

# A non-invasive MRI-based multimodal fusion deep learning model (MF-DLM) for predicting overall survival in bladder cancer: a multicentre retrospective study



Lingkai Cai,<sup>a,b,m</sup> Rongjie Bai,<sup>a,m</sup> Qiang Cao,<sup>a,m</sup> Weijie Sun,<sup>c,d,m</sup> Fei Wang,<sup>c</sup> Xiaotong Liu,<sup>e</sup> Bo Liang,<sup>g</sup> Meihua Jiang,<sup>h</sup> Gongcheng Wang,<sup>i</sup> Qiang Shao,<sup>j</sup> Xuping Jiang,<sup>k</sup> Chenghao Wang,<sup>a</sup> Chang Chen,<sup>a</sup> Zhengye Tan,<sup>a</sup> Qikai Wu,<sup>a</sup> Meiling Bao,<sup>j</sup> Hao Yu,<sup>a</sup> Pengchao Li,<sup>a</sup> Xiao Yang,<sup>a,\*\*</sup> and Qiang Lu<sup>a,\*</sup>

<sup>a</sup>Department of Urology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China

<sup>b</sup>Department of Urology, The Affiliated Wuxi People's Hospital of Nanjing Medical University, Wuxi, China

<sup>c</sup>Department of Computing Science, University of Alberta, Edmonton, Alberta, Canada

<sup>d</sup>Canadian VIGOUR Centre, University of Alberta, Edmonton, Alberta, Canada

<sup>e</sup>Department of Urology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou University, Zhengzhou, China

<sup>f</sup>Department of Radiology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

<sup>g</sup>Department of Radiology, Affiliated Hospital of Nanjing University of Traditional Chinese Medicine, Jiangsu Provincial Hospital of Traditional Chinese Medicine, Nanjing, China

<sup>h</sup>Department of Urology, The Affiliated Huaian No.1 People's Hospital of Nanjing Medical University, Huaian, China

<sup>j</sup>Department of Urology, Suzhou Hospital Affiliated of Nanjing Medical University, Suzhou, China

<sup>k</sup>Department of Urology, Yixing People's Hospital, Wuxi, China

<sup>i</sup>Department of Pathology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China

## Summary

**Background** Accurate prognosis prediction in bladder cancer (BCa) is crucial for personalized treatment. This study aimed to develop and validate a non-invasive model using magnetic resonance imaging (MRI) for predicting the overall survival (OS) in patients with BCa.

**Methods** This retrospective multicentre study included 1131 patients with BCa from eight institutions in China from June 2011 to March 2024. 871 patients were enrolled from one centre, who were randomly divided (8:2) into training ( $n = 697$ ) and internal validation ( $n = 174$ ) sets. For the external test set, 260 patients with BCa from seven centres were retrospectively included. We developed a multimodal fusion deep learning model (MF-DLM), leveraging a cross-attention mechanism to integrate four key preoperative data modalities: three-dimensional (3D) deep learning features using a modified 3D ResNet50 network, 3D radiomics features, morphological MRI features, and clinical features. Patients were stratified into low- and high-risk prognostic groups based on MF-DLM scores, and model interpretability was evaluated using Shapley additive explanations (SHAP) and Gradient-weighted class activation mapping (Grad-CAM).

**Findings** The median follow-up time for the training, validation, and external test sets are 38.0 months (interquartile ranges [IQR]: 22.0, 62.0), 40.5 months (IQR: 23.0, 71.0), and 38.5 months (IQR: 26.0, 50.0), respectively. The MF-DLM demonstrated excellent performance in predicting OS, achieving higher C-index values than pathological T stage (training: 0.902 vs. 0.793,  $p < 0.001$ ; validation: 0.864 vs. 0.757,  $p = 0.014$ ; external test: 0.841 vs. 0.760,  $p = 0.047$ ). In addition, MF-DLM-based low-risk group demonstrated significantly longer OS in the training, validation, and external test sets ( $p < 0.001$ ). In the adjuvant therapy (AT) cohort, high-risk patients had significantly worse prognosis compared with low-risk patients ( $p < 0.0001$ ). Additionally, high-risk pathological T3/4 patients exhibited no statistically significant OS difference between those who received AT and those who did not ( $p = 0.18$ ), whereas low-risk pathological T3/4 patients experienced significantly improved OS with AT ( $p = 0.0059$ ). Besides, the low-risk group had better OS than the high-risk group in neoadjuvant therapy cohort ( $p = 0.0032$ ).

eClinicalMedicine

2025;90: 103640

Published Online 13

November 2025

<https://doi.org/10.1016/j.eclim.2025.103640>

\*Corresponding author. Department of Urology, The First Affiliated Hospital of Nanjing Medical University, No. 300, Guangzhou Road, Gulou District, Nanjing City, China.

\*\*Corresponding author. Department of Urology, The First Affiliated Hospital of Nanjing Medical University, No. 300, Guangzhou Road, Gulou District, Nanjing City, China.

E-mail addresses: [doctorlvqiang@njmu.edu.cn](mailto:doctorlvqiang@njmu.edu.cn) (Q. Lu), [yangxiao2915@163.com](mailto:yangxiao2915@163.com) (X. Yang).

<sup>m</sup>These authors contributed equally to this work.

**Interpretation** The MF-DLM can reliably predict OS in patients with BCa and provide additional prognostic stratification beyond pathological T and N stages. Furthermore, MF-DLM-based risk groups can identify patients most likely to benefit from perioperative therapy.

**Funding** The Noncommunicated Chronic Diseases-National Science and Technology Major Project (2024ZD0525700); National Natural Science Foundation of China (82273152, 82503879), Jiangsu Province Hospital (the First Affiliated Hospital of Nanjing Medical University) Clinical Capacity Enhancement Project (JSPH-MA-2022-5), China Postdoctoral Science Foundation funded project (2024M761211), and the Nanjing Postdoctoral Science Foundation funded project (2024BHS210).

**Copyright** © 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Keywords:** Bladder cancer; MRI; Prognosis; Deep learning

## Research in context

### Evidence before this study

We searched PubMed on August 1, 2025, using the terms ("Bladder cancer") AND ("Prognosis" OR "Overall survival") AND ("Deep learning" OR "Artificial intelligence" OR "Radiomics"), without language restrictions. We identified nine studies utilizing artificial intelligence (AI) methods based on magnetic resonance imaging (MRI) or computed tomography (CT) for predicting bladder cancer prognosis. Among these, three studies employed deep learning models. Only two studies explored the combined role of deep learning and radiomics. One of these was magnetic resonance imaging (MRI)-based and conducted at a single-centre with a small sample size ( $n = 191$ ). Furthermore, none of the studies investigated feature interaction/fusion, nor did any examine the model's potential role in guiding bladder cancer treatment decisions.

### Added value of this study

To predict prognosis in bladder cancer, using the largest multicentre MRI cohort of bladder cancer, we developed a

multimodal fusion deep learning model (MF-DLM) that integrated three-dimensional (3D) deep learning features, 3D radiomics features, human-assessed morphological features, and clinical variables via an attention mechanism. Beyond demonstrating robust prognostic stratification, we assessed the model's potential in guiding perioperative treatment decision-making for bladder cancer and partially elucidated its underlying biological basis through correlating with histopathological characteristics.

### Implications of all the available evidence

Our findings indicated that the MF-DLM accurately predicted OS in bladder cancer. The model could identify a favorable-prognosis subgroup within muscle-invasive bladder cancer (MIBC) and select patients who were likely to benefit from perioperative adjuvant therapies. With prospective validation, MF-DLM might help identify patients most likely to respond to specific treatments, supporting personalized management of bladder cancer.

## Introduction

Bladder cancer (BCa) is one of the most prevalent malignant tumors in the urinary system. Clinical treatment is largely determined by the status of muscle invasion and prognosis.<sup>1</sup> BCa is classified as non-muscle invasive bladder cancer (NMIBC) and muscle-invasive bladder cancer (MIBC) based on the depth of tumor invasion. The primary treatment for NMIBC is transurethral resection of the bladder tumor (TURBT), which yields a five-year overall survival (OS) rate of approximately 90%.<sup>2</sup> However, patients with MIBC face up to a 50% risk of distant recurrence within three years after radical cystectomy, indicating a significantly reduced survival.<sup>1</sup>

Effective risk stratification of OS is crucial for tailoring appropriate treatment options for patients with BCa.<sup>1</sup> Current OS stratification primarily relies on the TNM stage system. However, prognostic variety

exists even among patients within the same T stage.<sup>3,4</sup> Magnetic resonance imaging (MRI), a non-invasive technique with superior contrast resolution compared to computed tomography (CT), is gradually used for preoperative diagnosis and staging of BCa.<sup>5-7</sup> While MRI-based morphological features have demonstrated associations with BCa prognosis, their potential for capturing the underlying biological characteristics of tumors remains relatively unexplored and limited.<sup>5,8</sup> Advancing beyond conventional visual assessment, radiomics offers a non-invasive approach to characterize tumor biology by automatically extracting and analyzing high-throughput quantitative data from medical images, enabling a more detailed and refined characterization compared to morphological assessment alone.<sup>9</sup>

Recent studies have further demonstrated the prognostic value of radiomics features in patients with

BCa.<sup>10,11</sup> Yang et al. confirmed that incorporating radiomics features significantly enhanced the predictive performance of clinical models for recurrence-free survival in BCa.<sup>12</sup> Nevertheless, radiomic features, derived from artificially defined quantitative metrics, may not fully capture tumor characteristics at a deeper level. In contrast, deep learning (DL), which operated directly on medical images, could effectively capture additional imaging patterns beyond known radiomics features, and demonstrate promising applications in cancer diagnosis and recurrence risk assessment.<sup>13–15</sup> For instance, DL models applied to CT imaging have demonstrated substantial predictive value for prognostic stratification in patients with BCa.<sup>16–18</sup>

However, previous prognostic studies on BCa have predominantly relied on unimodal modeling approaches, overlooking the potential interactions among DL, radiomics, and morphological features. Recent single-centre evidence indicates that decision-level fusion of MRI-based DL and radiomics models can improve the prediction of NMIBC recurrence.<sup>19</sup> Yet, this study did not explore the interactions among different features, thus neglecting their complex relationships. Moreover, previous studies clearly demonstrated that the feature-level fusion of DL and radiomics features could improve model performance.<sup>20–22</sup> Therefore, given the prognostic stratification value of MRI-based morphological features in BCa,<sup>5,8</sup> the optimal interaction framework between DL, radiomics, and morphological features warrants further investigation.

Collectively, to the best of our knowledge, this study presents the largest multicentre MRI cohort of BCa. We integrated information from multiple modalities, including three-dimensional (3D) DL features, 3D radiomics features, human-accessible morphological features, and clinical features, through a cross-attention mechanism to achieve non-invasive preoperative assessment of OS in patients with BCa.

## Methods

### Ethics statement

This study was approved by the Research Ethics Committee of The First Affiliated Hospital of Nanjing Medical University (approval number: 2024-SR-338). Due to the retrospective nature of the study, the requirement for informed consent was waived. The patients whose personal data are presented in this study provided written informed consent. The study was conducted and reported following the TRIPOD + AI guidelines.<sup>23</sup>

### Participants

This retrospective multicentre clinical study utilized MRI from 1131 patients with BCa across eight centres. From June 2011 to May 2024, 871 patients were enrolled in our centre. These patients were randomly

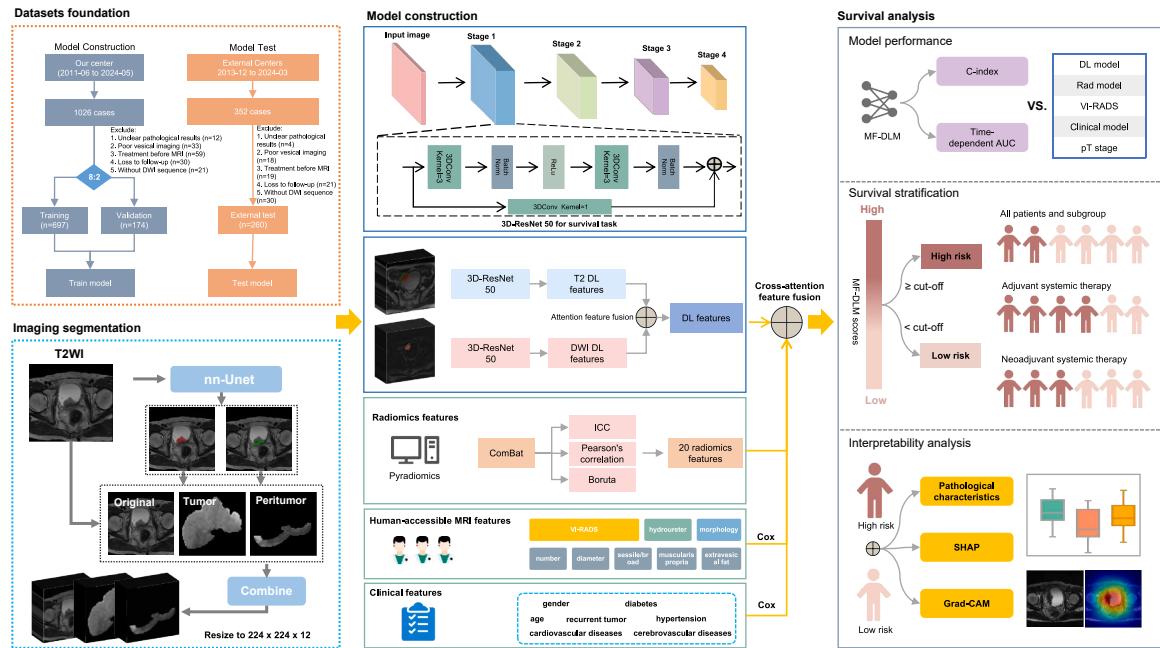
divided into a training set ( $n = 697$ ) and an internal validation set ( $n = 174$ ) with an 8:2 ratio. For the external test set, 260 patients with BCa from seven centres were retrospectively included, covering the period from December 2013 to March 2024. A flowchart illustrating the selection process of the study population is shown in Fig. 1.

### MRI examination and segmentation

MRI examinations were conducted with patients in the supine position. Before the MRI, patients were instructed to void 2 h before the examination and to limit fluid intake or further urination. The scan parameters are summarized in Supplementary Table S1. A skilled radiologist delineated the regions of interest (ROIs) by segmenting the bladder tumor layer by layer, extending the area 5 mm beyond the tumor margins while excluding the bladder cavity to define the peritumoral region. Furthermore, utilizing the nnU-Net,<sup>24</sup> a self-adaptive deep learning segmentation platform, we developed an automated model for the precise segmentation of bladder tumors and peritumoral regions. The 3D full-resolution segmentation model was trained using five-fold cross-validation on our centre dataset. The resulting automatically generated ROIs formed the basis for subsequent radiomic features extraction and predictive modeling. Details of the segmentation model's performance are provided in Supplementary Table S2.

### Radiomics, human-accessible MRI features, and clinical features

A rigorous radiomics features processing workflow was employed, which started with extracting 2030 features per imaging sequence from both the tumor and peritumoral regions, resulting in a total of 4060 features (Supplementary Fig. S1E). All features underwent Z-score normalization followed by 'ComBat' harmonization to address scanner-related batch effects (Supplementary Fig. S1A–D). Inter-observer reproducibility was assessed via intraclass correlation coefficient (ICC >0.75), resulting in 3476 reliable features for analysis (Supplementary Fig. S1F). Besides, radiomics features exhibiting collinearity (Pearson's correlation coefficient >0.8) were identified, and only one feature from each highly correlated pair was retained. A detailed description of the feature extraction process is provided in the Supplementary Materials (pp 2–3). Feature selection employing the Boruta algorithm identified 20 prognostic radiomics features (Supplementary Table S5). Subsequently, eight machine learning models were developed for OS prediction based on established methodologies. Model parameters were optimized through systematic grid searches, with comparative performance detailed in Supplementary Table S6 and Supplementary Fig. S2–4. Finally, a radiomics model was constructed based on the Random Survival Forest (RSF).



**Fig. 1:** Patient cohort and study design. Flow chart depicting patient selection and workflow diagram for developing the multimodal fusion deep learning model (MF-DLM).

In addition, two experienced doctors evaluated MRI based tumor morphological and clinical features. Then, univariate and multivariate Cox regression analyses of the human-accessible MRI morphological features and preoperative clinical features were conducted, respectively (*Supplementary Table S7-9*). Specific process for feature selection and model development are detailed in the Supplementary Materials (Page 3).

#### Multimodal fusion deep learning model (MF-DLM) conduction

In this study, we develop an end-to-end MF-DLM integrating DL, radiomics, morphological, and clinical features for survival prediction. The model was trained on an NVIDIA A40 GPU equipped with 48 GB of memory, operating under Ubuntu 20.04. Training and evaluation utilized version 2.3.1 of the PyTorch library.

The architecture employs dual parallel 3D ResNet50 backbones to extract features from T2-weighted imaging (T2WI) and diffusion-weighted imaging (DWI) sequences, dynamically fused through an adaptive attention module into a unified imaging representation.<sup>25</sup> The input consists of three-channel volumetric data containing original images, tumor regions, and peritumoral regions. To enhance multimodal integration, we use cross-attention modules in our fusion stage to align the DL-derived imaging features with radiomics, morphological, and clinical features, to capture complementary information from each modality. Finally, a multi-layer perception (MLP) subsequently

processes the concatenated feature vector comprising DL features, radiomics-based fusion features, and morphological- and clinical-based fusion features embeddings to generate the MF-DLM risk score. The MF-DLM risk score corresponds to the log hazard ratio estimated by the DeepSurv-based multimodal fusion model.<sup>15,26</sup> Detailed model information is provided in the Supplementary Materials (pp. 3–4 and *Supplementary Fig. S5*).

#### Statistical analysis

Statistical analyses were conducted using R (version 4.3.0; <http://www.R-project.org>). The performance of the final model was evaluated using the C-index. The time-dependent area under receiver operating characteristic (ROC) curves (t-AUC) was computed for designated time points. The best cutoff threshold for survival risk was established using the R package ‘survminer’ (version 0.5.0) in training set. The proportional hazards (PH) assumption was examined before the multivariable Cox regression to ensure that the variables fitted the PH assumption. SHAP (Shapley additive explanations) method was used to facilitate the interpretation of the model.<sup>27</sup> Gradient-weighted class activation mapping (Grad-CAM) visualizes critical areas on an image ‘focused’ by a model with the representation of attention heatmaps. The significance threshold was set at  $p < 0.05$ . OS is defined as the duration from surgery to death or the date of the last follow-up (*Supplementary Table S3*), and these data were reviewed as of June 15, 2025.

## Role of funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

## Results

### Patient characteristics

**Table 1** presents the detailed clinical and pathological features of patients in the training ( $n = 697$ ), validation ( $n = 174$ ), and external test ( $n = 260$ ) set from 8 centres. Of the 1131 patients, 953 are male (84.3%) and 188 are female (15.7%), with a mean age of  $67 \pm 12$  years. The median follow-up time for the training, validation, and external test sets are 38.0 months (interquartile ranges [IQR]: 22.0, 62.0), 40.5 months (IQR: 23.0, 71.0), and 38.5 months (IQR: 26.0, 50.0), respectively. The number of death events are 127 (18.2%), 31 (17.8%), and 38 (14.6%) in the respective sets. **Supplementary Table S11** shows the number of death events at different time points. No statistically significant differences were

observed in survival time or mortality across the three sets (**Table 1**, all  $p > 0.05$ ).

### MF-DLM architecture comparison

We develop a preoperative multimodal deep learning survival prediction model based on MRI, integrating DL features, radiomics features, human-accessible morphological features, and clinical features (**Fig. 1B**). To determine the optimal DL backbone, we conducted five-fold cross-validation on the training set comparing three DL feature extractors: 3D-ResNet 50, 3D-Vision Transformer, and 3D-DenseNet. The 3D-ResNet50 model achieves the highest mean C-index ( $0.838 \pm 0.014$ ) for predicting OS of BCa. This performance significantly outperforms the 3D-Vision Transformer ( $0.809 \pm 0.008$ ;  $p < 0.05$ , **Supplementary Table S11**) and is marginally higher than the 3D-DenseNet ( $0.814 \pm 0.021$ ;  $p = 0.073$ , **Supplementary Table S11**). Based on its superior performance, we select 3D-ResNet50 as the DL feature extractor for MF-DLM. Additionally, ablation study demonstrates that

Characteristics	Overall	Training set	Validation set	External test set	p value <sup>a</sup>
Frequency	1131	697	174	260	
Age, mean (SD)	$67 \pm 12$	$67 \pm 11$	$67 \pm 13$	$67 \pm 10$	0.972
Gender, NO. (%)					0.430
Male	953 (84.3)	595 (85.4)	144 (82.8)	214 (82.3)	
Female	178 (15.7)	102 (14.6)	30 (17.2)	46 (17.7)	
Recurrent tumor (%)					0.447
Absent	892 (78.9)	557 (79.9)	137 (78.7)	198 (76.2)	
Present	239 (21.1)	140 (20.1)	37 (21.3)	62 (23.8)	
Pathological T stage (%)					0.078
Ta	457 (40.4)	294 (42.2)	79 (45.4)	84 (32.3)	
T1	334 (29.5)	199 (28.6)	46 (26.4)	89 (34.2)	
T2	160 (14.1)	100 (14.3)	19 (10.9)	41 (15.8)	
T3	133 (11.8)	74 (10.6)	21 (12.1)	38 (14.6)	
T4	47 (4.2)	30 (4.3)	9 (5.2)	8 (3.1)	
Pathological N stage (%)					0.993
N0/Nx	1044 (92.3)	643 (92.3)	161 (92.5)	240 (92.3)	
$\geq N1$	87 (7.7)	54 (7.7)	13 (7.5)	20 (7.7)	
Pathological grade (%)					0.022
Low	322 (28.5)	205 (29.4)	59 (33.9)	58 (22.3)	
High	809 (71.5)	492 (70.6)	115 (66.1)	202 (77.7)	
Surgery modalities (%)					0.002
TURBT	598 (52.9)	387 (55.5)	98 (56.3)	113 (43.5)	
RC	533 (47.1)	310 (44.5)	76 (43.7)	147 (56.5)	
Lymphovascular invasion					0.097
Absent	1007 (89.0)	629 (90.2)	156 (89.7)	222 (85.4)	
Present	124 (11.0)	68 (9.8)	18 (10.3)	38 (14.6)	
OS, months M (Q <sub>1</sub> , Q <sub>3</sub> )	38.0 (23.0, 57.0)	38.0 (22.0, 62.0)	40.5 (23.0, 71.0)	38.5 (26.0, 50.0)	0.479
Status (%)					0.416
Alive	935 (82.7)	570 (81.8)	143 (82.2)	222 (85.4)	
Death	196 (17.3)	127 (18.2)	31 (17.8)	38 (14.6)	

SD: standard deviation; TURBT: transurethral resection of bladder tumor. RC: radical cystectomy; M (Q<sub>1</sub>, Q<sub>3</sub>): Median with interquartile range (first quartile, third quartile).

<sup>a</sup>p value was calculated by Chi-square test or t-test.

**Table 1:** Clinical characteristics of patients across the groups.

MF-DLM achieved a significantly higher C-index than the model without the Cross-attention module (0.838 vs. 0.782,  $p = 0.012$ ; [Supplementary Fig. S6](#)).

### Performance of MF-DLM in predicting OS

[Table 2](#) and [Fig. 2A–C](#) details the performance of the MF-DLM and the individual models in predicting OS of BCa. The results show that the MF-DLM exhibits outstanding performance for predicting OS in BCa, with C-indices of 0.902 (95% CI: 0.878–0.929), 0.864 (95% CI: 0.765–0.929), and 0.841 (95% CI: 0.771–0.895) in the training, internal validation, and external test sets, respectively. Besides, MF-DLM significantly outperforms pathological T (pT) stage across all sets (training: 0.902 vs. 0.793,  $p < 0.001$ ; validation: 0.864 vs. 0.757,  $p = 0.014$ ; external test: 0.841 vs. 0.760,  $p = 0.047$ ) as illustrated in [Fig. 2A–C](#).

The AUCs for the internal validation set are 0.944 (95% CI: 0.887, 1.000) for 1-year OS, 0.919 (95% CI: 0.787, 1.000) for 3-year OS, and 0.847 (95% CI: 0.681, 1.000) for 5-year OS ([Table 2](#), [Fig. 2E](#)). In the external test set, AUCs are 0.896 (95% CI: 0.771, 1.000), 0.862 (95% CI: 0.766, 0.959), and 0.836 (95% CI: 0.681, 0.990) for 1-, 3-, and 5-year OS, respectively ([Table 2](#), [Fig. 2F](#)). Additionally, brier scores are 0.051 (1-year), 0.082 (3-year), and 0.106 (5-year) in the validation set, and 0.018 (1-year), 0.088 (3-year), and 0.130 (5-year) in the external test set, indicating low prediction error probability for the MF-DLM.

In the internal validation set, a comparison of the MF-DLM and various modality models show that the C-index of the MF-DLM is higher than that of the other

multimodal features ([Table 2](#) and [Fig. 2B](#), DL model: 0.826 [95% CI: 0.739–0.891], Radiomic model: 0.822 [0.752–0.883], Vesical Imaging-Reporting and Data System [VI-RADS]: 0.799 [0.714–0.871], Clinical model: 0.564 [0.454–0.682]). Correspondingly, the external validation set exhibits analogous performance metrics ([Table 2](#) and [Fig. 2C](#); DL model: 0.817 [95% CI: 0.717–0.891], Radiomic model: 0.800 [0.737–0.858], VI-RADS: 0.755 [0.666–0.837], Clinical model: 0.578 [0.491–0.672]).

### Survival stratification of the MF-DLM in all patients

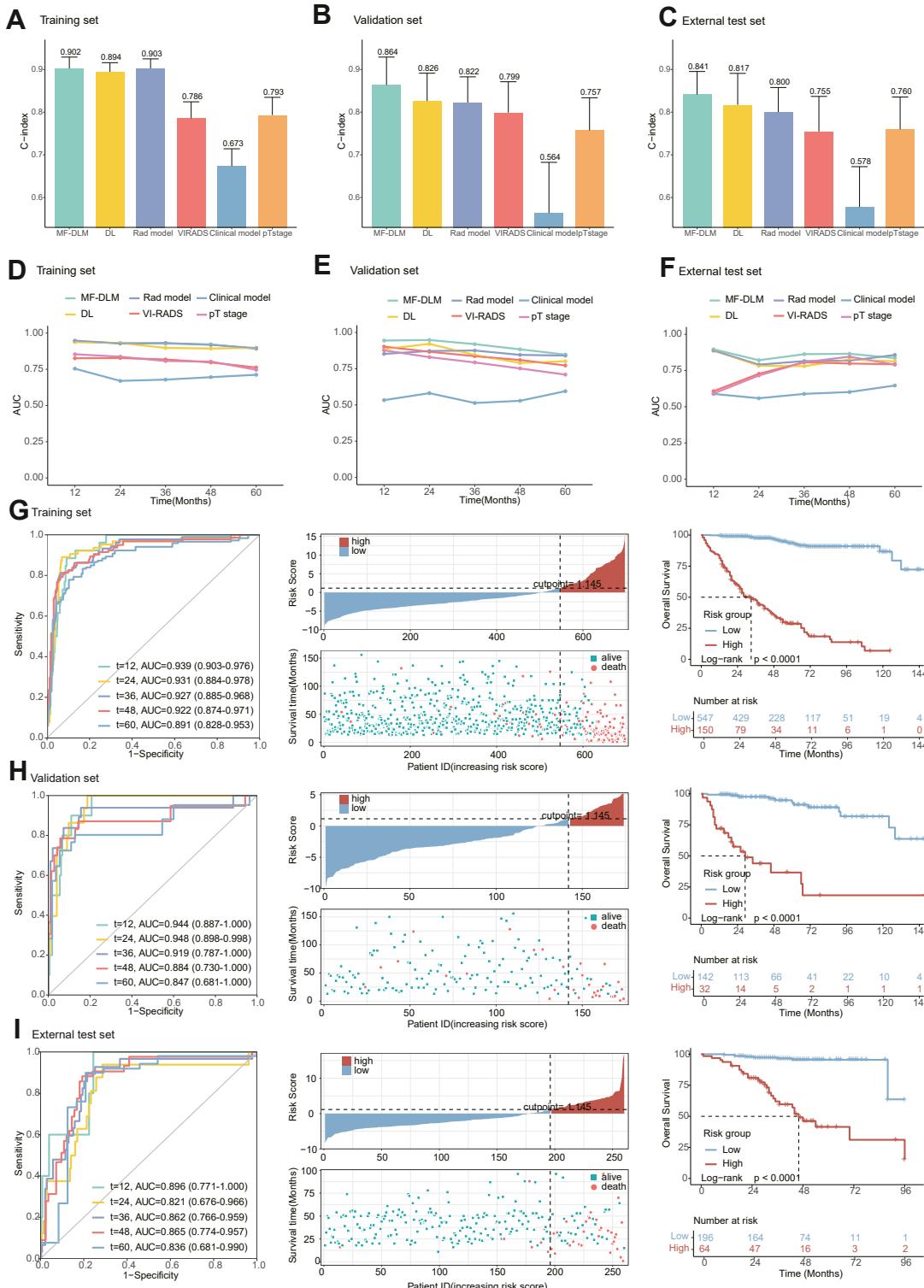
Patients were classified into low-risk and high-risk groups based on MF-DLM scores (threshold: 1.145). The low-risk group shows significantly longer OS across training, validation, and external test sets ([Fig. 2G–I](#),  $p < 0.001$ ). Furthermore, we demonstrate the capability of our risk groups derived from the MF-DLM to stratify patients across various external test sets, with significant prognostic differentiation observed in the majority of datasets ([Supplementary Fig. S8 A–E](#),  $p < 0.05$ ). MF-DLM risk stratification also significantly impacts OS across various clinical subgroups ([Supplementary Table S12](#) and [Supplementary Fig. S7](#)). Multivariate Cox regression analyses indicates that the high-risk group is an independent risk factor for OS ([Supplementary Table S13](#), hazard ratio [HR]: 12.95, 95% confidence interval [CI]: 8.67–19.33,  $p < 0.001$ ).

In the MIBC cohort, high-risk patients have significantly worse OS than low-risk patients ([Fig. 3A](#), HR:

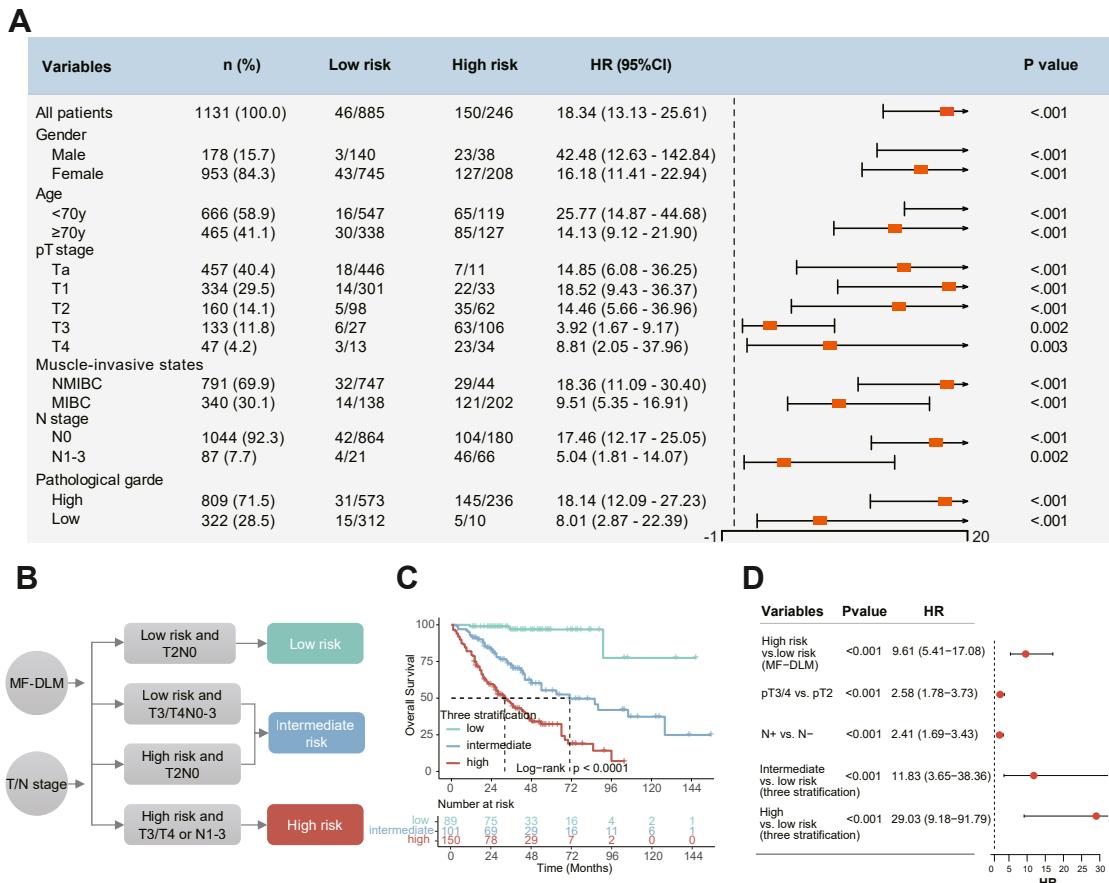
	Methods	1-year AUC	2-year AUC	3-year AUC	4-year AUC	5-year AUC	C-index
Training set (n = 697)	MF-DLM	0.939 (0.903–0.976)	0.931 (0.884–0.978)	0.927 (0.885–0.968)	0.922 (0.874–0.971)	0.891 (0.828–0.953)	0.902 (0.878–0.929)
	DL MRI model	0.920 (0.881–0.959)	0.917 (0.879–0.954)	0.904 (0.862–0.946)	0.899 (0.853–0.945)	0.897 (0.848–0.946)	0.894 (0.872–0.916)
	Radiomics model	0.948 (0.915–0.980)	0.928 (0.894–0.962)	0.931 (0.901–0.961)	0.919 (0.882–0.956)	0.896 (0.849–0.943)	0.903 (0.883–0.925)
	VI-RADS	0.826 (0.747–0.905)	0.828 (0.761–0.894)	0.818 (0.757–0.878)	0.797 (0.731–0.864)	0.762 (0.686–0.839)	0.786 (0.752–0.824)
	Clinical model	0.754 (0.634–0.874)	0.670 (0.577–0.763)	0.679 (0.596–0.761)	0.696 (0.616–0.775)	0.712 (0.634–0.791)	0.673 (0.627–0.714)
Validation set (n = 174)	MF-DLM	0.944 (0.887–1.000)	0.948 (0.898–0.998)	0.919 (0.787–1.000)	0.884 (0.730–1.000)	0.847 (0.681–1.000)	0.864 (0.765–0.929)
	DL MRI model	0.906 (0.843–0.970)	0.923 (0.856–0.991)	0.865 (0.730–1.000)	0.799 (0.620–0.978)	0.799 (0.628–0.970)	0.826 (0.739–0.891)
	Radiomics model	0.852 (0.733–0.970)	0.872 (0.765–0.979)	0.875 (0.773–0.978)	0.845 (0.740–0.950)	0.841 (0.733–0.948)	0.822 (0.752–0.883)
	VI-RADS	0.902 (0.824–0.981)	0.867 (0.751–0.983)	0.836 (0.680–0.992)	0.810 (0.643–0.976)	0.772 (0.593–0.952)	0.799 (0.714–0.871)
	Clinical model	0.533 (0.299–0.767)	0.581 (0.399–0.763)	0.513 (0.349–0.677)	0.529 (0.355–0.702)	0.595 (0.426–0.763)	0.564 (0.454–0.682)
External test set (n = 260)	MF-DLM	0.896 (0.771–1.000)	0.821 (0.676–0.966)	0.862 (0.766–0.959)	0.865 (0.774–0.957)	0.836 (0.681–0.990)	0.841 (0.771–0.895)
	DL MRI model	0.903 (0.823–0.982)	0.793 (0.637–0.949)	0.821 (0.704–0.938)	0.848 (0.742–0.954)	0.835 (0.693–0.976)	0.817 (0.717–0.891)
	Radiomics model	0.887 (0.812–0.963)	0.790 (0.658–0.923)	0.815 (0.707–0.922)	0.818 (0.716–0.919)	0.857 (0.737–0.978)	0.800 (0.737–0.858)
	VI-RADS	0.607 (0.256–0.959)	0.727 (0.550–0.905)	0.807 (0.681–0.932)	0.798 (0.675–0.921)	0.793 (0.652–0.934)	0.755 (0.666–0.837)
	Clinical model	0.589 (0.238–0.939)	0.559 (0.376–0.741)	0.589 (0.441–0.737)	0.602 (0.444–0.760)	0.646 (0.448–0.845)	0.578 (0.491–0.672)

MF-DLM: multimodal fusion deep learning model; MRI: magnetic resonance imaging; AUC: area under the curve; DL: deep learning; VI-RADS: Vesical Imaging Reporting and Data System.

**Table 2: Performance comparison of models for predicting the overall survival in the training, validation, and external test set.**



**Fig. 2:** Construction and evaluation of the MF-DLM based on multimodal features (A-C) The C-index values (main plot) and the AUC values at different time points (upper right corner) for MF-DLM and the individual models predicting OS in the training set (A), validation set (B), and external test set (C); (D-F) The AUC values at different time points for MF-DLM and the individual models predicting OS in the training set (D), validation set (E), and external test set (F); (G-I) The ROC curve (left), distribution of risk score and survival status (middle), Kaplan-Meier curve (right) of MF-DLM in training (G), validation (H) and external test set (I). MF-DLM: multimodal fusion deep learning model; AUC: area under the curve; ROC: receiver operating characteristics; OS: overall survival.



**Fig. 3:** The stratification effect of the MF-DLM-based two risk groups in different subgroups (A) Subgroup forest plot of risk groups based on the MF-DLM across different clinical and pathological characteristics. (B) Patients with MIBC were stratified into low-risk, intermediate-risk, and high-risk groups based on a combination of the MF-DLM-based risk groups and the pathological T and N stages. (C) Kaplan-Meier curves in patients stratified into low-risk, intermediate-risk, and high-risk groups. (D) Univariate Cox regression analysis of the MF-DLM-based risk groups, pT stage, pN stage, and triple stratification system. MF-DLM: multimodal fusion deep learning model; MIBC: muscle-invasive bladder cancer; OS: Overall survival.

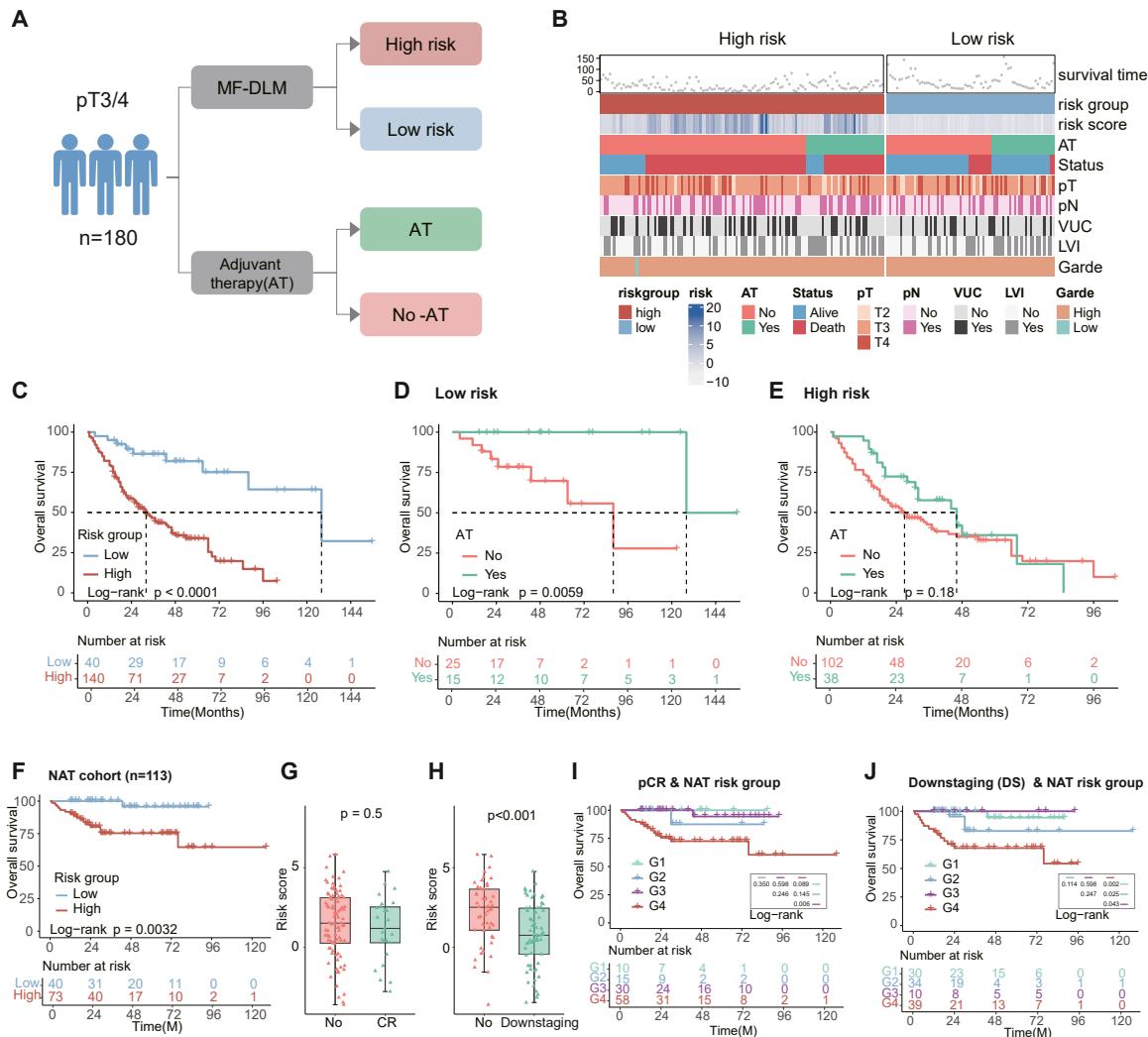
9.51, 95% CI: 5.35–16.91,  $p < 0.001$ ). Furthermore, we develop a prognostic risk score using a multivariable Cox regression that incorporates MF-DLM, pT stage, and pathological N (pN) stage, fulfilling the PH assumption (Supplementary Table S14). Patients were then stratified into low-, intermediate-, and high-risk groups based on tertiles of the risk scores. The three stratification system in MIBC shows significant prognostic differences (Fig. 3C–D,  $p < 0.001$ ). High-risk patients exhibit more substantial differences compared to low-risk patients when stratified by pT stage or pN stage alone (Fig. 3D, HR: 29.03, 95% CI: 9.18–91.79,  $p < 0.001$ ).

#### Survival stratification of the MF-DLM in the perioperative therapy cohorts

Fig. 4A–B details clinicopathological characteristics of the pathological T3/4 (pT3/4) sub-cohort. Further

stratification analysis within this pT3/4 population reveals that high-risk patients exhibited significantly poorer OS than low-risk patients (Fig. 4C,  $p < 0.001$ ). Notably, low-risk pT3/4 patients show improved OS with adjuvant therapy (AT) compared to those without (Fig. 4D,  $p = 0.0059$ ), whereas high-risk pT3/4 patients show no significant difference regardless of AT (Fig. 4E,  $p = 0.18$ ).

Additionally, a neoadjuvant risk score was established in an additional cohort of 113 patients with BCa treated with neoadjuvant therapy (NAT). The neoadjuvant low-risk group has better OS than the high-risk group (Fig. 4F,  $p = 0.0032$ ). Furthermore, subgroup analysis shows neoadjuvant low-risk patients have significantly better OS than high-risk patients in cohorts without pathological complete response (pCR) (Fig. 4I,  $p = 0.006$ ) or lacking downstaging (Fig. 4J,  $p = 0.043$ ). No significant OS difference is observed



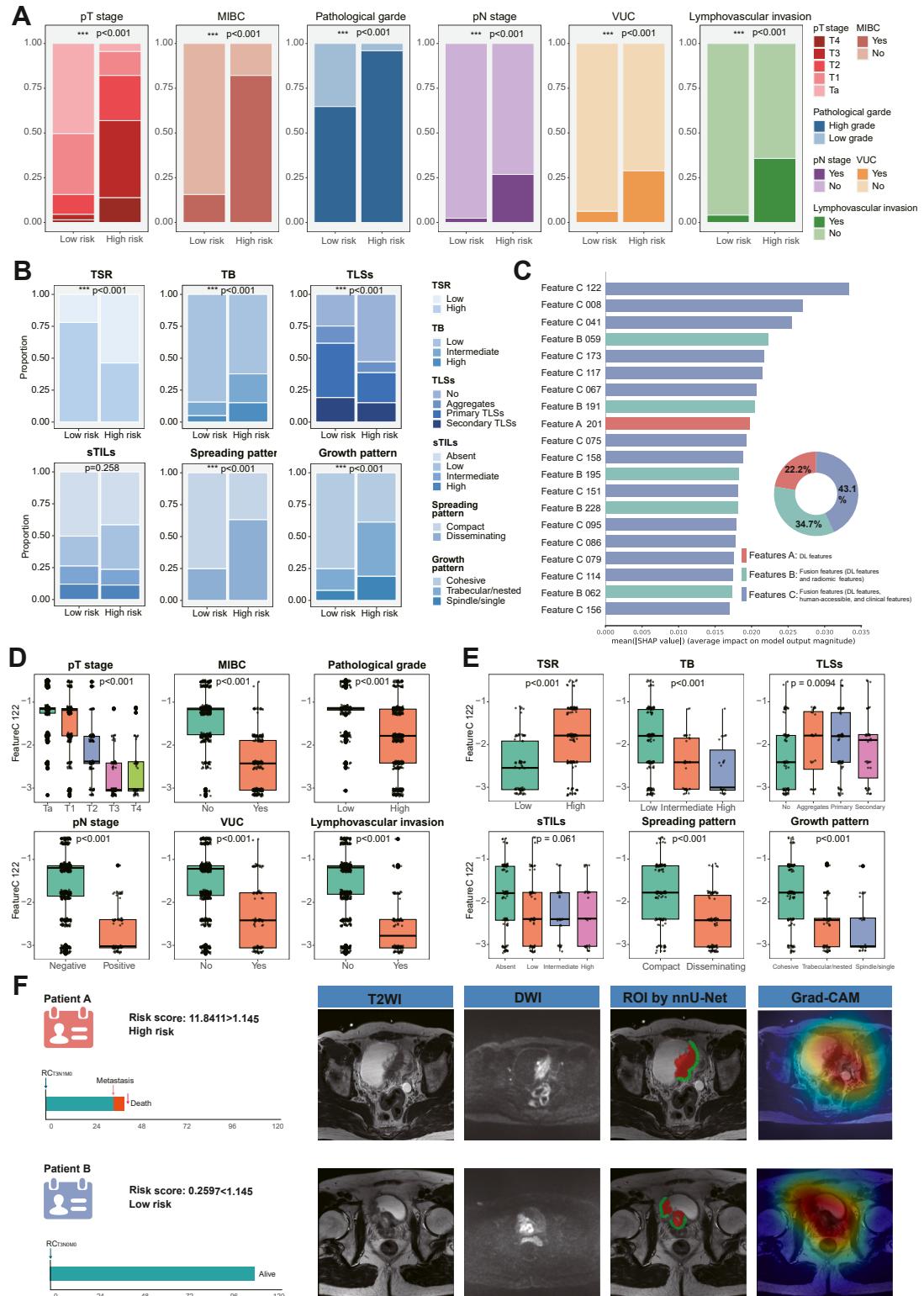
**Fig. 4:** The performance of the MF-DLM in perioperative treatment cohorts (A) The status of adjuvant therapy in the pT3/4 cohort. (B) Heatmap of risk stratification, prognosis, and pathological characteristics for patients in the pT3/4 cohort. (C) Kaplan-Meier curve of the MF-DLM-based risk group in pT3/4 cohort. (D, E) Kaplan-Meier curve of the MF-DLM risk scores for low-risk or high-risk patients in the pT3/4 cohort, comparing the adjuvant and non-adjuvant therapy groups. (F) Kaplan-Meier curves of MF-DLM-based risk groups in the neoadjuvant treatment cohort ( $n = 113$ ). (G, H) Boxplot of the MF-DLM risk scores in the status of pCR and downstaging. (I, J) Subgroups Kaplan-Meier curves of four-category groups combined the MF-DLM-based risk groups and the status of pCR (G1: pCR and low risk; G2: pCR and high risk; G3: non-pCR and low risk; G4: non-pCR and high risk) or downstaging (G1: pDS and low risk; G2: pDS and high risk; G3: non-pDS and low risk; G4: non-pDS and high risk). MF-DLM: multimodal fusion deep learning model; pT3/4: pathological T3/4; VUC: variant urothelial subtypes; LVI: lymphovascular invasion; pCR: pathological complete response; pDS: pathological downstaging.

between the two groups in the pCR (Fig. 4I,  $p = 0.350$ ) and downstaged cohort (Fig. 4J,  $p = 0.114$ ).

#### The interpretability of MF-DLM with pathological characteristics

To better understand MF-DLM, we evaluated its associations with pathological characteristics. The high-risk group shows significant correlations with higher T and lymph node metastasis, high-grade tumors, variant urothelial subtypes and lymphovascular invasion

(Fig. 5A,  $p < 0.001$ ). Furthermore, analysis of additional pathological characteristics (Supplementary Fig. S9) reveals that patients in the low-risk group exhibits higher tumor-stroma ratio (TSR) and greater proportions of tertiary lymphoid structures (TLSs). Conversely, the high-risk group demonstrates higher tumor budding (TB) with predominant trabecular/nested and spindle/single-cell growth patterns, along with a disseminating spreading pattern (Fig. 5B; all  $p < 0.001$ ).



**Fig. 5:** Interpretability analysis of the MF-DLM. (A) The bar charts of the pT stage, MIBC, pathological grade, pN stage, VUC, and lymphovascular invasion in two-risk groups based on MF-DLM. (B) The bar charts of the TSR, TB, TLSs, spreading pattern, growth patterns, and sTILs in two-risk groups. (C) Feature importance ranking for survival prediction task by the MF-DLM, based on SHAP. (D, E) The box plots illustrating the most significant MF-DLM-derived features across different pathological characteristics. (F) Typical high-risk and low-risk cases based on MF-DLM. MF-DLM: multimodal fusion deep learning model; TSR: Tumor-stroma ratio; TB: Tumor budding; TLSs: Tertiary lymphoid structures; sTILs: stromal tumor infiltrating lymphocytes; SHAP: Shapley additive explanations.

SHAP analysis elucidates the importance of model features. As illustrated in Fig. 5C, the cross-attention-based fusion features (combining DL features, morphological and clinical features) contributes most to the MF-DLM's survival prediction, accounting for 43.1% of the total feature importance. We further analyzed the top crucial MF-DLM feature. Fig. 5D–E demonstrate its correlation with pathological characteristics. This top MF-DLM feature shows significantly lower expression in subgroups with higher T stage, lymph node metastasis, high-grade tumors, variant urothelial subtypes, carcinoma in situ, and lymphovascular invasion (Fig. 5D; all  $p < 0.05$ ). Similarly, in tumors with low TSR, high TB, less TLSs, disseminating spread, and trabecular/nested or spindle/single-cell growth patterns, its expression is also significantly lower (Fig. 5E; all  $p < 0.05$ ). Supplementary Fig. S10 validates the MRI-histopathological correlation through a representative patient case with matched Hematoxylin and eosin (H&E) stained sections. Besides, two representative examples are presented by demonstrating the original images, nnU-Net-based ROI segmentations, and Grad-CAM visualizations (Fig. 5F).

## Discussion

In this retrospective multicentre study with 1131 patients, we develop and validate a multimodal fusion deep learning model that integrates MRI-based deep learning features, radiomic features, human-assessed morphological features, and clinical features for non-invasive prediction of OS. The model achieves C-index values of 0.902, 0.864, and 0.841 in the training, validation, and external test sets, respectively.

Medical images like CT and MRI are inherently 3D space, and employing 2D approaches for 3D tasks can result in the loss of essential structural information.<sup>28</sup> However, developing 3D deep learning models necessitates large datasets, which poses challenges in medical research due to limited data availability.<sup>29</sup> Unlike the study by Wei et al.,<sup>16</sup> which employed 2D CT images to predict OS in BCa, this study developed a 3D DL model using the largest reported dataset of BCa MRI and validated it across seven external centres. In addition, the peritumoral region reflects tumor microenvironment characteristics and invasion patterns<sup>30–32</sup> and multiple studies have shown that incorporating the peritumoral region could enhance predictive performance for tumor prognosis.<sup>33,34</sup> Thus, this study adopted a three-channel input format comprising the original image, the tumor region, and the peritumoral region.

The ResNet architectures effectively capture hierarchical image features and have been adopted in medical imaging for recognition and prognosis prediction.<sup>35</sup> In small-to-medium datasets, The ResNet shows strong generalization and superior performance than 3D-

Vision Transformer, and 3D-DenseNet.<sup>36</sup> In this study, we selected a modified 3D-ResNet50 as the DL feature extractor for survival tasks after a comparative evaluation of 3D-ResNet, 3D-Vision Transformer, and 3D-DenseNet, with the resulting MF-DLM demonstrating superior performance. Notably, the MF-DLM achieved a higher C-index compared to single-modality models, underscoring the value of multi-modal fusion. While the radiomics model demonstrated a slight advantage in predicting 5-year OS in the external test set, this isolated finding did not diminish the superiority of MF-DLM in comprehensive survival assessment, as it consistently outperformed all single-modality models across multiple timepoints.

Effective prognostic stratification, especially for patients with MIBC, is crucial for guiding clinical decision-making. TNM staging remains the primary tool for OS stratification in BCa. However, it does not account for tumor heterogeneity or biological characteristics, leading to prognostic variability among patients with the same stage.<sup>3,4</sup> As a result, traditional TNM staging falls short in supporting personalized treatment decisions. Notably, prior studies have not fully examined the distinct prognostic differences within the same BCa category.<sup>16,17,19</sup> In our study, we defined two risk groups based on the MF-DLM score, which achieved higher C-index scores than pT stage alone. Furthermore, the MF-DLM-driven risk groups demonstrated significant prognostic differences not only across the common clinical prognostic markers such as VI-RADS scores 2–4, but also within both pathological T and N stage subgroups. Given the poor prognosis of MIBC, we further integrated MF-DLM-driven risk groups with pathological T and N stages to establish a refined three stratification system (low-risk, intermediate-risk, and high-risk) for MIBC. This system revealed more pronounced OS difference between the high- and low-risk groups compared with stratification by pT or pN stage alone. In summary, these findings highlighted that the MF-DLM provided prognostic insights beyond pathological staging and offered a new perspective on prognostic stratification for BCa.

The MF-DLM also has potential to guide perioperative treatment decisions in bladder cancer. Perioperative systemic therapies, including neoadjuvant and adjuvant treatments, offered significant survival benefits for MIBC patients.<sup>1,37,38</sup> A recent meta-analysis with 1183 MIBC participants confirmed a significant OS advantage associated with cisplatin-based adjuvant chemotherapy.<sup>37</sup> However, identifying which pT3/4 patients benefit most from adjuvant therapy remains challenging.<sup>39</sup> Zhang et al.<sup>18</sup> evaluated a CT-based DL model in the adjuvant therapy setting, but their cohort comprised NMIBC patients undergoing TURBT and did not include a stratified analysis focused on the high-mortality-risk pT3/4 population. In our study, MF-DLM stratification revealed that low-risk pT3/4 patients

experienced greater OS benefit from adjuvant therapy compared with the broader pT3/4 population where high-risk patients derived no benefit. This suggested that the MF-DLM-based risk stratification could be a critical factor in determining adjuvant therapy efficacy. For neoadjuvant therapy, achieving pCR is strongly associated with OS,<sup>40</sup> yet no effective method currently exists to evaluate survival benefits in patients without pCR or downstaging. We found that lower MF-DLM scores correlated significantly with downstaging. Importantly, among patients without pCR or downstaging, those in the MF-DLM low-risk group still derived substantial survival benefits from neoadjuvant therapy. This finding indicated that the survival benefits of neoadjuvant therapy were not solely reflected in tumor reduction. In summary, MF-DLM-defined low-risk pT3/4 patients appeared to derive greater benefits from adjuvant therapy, while in the neoadjuvant setting, low-risk patients still achieve survival benefits even in the absence of pCR.

Although the model achieved excellent performance, the interpretability of deep learning models still poses challenges on their clinical application. To address this, we analyzed the relationship between the MF-DLM and histological characteristics. The high-risk group was significantly associated with adverse characteristics, such as higher pT and pN stages, high-grade tumors, variant urothelial subtypes, lymphovascular invasion and tumor budding. Distinct growth and spreading patterns in BCa reflect tumor biological behavior.<sup>41</sup> Patients with cohesive-compact tumors showed significantly better survival than those with diffusely infiltrating or trabecular/nested tumors.<sup>41</sup> Consistently, the MF-DLM high-risk group predominantly exhibited trabecular/nested and spindle/single-cell patterns with disseminated spreading, thereby validating its alignment with aggressive tumor behavior. Furthermore, the low-risk group exhibited higher TSR and TLSs, both of which are established biomarkers linked to treatment response.<sup>42,43</sup> TSR, which refers to the ratio of tumor components to stromal components, is critical in disease progression and treatment resistance. Low TSR was associated with worse prognosis and lower rates of pCR,<sup>42,44,45</sup> while mature TLSs correlate with improved immunotherapy response.<sup>43</sup> These findings provided a biological basis for the MF-DLM's ability to identify patients who could benefit from neoadjuvant and pT3/4 adjuvant therapies.

While leveraging established components like 3D ResNet50, multimodal integration, cross-attention fusion, and interpretability analysis (e.g., SHAP and Grad-CAM), our key innovation is a pipeline specifically optimized for BCa prognosis and validated across multiple centres. However, our study has limitations. Firstly, the model was trained and tested on retrospective data from a relatively homogeneous ethnic group and healthcare settings, which may introduce inherent

bias and unmeasured confounding factors. Secondly, we excluded the dynamic contrast-enhanced (DCE) sequence to maintain clinical applicability and data integrity, which could limit the model's ability to exploit perfusion dynamics, but also ensured that MF-DLM remained applicable to patients unable to undergo DCE for various reasons. Besides, although the median follow-up time exceeded three years, the reduced sample size limited the precision of the 5-year prognostic evaluation. Additionally, while we utilized SHAP as a post-hoc analysis method and Grad-CAM to explain MF-DLM, Grad-CAM generated heatmaps only for the image branch. Explainable models are necessary to resolve DL's "black box" limitations in future work. Finally, despite applying inter-centre batch harmonization, heterogeneity in acquisition protocols across institutions still challenged the model's generalizability. Nonetheless, by analyzing prognostic differences at distinct external sets, we confirmed substantial stratification efficacy of MF-DLM, lending potential support to its feasibility in broader clinical applications.

In conclusion, we developed and validated an artificial intelligence model using preoperative MRI to non-invasively predict OS in bladder cancer. In patients with MIBC, this model offered additional stratification value beyond the pathological T and N stages. Moreover, the MF-DLM can identify pT3/4 patients likely to benefit from adjuvant therapy, as well as MIBC patients likely to benefit from neoadjuvant therapy.

#### Contributors

Lingkai Cai: Writing – original draft, Visualization; Rongjie Bai, Qiang Cao: Formal Analysis, Data curation, Visualization; Weijie Sun, Fei Wang: Formal Analysis, Software, Visualization, Supervision; Xiaotong Liu, Bo Liang, Meihua Jiang, Gongcheng Wang, Qiang Shao, Xuping Jiang, Meiling Bao: Resources; Chenghao Wang, Chang Chen: Software, Visualization; Zhengye Tan, Qikai Wu, Hao Yu: Data curation, Formal Analysis; Pengchao Li: Supervision; Xiao Yang: Writing – review & editing, Funding acquisition, Methodology, Supervision; Qiang Lu: Writing – review & editing, Funding acquisition.

Xiao Yang and Qiang Lu directly accessed and verified the raw data, they accessed the raw data and confirmed their accuracy.

#### Data sharing statement

The data that support this study's findings are available from the corresponding authors with a signed data access agreement. Deidentified patient-level clinical and outcome data will be provided upon reasonable request. The MRI data are not publicly available because they contain sensitive information that could compromise patient privacy. The source code for the deep learning model is available online (<https://github.com/cai19cai/MF-DLM.git>).

#### Declaration of interests

The authors declare that they have no conflict of interest.

#### Acknowledgements

The authors thank the patients for providing personal information. This work was supported by the Noncommunicated Chronic Diseases-National Science and Technology Major Project (2024ZD0525700), the National Natural Science Foundation of China (82273152, 82503879), Jiangsu Province Hospital (the First Affiliated Hospital of Nanjing Medical University) Clinical Capacity Enhancement Project (JSPH-MA-2022-5), China Postdoctoral Science Foundation funded

project (2024M761211), and the Nanjing Postdoctoral Science Foundation funded project (2024BHS210).

#### Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.eclimn.2025.103640>.

#### References

- 1 Alfred WJ, Max Bruins H, Carrion A, et al. European association of urology guidelines on muscle-invasive and metastatic bladder cancer: summary of the 2023 guidelines. *Eur Urol*. 2024;85:17–31.
- 2 Babjuk M, Burger M, Capoun O, et al. European association of urology guidelines on non-muscle-invasive bladder cancer (Ta, T1, and carcinoma in situ). *Eur Urol*. 2022;81:75–94.
- 3 International Bladder Cancer Nomogram C, Bochner BH, Kattan MW, Vora KC. Postoperative nomogram predicting risk of recurrence after radical cystectomy for bladder cancer. *J Clin Oncol*. 2006;24:3967–3972.
- 4 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;67:93–99.
- 5 Zhuang J, Cai L, Sun H, et al. Vesical imaging reporting and data system (VI-RADS) could predict the survival of bladder-cancer patients who received radical cystectomy. *Sci Rep*. 2023;13:21502.
- 6 Panebianco V, Narumi Y, Altun E, et al. Multiparametric magnetic resonance imaging for bladder cancer: development of VI-RADS (vesical imaging-reporting and data system). *Eur Urol*. 2018;74:294–306.
- 7 Cao B, Li Q, Xu P, et al. Preliminary exploration of the application of vesical imaging-reporting and data system (VI-RADS) in post-treatment patients with bladder cancer: a prospective single-center study. *J Magn Reson Imaging*. 2022;55:275–286.
- 8 Yajima S, Yoshida S, Takahara T, et al. Usefulness of the inchworm sign on DWI for predicting pT1 bladder cancer progression. *Eur Radiol*. 2019;29:3881–3888.
- 9 Moore NS, McWilliam A, Aneja S. Bladder cancer radiation oncology of the future: prognostic modelling, radiomics, and treatment planning with artificial intelligence. *Semin Radiat Oncol*. 2023;33:70–75.
- 10 Woznicki P, Laqua FC, Messmer K, et al. Radiomics for the prediction of overall survival in patients with bladder cancer prior to radical cystectomy. *Cancers (Basel)*. 2022;14:4449.
- 11 Zhang X, Guo J, Wang L, et al. Recurrence risk prediction for non-muscle-invasive bladder urothelial carcinoma using diffusion and clinicopathology features. *Abdom Radiol (NY)*. 2025. <https://doi.org/10.1007/s00261-025-05023-6>.
- 12 Yang G, Bai J, Hao M, Zhang L, Fan Z, Wang X. Enhancing recurrence risk prediction for bladder cancer using multi-sequence MRI radiomics. *Insights Imaging*. 2024;15:88.
- 13 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.
- 14 Wang S, Dong D, Li L, et al. A deep learning radiomics model to identify poor outcome in COVID-19 patients with underlying health conditions: a multicenter study. *IEEE J Biomed Health Inform*. 2021;25:2353–2362.
- 15 Jiang X, Hoffmeister M, Brenner H, et al. End-to-end prognostication in colorectal cancer by deep learning: a retrospective, multicentre study. *Lancet Digit Health*. 2024;6:e33–e43.
- 16 Wei Z, Xv Y, Liu H, et al. A CT-based deep learning model predicts overall survival in patients with muscle invasive bladder cancer after radical cystectomy: a multicenter retrospective cohort study. *Int J Surg*. 2024;110:2922–2932.
- 17 Wang H, Zhang M, Miao J, et al. Deep learning signature based on multiphase enhanced CT for bladder cancer recurrence prediction: a multi-center study. *eClinicalMedicine*. 2023;66:102352.
- 18 Zhang M, Zhao Y, Hao D, et al. An interpretable CT-based deep learning model for predicting overall survival in patients with bladder cancer: a multicenter study. *NPJ Precis Oncol*. 2025;9:288.
- 19 Huang H, Huang Y, Kaggie JD, et al. Multiparametric MRI-based deep learning radiomics model for assessing 5-Year recurrence risk in non-muscle invasive bladder cancer. *J Magn Reson Imaging*. 2025;61:1442–1456.
- 20 Diao Z, Jiang H. A multi-instance tumor subtype classification method for small PET datasets using RA-DL attention module guided deep feature extraction with radiomics features. *Comput Biol Med*. 2024;174:108461.
- 21 Chen Z, Chen Y, Sun Y, et al. Predicting gastric cancer response to anti-HER2 therapy or anti-HER2 combined immunotherapy based on multi-modal data. *Signal Transduct Target Ther*. 2024;9:222.
- 22 Yang M, Ma J, Zhang C, et al. Multimodal data deep learning method for predicting symptomatic pneumonitis caused by lung cancer radiotherapy combined with immunotherapy. *Front Immunol*. 2024;15:1492399.
- 23 Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385:e078378.
- 24 Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnUNet: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 2021;18:203–211.
- 25 Dai Y, Gieseke F, Oehmcke S, Wu Y, Barnard K. Attentional feature fusion. 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA. 2021:3559–3568. <https://doi.org/10.1109/WACV48630.2021.00360>.
- 26 Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized treatment recommender system using a cox proportional hazards deep neural network. *BMC Med Res Methodol*. 2018;18:24.
- 27 Deng H, Eftekhari Z, Carlin C, et al. Development and validation of an explainable machine learning model for major complications after cytoreductive surgery. *JAMA Netw Open*. 2022;5:e2212930.
- 28 Zheng Y, Shi H, Hai B, Zhang J. A commentary on ‘A CT-based deep learning model predicts overall survival in patients with muscle invasive bladder cancer after radical cystectomy: a multicenter retrospective cohort study’. *Int J Surg*. 2024;110:5200–5201.
- 29 Singh SP, Wang L, Gupta S, Goli H, Padmanabhan P, Gulyas B. 3D deep learning on medical images: a review. *Sensors (Basel)*. 2020;20:5097.
- 30 Li C, Song L, Yin J. Intratumoral and peritumoral radiomics based on functional parametric maps from breast DCE-MRI for prediction of HER-2 and Ki-67 status. *J Magn Reson Imaging*. 2021;54:703–714.
- 31 Jiang W, Meng R, Cheng Y, et al. Intra- and peritumoral based radiomics for assessment of lymphovascular invasion in invasive breast cancer. *J Magn Reson Imaging*. 2024;59:613–625.
- 32 Xiao ML, Fu L, Wei Y, et al. Intratumoral and peritumoral MRI radiomics nomogram for predicting parametrial invasion in patients with early-stage cervical adenocarcinoma and adenosquamous carcinoma. *Eur Radiol*. 2024;34:852–862.
- 33 Li S, Wan X, Deng YQ, et al. Predicting prognosis of nasopharyngeal carcinoma based on deep learning: peritumoral region should be valued. *Cancer Imaging*. 2023;23:14.
- 34 Huang Y, Qian H. Advancing hepatocellular carcinoma management through peritumoral radiomics: enhancing diagnosis, treatment, and prognosis. *J Hepatocell Carcinoma*. 2024;11:2159–2168.
- 35 Kong C, Yan D, Liu K, Yin Y, Ma C. Multiple deep learning models based on MRI images in discriminating glioblastoma from solitary brain metastases: a multicentre study. *BMC Med Imaging*. 2025;25:171.
- 36 Raghu M, Unterthiner T, Kornblith S, Zhang C, Dosovitskiy A. Do vision transformers see like convolutional neural networks? *Adv Neural Inf Process Syst*. 2021;34. <https://doi.org/10.48550/arXiv.2108.08810>.
- 37 Advanced Bladder Cancer Meta-analysis Collaborators G. Adjuvant chemotherapy for muscle-invasive bladder cancer: a systematic review and meta-analysis of individual participant data from randomised controlled trials. *Eur Urol*. 2022;81:50–61.
- 38 Yin M, Joshi M, Meijer RP, et al. Neoadjuvant chemotherapy for muscle-invasive bladder cancer: a systematic review and two-step meta-analysis. *Oncologist*. 2016;21:708–715.
- 39 Taubert H, Eckstein M, Epple E, et al. Immune cell-associated protein expression helps to predict survival in muscle-invasive urothelial bladder cancer patients after radical cystectomy and optional adjuvant chemotherapy. *Cells*. 2021;10:159.
- 40 Voskuilen CS, Oo HZ, Genitsch V, et al. Multicenter validation of histopathologic tumor regression grade after neoadjuvant

- chemotherapy in muscle-invasive bladder carcinoma. *Am J Surg Pathol.* 2019;43:1600–1610.
- 41 Eckstein M, Matek C, Wagner P, et al. Proposal for a novel histological scoring system as a potential grading approach for muscle-invasive urothelial bladder cancer correlating with disease aggressiveness and patient outcomes. *Eur Urol Oncol.* 2024;7:128–138.
- 42 Liu L, Xu L, Wu D, et al. Impact of tumour stroma-immune interactions on survival prognosis and response to neoadjuvant chemotherapy in bladder cancer. *EBioMedicine.* 2024;104: 105152.
- 43 Teillaud JL, Houel A, Panouillet M, Riffard C, Dieu-Nosjean MC. Tertiary lymphoid structures in anticancer immunity. *Nat Rev Cancer.* 2024;24:629–646.
- 44 Friedman G, Levi-Galibov O, David E, et al. Cancer-associated fibroblast compositions change with breast cancer progression linking the ratio of S100A4(+) and PDPN(+) CAFs to clinical outcome. *Nat Cancer.* 2020;1:692–708.
- 45 Lou E, Vogel RI, Hoostal S, et al. Tumor-stroma proportion as a predictive biomarker of resistance to platinum-based chemotherapy in patients with ovarian cancer. *JAMA Oncol.* 2019;5:1222–1224.