Retrospective Cohort Study

INTERNATIONAL JOURNAL OF SURGERY

OPEN

A CT-based deep learning model predicts overall survival in patients with muscle invasive bladder cancer after radical cystectomy: a multicenter retrospective cohort study

Zongjie Wei, MD[a], Yingjie Xv, MD[a], Huayun Liu, MM[a], Yang Li, MD[a], Siwen Yin, MM[c], Yongpeng Xie, MD[a], Yong Chen, MD[c], Fajin Lv, MD[b], Qing Jiang, MD[d], Feng Li, MD[e,*], Mingzhao Xiao, MD[a,*]

Background: Muscle invasive bladder cancer (MIBC) has a poor prognosis even after radical cystectomy (RC). Postoperative survival stratification based on radiomics and deep learning (DL) algorithms may be useful for treatment decision-making and follow-up management. This study was aimed to develop and validate a DL model based on preoperative computed tomography (CT) for predicting postcystectomy overall survival (OS) in patients with MIBC.

Methods: MIBC patients who underwent RC were retrospectively included from four centers, and divided into the training, internal validation, and external validation sets. A DL model incorporated the convolutional block attention module (CBAM) was built for predicting OS using preoperative CT images. The authors assessed the prognostic accuracy of the DL model and compared it with classic handcrafted radiomics model and clinical model. Then, a deep learning radiomics nomogram (DLRN) was developed by combining clinicopathological factors, radiomics score (Rad-score) and deep learning score (DL-score). Model performance was assessed by C-index, KM curve, and time-dependent ROC curve.

Results: A total of 405 patients with MIBC were included in this study. The DL-score achieved a much higher C-index than Rad-score and clinical model (0.690 vs. 0.652 vs. 0.618 in the internal validation set, and 0.658 vs. 0.601 vs. 0.610 in the external validation set). After adjusting for clinicopathologic variables, the DL-score was identified as a significantly independent risk factor for OS by the multivariate Cox regression analysis in all sets (all $P < 0.01$). The DLRN further improved the performance, with a C-index of 0.713 (95% CI: 0.627–0.798) in the internal validation set and 0.685 (95% CI: 0.586–0.765) in external validation set, respectively.

Conclusions: A DL model based on preoperative CT can predict survival outcome of patients with MIBC, which may help in risk stratification and guide treatment decision-making and follow-up management.

Keywords: deep learning, muscle invasive bladder cancer, survival prediction, tomography, X-ray computed

Introduction

Bladder cancer is the 10th most common malignant tumor worldwide, characterized by high morbidity and mortality[1]. Muscle invasive bladder cancer (MIBC) refers to the depth of tumor infiltration invading the muscularis propria (stage T2) or the peripheral tissues of the bladder (stages T3 and T4), and it accounts for about 25–30% of new cases of bladder cancer[2,3]. Despite the rapid development of current treatment options for bladder cancer, the prognosis for patients with MIBC remains poor[4]. The existing prognostic stratification methods are mainly based on the established TNM tumor staging system[5,6].

[a]Department of Urology, [b]Department of Radiology, The First Affiliated Hospital of Chongqing Medical University, [c]Department of Urology, Chongqing University Fuling Hospital, [d]Department of Urology, The Second Affiliated Hospital of Chongqing Medical University and [e]Department of Urology, Chongqing University Three Gorges Hospital, Chongqing, People's Republic of China

Zongjie Wei and Yingjie Xv contributed equally to this work.

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

*Corresponding authors. Address: Department of Urology, The First Affiliated Hospital of Chongqing Medical University, No. 1 Youyi Road, Yuzhong District, Chongqing 400016, People's Republic of China. Tel.: +861 360 839 9433. E-mail: mingzhaoxiao@cqmu.edu.cn (M. Xiao); Department of Urology, Chongqing University Three Gorges Hospital, No. 165, Xincheng Road, Wanzhou District, Chongqing 404100, People's Republic of China. Tel.: +861 399 666 5888. E-mail: 18581354134@163.com (F. Li).

International Journal of Surgery (2024) 110:2922–2932

Received 28 November 2023; Accepted 31 January 2024

Supplemental Digital Content is available for this article. Direct URL citations are provided in the HTML and PDF versions of this article on the journal's website, www.lww.com/international-journal-of-surgery.

Published online 13 February 2024

http://dx.doi.org/10.1097/JS9.0000000000001194

However, prognostic differences have been observed even in patients with the same stage. Therefore, accurate prognostic risk stratification of patients with MIBC would be helpful in guiding therapeutic decision-making and developing follow-up strategies.

Computed tomography (CT) is commonly used for the diagnosis and staging of bladder cancer, but the results are usually based on the experience of radiologist with low interobserver consistency[7,8]. Radiomics, a novel quantitative image analysis method, can provide a noninvasive tool for identifying the biological behavior of tumors by automatically extracting and analyzing high-throughput quantitative data from medical images (e.g. CT, MRI, or positron emission tomography (PET))[9]. Radiomics has proven to be promising in the automatic diagnosis and prognostic prediction of colorectal tumors[10], hepatocellular carcinoma[11,12], breast lesions[13] and some other diseases. With the development of computer vision, deep learning (DL) can automatically extract deeper and more comprehensive features and develop model from raw images through multilevel convolution[14–16]. In contrast to radiomics, DL does not require radiologists to manually label region of interest (ROI) regions and explicitly specify rules of features. DL has even achieved excellent performance in medical image analysis tasks for various tumors for nearly a decade[17–19]. In addition, many studies have also found that attention-based convolutional neural network (CNN), which can focus the network's attention on objects of interest, can achieve advanced performance in many tasks[20,21]. These technological advances in image analysis show promise in identifying tumor categories, differentiating pathological grades, predicting treatment efficacy, and may be helpful in risk stratification of patients with MIBC using preoperative enhanced CT and clinical data. Recently, radiomics and DL technologies based on CT and MRI have been shown to be useful in predicting muscular infiltration[22,23], tumor grade[24], and postoperative recurrence[25,26] in bladder cancer. However, few studies have focused on survival prediction, which is very important for decision-making on treatment options. These previous studies were single-center studies and just relied on handcrafted radiomics features[26–28]. To the best of our knowledge, there is currently no validated DL model based on preoperative enhanced CT image for predicting overall survival (OS) outcome in patients with MIBC.

The aims of this study were to develop and validate an attention-based DL model to predict OS in patient with MIBC using preoperative CT images. In addition, we constructed a deep learning radiomics nomogram (DLRN) to further enhance the efficacy of the model by integrating deep learning score, radiomics score, and clinicopathological factors.

HIGHLIGHTS

- We developed and validated an attention-based deep learning (DL) model to predict overall survival in patient with muscle invasive bladder cancer using preoperative computed tomography images.
- The proposed DL model showed promising performance and outperformed classic handcrafted radiomics model.
- A deep learning radiomics nomogram (DLRN), integrating DL-score, radiomics score, and clinicopathological factors, achieved better performance.

Patients and methods

Patients

This study was approved by the Ethics Committee of our institutions, and the requirement for informed consent was waived. The study was registered on the ClinicalTrials.gov network. This study has been reported in line with the strengthening the reporting of cohort studies in surgery (STROCSS) (Supplemental Digital Content 1, http://links.lww.com/JS9/B907) criteria[29] and in according to the Declaration of Helsinki.

We retrospectively analyzed the data of patients who underwent radical cystectomy with pathologically confirmed MIBC from four independent hospitals. The inclusion criteria were (1) pathologically confirmed MIBC; (2) underwent radical cystectomy; (3) contrast-CT scans available within 2 weeks before surgery. The exclusion criteria were (1) received neoadjuvant chemotherapy; (2) pathologically confirmed nonurothelial carcinoma; (3) with other malignant neoplasms; (4) missing CT data or poor image quality; and (5) incomplete clinical and follow-up data. Finally, a total of 317 patients were enrolled in this study from center 1 and center 2 between June 2011 and June 2022. These patients were then randomly divided into a training set ($n = 221$) and an internal validation set ($n = 96$) by a 7:3 ratio for model construction. In external validation set, we retrospectively included 88 MIBC patients from center 3 and center 4 between October 2013 and August 2020. A flowchart of study population selection process is shown in Figure 1A. Clinical and pathological data were collected from the electronic medical record, including age, sex, history of NMIBC status, number of tumors, pathologic T stage, pathologic N stage, pathological grade, and lymphovascular invasion. Pathologic tumor staging was performed according to the American Joint Committee on Cancer Staging Manual (8th edition).

In this study, the primary endpoint was OS, defined as the time from surgery to death from any cause. Postoperative follow-up was performed according to the international guidelines for bladder cancer[8,30]. The overall workflow of this study is summarized in Figure 1.

Image collection and preprocess

The CT scan images were downloaded from Picture Archiving and Communication Systems (PACS) and saved as the original Digital Imaging and Communications in Medicine (DICOM) format. The nephrographic phase (NP) CT images, the most common imaging for tumor identification in bladder cancer, was chosen for subsequent analysis. The details of the CT image acquisition parameters for the different centers are shown in Supplementary Table S1 (Supplemental Digital Content 2, http://links.lww.com/JS9/B908).

The ROI of the tumor were manually segmented on NP CT images by a urological radiologist with 5 years of experience via the ITK-SNAP software (version 3.6.0, http://www.itksnap.org). All ROIs were examined by a senior radiologist with more than 15 years of experience in image processing for bladder cancer. The two-dimensional (2D) slice with the largest tumor on z-axis was selected as the input data per patient. All images were filtered with a window of [-50, 150] HU. We cropped the bounding box focused on the tumor with an expansion of five pixels at the outer edge of the ROIs. Then, all cropped images were resized to 224×224 and the pixel values were normalized to [0, 1]. To

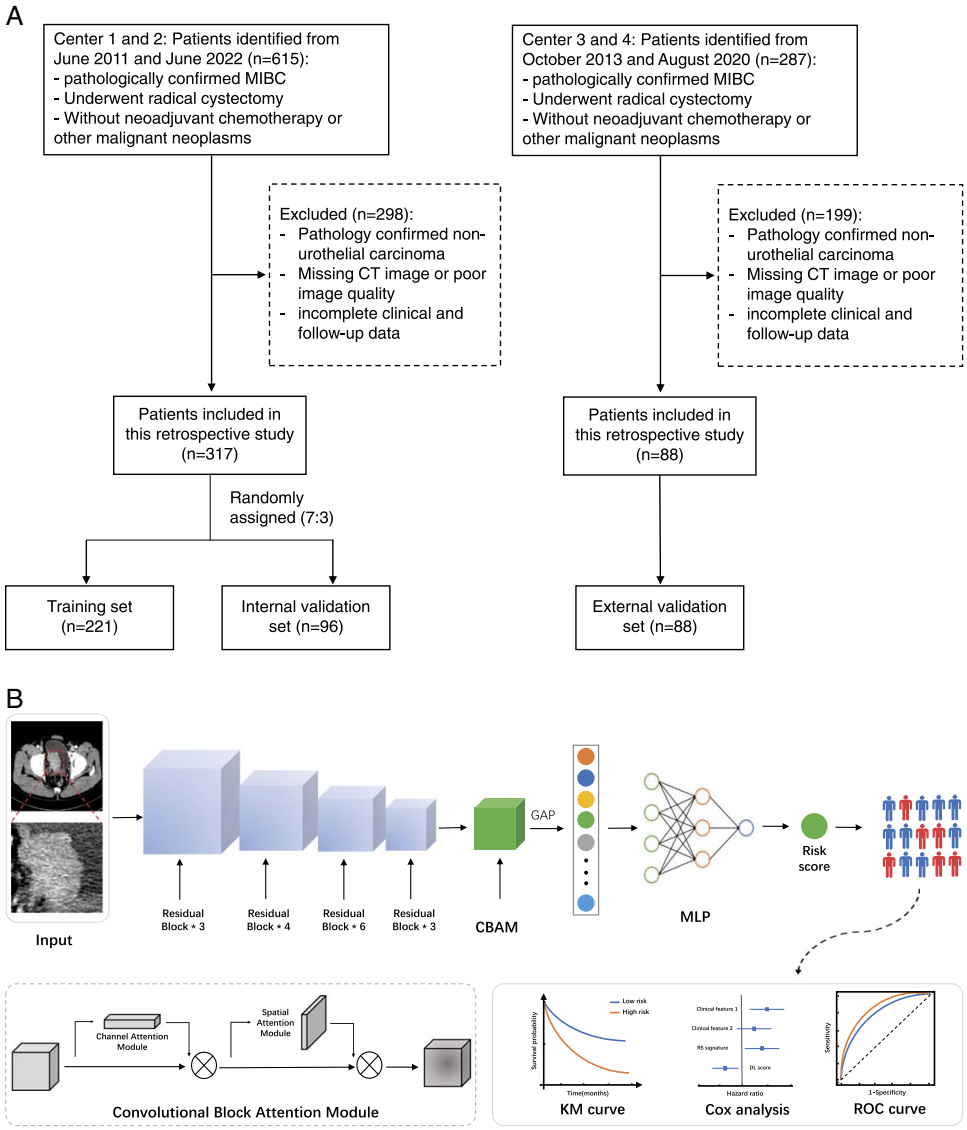


Figure 1. The overall workflow of this study. The detail flowchart of patient recruitment process (A). The schematic diagram of CBAM-based convolutional neural network architecture, and the architecture of CBAM is shown in the lower left corner (B). CBAM, convolution block attention module; DL, deep learning; GAP, global average pooling layer; MLP, multilayer perception; ROI, region of interest; ROC, receiver operating characteristic.

improve the performance of the model and prevent network overfitting, random data augmentation methods, including flipping, rotating and scaling, were used in the training set.

DL model architecture

To develop a DL model for predicting OS of bladder cancer, a convolutional neural network (CNN) of pretrained ResNet-50 architecture was applied as a backbone network. A convolutional block attention module (CBAM) was added behind the last convolutional layer to improve the model's perceptual ability of the tumor region[31]. CBAM consists of channel and spatial attention modules that help to enhance the feature representation of different channels and extract critical information about different locations in space, respectively. The more details of ResNet-50 and CBAM can be found in Supplementary Material (Supplemental Digital Content 2, http://links.lww.com/JS9/B908). Then, the feature maps of the CT images were obtained by global average pooling. A series of multilayer perceptron (MLP) with a hidden layer of 256 neurons, activation function (ReLU) layer and batch normalization layer constituted the top layer of the neural network. The output layer consisted of only one neuron, which served as a prognostic risk score for the patient.

The DL model was trained on a Nvidia GeForce RTX 3090 GPU with the Pytorch framework. The weights and biases of the model were initialized using the weights of the pretrained model from ImageNet. To calculate the difference in survival risk between patients during optimization, we customized the loss function based on Cox partial likelihood, inspired by the DeepSurv model[32]. Further details about the loss function can be found in the Supplementary Materials (Supplemental Digital Content 2, http://links.lww.com/JS9/B908). For the training stage, the initial learning rate was initially set to 0.0001 and

Table 1

The baseline clinical characteristics of patients in three sets.

Characteristic	All patients ($n=405$)	Training set ($n=221$)	Internal validation set ($n=96$)	External validation set ($n=88$)	*P*
Age					0.719
≤65 years	184 (45.4%)	104 (47.1%)	43 (44.8%)	37 (42.0%)	
>65 years	221 (54.6%)	117 (52.9%)	53 (55.2%)	51 (58.0%)	
Sex					0.123
Female	45 (11.1%)	22 (10.0%)	8 (8.3%)	15 (17.0%)	
Male	360 (88.9%)	199 (90.0%)	88 (91.7%)	83 (83.0%)	
History of NMIBC					0.066
No	333 (82.2%)	186 (84.2%)	82 (85.4%)	65 (73.9%)	
Yes	72 (17.8%)	35 (15.8%)	14 (14.6%)	23 (26.1%)	
Number of tumors					0.294
Single	249 (61.5%)	134 (60.6%)	55 (57.3%)	60 (68.2%)	
Multiple	156 (38.5%)	87 (39.4%)	51 (42.7%)	28 (31.8%)	
pT stage					0.849
T2	268 (66.2%)	148 (67.0%)	64 (66.7%)	56 (63.7%)	
T3	104 (25.7%)	50 (22.6%)	25 (26.0%)	26 (29.5%)	
T4	33 (8.1%)	23 (10.4%)	7 (7.3%)	6 (6.8%)	
pN stage					0.403
N0/Nx	355 (87.8%)	194 (87.8%)	87 (90.6%)	74 (84.1%)	
≥N1	50 (12.3%)	27 (12.2%)	9 (9.4%)	14 (15.9%)	
Pathological grade					0.446
Low	29 (7.2%)	16 (7.2%)	9 (9.4%)	4 (4.5%)	
High	376 (92.8%)	205 (92.8%)	87 (90.6%)	84 (95.5%)	
Lymphovascular invasion					0.734
Absent	319 (78.8%)	177 (80.1%)	75 (78.1%)	67 (76.1%)	
Present	86 (21.2%)	44 (19.9%)	21 (21.9%)	21 (23.9%)	
Median OS, months	53.0 (45.5-60.5)	54.4 (39.9–68.8)	58.0 (42.4–73.5)	48.3 (37.9–58.6)	0.588

OS, overall survival; pN, pathological N stage; pT, pathological T stage.

decayed to half when the C-index of the internal validation set did not improve further for 10 consecutive times. The model was trained for a maximum of 50 epochs using the Adam optimizer with a batch size of 32 and the early stopping's patience of 20. The model with the highest C-index in the validation was finally selected. Finally, the heatmaps of suspicious regions were generated using a gradient-weighted class activation map (Grad-CAM) to visualize significant lesion regions identified by the model[33].

Radiomics and DLRN model construction

Radiomics features were extracted from the 3D ROIs of each patient's CT images using the Pyradiomics package (version 2.2.0) in Python. In the training set, radiomic features were selected by using inter and intra-correlation coefficients (ICCs) analysis, univariate Cox analysis, spearman rank correlation test and least absolute shrinkage and selection operator (LASSO) Cox regression. The radiomics score (Rad-score) was calculated by combining the selected features and corresponding weights in multivariate Cox regression analysis. More details of radiomics features extraction and radiomics score construction can be found in Supplementary Material (Supplemental Digital Content 2, http://links.lww.com/JS9/B908).

Univariate Cox regression analysis was used to assess the relationship between clinicopathological characteristics and OS in the training set. Combining independent clinicopathological factors, Rad-score and DL-score, a multivariate Cox proportional risk regression was used to develop a DLRN.

Statistical analysis

Harrell's consistency index (C-index), time-dependent receiver operating characteristic (ROC) analyses were used to assess and test the prognostic performance of the proposed models in all cohorts. The optimal cut-off point for prognostic score was calculated using X-tile software in the training set, and patients were risk stratified into high-risk and low-risk group. Kaplan–Meier survival analysis and the log-rank test were performed to assess the difference in prognosis between two groups. Subgroup analyses were performed to determine the relationship between the two groups and subgroups defined by sex, age, history of NMIBC status, number of tumors, pathologic T stage, pathologic N stage, pathological grade, and lymphovascular invasion status. The Cox proportional hazards model was used for univariate and multivariate analysis. Calibration curves were used assess the agreement between predicted outcomes of nomogram and actual outcomes.

Continuous variables were presented as the median and interquartile range (IQR), and categorical variables were presented as frequencies and percentages. Comparisons of baseline differences were conducted using χ^2 test for categorical variables and one-way ANOVA for continuous variables, respectively. Data analyses were performed using R (version 4.1.2; R Foundation for Statistical Computing; https://www.r-project.org) and the SPSS software (version 23.0; SPSS Inc.). All analyses were considered statistically significant with a two-sided P-value < 0.05.

Table 2

The performance of models in predicting OS in MIBC patients.

Cohorts	Models	C-index	1-years AUC	3-years AUC	5 years AUC
Training set	Clinical model	0.668 (0.612–0.723)	0.674 (0.533–0.814)	0.715 (0.639–0.790)	0.690 (0.589–0.791)
	Rad-score	0.669 (0.610–0.724)	0.740 (0.618–0.862)	0.713 (0.636–0.791)	0.711 (0.607–0.815)
	DL-score	0.727 (0.673–0.778)	0.767 (0.649–0.885)	0.771 (0.699–0.842)	0.752 (0.654–0.851)
	Combined model	0.766 (0.718–0.813)	0.803 (0.698–0.907)	0.821 (0.758–0.883)	0.807 (0.712–0.903)
Internal validation set	Clinical model	0.618 (0.531–0.706)	0.661 (0.496–0.826)	0.611 (0.468–0.754)	0.686 (0.549–0.822)
	Rad-score	0.652 (0.566–0.741)	0.743 (0.565–0.921)	0.665 (0.536–0.793)	0.693 (0.540–0.847)
	DL-score	0.690 (0.605–0.774)	0.740 (0.590–0.890)	0.693 (0.551–0.835)	0.778 (0.645–0.911)
	Combined model	0.713 (0.627–0.798)	0.792 (0.677–0.906)	0.716 (0.582–0.849)	0.815 (0.691–0.939)
External validation set	Clinical model	0.610 (0.470–0.698)	0.771 (0.564–0.978)	0.714 (0.598–0.830)	0.558 (0.396–0.720)
	Rad-score	0.601 (0.493–0.694)	0.689 (0.446–0.931)	0.621 (0.490–0.753)	0.651 (0.479–0.823)
	DL-score	0.658 (0.556–0.739)	0.663 (0.438–0.886)	0.713 (0.592–0.833)	0.812 (0.688–0.937)
	Combined model	0.685 (0.586–0.765)	0.843 (0.719–0.967)	0.756 (0.637–0.873)	0.754 (0.614–0.893)

AUC, an area under the receiver operating characteristic curve; DL-score, deep learning score; OS, overall survival; Rad-score, radiomics score.

Results

Patient characteristics

The baseline clinical characteristics of patients are shown in Table 1. Among these 405 patients enrolled in this study, the median age was 67.0 years (IQR, 59.0–73.0), 360 (88.9%) patients was men and 268 (66.2%) patients had pT2 stage. The clinical characteristics were similar and no statistically significant difference was observed in different sets. The median follow-up period was 48.4 months (IQR, 34.6–68.1) in all patients. The median OS were 54.4 months (95% CI: 39.9–68.8) in the training set, 58.0 months (95% CI: 42.4–73.5) in the internal validation set and 48.3 months (95% CI: 37.9–58.6) in the external validation set, respectively.

Performance of the DL model

We trained an attention-based DL model to predict OS after radical cystectomy in MIBC patients using preoperative CT images. The output of the model was used as deep learning computed risk score (DL-score). The C-index of the DL-score achieved 0.727 (95% CI: 0.673–0.778) in the training set, 0.690 (95% CI: 0.605–0.774) in the internal validation set, and 0.658 (95% CI: 0.556–0.739) in the external validation cohort (Table 2). The DL-score was able to accurately predict 1, 3, 5 years OS, with AUC of 0.740 (95% CI: 0.590–0.890), 0.693 (95% CI: 0.551–0.835) and 0.778 (95% CI: 0.645–0.911) in the internal validation set, and 0.663 (95% CI: 0.438–0.890), 0.713 (95% CI: 592–0.833), and 0.812 (95% CI: 0.688–0.937) in the external validation set, respectively (Fig. S1, Supplemental Digital Content 2, http://links.lww.com/JS9/B908). Patients were stratified into high-risk and low-risk groups based on a DL-score with a cutoff value of −0.06, which was obtained in the training set by X-tile software. Patients in the high-risk group had worse OS compared to the patients in the low-risk group in the train set [hazard ratio (HR), 5.340 (95% CI: 3.542–8.050), $P < 0.0001$], in the internal validation set [HR, 4.744 (95% CI: 2.348–9.587), $P < 0.0001$], and in the external validation set [HR, 2.559 (95% CI: 1.374–4.767), $P = 0.0022$] (Fig. 2A).

DL-score was statistically independent of potential confounders

We then assessed the association between DL-score and OS outcomes in patients with MIBC using univariate and multivariate Cox models. From the results of univariate Cox regression analysis, age, pathological T stage, pathological N stage, Rad-score, and DL-score were significantly associated with OS in the training set (Table S2, Supplemental Digital Content 2, http://links.lww.com/JS9/B908). The results of multivariate Cox regression analysis demonstrated that, adjusting for clinicopathologic variables (age, pathologic T stage, and pathologic N stage), DL-score remained an significantly independent prognostic factor of OS in the training set ($P < 0.001$), internal validation set ($P = 0.001$), and external validation set ($P = 0.002$) (Fig. 2B and Table S3, Supplemental Digital Content 2, http://links.lww.com/JS9/B908).

DL-score performance in different subgroups and model interpretability

The subgroup analysis showed significant differences in OS between high and low-risk group in all subgroups defined by sex, age, history of NMIBC status, numbers of tumor, pathologic T stage, pathologic N stage, pathological grade, and lymphovascular invasion (all $P < 0.01$) (Fig. 3A). Additionally, even in the subgroups identified by the Rad-score, the DL-score was able to accurately identify high-risk and low-risk patients (Fig. 3B). Furthermore, the OS of patients with pT2 stage in the subgroup of DL-score defined high-risk was significantly shorter compared with that patients with pT3 or pT4 stage in low-risk [HR, 2.194 (95% CI: 1.383–3.479), $P < 0.001$] (Fig. S2, Supplemental Digital Content 2, http://links.lww.com/JS9/B908). The OS of patients with pN0 or Nx stage in the subgroup of DL-score defined high-risk was significantly shorter compared with that patients with pN1 or pN2 or pN3 stage in low-risk (HR, 2.18 [95% CI, 1.087-4.371], $P = 0.024$) (Fig. S3, Supplemental Digital Content 2, http://links.lww.com/JS9/B908).

The risk score distribution computed by the DL model and the gradient-weighted class activation map of the original image are shown in Figure 4. The red area was mainly focused on the basal part and the whole tumor area, implying that the DL model focuses on these areas to extract valuable information.

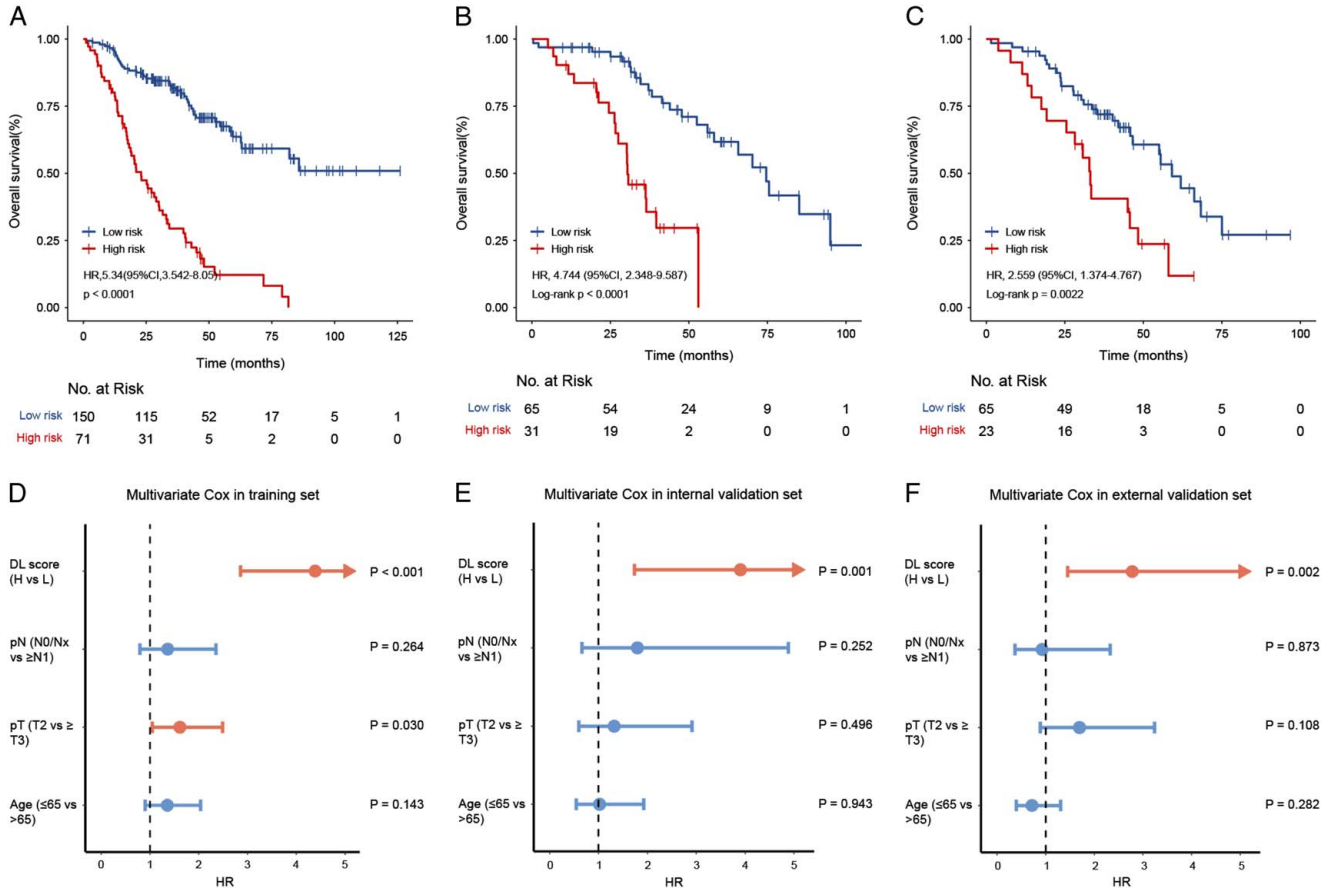


Figure 2. Kaplan–Meier curves shown survival difference between high-risk group and low-risk group stratified by DL-score in the training cohort (A), internal validation set (B) and external validation set (C). Forest plots of the multivariable Cox regression analysis of DL-score in the training cohort (D, internal validation set (E) and external validation set (F).

DL-score outperformed Rad-score and clinical model

Twelve radiomics features were screened to build the radiomics score (Rad-score) (Fig. S4, Supplemental Digital Content 2, http://links.lww.com/JS9/B908). The Kaplan–Meier curves and time-dependent ROC curves according to the Rad-score in the training, internal validation and external validation set are shown in Figure S5 (Supplemental Digital Content 2, http://links.lww.com/JS9/B908) and Figure S6 (Supplemental Digital Content 2, http://links.lww.com/JS9/B908). The clinical model was built by combing age, pathologic T stage, and pathologic N stage using multivariate Cox regression analyses. The Rad-score exhibited a higher C-index (0.652, 95% CI: 0.566–0.741) than clinical model (0.618, 95% CI: 0.531–0.706) in internal validation set, but was still much lower than DL-score (0.690, 95% CI: 0.605–0.774) (Table 2). However, the C-index of Rad-score in the external validation set was only 0.601 (95% CI: 0.493–0.694), which was slightly lower than the clinical model (0.610, 95% CI: 0.470–0.698). Compared with DL-score, the Rad-score and clinical model had a much lower time-dependent AUC at 5 years, ranging from 0.651 (95% CI: 0.479–0.823) to 0.693 (95% CI: 0.540–0.847) and 0.558 (95% CI: 0.396–0.720) to 0.686 (95% CI: 0.549–0.822) in validation set, respectively (Fig. S7, Supplemental Digital Content 2, http://links.lww.com/JS9/B908).

Performance of the DLRN

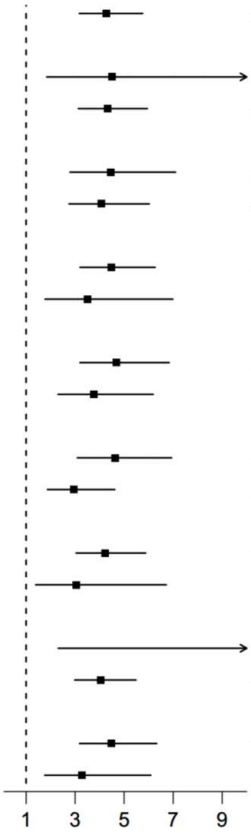
The DLRN was built by integrating DL-score, Rad-score, and clinicopathologic risk factors for independently prediction of OS in a training cohort (Fig. 5A). The DLRN achieved superior prognostic prediction performance compared with single-modality model, with a C-index of 0.713 (95% CI: 0.627–0.798) in the internal validation set and 0.685 (95% CI: 0.586–0.765) in the external validation set. The ROC curve at 1, 3, and 5 years are shown in Figure 5B. There was good agreement between the nomogram estimations and actual observation in the training, internal validation, and external validation set (Fig. 5C).

Discussion

In this multicenter study, we proposed a DL model based on attention mechanism to predict OS in MIBC patients after radical cystectomy through preoperative CT images. Furthermore, DL model outperformed classic handcrafted radiomics model and clinicopathologic feature-based model. Our results suggest that DL computed risk scores can provide added prognostic value beyond traditional clinicopathologic features. Traditionally, more aggressive radical cystectomy and pelvic lymph node dissection or in combination with neoadjuvant chemotherapy are the mainstay of treatment for MIBC. In recent years, with the advancements in

A

Variable	Events/patients (n) All patients	Low risk	High risk	HR(95% CI)	P value
All patients	188/405	94/280	94/125	4.27(3.17–5.76)	<0.001
Gender					
Female	23/45	15/37	8/8	4.51(1.84–11.03)	0.001
Male	165/360	79/243	86/117	4.32(3.13–5.95)	<0.001
Age					
≤65	79/184	44/136	35/48	4.45(2.79–7.1)	<0.001
>65	109/221	50/144	59/77	4.07(2.75–6.03)	<0.001
History of NMIBC					
No	150/333	80/241	70/92	4.48(3.2–6.27)	<0.001
Yes	38/72	14/39	24/33	3.52(1.77–6.99)	<0.001
Multiple					
No	122/249	61/174	61/75	4.68(3.2–6.84)	<0.001
Yes	66/156	33/106	33/50	3.77(2.3–6.19)	<0.001
pT stage					
pT2	105/268	60/203	45/65	4.63(3.09–6.94)	<0.001
≥pT2	83/137	34/77	49/60	2.94(1.87–4.62)	<0.001
pN stage					
pN0/Nx	157/355	85/256	72/99	4.23(3.04–5.88)	<0.001
≥pN1	31/50	9/24	22/26	3.05(1.39–6.72)	0.006
Pathological grade					
Low	7/29	3/22	4/7	13.33(2.31–76.83)	0.004
High	181/376	91/258	90/118	4.04(2.98–5.49)	<0.001
Lymphovascular invasion					
No	141/319	73/229	68/90	4.48(3.18–6.33)	<0.001
Yes	47/86	21/51	26/35	3.27(1.76–6.09)	<0.001



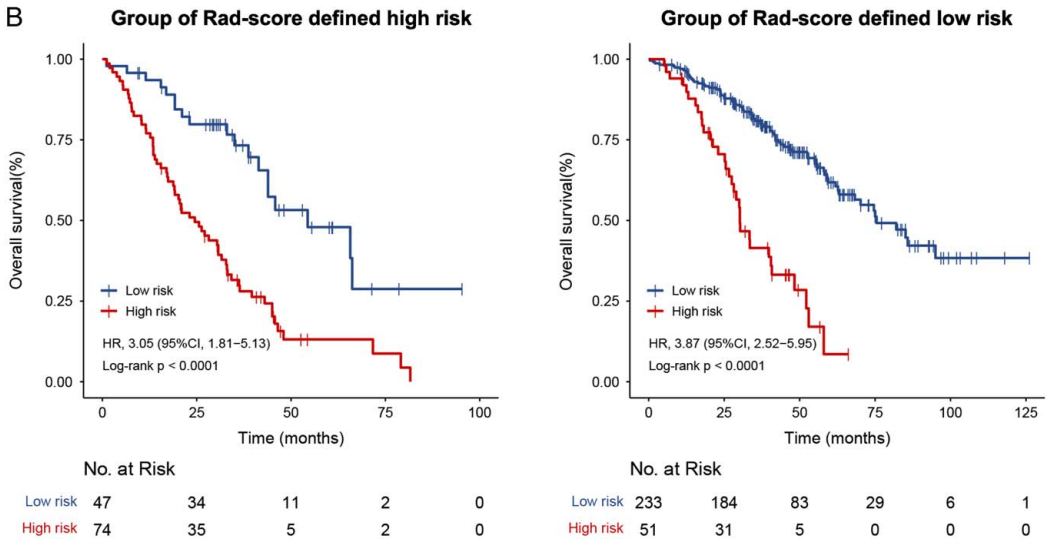


Figure 3. HR of OS for all patients according to the DL-score in different subgroup (A). Kaplan–Meier survival analysis for OS of the DL-score in the subgroups stratified by Rad-score (B). HR, hazard ratio.

immunotherapy and targeted therapeutic agents, there are more available treatment options for MIBC patients. Personalized treatment decisions and follow-up management are mainly based on patients' prognostic risk stratification. However, most current prognostic models are primarily based on established TNM tumor staging system without taking into account other risk factors such as imaging and genetic information[5,34].

Radiomics can provide noninvasive tools to identify tumor biological behavior by extracting and analyzing high-throughput quantitative data from medical images. In recent years, an

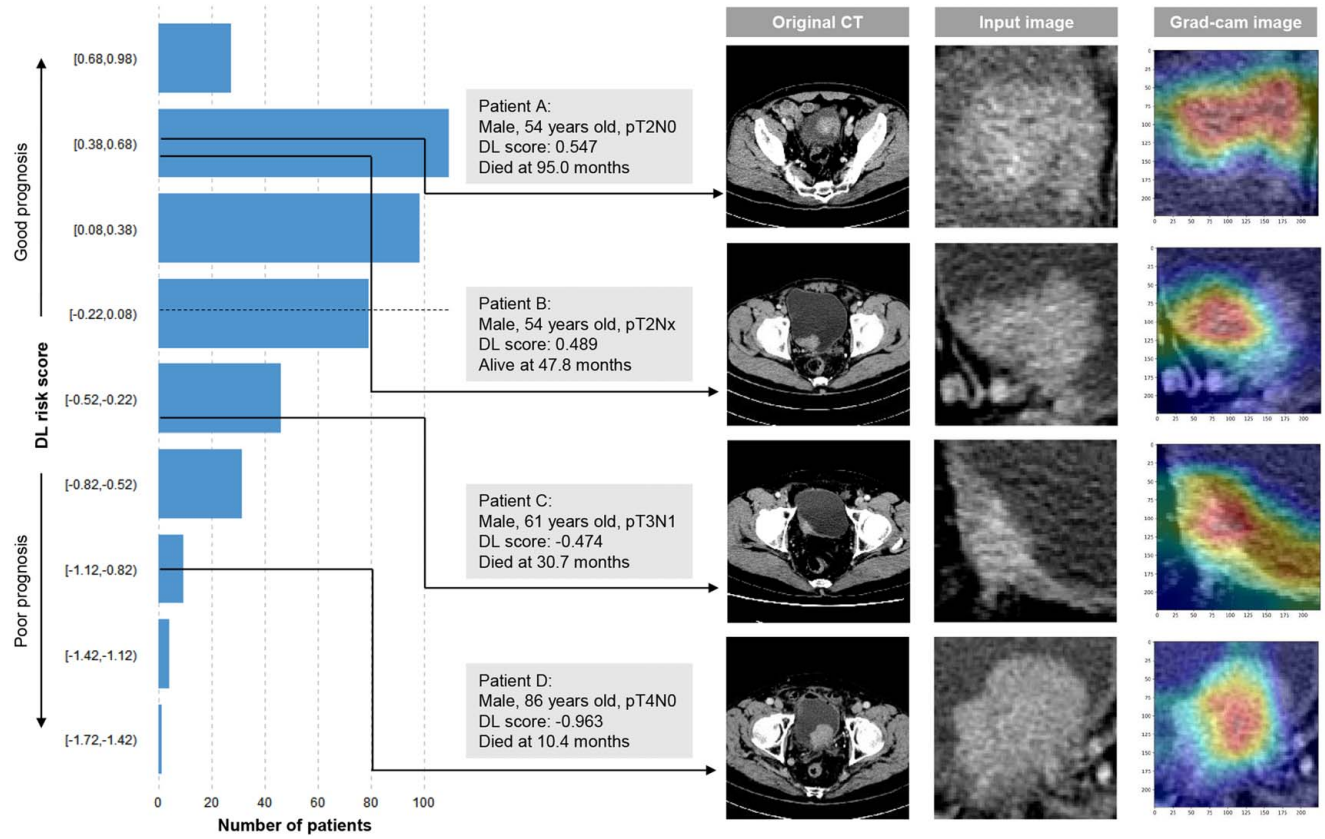


Figure 4. The risk score distribution computed by the DL model, and some representative examples of original CT images, cropped input CT images and corresponding gradient-weighted class activation maps. The Grad-CAM results shown that the attention-based deep learning model highly focused on the basal part of tumor and the whole tumor area in the images.

increasing number of radiomics models based on CT or MRI have been utilized to predict OS and progression-free survival in patients with BCa. Zhang *et al.*[26] constructed radiomics signature from diffusion-weighted imaging (DWI) images to predict progression-free survival (PFS) in MIBC patients and radiomics signature alone achieved a C-index of 0.612 in the test data set. In another study, the combination of radiomics features extracted from tumor and lymph node regions and clinical parameters to predict OS in bladder cancer patients achieved a mean area under the ROC curve of 0.785 integrated over 1 to 7 years after RC[28]. Our study found that radiomics model based on handcrafted radiomics features was able to predict OS in MIBC patients, with comparable predictive performance to previous studies [C-index of 0.652 (95% CI: 0.566–0.741) in the internal validation set and 0.601 (95% CI: 0.493–0.694) in the external validation set]. However, handcrafted radiomics features, limited by preselection of morphological and textural features, may not capture all useful information from radiological images.

DL, an advanced machine learning technique, can directly extract more comprehensive and in-depth information from raw images and build end-to-end predictive models. DL shows a promising future in predicting tumor grade, treatment response, and prognosis among many kinds of tumors[17,35]. However, few studies have directly compared DL models with traditional radiomics models. Our study found that DL computed risk scores outperformed traditional handcrafted radiomics model and can further differentiate high-risk and low-risk patients from subgroups classified by radiomics model. More importantly, the prognosis in MIBC patients with high TNM stage may not always be worse than that in patients with low stage. In this study, DL model can accurately identify patients with favorable prognosis in high stage disease and worse prognosis in low stage disease. This suggests that DL algorithms can extract prognostic information from enhanced CT images, which complements clinical prognostic features.

In a previous study, a combined model incorporating DL features, radiomics features, and clinical factors was built to predict survival after radical surgery for bladder cancer[27]. However, the study's patient cohort was relatively small and included only a single medical center. Notably, like many previous studies, this study constructed a binary classification to predict survival prognosis, which would have lost much useful survival information. Survival prediction research, which differs from traditional classification tasks, need to use both survival status and survival time information due to the presence of censored data. In the present study, we used ResNet-50 as the network backbone for feature extraction, which has been widely used in other medical tasks. We then customized the loss function based on Cox partial likelihood, and constructed an end-to-end DL models to predict survival outcome. Previous studies suggested that attentional mechanisms, such as squeeze-and-excitation (SE) module and CBAM, may improve the prediction performance of convolutional neural networks[20,21]. Our results found that the

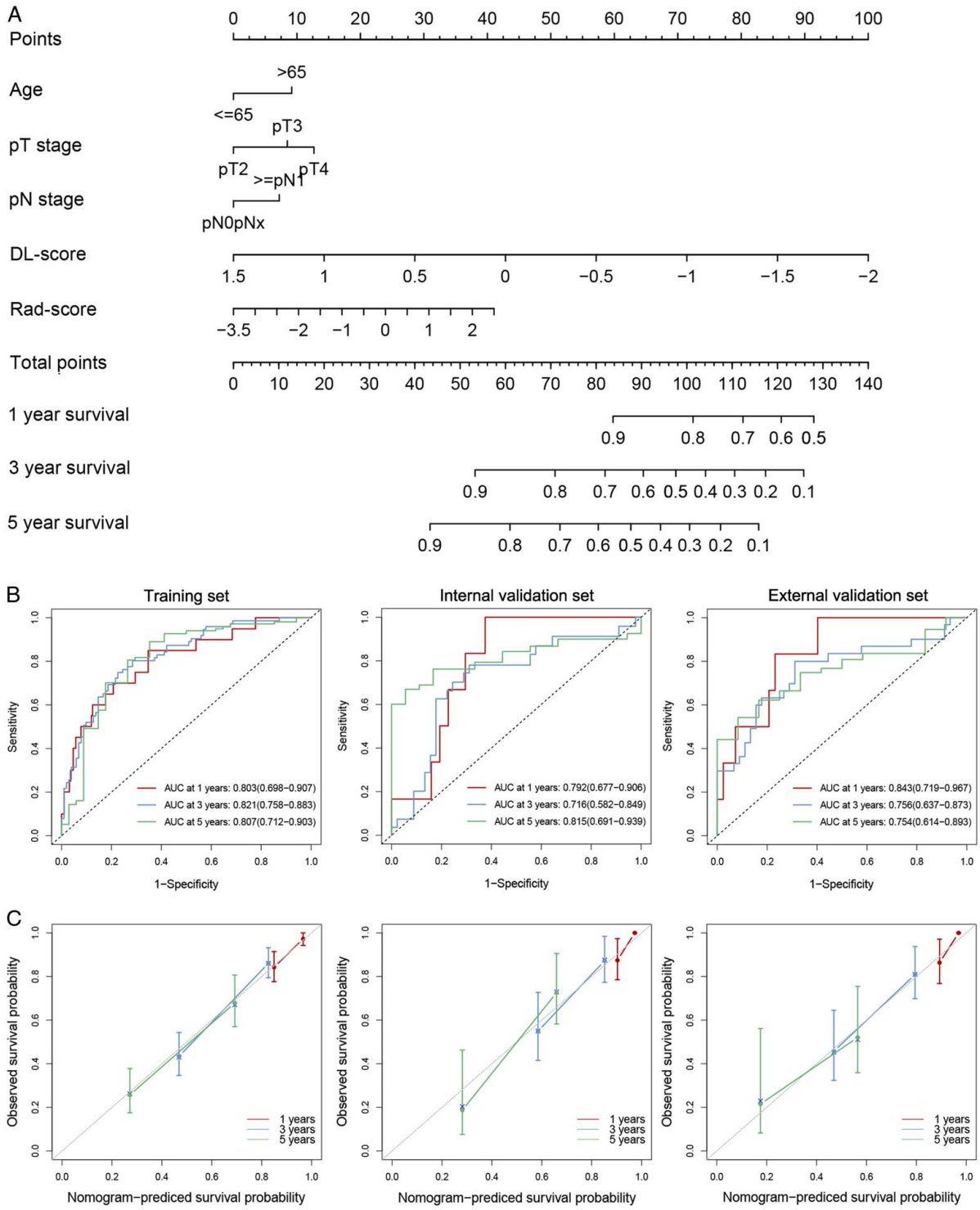


Figure 5. The deep learning radiomics nomogram (DLRN), integrating DL-score, Rad-score, age, pathological T stage and pathological T stage, was developed in the training set (A). Time-dependent ROC curve at 1, 3, and 5 years was used to assess the prognostic performance of DLRN in the training, internal validation and external validation set (B). The calibration curve showed good agreement between the nomogram estimations and actual observation in the training, internal validation and external validation set (C).

CBAM-based DL model outperformed the traditional radiomics model and clinical model in predicting postcystectomy prognosis. The results suggest that CT-based DL algorithms might have a predictive value in the outcome prediction of patients with MIBC.

In the current literature, tumor grading, depth of tumor infiltration, lymph node status, histological variation, and clinical features have been identified as important predictors of OS after radical cystectomy[5]. Previous studies have shown that the

inclusion of additional prognostically relevant factors may contribute to the construction of a more robust model[36,37]. In this study, a combined model was developed by combining clinicopathological factors, Rad-score, and DL-score through nomogram and achieved best efficacy. Considering the advanced age of most bladder tumor cases, comorbidity and performance assessment holds significant importance in guiding treatment decisions and predicting survival outcomes[38]. Future artificial intelligence models may consider incorporating these factors, such as Age-adjusted Charlson Comorbidity Index, Preoperative Score to Predict Postoperative Mortality, and the American Society of Anesthesiologists (ASA) score.

This study has several limitations. Firstly, although our model was validated in an independent external cohort, all included patients were from the same region in southwest China. In addition, all data were collected retrospectively with small sample sizes, which inevitably introduces selection bias and hidden confounders. Therefore, a large prospective cross-regional and cross-national data are needed for future validation. Moreover, the study could not evaluate the impact of neoadjuvant therapy on OS due to the exclusion of this patient subgroup. Secondly, we only selected the largest section image of the tumor as input data to the DL model, which may lose some important three-dimensional morphological information inside the tumor. Thirdly, manual delineation of tumor areas requires professional radiologists and the results were highly influenced by subjective experience. In the future, we will investigate the application of automated segmentation techniques and newer DL models in bladder cancer. Fourthly, future studies may consider incorporating other prognostic factors, such as genetics and pathologic data, to improve the model's prognostic predictive ability for patients with MIBC.

Conclusion

In conclusion, we developed and validated a preoperative CT-based DL model to predict OS of MIBC patients who underwent radical cystectomy. The DLRN model showed better performance after combining DL-score, Rad-score, and clinicopathological factors. Our model can assist clinicians to risk stratify MIBC patients and may be helpful in guiding personalized treatment decision-making process and follow-up strategies.

Ethical approval

This study was approved by the Ethics Committee of The First Affiliated Hospital of Chongqing Medical University (Approval No. 2022-K508), and the requirement for informed consent was waived due to the retrospective nature of the study.

Consent

Not applicable.

Sources of funding

This study was supported by the Chongqing Talent Program (grant number: CQYC202003).

Author contribution

M.X., F.L., and Q.J.: conception and design; Z.W., H.L., S.Y., Y.L., and Y.C.: data collection; Z.W. and Y.X.: data analysis and interpretation; Z.W. and Y.X.: deep learning methods; Z.W. and Y.X.: manuscript writing. M.X., F.L., and Y.X.: manuscript review. All authors contributed to the article and approved the submitted version.

Conflicts of interest disclosure

The authors declare that there are no conflicts of interest.

Research registration unique identifying number (UIN)

1. Name of the registry: AI-BLCA.
2. Unique identifying number or registration ID: NCT06092450.
3. Hyperlink to your specific registration (must be publicly accessible and will be checked): https://clinicaltrials.gov/study/NCT06092450?term=NCT06092450&rank=1.

Guarantor

Mingzhao Xiao and Feng Li.

Data availability statement

The original data are not publicly available, but are available from the corresponding author upon reasonable request after IRB approval in accordance with the institute's policy.

Provenance and peer review

Not commissioned, externally peer-reviewed.

References

[1] Lobo N, Afferi L, Moschini M, *et al*. Epidemiology, screening, and prevention of bladder cancer. Eur Urol Oncol 2022;5:628–39.
[2] Lenis AT, Lec PM, Chamie K, *et al*. Bladder cancer: a review. JAMA 2020;324:1980–91.
[3] Flaig TW, Spiess PE, Agarwal N, *et al*. Bladder cancer, Version 3.2020, NCCN clinical practice guidelines in oncology. J Natinl Comprehens Cancer Network : JNCCN 2020;18:329–54.
[4] Patel VG, Oh WK, Galsky MD. Treatment of muscle-invasive and advanced bladder cancer in 2020. CA Cancer J Clin 2020;70:404–23.
[5] Bochner BH, Kattan MW, Vora KC. Postoperative nomogram predicting risk of recurrence after radical cystectomy for bladder cancer. J Clin Oncol 2006;24:3967–72.
[6] Amin MB, Greene FL, Edge SB, *et al*. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. CA Cancer J Clin 2017;67:93–9.
[7] Tsili AC, Varkarakis I, Pasoglou V, *et al*. CT of the urinary tract revisited. Eur J Radiol 2023;160:110717.
[8] Witjes JA, Bruins HM, Cathomas R, *et al*. European association of urology guidelines on muscle-invasive and metastatic bladder cancer: summary of the 2020 guidelines. Eur Urol 2021;79:82–104.
[9] Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology 2016;278:563–77.
[10] Shin J, Seo N, Baek SE, *et al*. MRI radiomics model predicts pathologic complete response of rectal cancer following chemoradiotherapy. Radiology 2022;211986:351–8.

[11] Yang L, Gu D, Wei J, *et al*. A Radiomics nomogram for preoperative prediction of microvascular invasion in hepatocellular carcinoma. Liver Cancer 2019;8:373–86.
[12] Chen M, Cao J, Hu J, *et al*. Clinical-radiomic analysis for pretreatment prediction of objective response to first transarterial chemoembolization in hepatocellular carcinoma. Liver Cancer 2021;10:38–51.
[13] 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 Network Open 2020;3:e2028086.
[14] Wang K, Ren Y, Ma L, *et al*. Deep learning-based prediction of treatment prognosis from nasal polyp histology slides. Int Forum Allergy Rhinol 2023;13:886–98.
[15] Borhani S, Borhani R, Kajdacsy-Balla A. Artificial intelligence: a promising frontier in bladder cancer diagnosis and outcome prediction. Crit Rev Oncol Hematol 2022;171:103601.
[16] Uhm KH, Jung SW, Choi MH, *et al*. Deep learning for end-to-end kidney cancer diagnosis on multi-phase abdominal computed tomography. NPJ Precision Oncol 2021;5:54.
[17] Jiang X, Zhao H, Saldanha OL, *et al*. An MRI deep learning model predicts outcome in rectal cancer. Radiology 2023;307:e222223.
[18] Huang B, Sollee J, Luo YH, *et al*. Prediction of lung malignancy progression and survival with machine learning based on pre-treatment FDG-PET/CT. EBioMedicine 2022;82:104127.
[19] Zhong L, Dong D, Fang X, *et al*. A deep learning-based radiomic nomogram for prognosis and treatment decision in advanced nasopharyngeal carcinoma: a multicentre study. EBioMedicine 2021;70:103522.
[20] Mao N, Zhang H, Dai Y, *et al*. Attention-based deep learning for breast lesions classification on contrast enhanced spectral mammography: a multicentre study. Br J Cancer 2023;128:793–804.
[21] Hu C, Chen W, Li F, *et al*. Deep learning radio-clinical signatures for predicting neoadjuvant chemotherapy response and prognosis from pretreatment CT images of locally advanced gastric cancer patients. Int J Surg 2023;109:1980–92.
[22] Zhang G, Wu Z, Zhang X, *et al*. CT-based radiomics to predict muscle invasion in bladder cancer. Eur Radiol 2022;32:3260–8.
[23] Wang H, Xu X, Zhang X, *et al*. Elaboration of a multisequence MRI-based radiomics signature for the preoperative prediction of the muscle-invasive status of bladder cancer: a double-center study. Eur Radiol 2020; 30:4816–27.
[24] Zhang G, Xu L, Zhao L, *et al*. CT-based radiomics to predict the pathological grade of bladder cancer. Eur Radiol 2020;30:6749–56.
[25] Qian J, Yang L, Hu S, *et al*. Feasibility study on predicting recurrence risk of bladder cancer based on radiomics features of multiphase CT images. Front Oncol 2022;12:899897.
[26] Zhang S, Song M, Zhao Y, *et al*. Radiomics nomogram for preoperative prediction of progression-free survival using diffusion-weighted imaging in patients with muscle-invasive bladder cancer. Eur J Radiol 2020;131:109219.
[27] Sun D, Hadjiiski L, Gormley J, *et al*. Survival prediction of patients with bladder cancer after cystectomy based on clinical, radiomics, and deep-learning descriptors. Cancers 2023;15:4372.
[28] Woźnicki P, Laqua FC, Messmer K, *et al*. Radiomics for the prediction of overall survival in patients with bladder cancer prior to radical cystectomy. Cancers 2022;14:4449.
[29] Mathew G, Agha R, Albrecht J, *et al*. STROCSS 2021: strengthening the reporting of cohort, cross-sectional and case-control studies in surgery. Int J Surg 2021;96:106165.
[30] Flaig TW, Spiess PE, Abern M, *et al*. NCCN Guidelines® Insights: Bladder Cancer, Version 2.2022. J Natinl Comprehens Cancer Network: JNCCN 2022;20:866–78.
[31] Woo S, Park J, Lee J-Y, *et al*. Cbam: convolutional block attention module. Proceedings Eur Conf Computer Vision (ECCV) 2018:3–19.
[32] 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.
[33] Selvaraju RR, Cogswell M, Das A, *et al*. Grad-cam: visual explanations from deep networks via gradient-based localization. Proceedings IEEE Int Conference Computer Vision 2017:618–26.
[34] Bhambhvani HP, Zamora A, Shkolyar E, *et al*. Development of robust artificial neural networks for prediction of 5-year survival in bladder cancer. Urol Oncol 2021;39:193.e197–12.
[35] She Y, He B, Wang F, *et al*. Deep learning for predicting major pathological response to neoadjuvant chemoimmunotherapy in non-small cell lung cancer: A multicentre study. EBioMedicine 2022;86:104364.
[36] Gui CP, Chen YH, Zhao HW, *et al*. Multimodal recurrence scoring system for prediction of clear cell renal cell carcinoma outcome: a discovery and validation study. The Lancet Digital Health 2023;5:e515–24.
[37] Wang R, Dai W, Gong J, *et al*. Development of a novel combined nomogram model integrating deep learning-pathomics, radiomics and immunoscore to predict postoperative outcome of colorectal cancer lung metastasis patients. J Hematol Oncol 2022;15:11.
[38] Selvi I, Arik AI, Baydilli N, *et al*. Evaluation of comorbidity indices in determining the most suitable candidates for uro-oncological surgeries in elderly men. Cent European J Urol 2021;74:24–38.