basis exploration

As shown in Fig. 6A, significant differences in gene expression were found between 17 patients with low probability and 7 patients with high probability of being PD HCC according to this model. The DEGs were categorized into two main clusters, which was highly consistent with the classification of this model. Figure 6B visualized the number and magnitude of DEGs. In Fig. 6C, molecular function analysis revealed the significant enhancement of gated channel activity, ligand-gated channel activity, growth factor activity, etc. Cellular Components analysis showed that DEGs were most enriched in neuronal cell body and perikaryon. Moreover, tumors with high probability are more vulnerable to some critical biological process that results in tumors being more heterogeneous, including modulation of chemical synaptic transmission, response to metal ions, regulation of mesodermal cell differentiation, and regulation of epithelial cell differentiation (Fig. 6CD).

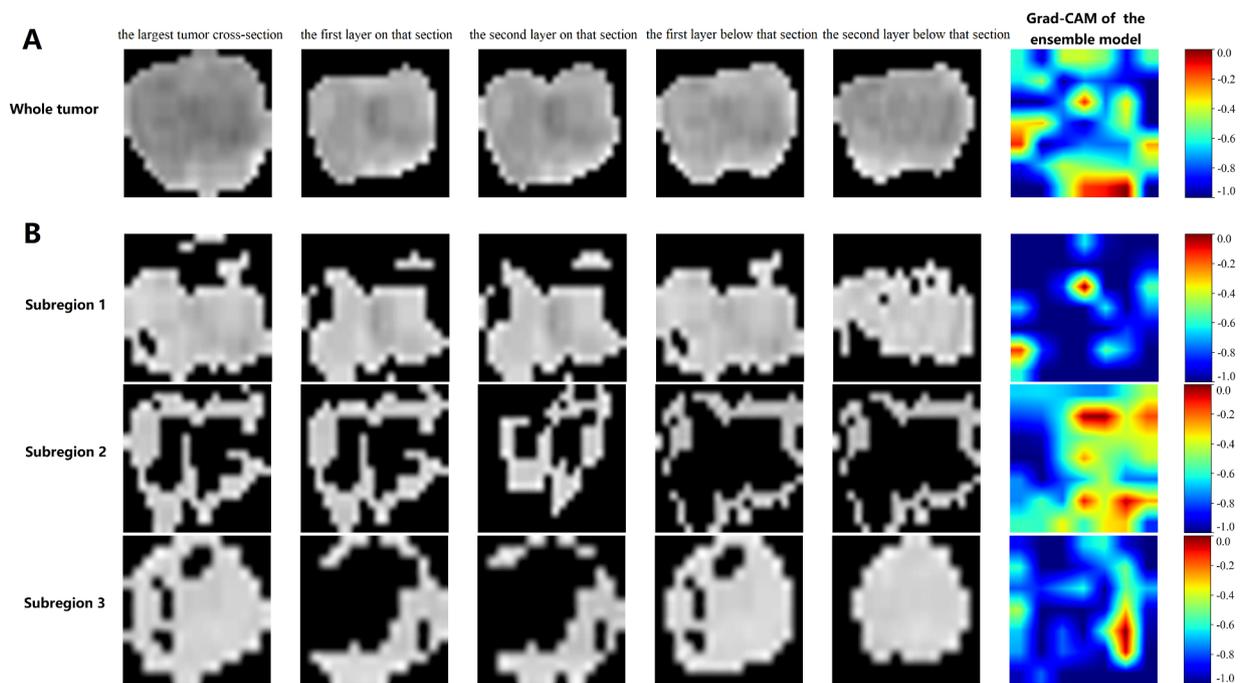


Fig. 4 The attention regions of an example case from the external test set. **(A)** Whole-tumor 2.5D deep learning model; **(B)** 2.5D feature-fusion deep learning model based on intratumoral heterogeneity; From left to right: the largest tumor section, the first and second layers on that section, and the first and second layers below it. The heat map generated by Grad-CAM highlights the areas of concern for the model, allowing us to see which parts are most important for model decisions. physicians can increase their trust in the model by looking at heat maps and understanding why the model makes a particular prediction

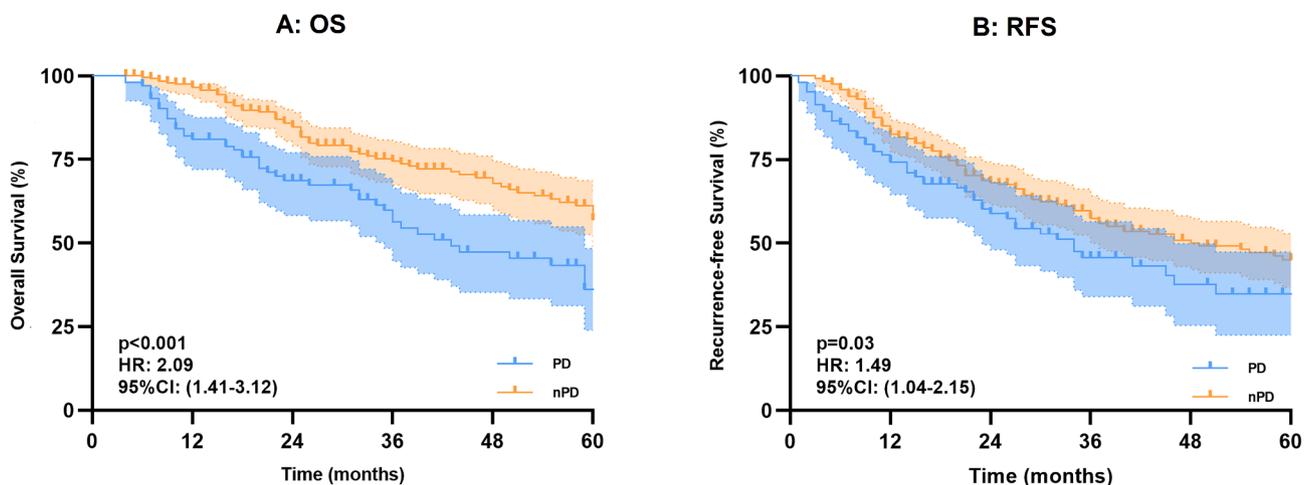


Fig. 5 Prognostic value of the 2.5D feature-fusion signature in hepatocellular carcinoma. **(A)** Overall Survival; **(B)** Recurrence-free Survival. These Figures demonstrated the OS and RFS could be significantly stratified by the signature of the 2.5D feature-fusion model

Discussion

Preoperative definition of the histopathologic grade in HCC is one of the most critical factors for optimal management strategies. The primary objective of this study is to investigate the heterogeneity patterns within the HCC regions using voxel-level radiomics features and K-means unsupervised clustering, then to establish subregional 2.5D and 3D deep learning feature-fusion models to

improve the performance of predicting differentiation grade of HCC. In this study, the 2.5D feature-fusion model demonstrated superior performance in identifying patients with PD HCC and was well validated across the other three sets. Moreover, Kaplan-Meier curves showed that this model’s signature had good survival stratification.

Table 3 Univariable and multivariable Cox proportional hazards analyses for OS and RFS

Variable	Overall Survival			Recurrence-free Survival				
	Univariable Analysis		Multivariable Analysis	Univariable Analysis		Multivariable Analysis		
Signature from ITH model	1.90 (1.33, 2.71)	< 0.001	2.04 (1.42, 2.92)	< 0.001	1.44 (1.03, 2.02)	0.03	1.49 (1.06, 2.11)	0.02
Age (> 65 years)	1.10 (0.71, 1.71)	0.67	1.04 (0.65, 1.66)	0.87	1.04 (0.71, 1.54)	0.84	0.92 (0.61, 1.40)	0.70
Gender (male)	1.10 (0.68, 1.78)	0.69	1.19 (0.73, 1.96)	0.49	1.15 (0.73, 1.80)	0.55	1.12 (0.70, 1.77)	0.64
HBV (Yes/No)	0.97 (0.65, 1.43)	0.87	0.92 (0.60, 1.40)	0.69	0.89 (0.62, 1.27)	0.52	0.89 (0.61, 1.30)	0.55
Cirrhosis (Yes/No)	1.13 (0.80, 1.59)	0.51	1.29 (0.89, 1.88)	0.18	1.03 (0.75, 1.41)	0.87	1.10 (0.79, 1.55)	0.56
INR (> 1.5 ratio)	2.61 (0.36, 18.85)	0.34	4.73 (0.63, 35.76)	0.13	3.13 (0.77, 12.68)	0.11	4.42 (1.04, 18.75)	0.04
Tumor grade (PD/nPD)	1.62 (1.09, 2.39)	0.02	1.28 (0.79, 2.08)	0.31	1.36 (0.93, 1.98)	0.12	1.21 (0.77, 1.90)	0.42
Multiple tumor (Yes/No)	1.08 (0.64, 1.83)	0.77	1.19 (0.70, 2.04)	0.52	1.38 (0.87, 2.20)	0.17	1.43 (0.89, 2.28)	0.14
Pseudocapsule (Yes/No)	0.74 (0.51, 1.07)	0.11	0.68 (0.46, 0.99)	0.05	0.89 (0.64, 1.25)	0.51	0.81 (0.57, 1.15)	0.24
Tumor size (> 50 mm)	1.35 (0.95, 1.91)	0.09	1.56 (1.08, 2.26)	0.02	1.13 (0.98, 1.30)	0.11	1.17 (1.01, 1.36)	0.04

Hazard ratio (HR), 95% CI and *p*-value were reported to measure the effects. HBV: hepatitis B virus; INR: international normalized ratio

The application of deep learning and radiomics for predicting tumor status of HCC shows promise. Several recent publications have investigated the potential of radiomic features to classify tumor differentiation in HCC [6, 24–27]. Despite the success of these studies, their attention was focused on improving predictive capabilities rather than exploring the underlying biological basis or investigating the clinical utility. Currently, the research investigating ITH has been relatively limited [14, 28, 29]. This study introduced an effective automated method to reveal the underlying patterns of ITH in HCC by utilizing voxel-level radiomics features from dynamic contrast-enhanced MRI images. Subsequently, the feature-fusion model was constructed by extracting deep learning hidden features from three subregions. This study's 2.5D feature-fusion model demonstrated superior discriminative ability compared to the whole-tumor model. This model also exhibited high AUC values, ranging from 0.78 to 0.81, in both the external and the temporal test sets. This fusion strategy, which integrates deep learning hidden features from multiple subregions, may result in a richer and more representative feature than a simple whole-tumor approach. Improving features' quality and quantity contributes to constructing a more efficient and targeted model.

As a 3D spital imaging, MRI images trained by a 3D network might obtain more critical information than the 2D or 2.5D approach [30–32]. Contrary to expectations, the 3D feature-fusion model in this study did not perform superior to the 2.5D feature-fusion model. A recent systematic review also indicated that 2D networks generally perform better when compared to baseline models [33]. Compared to 2D deep convolutional neural networks (DCNN), the training of 3D DCNN typically requires larger datasets, and the limited availability of pre-trained 3D networks further exacerbates this issue [22, 34]. Furthermore, the subregional images are discontinuous, which may result in the struggle of 3D models to learn global features. These factors potentially influence

the performance of the feature-fusion model based on a 3D network.

The model's potential application value needs further elucidation to understand its significance and implications for clinical decision-making. First, we employed visualization methods to identify and output the regions deemed important by the network to investigate the prediction process of this model [20]. Subregional deep learning may provide an intuitive understanding of tumor heterogeneity [16]. We found the 2.5D feature-fusion model selected one feature from Subregion 2, and the hotspots were more widespread in Subregion 2. Second, although the model does not serve as a definitive diagnostic criterion for classifying tumor differentiation, the probabilities it generates can assist clinicians in optimizing treatment strategies to a large extent [35]. For patients predicted to have a higher probability of PD HCC, further molecular testing is recommended to evaluate the suitability and choices for neoadjuvant therapy. Conversely, invasive biopsies and subsequent expensive testing can largely be avoided in patients predicted to have non-PD HCC.

Moreover, there is a growing recognition that genetic variation plays a pivotal role in tumor biology. In this study, genetic analysis may have the potential to provide a deeper understanding of the molecular basis of tumor differentiation from the perspective of ITH. By integrating genetic information with imaging data, we explored the biological significance underlying the feature-fusion model based on ITH. Genetic analysis revealed that tumors with a high probability of PD were more susceptible to several critical biological processes, including regulation of epithelial cell differentiation and mesodermal cell differentiation. Growth factor activity, a classic molecular function associated with tumor, was also significantly enriched. Moreover, the response to metal ions was also a significant factor influencing the degree of tumor differentiation [36, 37]. A recently published study suggested ferroptosis was identified as a vulnerability and

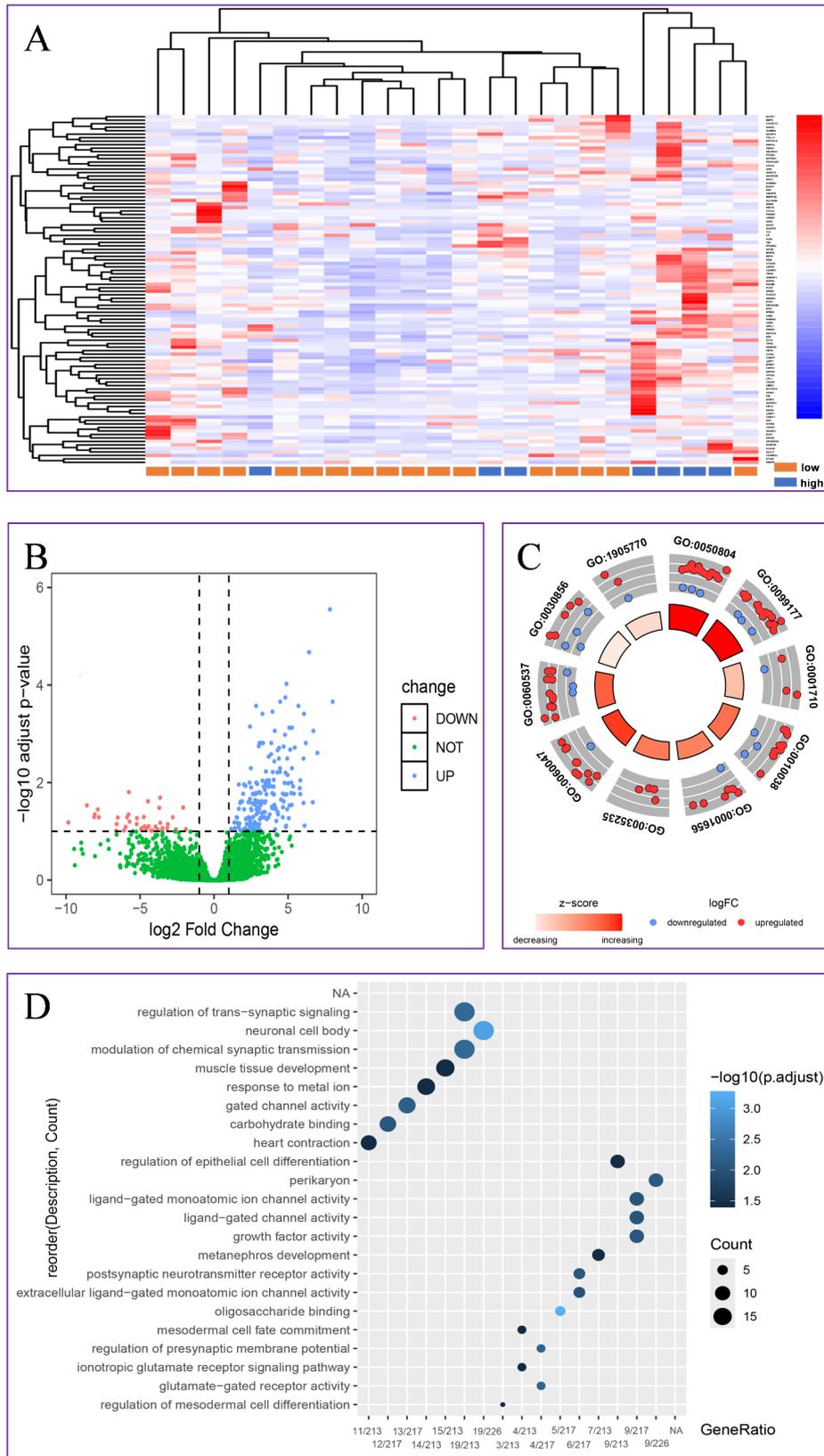


Fig. 6 (See legend on next page.)

(See figure on previous page.)

Fig. 6 The genetic analysis for exploring the underlying biological basis. **(A)** Heatmap of differentially expressed genes in tissues categorized as low and high probability of being poorly differentiated HCC. The genes on the right side of the heatmap cannot be displayed clearly, so you can contact us if you need more information; **(B)** Volcano plot of gene expression profiles in samples from the two groups. **(C)** Bubble plot for functional enrichment analysis in differentially expressed genes, including biological processes, cellular components and molecular functions. **(D)** Circle plot of the distribution of differential genes in enrichment analyses

potential therapeutic target of high-ITH breast tumors [15]. The response to metal ions involves the regulation of cell fate and immune status, especially in tumors with high ITH, where homeostatic regulation of metal ions may have different effects on different subsets of tumor cells [38, 39]. This provides the clues of correlation between imaging textures responding to ITH and potential biological basis. A lenient adjusted p -value threshold (<0.1) was adopted for this exploratory analysis to prioritize hypothesis generation given the small sample size of available data. Future studies with larger cohorts will validate these associations at stricter thresholds (<0.05 or <0.01).

ITH is a significant concept in tumor biology [40, 41]. In further research, the heterogeneity patterns and the identification of high-risk regions are expected to be thoroughly investigated to provide significant tumor characteristics. Moreover, strengthening the interdisciplinary collaboration between medical images and fields such as clinical pathology, molecular biology, and bioinformatics will enable a deeper insight of ITH and provide richer information for predicting tumor states. Future work will focus on correlating radiomic sub-regions with pathological analysis, molecular biology, and bioinformatics to establish direct links between imaging phenotypes, genomic heterogeneity, and immune micro-environmental changes.

This study has several limitations. First, this is a retrospective analysis; thus, patient selection bias and potential bias in tumor grade distribution are inevitable. Future prospective large-scale studies remain necessary. Second, to alleviate the computational demand caused by a large number of voxels, we resampled these images to a voxel size of $2 \times 2 \times 2$ mm, which shortened the processing time but potentially sacrificed some spatial detail. This resampling approach may affect the model's ability to capture fine structural changes. However, our external validation results indicate that even at this voxel size, the model maintains high accuracy and stability. Future research could explore the impact on model performance using original resolution data or finer voxel sizes. Third, only the portal venous phase sequence was adopted in this study, but the combination of multi-modality imaging is a future direction for predictive models in medical fields. Future research should include multi-modality imaging data to comprehensively evaluate the model's predictive performance based on ITH.

Conclusion

In conclusion, the 2.5D feature-fusion model based on ITH showed good performance for predicting preoperative histopathologic grade of HCC. This feature-fusion model, which non-invasively assesses ITH, may serve as a promising strategy for satisfactory preoperative risk prediction in HCC and provide insights for future studies related to ITH in HCC.

Abbreviations

AUC	Area under the curve
DCNN	Deep convolutional neural networks
DEGs	Differential expression genes
GO	Gene ontology
HCC	Hepatocellular carcinoma
ITH	Intratumoral heterogeneity
MD	Moderately differentiated
MRI	Magnetic resonance imaging
NRI	Net reclassification index
OS	Overall survival
PD	Poorly differentiated
PVP	Portal venous phase
RFS	Recurrence-free survival
WD	Well differentiated

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12885-025-13781-1>.

Supplementary Material 1
Supplementary Material 2
Supplementary Material 3
Supplementary Material 4
Supplementary Material 5
Supplementary Material 6
Supplementary Material 7

Acknowledgements

We appreciated the help of Beijing Baiyang Guoxin Intelligent Technology Co., Ltd. in automated tumor segmentation.

Author contributions

Shaoming Song, Gong Zhang, and Zhiyuan Yao conceived and designed the protocol; Ruiqiu Chen, Kai Liu, Tianchen Zhang, and Guineng Zeng collected the data; Shaoming Song, Gong Zhang, and Zhiyuan Yao analyzed the data and wrote the manuscript; Zizheng Wang and Rong Liu critically revised the manuscript; Zizheng Wang and Rong Liu provided financial support. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Science and Technology Major Project (Grant no. 2021ZD0113301), AI+Health Collaborative Innovation Cultivation Project of Beijing City (Grant no. Z221100003522005), and Youth independent innovation science foundation of Chinese PLA General Hospital (Grant no. 22QNFC145).

Data availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations**Ethics approval and consent to participate**

The present study was approved by the Ethics Committee of Chinese PLA general hospital, and the need for patient consent was waived due to retrospective studies. All methods were carried out according to relevant guidelines and regulations.

Consent for publication

Not applicable. All data has been anonymized.

Competing interests

The authors declare no competing interests.

Author details

¹The First School of Clinical Medicine, Lanzhou University, Lanzhou 730000, China

²Faculty of Hepatopancreatobiliary Surgery, The First Medical Center of Chinese PLA General Hospital, Beijing 100853, China

³Key Laboratory of Digital Hepatobiliary Surgery, Chinese PLA General Hospital, Beijing 100853, China

⁴Department of Hepatobiliary Surgery, The Sixth Medical Center of Chinese PLA General Hospital, Beijing 100048, China

⁵Nankai University School of Medicine, Nankai University, Tianjin 300071, China

⁶Department of Hepatobiliary Surgery, Senior Department of Hepatology, The Fifth Medical Center of Chinese PLA General Hospital, Beijing 100039, China

Received: 22 October 2024 / Accepted: 20 February 2025

Published online: 18 March 2025

References

- Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209–49.
- McGlynn KA, Petrick JL, El-Serag HB. Epidemiology of hepatocellular carcinoma. *Hepatology*. 2021;73(Suppl 1):4–13.
- Chen ZH, Zhang XP, Feng JK, et al. Actual long-term survival in hepatocellular carcinoma patients with microvascular invasion: a multicenter study from China. *Hepatology*. 2021;73(3):642–50.
- Xu XF, Xing H, Han J, et al. Risk factors, patterns, and outcomes of late recurrence after liver resection for hepatocellular carcinoma: a multicenter study from China. *JAMA Surg*. 2019;154(3):209–17.
- Llovet JM, Pinyol R, Yarchoan M, et al. Adjuvant and neoadjuvant immunotherapies in hepatocellular carcinoma. *Nat Rev Clin Oncol*. 2024;21(4):294–311.
- Mao Y, Wang J, Zhu Y, et al. Gd-EOB-DTPA-enhanced MRI radiomic features for predicting histological grade of hepatocellular carcinoma. *Hepatobiliary Surg Nutr*. 2022;11(1):13–24.
- Ren Z, Lan Q, Zhang Y, et al. Exploring simple triplet representation learning. *Comput Struct Biotechnol J*. 2024;23(1):1510–21.
- Ren Z, Wang S, Zhang Y. Weakly supervised machine learning. *CAAI Trans Intell Technol*. 2023;8(3):549–80.
- Ren Z, Zhang Y, Wang SA. Hybrid framework for lung cancer classification. *Electronics*. 2022;11(10):1614.
- Wei J, Ji Q, Gao Y, et al. A multi-scale, multi-region and attention mechanism-based deep learning framework for prediction of grading in hepatocellular carcinoma. *Med Phys*. 2022;50(4):2290–302.
- Gatenby RA, Grove O, Gillies RJ. Quantitative imaging in cancer evolution and ecology. *Radiology*. 2013;269(1):8–15.
- Lloyd MC, Cunningham JJ, Bui MM, et al. Darwinian dynamics of intratumoral heterogeneity: not solely random mutations but also variable environmental selection forces. *Cancer Res*. 2016;76(11):3136–44.
- Losic B, Craig AJ, Villacorta-Martin C, et al. Intratumoral heterogeneity and clonal evolution in liver cancer. *Nat Commun*. 2020;11(1):291.
- Shi Z, Huang X, Cheng Z, et al. MRI-based quantification of intratumoral heterogeneity for predicting treatment response to neoadjuvant chemotherapy in breast cancer. *Radiology*. 2023;308(1):e222830.
- Su GH, Xiao Y, You C, et al. Radiogenomic-based multiomic analysis reveals imaging intratumor heterogeneity phenotypes and therapeutic targets. *Sci Adv*. 2023;9(40):eadf0837.
- Yang Y, Han Y, Zhao S, et al. Spatial heterogeneity of edema region uncovers survival-relevant habitat of glioblastoma. *Eur J Radiol*. 2022;154(1):10423.
- Calinski T. J Harabasz 1974 A dendrite method for cluster analysis. *J C S I* 3 1 1–27.
- Computational RJo, Mathematics A. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. 1987.
- Halkidi M, Batistakis Y, Vazirgiannis M. J.J.o.I.I.S. On Clustering Validation Techniques. 2001.
- Kim Y, Kim YG, Park JW, et al. A CT-based deep learning model for predicting subsequent fracture risk in patients with hip fracture. *Radiology*. 2024;310(1):e230614.
- Li Y, Wei D, Liu X, et al. Molecular subtyping of diffuse gliomas using magnetic resonance imaging: comparison and correlation between radiomics and deep learning. *Eur Radiol*. 2022;32(2):747–58.
- Wang W, Liang H, Zhang Z, et al. Comparing three-dimensional and two-dimensional deep-learning, radiomics, and fusion models for predicting occult lymph node metastasis in laryngeal squamous cell carcinoma based on CT imaging: a multicentre, retrospective, diagnostic study. *EClinicalMedicine*. 2024;67(102385).
- Erickson BJ, Kirk S, Lee Y, et al. The Cancer genome atlas liver hepatocellular carcinoma collection (TCGA-LIHC) (Version 5) [Data set]. *The Cancer Imaging Archive*; 2016.
- Han YE, Cho Y, Kim MJ, et al. Hepatocellular carcinoma pathologic grade prediction using radiomics and machine learning models of gadoteric acid-enhanced MRI: a two-center study. *Abdom Radiol (New York)*. 2023;48(1):244–56.
- Hu X, Li C, Wang Q, et al. Development and external validation of a radiomics model derived from preoperative gadoteric acid-enhanced MRI for predicting histopathologic grade of hepatocellular carcinoma. *Diagnostics (Basel)*. 2023;13(3).
- Liu HF, Lu Y, Wang Q, et al. Machine learning-based CEMRI radiomics integrating LI-RADS features achieves optimal evaluation of hepatocellular carcinoma differentiation. *J Hepatocellular Carcinoma*. 2023;10:2103–15.
- Liu HF, Wang M, Wang Q, et al. Multiparametric MRI-based intratumoral and peritumoral radiomics for predicting the pathological differentiation of hepatocellular carcinoma. *Insights Imaging*. 2024;15(1):97.
- Nie P, Zhang J, Miao W, et al. Incremental value of radiomics-based heterogeneity to the existing risk criteria in predicting recurrence of hepatocellular carcinoma after liver transplantation. *Eur Radiol*. 2023;33(9):6608–18.
- Yuan J, Wu M, Qiu L, et al. Tumor habitat-based MRI features assessing early response in locally advanced nasopharyngeal carcinoma. *Oral Oncol*. 2024;158(106980).
- Wei H, Zheng T, Zhang X, et al. Deep learning-based 3D quantitative total tumor burden predicts early recurrence of BCLC A and B HCC after resection. *Eur Radiol*. 2025;35(1):127–139.
- Dong SY, Wang WT, Chen XS, et al. Microvascular invasion of small hepatocellular carcinoma can be preoperatively predicted by the 3D quantification of MRI. *Eur Radiol*. 2022;32(6):4198–209.
- Li H, Jia Y, Zhu H, et al. Multi-level feature extraction and reconstruction for 3D MRI image super-resolution. *Comput Biol Med*. 2024;171(108151).
- Demircioglu A. Are deep models in radiomics performing better than generic models? A systematic review. *Eur Radiol Exp*. 2023;7(1):11.
- Yu Q, Ning Y, Wang A, et al. Deep learning-assisted diagnosis of benign and malignant parotid tumors based on contrast-enhanced CT: a multicenter study. *Eur Radiol*. 2023;33(9):6054–65.
- 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).
- Guo W, Zhang S, Chen Y, et al. An important role of the hepcidin-ferroportin signaling in affecting tumor growth and metastasis. *Acta Biochim Biophys Sin (Shanghai)*. 2015;47(9):703–15.
- Manz DH, Blanchette NL, Paul BT, et al. Iron and cancer: recent insights. *Ann N Y Acad Sci*. 2016;1368(1):149–61.

38. Liu Y, Wang Y, Song S, et al. Cancer therapeutic strategies based on metal ions. *Chem Sci*. 2021;12(37):12234–47.
39. Xu L, Peng M, Gao T, et al. Nanoenabled intracellular metal ion homeostasis regulation for tumor therapy. *Adv Sci (Weinh)*. 2024;11(7):e2306203.
40. Chen J, Kaya NA, Zhang Y et al. A multimodal atlas of hepatocellular carcinoma reveals convergent evolutionary paths and 'bad apple' effect on clinical trajectory. *J Hepatol*. 2024;81(4):667–678.
41. Nguyen PHD, Ma S, Phua CZJ, et al. Intratumoural immune heterogeneity as a hallmark of tumour evolution and progression in hepatocellular carcinoma. *Nat Commun*. 2021;12(1):227.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.