BMC Cancer

RESEARCH Open Access

Preoperative clinical radiomics model based on deep learning in prognostic assessment of patients with gallbladder carcinoma

Zhechuan Jin[1,2], Chen Chen[1], Dong Zhang[1], Min Yang[3,4], Qiuping Wang[3], Zhiqiang Cai[5], Shubin Si[5], Zhimin Geng[1*] and Qi Li[1*]

Abstract

Objective We aimed to develop a preoperative clinical radiomics survival prediction model based on the radiomics features via deep learning to provide a reference basis for preoperative assessment and treatment decisions for patients with gallbladder carcinoma (GBC).

Methods A total of 168 GBC patients who underwent preoperative upper abdominal enhanced CT from one high-volume medical center between January 2011 to December 2020 were retrospectively analyzed. The region of interest (ROI) was manually outlined by two physicians using 3D Slicer software to establish a nnU-Net model. The DeepSurv survival prediction model was developed by combining radiomics features and preoperative clinical variables.

Results A total of 1502 radiomics features were extracted from the ROI results based on the nnU-Net model and manual segmentation, and 13 radiomics features were obtained through the 4-step dimensionality reduction methods, respectively. The C-index and AUC of 1-, 2-, and 3-year survival prediction for the nnU-Net based clinical radiomics DeepSurv model was higher than clinical and nnU-Net based radiomics DeepSurv models in the training and testing sets, and close to manual based clinical radiomics DeepSurv model. Delong-test was performed on the AUC of 1-, 2-, and 3-year survival prediction for the two preoperative clinical radiomics DeepSurv prediction models in the testing set, and the results showed that the two models had the same prediction efficiency (all $P > 0.05$).

Conclusions By using the DeepSurv model via nnU-Net segmentation, postoperative survival outcomes for individual gallbladder carcinoma patients could be assessed and stratified, which can provide references for preoperative diagnosis and treatment decisions.

Keywords Gallbladder carcinoma, Radiomics, Nnu-net, Deep learning, Prognosis

*Correspondence:
Zhimin Geng
gengzhimin@mail.xjtu.edu.cn
Qi Li
docliqi@163.com
[1]Department of Hepatobiliary Surgery, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, Shaanxi, China
[2]Department of General Surgery, Beijing Anzhen Hospital, Capital Medical University, Beijing 100029, China
[3]Department of Radiology, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, China
[4]Department of Radiology, Norinco General Hospital, Xi'an 710065, China
[5]Department of Industrial Engineering, School of Mechanical Engineering, Northwestern Polytechnical University, Xi'an710072, Shaanxi, China

Introduction

Gallbladder carcinoma (GBC) is the most prevalent biliary tract malignancy and ranks sixth among all digestive tract malignancies, which is characterized by a low early diagnosis rate, a high degree of malignancy, and a poor prognosis [1, 2]. The incidence rate of GBC has been increasing worldwide in recent years, and it is (1.00 ~ 1.30)/100,000 in China [3]. Despite advancements in diagnosis and treatment, the survival of GBC has remained stagnant over the last 10 years. Patients with GBC can benefit from a preoperative evaluation of their prognosis, which will assist in selecting effective treatment options [4]. There is an urgent need for comprehensive, accurate, and scientific ways to assess and analyze the prognosis of patients with GBC preoperatively.

Radiomics, gradually being applied to the diagnosis of the disease and the prognosis assessment of patients with malignancies, can effectively compensate for the shortcomings of the human eye to identify signs and empirical judgments from traditional imaging examinations [5–7]. However, in procedures of radiomics research, it is a critical step to outline and segment the region of interest (ROI) with the characteristics of time-consuming and susceptible. In order to address the limitations of the manual ROI segmentation approach to optimize the procedures, Isensee et al. [8] proposed the nnU-Net model, as a semantic segmentation model of biomedical 3D images to realize automatic segmentation of ROI, which can optimize research procedures with the advantages of convenience, better consistency and efficiency, and strong generalizability.

Cox proportional hazards deep neural network (DeepSurv)is a prognostic model for survival analysis that aims to investigate the associations between covariates and the effectiveness of various treatments [9]. Compared to traditional survival prediction models, DeepSurv can avoid making overly hypothetical predictions for sample data, and allow for a more comprehensive use of the clinicopathological variables to improve the predictive efficacy.

Nevertheless, there are no reports about Deepsurv prediction models based on preoperative radiomics features from nnU-Net segmentation and clinical variables for prognostic assessment in patients with GBC after curative-intent resection. In this study, we aimed to establish a preoperative clinical radiomics Deepsurv prediction model based on the radiomics features via deep learning to provide a reference basis for preoperative assessment and treatment decisions.

Methods

Patients and design

The study included all histologically confirmed GBC patients treated at the First Affiliated Hospital of Xi'an Jiaotong University between 2011 and 2020.The inclusion criteria were as follows: (1) postoperative pathologically confirmed GBC; (2) margin status recorded microscopically negative or positive (R0/R1); (3) preoperative underwent enhanced CT scan of the upper abdomen in-hospital; (4) data on clinicopathological characteristics and follow-up were available. The exclusion criteria were as follows: (1) patients who received neoadjuvant therapy or other treatments for malignant tumors before surgery; (2) patients with distant metastases; (3) patients who died perioperatively (postoperative survival was less than 30 days). In total, 168 patients met the inclusion/exclusion criteria and were included in the study. Through January 2022, all included patients were evaluated using the 8th edition AJCC staging system.

Follow-up

For all participants in the study, routine follow-up was performed in outpatient and telephone settings. During the first year following surgery, liver function, tumor biomarkers (CEA, CA19-9, CA125), and ultrasound, contrast-enhanced CT or MRI examination were reviewed every 2–3 months, and over a one-year period, follow-ups were conducted once every 3–6 months. We calculated the overall survival (OS) from the date of radical resection to the date of death or last follow-up of the patient, as well as clinical evidence of tumor recurrence. The follow-up ended in January 2022.

Image acquisition

Due to the profile of GBC lesions was susceptible to contrast agents during the arterial phase, this study focused only on the portal vein phase images. Each patient underwent an abdominal CT enhancement examination within one month preoperatively. The scanning equipment was a 256-detector row CT (Philips Healthcare, Netherlands). CT scan parameters included the following. Tube voltage was set at 120 kVp. Tube current was 250 mAs, with tube current modulation (TCM) activated at a DoseRight Index 23. Slice thickness and reconstruction interval were both 5.0 mm. Image reconstruction was performed using iDose 4 at level 3 with a Standard (B) filter. The display method for image windowing parameters was s configured with a window level of 40 and a window width of 300. The contrast agent used was Iodixanol (Visipaque) 320 mgI/mL, administered at a dose calculated as injection volume = weight (kg) × (1.2–1.5) mL/kg, with an injection rate of 2.7–2.8 mL/s. CT scans were performed in the arterial phase (25–35 s post-injection) and the portal vein phase (65–70 s post-injection) following administration of the non-ionic contrast agent. The study's schematic workflow was presented in Fig. 1.

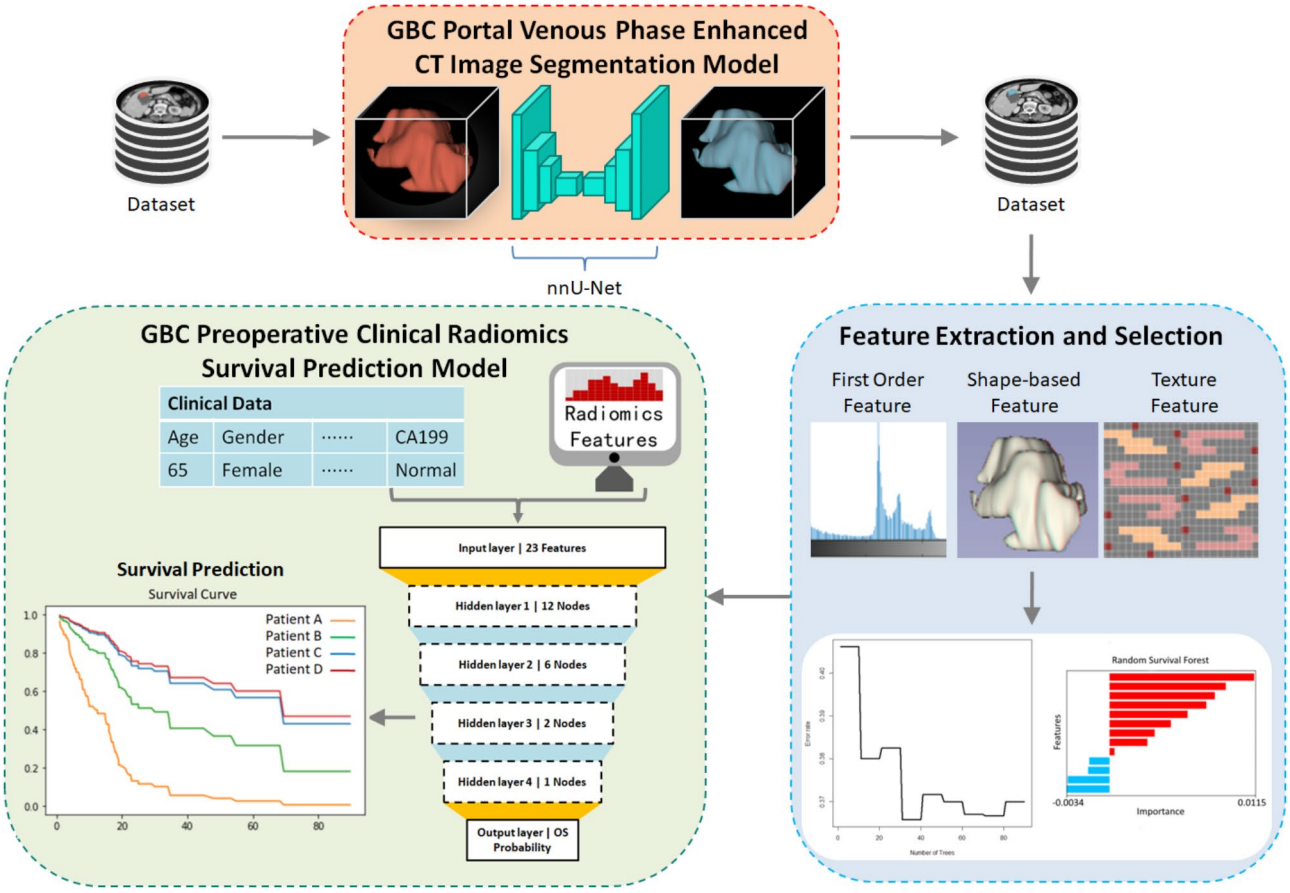

Fig. 1 Schematic workflow of the study

Image segmentation and extraction of features

The segmentation process of the region of interest (ROI) was conducted by two physicians manually (A radiologist as reader 1 with 5 years of abdominal imaging experience and a radiologist as reader 2 with 20 years of abdominal imaging experience reviewed all CT scans to explore the characteristics of each image.). Specifically, reader 1 used 3D Slicer software version 4.11 (https://www.slicer.org/) to label the lesion area, and reader 2 reviewed the segmentation results. Any inconsistent differences were resolved through consultation.

Subsequently, a nnU-Net (https://github.com/MIC-DKFZ/nnUNet) ROI segmentation model was developed based on a training set ($N=126$) and a testing set ($N=42$) in a 7:3 ratio (Baseline characteristics and overall survival comparison between the training and testing sets shown in Supplementary Tables 1 and Supplementary Fig. 1). Then, the model was inferred in the training and testing sets to obtain the result of automatic ROI segmentation. Based on these ROI results, the Pyradiomics package (version 3.0.1) was used to extract the radiomics features, which included 14 shape features, 18 first-order statistical features, 75 texture features (24 GLCM features, 16 GLSZM features, 16 GLRLM features, 5 NGTDM5, SGLDM14) and 1395 higher-order features based on image filter transformation, for a total of 1502 radiomics features.

Then significant radiomics features were screened by 4-step dimensionality reduction methods, including variance method, Pearson analysis, univariate Cox analysis, and random survival forest. And overall pipeline's robustness was assessed by 5-fold cross-validation, and the predictive performance of the segmentation models was assessed by the Dice similarity coefficient and intraclass correlation coefficient (ICC) [10]. The ICC was calculated to assess the consistency between manual and segmentation model-based ROI. Specifically, a two-way random effects model was used to compute the ICC for the radiomics features extracted from the two ROI delineation methods. The ICC calculation was performed using R version 4.1.1 with the irr package (version 0.84.1). An ICC > 0.75 indicated a satisfactory agreement and generalization ability was evaluated by 5-fold cross-validation.

Deepsurv prediction model development and assessment

DeepSurv was established by Python version 3.7.10, the TensorFlow version 1.15.5 deep learning framework, and the TFDeepsurv toolkit version 2.1.0. After adjusting the

hyperparameters according to the training curves, and determining the interval range of each hyperparameter, Bayesian optimization was used as the search strategy for the hyperparameters. The number of search rounds was 3000, and the model was iterated 2000 epochs at each search to determine the regionally optimal solution for each hyperparameter. The predictive ability of the survival prediction model was assessed using C-index, ROC curves and Delong-test, respectively. In addition, the mean and standard deviation of the C-index and the area under the ROC curves (AUC) were estimated through bootstrap. The code was available for use at https://github.com/chickenburge/GBC-prognostic-assessment.

Statistical analysis

All statistical analyses were performed using SPSS version 25, R version 4.1.1, scikit-learn tool library version 0.24.2 of Python version 3.7.10. Normally distributed measurement data were expressed as mean ± standard deviation. The differences between groups were evaluated using Student's t-test (two-sided). Data with skewed distributions were expressed as medians (ranges), and Mann-Whitney tests were conducted to determine the differences. Analyses of categorical variables were conducted using the χ^2 test. Survival analysis was conducted using the Kaplan-Meier method and the Log-rank test. Variables with $P<0.05$ were considered statistically significant.

Results

CT segmentation model of the portal phase of GBC

A 5-fold cross-validation was conducted for the 3D nnU-Net full-resolution segmentation model (with a high image resolution) in the training set, with 1000 epochs of iteration in each fold. Based on the decreasing trend of the loss function curves in red in the training set and the loss function curves in blue in the testing set, indicated that the five segmentation models fit well (Supplementary Fig. 2).

The Dice similarity coefficients for ROI of manual segmentation and nnU-Net model segmentation were (0.92 ± 0.08) and (0.74 ± 0.15) in the training and testing sets, respectively. The results of the Dice similarity coefficient analysis and the legend of the ROI segmentation results (Fig. 2) indicated that the predictive performance of the portal vein phase segmentation for GBC was satisfactory, which could meet the requirements of subsequent radiomics and survival prediction research.

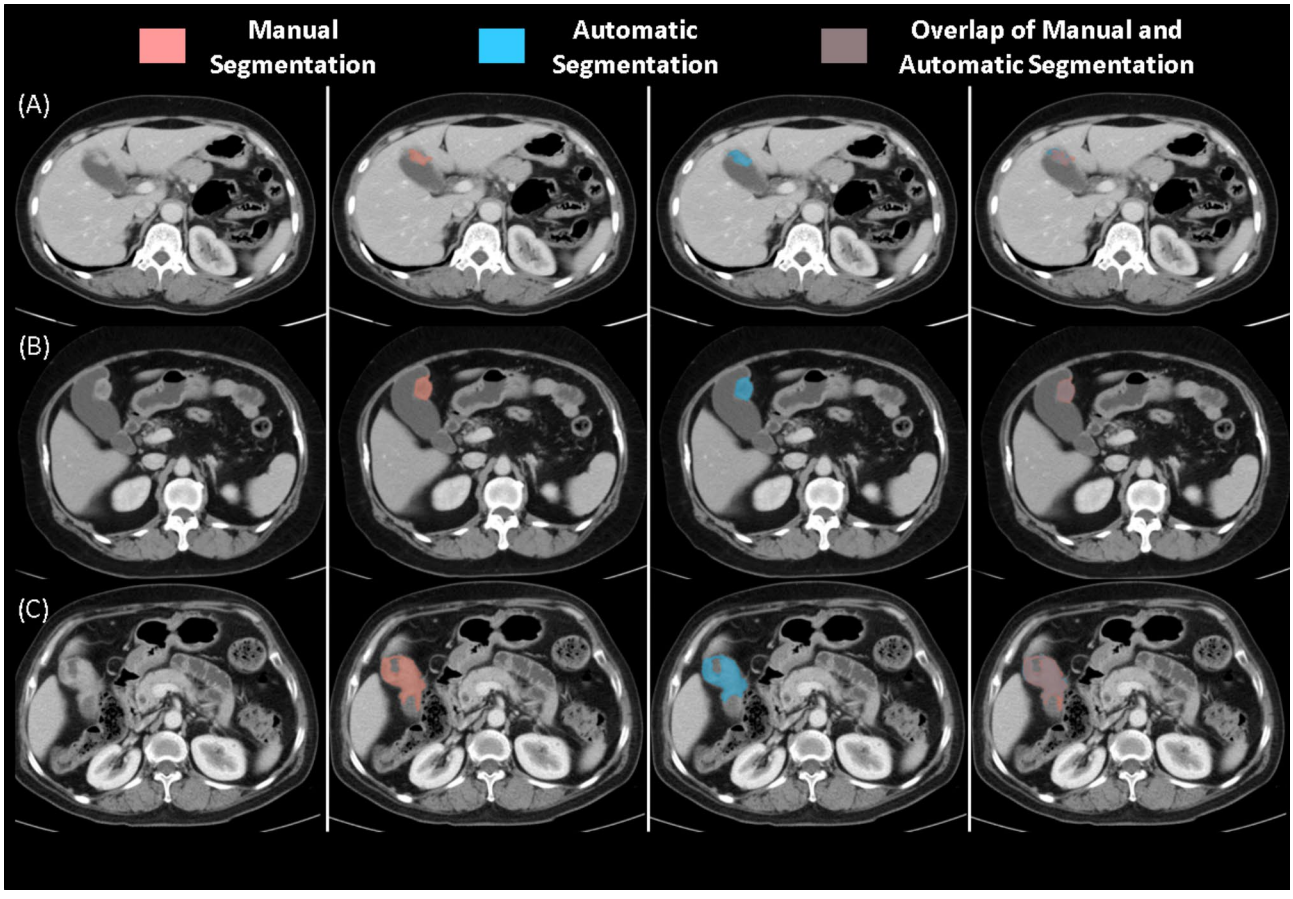

Fig. 2 Comparison of the results of nnU-Net and manual segmentation methods in 3 different patients

Extraction and screening of radiomics features

1502 radiomics features were extracted from the ROI of manual and nnU-Net segmentation, respectively. Consistency analysis was performed on the extracted radiomics features from both methods, and the median ICC value of all features was calculated to be 0.87 (IQR:0.47, 0.94).

In the nnU-Net model segmentation, 13 radiomics features were obtained by a 4-step dimensionality reduction method. The 13 radiomics features include 3 shape features and 10 higher-order features, and the final radiomics feature screening results based on the random forest model are shown in Table 1, the features are listed in descending order of importance, the most important omics feature is original_shape_MinorAxisLength, the importance is 0.01140, and the least important feature is logarithm_firstorder_Kurtosis. The importance is -0.00339.

In the manual segmentation, 13 radiomics features were also obtained through the 4-step dimensionality reduction methods. The 13 radiomics features included 2 shape features and 11 higher-order features, and the final radiomics features screening results were shown in Table 1.

Construction of survival prediction models

Based on the above results, we developed four Deepsurv prediction models including clinical model, radiomics model and two clinical radiomics models established based on the 10 preoperative clinical variables and 13 radiomics features from the nnU-Net model and manual segmentation, respectively.

3000 rounds of the hyperparametric search were conducted by a Bayesian optimized parametric search strategy. The specific hyperparameters of the different Deepsurv prediction models are shown in Supplementary Table 2. The C-index and its 5-fold cross validation (Supplementary Tables 3 and Supplementary Table 4), the AUC of 1-, 2-, and 3-year survival prediction for the nnU-Net based clinical radiomics model were higher than clinical and nnU-Net based radiomics prediction models in the training and testing sets, and close to manual based clinical radiomics model (Fig. 3; Table 2).

The Delong-test was conducted to compare statistical differences of the AUC of 1-year, 2-year and 3-year survival prediction for the two preoperative clinical radiomics DeepSurv prediction models via nnU-Net model and manual segmentation in the testing set. The results showed that the Z-values of the AUC of 1-, 2-, and 3-year survival prediction between the two models were 1.53, 1.42 and 0.49, respectively (all $P > 0.05$). The results

Table 1 Screening results and importance degree of radiomics features

Image segmentation methods	Radiomics features	Importance degree
nnU-Net segmentation	original_shape_MinorAxisLength	0.01140
	original_shape_Maximum2DDiameterSlice	0.00917
	gradient_glrlm_RunLengthNonUniformity	0.00831
	wavelet-HLL_firstorder_Skewness	0.00763
	gradient_firstorder_TotalEnergy	0.00615
	original_shape_Maximum3DDiameter	0.00484
	log-sigma-3-0-mm-3D_gldm_LargeDependenceLowGrayLevelEmphasis	0.00354
	log-sigma-3-0-mm-3D_ngtdm_Busyness	0.00299
	wavelet-LHL_glszm_SizeZoneNonUniformity	0.00039
	wavelet-HLH_glszm_SizeZoneNonUniformity	-0.00161
	wavelet-LHH_gldm_LargeDependenceLowGrayLevelEmphasis	-0.00169
	wavelet-HHH_glrlm_LongRunEmphasis	-0.00329
	logarithm_firstorder_Kurtosis	-0.00339
Manual segmentation	log-sigma-3-0-mm-3D_glszm_LargeAreaHighGrayLevelEmphasis	0.01954
	square_firstorder_TotalEnergy	0.01162
	wavelet-HLL_firstorder_Kurtosis	0.00571
	gradient_glszm_ZoneVariance	0.00533
	original_shape_Maximum2DDiameterRow	0.00505
	original_shape_Maximum3DDiameter	0.00338
	wavelet-LLL_firstorder_Minimum	0.00234
	log-sigma-3-0-mm-3D_firstorder_Kurtosis	0.00019
	wavelet-HLL_firstorder_Skewness	0.00013
	log-sigma-3-0-mm-3D_gldm_LargeDependenceLowGrayLevelEmphasis	-0.00113
	wavelet-HLH_glszm_GrayLevelNonUniformity	-0.00136
	gradient_firstorder_Skewness	-0.00161
	wavelet-HLL_glszm_GrayLevelNonUniformity	-0.00299

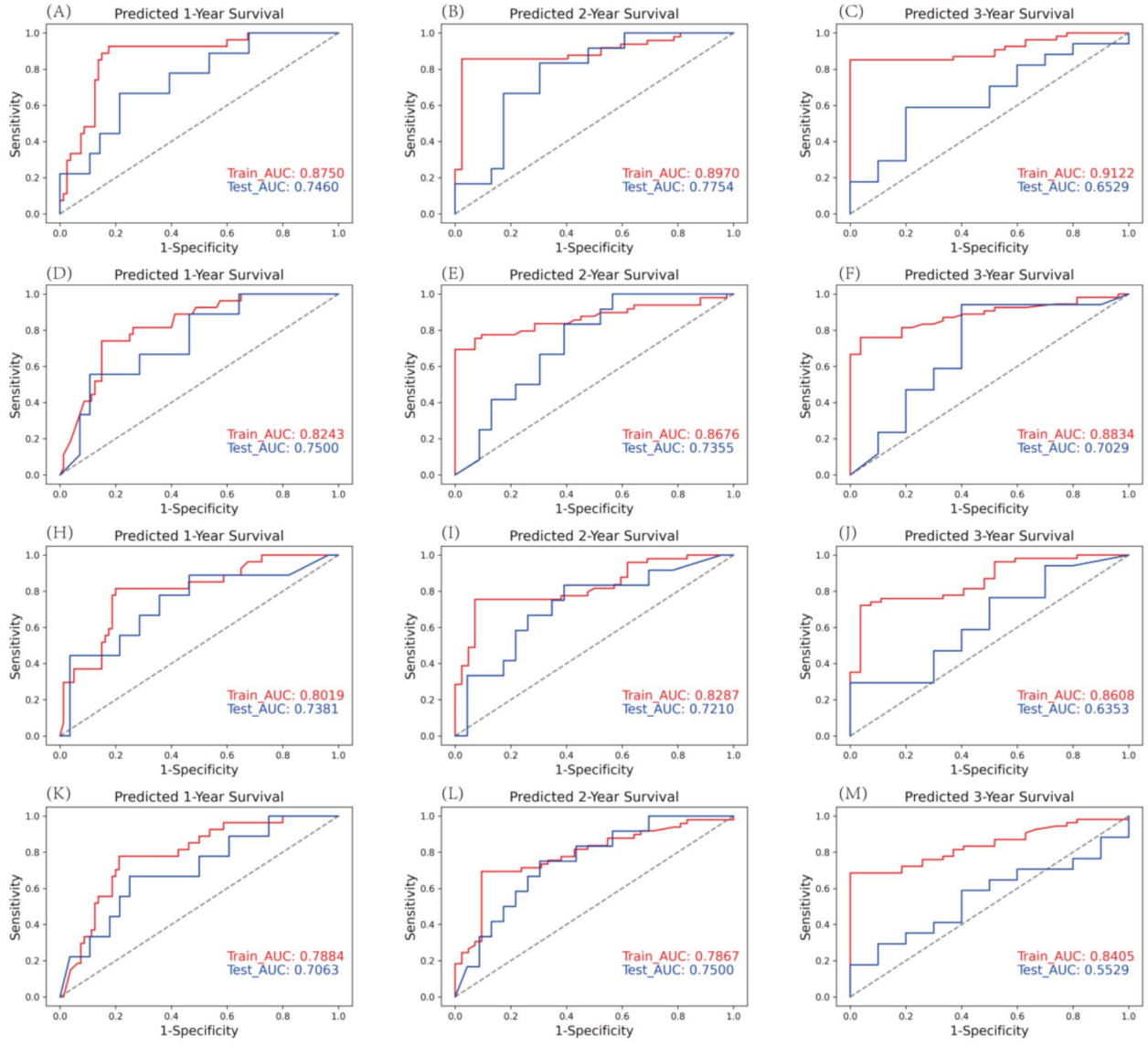

Fig. 3 Receiver operating characteristic curves for different survival prediction models predicting overall survival at 1-, 2- and 3-year. (**A-C**). nnU-Net based clinical radiomics Deepsurv model; (**D-F**). Manual based clinical radiomics Deepsurv model; (**H-J**) Clinical Deepsurv model; (**K-M**). nnU-Net based radiomics Deepsurv model

showed that the DeepSurv prediction models based on nnU-Net and manual segmentation had the same prediction efficiency.

Predictive value of radiomics features and preoperative clinical features

In order to validate the predictive value of preoperative clinical and radiomics features on individual patient survival, four patients with close preoperative clinical variables were selected in the testing set (Table 3).

The Clinical Deepsurv model relies on preoperative clinical variables (including age, gender, total bilirubin levels, tumor marker levels, Gallstones, lymph node status, liver invasion, and other organ invasion in Supplementary Table 1) to predict postoperative survival probability over time. Due to the similarity in the clinical pathological features of four patients (Table 3), the survival curves predicted by the model for these four patients are overlapping, resulting in a single combined survival curve (Fig. 4A).

The nnU-Net based radiomics Deepsurv model predicts survival based on 13 radiomics features extracted from preoperative portal vein phase images. The model produces two overlapping survival curves for the four patients (one for Patient A, and another for Patients B, C, and D) (Fig. 4B), indicating that it performs better than

Table 2 The C-index and AUC of different survival prediction models

	nnU-Net based clinical radiomics Deepsurv model		Manual based clinical radiomics Deepsurv model		Clinical Deepsurv model		nnU-Net based radiomics Deepsurv model	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing
C-index	0.82(0.82 ± 0.03)	0.74(0.74 ± 0.05)	0.77(0.77 ± 0.03)	0.71(0.71 ± 0.05)	0.77(0.76 ± 0.03)	0.69(0.69 ± 0.07)	0.75(0.75 ± 0.03)	0.67(0.67 ± 0.06)
5-fold cross validation	0.79 ± 0.01	0.70 ± 0.01	0.78 ± 0.04	0.69 ± 0.04	0.77 ± 0.01	0.66 ± 0.05	0.75 ± 0.01	0.67 ± 0.02
1-year survival AUC	0.88(0.90 ± 0.02)	0.75(0.82 ± 0.05)	0.82(0.88 ± 0.02)	0.75(0.71 ± 0.06)	0.80(0.85 ± 0.03)	0.74(0.73 ± 0.05)	0.79(0.81 ± 0.03)	0.71(0.69 ± 0.06)
2-year survival AUC	0.90(0.89 ± 0.03)	0.78(0.82 ± 0.05)	0.87(0.88 ± 0.03)	0.74(0.73 ± 0.07)	0.83(0.83 ± 0.03)	0.72(0.75 ± 0.05)	0.79(0.81 ± 0.04)	0.75(0.71 ± 0.07)
3-year survival AUC	0.91(0.91 ± 0.02)	0.65(0.77 ± 0.07)	0.88(0.90 ± 0.02)	0.70(0.74 ± 0.06)	0.86(0.87 ± 0.03)	0.64(0.68 ± 0.07)	0.84(0.84 ± 0.03)	0.55(0.65 ± 0.08)

the preoperative clinical Deepsurv model, although it still cannot provide highly accurate predictions.

The nnU-Net based clinical radiomics Deepsurv model integrates both preoperative clinical features and radiomics features, resulting in four distinct survival curves. Patients C and D have similar survival outcomes and relatively good prognosis, while Patient A has the worst survival condition. The survival status of Patients B falls in between those of Patients C, D, and A (Fig. 4C), which is consistent with the actual survival outcomes of the four patients (Table 3). The results indicated that the preoperative clinical radiomics model could assess and stratify postoperative survival outcomes for individual patient. Hence, preoperative radiomics features can complement and imply postoperative pathological and clinical information.

Discussion

Many studies have examined the factors that affect postoperative survival in GBC, including age, gender, preoperative bilirubin levels, tumor biomarker levels, pathological TNM stage, tumor differentiation, number of lymph nodes dissected and postoperative complications [11–13], and also focused on combining preoperative and postoperative variables to construct survival prediction model. However, few studies discussed the predictive value of preoperative clinical variables and information for patients with GBC after radical resection, which may be attributed to the difficulties of preoperative information to fully characterize the biology and extent of disease progression compared to postoperative pathological features. However, as a radiological examination, CT provides a comprehensive picture of the tumor status preoperatively, such as tumor size and degree of lesion invasion related to disease progression, as well as tumor texture associated with pathology [14]. Therefore, it is imperative to mine and integrate preoperative information for prognostic assessment.

For medical radiology, the advent of radiomics paved the way for a major leap from qualitative to quantitative analysis, which is an effective method of enhancing clinical advantages. In the study, radiomic features could complement empirical qualitative analyses of enhanced CT and other preoperative clinical data in providing important prognostic information, thus improving the predictive ability of survival prediction models. Radiomics research relies on accurate ROI segmentation, and segmenting medical images is one of the main challenges [15]. Manual segmentation, as the most used method for ROI segmentation in radiomics research with the limitation of being time-consuming and labor-intensive, is difficult to implement in large sample datasets and avoid individual subjective interference factors. In this study, we constructed a semantic segmentation model

Table 3 Clinicopathological features of four patients

	Patient A	Patient B	Patient C	Patient D
Preoperative variables				
Age (year)	66	76	68	76
Gender	Male	Male	Female	Male
TBIL (umol/L)	11.20	12.40	16.70	13.10
CEA (ng/ml)	0.20	2.11	2.33	1.18
CA19-9 (U/ml)	4.00	16.48	48.93	7.00
CA-125 (U/ml)	7.62	14.90	7.20	15.26
Gallstones	No	No	No	Yes
Image lymph node status	No	No	No	No
Image liver invasion	No	No	No	No
Image other organs invasion	No	No	No	No
Postoperative pathological examination				
AJCC 8th edition T stage	3	3	3	3
AJCC 8th edition N stage	0	2	0	1
Margin status	R0	R0	R0	R0
Tumor differentiation	Poor	Moderate	Moderate	Poor
Histopathological classification	Adenocarcinoma	Adenocarcinoma	Adenocarcinoma	Adenocarcinoma
Survival status	death	death	death	death
Overall survival (month)	9	23	31	31

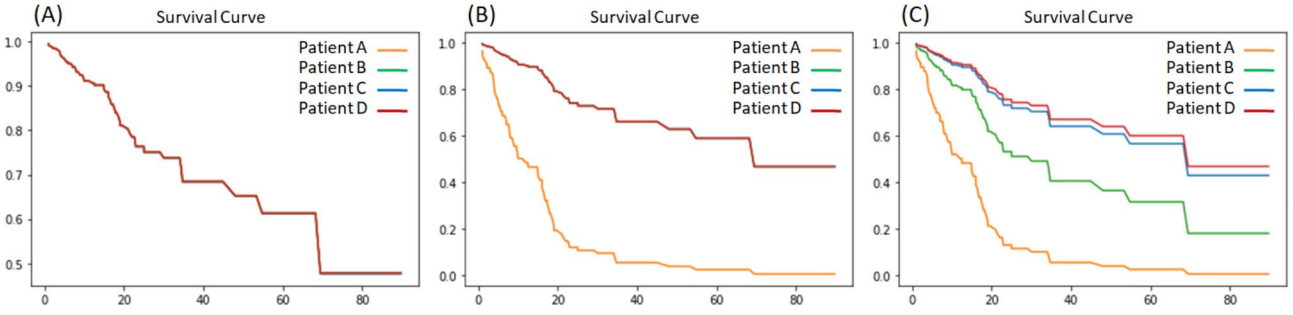

Fig. 4 Survival curves of different survival prediction models for postoperative survival prediction in four patients. (**A**). Clinical Deepsurv model; (**B**). nnU-Net based radiomics Deepsurv model; (**C**). nnU-Net based clinical radiomics Deepsurv model

via the nnU-Net deep neural network for the portal vein phase. The mean Dice similarity coefficient of the nnU-Net segmentation model was (0.74 ± 0.15) in the testing set, which indicated the model was with a satisfactory performance. Delong-test showed that the prediction performance of the ROI results segmented via the nnU-Net model was close to manual segmentation. Shu et al. [16] developed 2 segmentation models for pituitary adenomas in MRI T1-weighted phase images via the nnU-Net model, and the Dice similarity coefficients were both higher than 0.80. Song et al. [17] also achieved a Dice similarity coefficient of 0.82 based on the nnU-Net segmentation model for lung nodules. However, a colorectal tumor segmentation model via the nnU-Net model with the Dice similarity coefficient of 0.74 was close to the results of our study [18]. Cavity organ tumors always have irregular shapes and are more difficult to segment than solid organ tumors, which increases the limitations and difficulty of segmenting. Therefore, a supplemental sample size will be required to examine the advantages and disadvantages of different deep learning segmentation models.

13 radiomics features after 4-step dimensionality reduction methods were obtained by screening from 1502 radiomics features in the study, including three shape features, namely the short-axis length, maximum two-dimensional diameter (cross-section), and maximum three-dimensional diameter of the ROI, which prompted the potential information including the tumor shape, tumor size and the extent invasion of surrounding organs of the lesion. The other 10 higher-order features were mostly textural features based on images derived under different filters, which were difficult to interpret their physical meaning through conventional experience. Kuess et al. [19] found that the textural features of the MRI images correlated with the pathological nature of the prostate carcinoma lesions from 25 patients. Aydos et al. [20] also analyzed the correlation between PET-CT images and tumor histopathological findings in 90 patients of non-small cell lung cancer, and it found

that textural features correlated with tumor tissue type, degree of tumor differentiation and degree of tissue necrosis, and could predict tumor type and mediastinum lymph node metastasis. Gupta et al. [21] found a correlation between the medium texture scale and tumor pathological differentiation in patients with GBC, providing a new direction on tumor differentiation for preoperative assessment. Therefore, preoperative radiomics features can supplement potential pathological and prognostic information.

At present, there is a paucity of research on the prognostic assessment for GBC based on radiomics. Liu et al. [22] established a nomogram prediction model based on clinicopathological and radiomic feature with a 1-, 3-year AUC of 0.73 for postoperative survival prediction. In contrast, our study achieved AUCs of 0.82 ± 0.05 and 0.77 ± 0.07 for 1-year and 3-year postoperative survival predictions, respectively, and a C-index of 0.74 ± 0.05, in the test set (Table 2). The survival prediction model based on deep learning in our study outperforms previous traditional machine learning methods and can fulfill more individualized survival predictions. Meng et al. [23] integrated preoperative contrast-enhanced CT with preoperative and postoperative clinicopathological information to develop a Deep Learning Radiomic Nomogram incorporating clinical, radiomics, and deep learning signatures, achieving a C-index of 0.736. In this study, comparable predictive performance was achieved using only preoperative information from patients. The sample size remains a major limitation in deep learning research. In future studies with larger sample sizes, research approaches leveraging deep learning features hold significant promise. Moreover, Xiang et al. [24] developed a nomogram prediction model based on radiomics signature, pathological T stage, N stage and tumor differentiation grade for the assessment of postoperative recurrence of GBC, which has limited value for clinical treatment decisions preoperatively despite its good predictive ability.

Compared to reported studies on GBC survival prediction based on radiomics, we developed a nnU-Net segmentation model and Deepsurv prediction model based on the portal vein phase CT images, more convincing and innovative, and the C-index of the nnU-Net based clinical radiomics model was higher than clinical and radiomics prediction models in the training and testing sets, and close to manual based clinical radiomics model, with the best prognostic assessment ability. By comparing four patients with similar preoperative clinical variables, we found that the prognosis of the patients was still different owing to pathological features such as N stage and tumor differentiation variations (Table 3). We found that the preoperative nnU-Net based clinical radiomics model could assess and stratify postoperative survival outcomes for individual patient, but the preoperative clinical Deepsurv prediction model was unable to distinguish prognosis for different patient in the study cohort. Therefore, when radiomics features were added to the clinical radiomics model, the prediction ability of the clinical radiomics model was correspondingly improved to discriminate prognosis for individual patient.

There were several limitations in our study. It was a single-center retrospective study with a small sample that would need to be validated with a prospective study. Furthermore, we chose only portal vein phase images for analysis in this study, and we were unable to combine arterial phase images with portal vein phase images. Additionally, it still needs to be further explored the predictive ability of clinical radiomics survival models based on different machine learning algorithms to provide decision support for the clinical diagnosis and treatment of GBC by a multicenter, prospective, and large-scale study.

Conclusion

In summary, this study confirmed that the preoperative nnU-Net based clinical radiomics DeepSurv model for GBC performed better than both the preoperative clinical and the radiomics DeepSurv prediction models, which has a certain survival predictive value for GBC patients after radical resection and provides a reference for preoperative diagnosis and treatment decisions.

Abbreviations

GBC	Gallbladder carcinoma
ROI	Region of interest
DeepSurv	Deep neural network
ICC	Intraclass correlation coefficient

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-025-13711-1.

Supplementary Material 1

Acknowledgements

No further acknowledgments.

Author contributions

Z. Geng and Q. Li conceived and designed the experiments. Z. Jin and Q. Li performed the experiments. C. Chen, D. Zhang, and M. Yang offered, interpreted, and analyzed the data. Z. Jin, Q. Li, Q. Wang, Z. Cai and S. Si conducted statistical analysis and interpreted the results of the analysis. Z. Jin and Q. Li wrote the paper. Z. Geng reviewed the manuscript. All authors read and approved the final manuscript.

Funding

Supported by the National Natural Science Foundation of China (No. 62076194); Key Research and Development Program of Shaanxi Province (No.2021-SF-016 and 2022-SF-606); and Institutional Foundation of the First Affiliated Hospital of Xi'an Jiaotong University (No.2024-QN-15).

Data availability

The database of the study is available in the corresponding author and will be sent when requested by e-mail.

Declarations

Ethics approval and consent to participate
The study was approved by the ethics committee of the First Affiliated Hospital of Xi'an Jiaotong University (No: XJTU1AF2022LSK-089), Xi'an, China. Written informed consent was obtained from all included patients and their families before study enrollment.

Consent for publication
Not applicable.

Statement
All methods were carried out in accordance with relevant guidelines and regulations in the study.

Competing interests
The authors declare no competing interests.

Received: 4 January 2024 / Accepted: 11 February 2025
Published online: 25 February 2025

References

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68:394–424.
2. Hundal R, Shaffer EA. Gallbladder cancer: epidemiology and outcome. Clin Epidemiol. 2014;6:99–109.
3. Huang J, Patel HK, Boakye D, et al. Worldwide distribution, associated factors, and trends of gallbladder cancer: a global country-level analysis. Cancer Lett. 2021;521:238–51.
4. Lindnér P, Holmberg E, Hafström L. Gallbladder cancer - no improvement in survival over time in a Swedish population. Acta Oncol. 2018;57:1482–9.
5. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer. 2012;48:441–6.
6. Xu X, Zhang HL, Liu QP, et al. Radiomic analysis of contrast-enhanced CT predicts microvascular invasion and outcome in hepatocellular carcinoma. J Hepatol. 2019;70:1133–44.
7. Yu Y, Tan Y, Xie C, et al. Development and validation of a preoperative magnetic resonance imaging radiomics-based signature to predict axillary lymph node metastasis and disease-free survival in patients with early-stage breast cancer. JAMA Netw Open. 2020;3:e2028086.
8. Isensee F, Jaeger PF, Kohl S, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods. 2021;18:203–11.
9. Katzman JL, Shaham U, Cloninger A, et al. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Med Res Methodol. 2018;18:24.
10. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15:155–63.
11. Goetze TO. Gallbladder carcinoma: prognostic factors and therapeutic options. World J Gastroenterol. 2015;21:12211–7.
12. Oven UB, Bilici A, Seker M, et al. Prognostic factors for operated gallbladder cancer. J Gastrointest Cancer. 2019;50:451–7.
13. Kayahara M, Nagakawa T, Nakagawara H, et al. Prognostic factors for gallbladder cancer in Japan. Ann Surg. 2008;248:807–14.
14. An H, Wang Y, Wong E, et al. CT texture analysis in histological classification of epithelial ovarian carcinoma. Eur Radiol. 2021;31:5050–8.
15. Kumar V, Gu Y, Basu S, et al. Radiomics: the process and the challenges. Magn Reson Imaging. 2012;30:1234–48.
16. Shu X, Zhou Y, Li F, et al. Three-dimensional semantic segmentation of pituitary adenomas based on the deep learning framework-nnU-Net: a clinical perspective. Micromachines (Basel). 2021;12:1473.
17. Song J, Huang SC, Kelly B, et al. Automatic lung nodule segmentation and intra-nodular heterogeneity image generation. IEEE J Biomed Health Inf. 2022;26:2570–81.
18. Zheng S, Lin X, Zhang W, et al. MDCC-Net: multiscale double-channel convolution U-Net framework for colorectal tumor segmentation. Comput Biol Med. 2021;130:104183.
19. Kuess P, Andrzejewski P, Nilsson D, et al. Association between pathology and texture features of multi parametric MRI of the prostate. Phys Med Biol. 2017;62:7833–54.
20. Aydos U, Ünal ER, Özçelik M, et al. Texture features of primary tumor on (18) F-FDG PET images in non-small cell lung cancer: the relationship between imaging and histopathological parameters. Rev Esp Med Nucl Imagen Mol (Engl Ed). 2021;40:343–50.
21. Gupta P, Rana P, Ganeshan B, et al. Computed tomography texture-based radiomics analysis in gallbladder cancer: initial experience. Clin Exp Hepatol. 2021;7:406–14.
22. Liu Z, Zhu G, Jiang X, et al. Survival prediction in gallbladder cancer using CT based machine learning. Front Oncol. 2020;10:604288.
23. Meng F, Zhang J, Guo Y, et al. Contrast-enhanced CT-based deep learning radiomics nomogram for the survival prediction in gallbladder cancer. Acad Radiol. 2024;31:2356–66.
24. Xiang F, Liang X, Yang L, et al. Contrast-enhanced CT radiomics for prediction of recurrence-free survival in gallbladder carcinoma after surgical resection. Eur Radiol. 2022;32:7087–97.

Publisher's note