

Review

# Application Value of Radiomics-Based Machine Learning for Preoperative Risk Stratification of Bladder Cancer: Systematic Review and Meta-Analysis

Zirong He<sup>1\*</sup>, MM; Yinghua Liu<sup>1\*</sup>, MD; Qin Jiang<sup>1</sup>, MM; Fang Hou<sup>1,2\*</sup>, MD; Bing Xu<sup>1\*</sup>, MM

<sup>1</sup>Department of Pediatric Surgery, Sichuan Provincial People's Hospital, School of Medicine, University of Electronic Science and Technology of China, Chengdu, Sichuan, China

<sup>2</sup>Sichuan Provincial People's Hospital East Sichuan Hospital & Dazhou First People's Hospital, Dazhou, China

\*these authors contributed equally

## Corresponding Author:

Bing Xu, MM

Department of Pediatric Surgery, Sichuan Provincial People's Hospital, School of Medicine, University of Electronic Science and Technology of China

No. 32 West Second Section, First Ring Road, Qingyang District

Chengdu, Sichuan 610072

China

Phone: 86 18981838032

Email: [yourxubing@126.com](mailto:yourxubing@126.com)

## Abstract

**Background:** Some researchers have explored the application of radiomics-based machine learning to detect preoperative muscle invasion, high-grade tumors, human epidermal growth factor receptor 2 expression, and other risk factors for bladder cancer. However, systematic evidence proving its effectiveness remains lacking.

**Objective:** This study aimed to evaluate the performance of radiomics-based machine learning in preoperative risk stratification for patients with bladder cancer. These findings could contribute to advancing the development or updating of intelligent risk assessment tools for bladder cancer.

**Methods:** The Embase, Cochrane Library, PubMed, and Web of Science databases were systematically retrieved for publicly available studies on the effectiveness of radiomics-based machine learning (ML) in the preoperative risk stratification of bladder cancer up to October 17, 2025. The risk of bias in the included studies was evaluated using the Prediction Model Risk of Bias Assessment Tool for Artificial Intelligence. The overall quality of the studies was quantified using the Radiomics Quality Scoring tool. The certainty of the evidence was graded using the Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) framework. Subgroup analyses were conducted according to the type of imaging source and modeling method.

**Results:** This meta-analysis ultimately incorporated 57 studies with a total of 11,933 participants. These studies primarily used radiomics-based ML to identify muscle invasion (n=34) and high-grade tumors (n=16). Additionally, the methodology was used to evaluate human epidermal growth factor receptor 2 positive expression (n=3), Ki-67 expression (n=2), and lymph node staging (n=2) preoperatively in bladder cancer. In the validation sets, the pooled area under the receiver operating characteristic curve (AUROC) for identifying muscle invasion was 0.893 (95% CI 0.840-0.948), 0.916 (95% CI 0.891-0.942), and 0.840 (95% CI 0.737-0.958) for computed tomography (CT)-, magnetic resonance imaging (MRI)-, and ultrasound-based radiomics, respectively. The AUROC was 0.874 (95% CI 0.852-0.896) and 0.921 (95% CI 0.867-0.979) for models integrating clinical features with CT- or MRI-based radiomics, respectively. The pooled AUROC for diagnosing high-grade tumors was 0.874 (95% CI 0.775-0.985), 0.846 (95% CI 0.663-1.000), and 0.750 (95% CI 0.636-0.884) for CT-, MRI-, and ultrasound-based radiomics, respectively. Furthermore, the AUROC was 0.919 (95% CI 0.774-1.000) for MRI-based radiomics combined with clinical features.

**Conclusions:** This is the first systematic review to comprehensively evaluate the role of radiomics in preoperative risk stratification for bladder cancer. It provides evidence to inform the development and refinement of future ML-based tools for image analysis in this setting. However, this evidence faces significant challenges, including methodological shortcomings and a high risk of bias and low GRADE level, which preclude its readiness for clinical translation. Future studies should

standardize the methodological workflows in radiomics, conduct multicenter research, and thoroughly evaluate and discuss the validity of external validation.

**Trial Registration:** PROSPERO CRD42024561649; <https://www.crd.york.ac.uk/PROSPERO/view/CRD42024561649>

*J Med Internet Res* 2026;28:e81084; doi: [10.2196/81084](https://doi.org/10.2196/81084)

**Keywords:** radiomics; bladder cancer; muscle invasion; pathologic grading; risk classification

## Introduction

Bladder cancer is the tenth most common cancer globally. It ranks fourth in new cancer cases in men and is the eighth leading cause of cancer death [1,2]. In the United States, an estimated 83,190 new cases of bladder cancer occurred in 2024 (63,070 in men and 20,120 in women), resulting in 16,840 deaths (12,290 in men and 4550 in women) [2]. Thus, bladder cancer has become a serious social burden.

Transurethral resection of bladder tumors (TURBT), followed by pathological analysis, provides the basis for diagnosing, staging, and treating bladder cancer [3]. Nonetheless, this procedure is imprecise in assessing muscle invasion. The absence of detrusor muscle in the specimen is linked to a significantly elevated risk of residual lesions, early recurrence, and understaging of tumor [3,4]. TURBT also carries a few operational risks. A recent study of the learning curve for TURBT indicates that urologists should perform at least 100 TURBT procedures before achieving acceptable oncological outcomes and meeting the minimum requirements for surgical success. Furthermore, the study found that the first 45 TURBT procedures resulted in the worst outcomes [5]. Repeat transurethral resection provides a second chance in such cases. While repeat transurethral resection holds potential for improving tumor prognosis, evidence demonstrating its capacity to confer a survival benefit remains inconclusive [6,7]. Like TURBT, cystoscopy is an invasive procedure with risks of insufficient sampling and understaging. This creates a need for noninvasive, accurate diagnostic alternatives [8,9].

Magnetic resonance imaging (MRI) and computed tomography (CT) have been used to provide additional staging information. Nevertheless, neither technique can accurately evaluate microscopic infiltrates. Both of them aim to confirm or exclude locally advanced disease (stage  $\geq$ T3b) [10]. Therefore, exploring new preoperative techniques that can effectively identify the risk of muscle invasion and pathological grading of bladder cancer is clinically important.

In recent years, radiomics has received significant attention from researchers. Like therapies inspired by molecular biology, radiomics shows great promise in advancing precision medicine. This technique can predict outcomes individually or in combination with genomic, comorbidity, clinical, or demographic data. The process involves acquiring images, identifying volumes of interest (ie, those with possible prognostic value), segmenting the volume (ie, depicting the boundaries of the volume with computer-assisted contouring), and extracting and qualifying descriptive features from the volume. These features are then used

to populate searchable databases, which are subsequently mined to develop classifier models. Radiomics shows promise as a quantitative imaging biomarker for both characterizing bladder cancer and predicting its prognosis [11-14]. At the time of this writing, high-throughput computing enables the rapid extraction of numerous quantitative features from tomographic images (eg, CT, magnetic resonance, or positron emission tomography). Radiomics transforms medical images into quantifiable data for analysis. The motivation stems from the wealth of pathophysiological information embedded in medical images, which quantitative analysis can unlock to enhance decision-making [11,13,15]. This algorithmic assistance may provide more accurate histopathological diagnoses, save time, and increase clinician confidence, contributing to improved prognosis [12,16-18]. Recently, as radiomics has developed in oncological diagnosis and treatment, researchers have examined the application of radiomics-based machine learning (ML) in identifying muscle invasion [18-21] and high-grade tumors [15], as well as in determining human epidermal growth factor receptor 2 (HER2) and Ki-67 expression and lymph node (LN) staging for bladder cancer. Compared with genomic biomarkers, artificial intelligence (AI) systems leveraging digitized images offer a more cost-effective and readily scalable solution. These tools can continuously learn from new data, thereby enhancing their predictive performance and, consequently, their value to health care over time [22]. However, comprehensive and systematic evidence of its effectiveness remains lacking.

To address this gap, we conducted a meta-analysis aimed at systematically evaluating the application of radiomics-based ML to preoperative risk stratification in bladder cancer and quantitatively summarizing its diagnostic accuracy. The findings may establish an evidence base for developing and refining future image-based, AI-assisted tools of risk stratification.

## Methods

### Study Registration

This study followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines (Checklist 1). The review was prospectively registered with PROSPERO (International Prospective Register of Systematic Reviews) under the title “Assessment of Preoperative Risk Stratification for Bladder Cancer Using Machine Learning Based on Radiomics: A Systematic Review and Meta-Analysis” (CRD42024561649).

## Eligibility Criteria

### Inclusion Criteria

The inclusion criteria were as follows:

1. Patients with bladder cancer diagnosed by biopsy;
2. A complete ML model covering radiomics features was constructed to identify muscle invasion, high-grade tumors, HER2, LN staging, and Ki-67 expression;
3. Studies reported in English.

### Exclusion Criteria

The exclusion criteria were as follows:

1. Unpublished conference abstracts;
2. Studies with image segmentation only and without complete ML models for identifying muscle invasion, high-grade tumors, HER2, LN staging, and Ki-67 expression;
3. Studies that only performed texture analysis, with no ML classifiers constructed based on extracted textures;
4. Lack of metrics (eg, sensitivity or recall, area under the receiver operating characteristic curve [AUROC], specificity, calibration curves, accuracy, confusion matrix, precision, and  $F_1$ -score) for predicting the accuracy of ML models.

## Search Strategy and Data Source

The Cochrane Library (CENTRAL), Embase, PubMed, and Web of Science were retrieved up to June 17, 2024. This investigation did not perform simultaneous searches across multiple databases on a single platform. Searches were not extended to dedicated conference abstract databases or web sources beyond the primary databases. The authors of unpublished conference abstracts were not contacted to obtain full study details. To mitigate the risk of omitting relevant studies, the reference lists of eligible articles and pertinent review papers were manually scrutinized. No search filters were applied. The search strategy integrated both controlled vocabulary, such as Medical Subject Headings and Emtree terms, and free-text keywords to optimize sensitivity. Subject headings included “Urinary Bladder Neoplasms” and “machine learning.” Search strategies were tailored to the specific syntax of each database and combined using Boolean operators. No restrictions were applied concerning publication date or geographical location. Before the final data analysis, an updated search was performed in all specified databases on October 17, 2025 ([Multimedia Appendix 1](#)).

## Selection of Studies and Data Extraction

EndNote was used to import the retrieved articles. After eliminating duplicate records, the remaining articles were reviewed by titles and abstracts. Exclusions were categorized as follows: meta-analyses or reviews, replies or letters, case reports, animal experiments, registry or clinical trial protocols, non-English articles, and preprints. Initially relevant articles were screened by full text to determine eligible studies.

Before data extraction, a spreadsheet was created. The collected information encompassed DOI, title, publication

year, first author, country, study type, purpose of the task, patient source, image source, recording of a complete image acquisition protocol, number of researchers involved, whether preliminary experiments tested different imaging parameters, radiomic region of interest (ROI) segmentation software, total number of outcome events, whether test-retest experiments were performed, total number of cases, training set size, number of outcome events in the training set, validation set generation method, validation set size, number of outcome events in the validation set, variable selection method, model types used, overfitting assessment method, public availability of data and code, mean age, sex, specimen source, AUROC (95% CI), number of true negatives, number of true positives, specificity, sensitivity, precision, accuracy, and  $F_1$ -score.

Results from each dataset within a study were included only once. When multiple studies published by the same author over different years were suspected of having overlapping datasets, only the study with the larger sample size was included. When multiple models were present, the model demonstrating optimal performance in the validation set was selected for inclusion. If a single dataset contained multiple validation sets, all were incorporated into the analysis.

Two researchers (ZH and YL) screened the articles and extracted the data separately. Their results were cross-checked. Any discrepancies were addressed with the help of a third researcher (BX).

## Assessment of Risk of Bias and Quality

The Prediction Model Risk of Bias Assessment Tool for Artificial Intelligence (PROBAST-AI) provides a framework to critically appraise the risk of bias (ROB) and applicability of ML-based multivariable prediction models [23,24]. The assessment for ROB comprises four domains: participants and data sources, predictors, outcome, and analysis (1=strongly disagree, 2=somewhat disagree, 3=I don't know, 4=somewhat agree, 5=strongly agree). A domain was judged as high risk if any signaling question was rated “no/probably no,” and as low risk only if all questions were answered “yes/probably yes.” An “I don't know” response yielded an unclear risk designation.

Furthermore, the quality of the eligible investigations was appraised using the Radiomics Quality Score (RQS), which ranges from -8 to 36 [25]. This RQS primarily considers image protocol quality, multiple image segmentation methods, study of modalities, image acquisition time, feature reduction, and model construction using radiomic and nonradiomic features (prognosis and molecular subtyping). It also considers the detection and discussion of radiomic and biological correlations, cutoff value analysis, calibration statistics, discrimination statistics, validation, prospective studies registered in trial databases, comparison with the “gold standard,” cost-effectiveness analysis, potential clinical utility, and open science and data.

Two investigators independently evaluated the ROB and methodological quality and then performed a cross-check of

their assessments. Any discrepancies were adjudicated with the help of a third researcher (BX).

## Synthesis Methods

A meta-analysis was performed on AUROC, a metric for the accuracy of ML models. The analysis required the SE or 95% CI of the AUROC. However, some of the included studies lacked a 95% CI and SE. In this scenario, SE was estimated by referring to the study by Debray et al [26]. A random-effects model was used for the meta-analysis to account for potential heterogeneity among the primary studies due to variations in model parameter tuning and predictor selection. Furthermore, the Hartung-Knapp-Sidik-Jonkman method was applied [27].

Furthermore, a meta-analysis was performed using a bivariate mixed effects model to pool sensitivity and specificity. This analysis was based on diagnostic contingency tables. Nevertheless, these tables were not directly provided in some studies. Then, the required values were calculated via reported sensitivity, specificity, precision, and sample sizes. Subsequently, subgroup analyses were conducted based on the dataset (training and validation sets), image source, and model type. A meta-analysis of AUROC values was conducted using packages “*meta*” and “*metafor*” in R (v4.5.2; R Foundation for Statistical Computing). A bivariate mixed effects model was implemented using the “*midas*” package in Stata (v15.0; StataCorp LLC).

$$SE(c) \approx \sqrt{\frac{c(1-c) \left[ 1 + \frac{n^*(1-c)}{2-c} + \frac{m^*c}{1+c} \right]}{mn}}$$

Parameter definitions:  $c$  denotes the  $c$ -statistic;  $n$  represents the number of observed events; and  $m$  corresponds to the total sample size  $n^* = m^* = \frac{m+n}{2} - 1$ .

Finally, subgroup analyses were conducted within each task category based on case number, ML type (logistic regression [LR] vs other ML), and modeling variables (radiomics features alone vs radiomics features+clinical characteristics).

## Certainty of Evidence

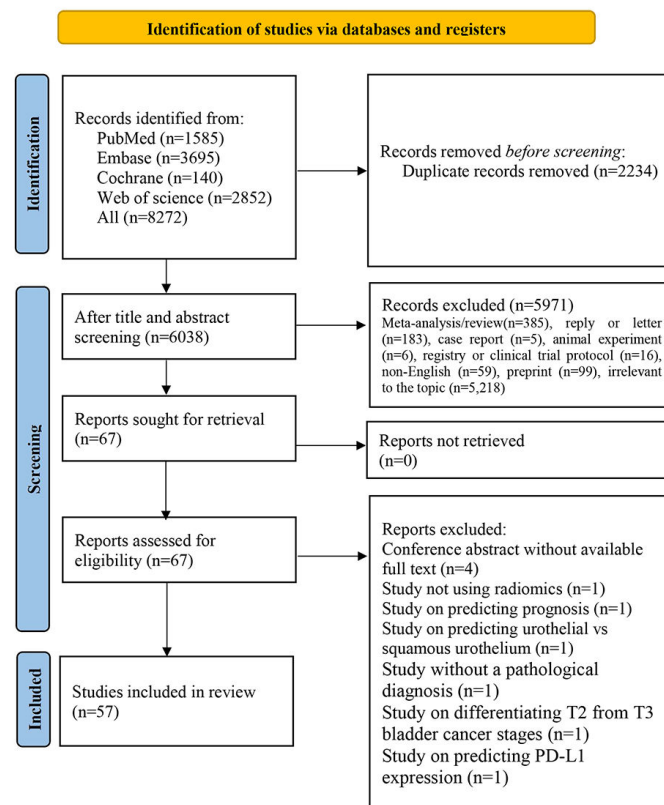
The certainty of evidence for the AUROC estimates derived from various subgroup analyses was evaluated using the Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) framework. Two investigators independently performed the evidence grading and a cross-verification of their assessments. Any discrepancies were adjudicated by a third investigator (BX).

## Results

### Study Selection

The database search yielded 8272 articles. After removing 2234 duplicates, 6038 articles remained and were reviewed based on their titles and abstracts. This process resulted in the exclusion of 5971 ineligible records. Overall, 67 articles remained and were reviewed by full texts, leading to the removal of 10 for irrelevance, including 4 non-peer-reviewed conference papers, one study lacking radiomics analysis, one focused on differentiating pure urothelial carcinoma from urothelial carcinoma with squamous differentiation, one without radiomics, one enrolling participants without a pathological diagnosis, one differentiating T2 from T3 bladder cancer stages, and one predicting programmed death-ligand 1 expression. Ultimately, the meta-analysis incorporated 57 articles [28-84] (Figure 1).

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) of the literature selection process.



### Characteristics of Studies

Fifty-seven [28-84] eligible studies, published between 2017 and 2025, encompassed data on 11,933 individuals with bladder cancer. Most were case-control studies, including 40 [28,33-36,38-41,43-46,48,49,51,53-57,61-76,79,82,84] single-center and 17 [29-32,37,42,47,50,52,58-60,77,78,80,81,83] multicenter studies. Thirty-four studies primarily investigated muscle invasion; 20 [30,32,33,41-43,46,51,52,54,57-60,63,64,69,71,78,83], 13 [28,29,31,34,45,48,50,62,75,79-82], and one [65] used MRI-, CT-, and ultrasound-based radiomics, respectively. Sixteen studies [36,38-40,44,47,49,53,56,61,66,67,70,72,73,84] focused on discriminating high-grade tumors. Three studies [55,74,77] examined HER2-positive expression, 2 [35,76] examined Ki-67 expression, and 2 [37,68] examined LN staging in bladder cancer. ITK-SNAP (University of Pennsylvania) was the most frequently used software for ROI segmentation. Other software platforms included MATLAB (R2012b; Matrix Laboratory; MathWorks), the open-source Medical Imaging Interaction Toolkit, and a computer-assisted visualization and analysis software system. Fifty-three studies explicitly described the generation of a validation; among these, 37 [28,34-37,39-41,43-46,48,49,51,53-57,61-66,68-72,74-76,79,82,84] used internal validation techniques such as random sampling or cross-validation and 16 [29-32,42,47,50,52,58-60,77,78,80,81,83] studies used external validation. A total of 10 distinct model types were evaluated (Table 1).

**Table 1.** Basic information.

Studies	Country	Study type	Source of the patient	Age (years)	Gender (n)	Task	Radiomics source	Specimens source	Number of image researcher	ROI <sup>a</sup> segmentation software	The total number of cases (N)	Number of cases in the training set (n)	Generation method of validation set	Number of cases in validation set	Model type
Zhang et al (2024) [28]	China	Case-control	Single center	<ul style="list-style-type: none"> <li>Total, mean (SD): 67.821 (10.468)</li> <li>Training set, mean (SD): 68.066 (11.184)</li> <li>Validation set, mean (SD): 67.254 (8.64)</li> </ul>	<ul style="list-style-type: none"> <li>Total: F: 36; M: 160;</li> <li>Training set: F: 20; M: 117;</li> <li>Validation set: F: 16; M: 43</li> </ul>	Muscle invasion status	CT <sup>d</sup>	Pathology (TURBT <sup>e</sup> or a radical surgical specimen) +CT	2	ITK-SNAP program (version 4.0.1; University of Pennsylvania)	196	137	Random sampling	59	LR <sup>f</sup> , LASSO <sup>g</sup>
Zhang et al (2021) [29]	China	Case-control	Multicenter	<ul style="list-style-type: none"> <li>Training set, median (IQR): 65 (56-72);</li> <li>Internal validation, median (IQR): 68 (61-74);</li> <li>External validation, median (IQR): 65 (59-77)</li> </ul>	<ul style="list-style-type: none"> <li>Training set: F: 75; M: 218;</li> <li>Internal validation: F: 13; M: 60;</li> <li>External validation: F: 13; M: 62</li> </ul>	Muscle invasion status	CT	Pathology (TURBT or a radical surgical specimen) +CT	2	Deepwise Research Platform (Deepwise Inc)	Center 1: 366 Center 2: 75	293 (Development group+adjustment Group)	External validation	73+75 (Internal+external)	DL <sup>h</sup>
Ye et al (2023) [30]	China	Case-control	Multicenter	<ul style="list-style-type: none"> <li>Training set, median (IQR): 67 (59-75);</li> <li>Internal validation, median (IQR): 69 (62-71);</li> <li>External validation, median (IQR): 64 (55-73)</li> </ul>	<ul style="list-style-type: none"> <li>Training set: F: 20; M: 109;</li> <li>Internal validation: F: 0; M: 30;</li> <li>External validation: F: 8; M: 47</li> </ul>	Muscle invasion status	MRI <sup>i</sup>	Pathology (TURBT or a radical surgical specimen) +MRI	1	ITK-SNAP	Center 1: 160; Center 2: 55	— <sup>j</sup>	External validation	Internal validation: 25 (31 tumor lesions; MIBC <sup>k</sup> : 4)	SVM <sup>l</sup>

Studies	Country	Study type	Source of the patient	Age (years)	Gender (n)	Task	Radiomics source	Specimens source	Number of image researcher	ROI <sup>a</sup> segmentation software	The total number of cases (N)	Number of cases in the training set (n)	Generation method of validation set	Number of cases in validation set	Model type
Ren et al (2023) [31]	China	Case-control	Multicenter	<ul style="list-style-type: none"> <li>• Training set, median (IQR): F: 57; M: 131; 70.5 (37-90)</li> <li>• Validation set, F: 35; M: 46</li> </ul>	<ul style="list-style-type: none"> <li>• Training set: F: 57; M: 131; 70.5 (37-90)</li> <li>• Validation set: F: 35; M: 46</li> </ul>	Muscle invasion status	CT	Pathology +within 30 days CT	1	ITK-SNAP	269	188	External validation	81	LR
Qureshi et al (2024) [32]	United States	Case-control	Multicenter	72 (36-88)	—	Muscle invasion status	MRI	Pathology +MRI	0	MATLAB (MATrix Laboratory; MathWorks)	—	—	External validation	—	NB <sup>m</sup> , SVM, DT <sup>n</sup> , KNN <sup>o</sup> , LR
Özdemir et al (2023) [33]	Türkiye	Case-control	Single center	—	—	Muscle invasion status	MRI	Pathology +MRI	1	—	60	—	—	—	LR
Chen et al (2022) [34]	China	Case-control	Single center	<ul style="list-style-type: none"> <li>• Training set, mean (SD): MIBC: F: 26; M: 6; NMIBC: 68.1250 (10.7065); NMIBC<sup>p</sup>: 66.7216 (12.9605)</li> <li>• Validation set: MIBC: F: 11; M: 0; NMIBC: 66.7216 (12.9605)</li> <li>• Validation set: MIBC: F: 26; M: 7</li> </ul>	<ul style="list-style-type: none"> <li>• Training set: MIBC: F: 26; M: 6; NMIBC: 68.1250 (10.7065); NMIBC<sup>p</sup>: 66.7216 (12.9605)</li> <li>• Validation set: MIBC: F: 11; M: 0; NMIBC: 66.7216 (12.9605)</li> <li>• Validation set: MIBC: F: 26; M: 7</li> </ul>	Muscle invasion status	CT	Pathology (TURBT or a radical surgical specimen within+4 weeks MRI)	2	ITK-SNAP	173	129	Random sampling	44	DL
Zheng et al (2021) [35]	China	Case-control	Single center	<ul style="list-style-type: none"> <li>• Training set, mean (SD): &lt;65 years: 42 (33.6); ≥65 years: 83 (66.4)</li> <li>• Validation set, mean (SD):</li> </ul>	<ul style="list-style-type: none"> <li>• Training set: F: 19; M: 106</li> <li>• Validation set: F: 13; M: 41</li> </ul>	Ki-67 expression	MRI	Pathology (TURBT or a radical surgical specimen) +MRI	2	ITK-SNAP	179	125	Random sampling	54	SMOTE-LASSO <sup>i</sup>

Studies	Country	Study type	Source of the patient	Age (years)	Gender (n)	Task	Radiomics source	Specimens source	Number of image researcher	ROI <sup>a</sup> segmentation software	The total number of cases (N)	Number of cases in the training set (n)	Generation method of validation set	Number of cases in validation set	Model type
Ye et al (2023) [36]	China	Cohort study	Single center	<65 years: 24 (44.4); ≥65 years: 30 (55.6) • Training set, mean (SD): F: 10; M: 54; 66.1 (11.2); • Validation set: F: 5; M: 23	—	Histological grade	MRI	Pathology +MRI	2	ITK-SNAP	92	64	Random sampling	28	LR
Starmans et al (2022) [37]	Netherlands	Case-control	Multicenter	66.1 (11.2)	—	Lymph node staging	CT	Pathology +CT	1	WORC toolbox (Erasmus MC)	209	—	Cross validation	—	DL
Sarkar et al (2023) [38]	United States	Case-control	Single center	—	—	Histological grade	CT	—	2	—	100	—	—	—	NB
Li et al (2023) [39]	China	Case-control	Single center	—	—	Histological grade	MRI	—	2	ITK-SNAP	169	118	Random sampling	51	LR
Deng et al (2023) [40]	China	Case-control	Single center	• Training set, mean (SD): low grade: F: 61.47 (11.34); high grade: F: 68.79 (9.87) • Validation set, mean (SD): F: 7; M: 25	—	Histological grade	CT	Pathology +CT	2	Darwin Research Platform (Yizhun Medical AI)	105	73	Random sampling	32	SVM, KNN, GBDT <sup>†</sup> , LR, RF <sup>‡</sup> , XGBoost <sup>†</sup>
Chen et al (2023) [41]	China	Case-control	Single center	72.09 (9.035)	—	Muscle invasion status	MRI, pathology	Pathology +MRI	2	Darwin Research Platform	445	312	Random sampling	133	SVM, KNN, Decision tree, RF, XGBoost, GBDT
Yu et al (2024) [42]	China	Case-control	Multicenter	—	—	Muscle invasion status	MRI, pathology	Pathology (TURBT or a radical surgical specimen) +MRI	2	—	436	404	External validation	32	DL

Studies	Country	Study type	Source of the patient	Age (years)	Gender (n)	Task	Radiomics source	Specimens source	Number of image researcher	ROI <sup>a</sup> segmentation software	The total number of cases (N)	Number of cases in the training set (n)	Generation method of validation set	Number of cases in validation set	Model type
Xu et al (2019) [43]	China	Case-control	Single center	<ul style="list-style-type: none"> <li>Total, mean (range): 66.1 (37-93);</li> <li>Training set: F: 27; M: 104</li> <li>Validation set: F: 22; M: 65</li> <li>Validation set, mean (range): 65.8 (38-86);</li> <li>Validation set, mean (range): 66.5 (37-93)</li> </ul>	<ul style="list-style-type: none"> <li>Total F: 49; M: 169</li> <li>Training set: F: 27; M: 104</li> <li>Validation set: F: 22; M: 65</li> </ul>	Muscle invasion status	MRI	Pathology (TURBT) or a radical surgical specimen +MRI	2	ITK-SNAP	218	136	Random sampling	87	RF
Zhou et al (2019) [44]	China	Case-control	Single center	<ul style="list-style-type: none"> <li>Mean (SD): 64.75 (4.74)</li> <li>Median (IQR): 68 (12)</li> </ul>	<ul style="list-style-type: none"> <li>F: 42; M: 66</li> <li>F: 52; M: 317</li> </ul>	Histological grade	MRI	Pathology (TURBT)	2	ITK-SNAP	108	72	Random sampling	36	LR
Yang et al (2021) [45]	China	Case-control	Single center	<ul style="list-style-type: none"> <li>Training set, n (%): &lt;65 years: 48 (37.2); ≥65 years: 81 (62.8);</li> <li>Validation set, n (%): &lt;65 years: 20 (35.7); ≥65 years: 36 (64.3)</li> </ul>	<ul style="list-style-type: none"> <li>Training set: F: 20; M: 109;</li> <li>Validation set: F: 12; M: 44</li> </ul>	Muscle invasion status	CT	Pathology (a radical surgical specimen) +CT	2	—	369	—	Random sampling	—	DL-CNN <sup>u</sup>
Zheng et al (2021) [46]	China	Case-control	Single center	<ul style="list-style-type: none"> <li>Training set, median (IQR): 66 (59-74)</li> <li>Validation set, median (IQR): 65 (58-72)</li> </ul>	<ul style="list-style-type: none"> <li>Training set: F: 469; M: 0</li> <li>Validation set: F: 131; M: 44</li> </ul>	Muscle invasion status	MRI	Pathology +MRI	2	ITK-SNAP	185	129	Random sampling	56	LASSO, RF, SVM
Song et al (2023) [47]	China	Case-control	Multicenter	<ul style="list-style-type: none"> <li>Training set, median (IQR): 66 (59-74)</li> <li>Validation set, median (IQR): 65 (58-72)</li> </ul>	<ul style="list-style-type: none"> <li>Training set: F: 469; M: 0</li> <li>Validation set: F: 131; M: 44</li> </ul>	Histological grade	CT	—	2	ITK-SNAP	688	469	External validation	219	LR, NB, SVM, KNN, RF, DT, XGBoost, LightGBM, GBDT, AdaBoost <sup>v</sup> , ANN <sup>w</sup> , DL
Cui et al (2022) [48]	China	Case-control	Single center	<ul style="list-style-type: none"> <li>Training set, mean (SD):</li> </ul>	<ul style="list-style-type: none"> <li>Training set: NMBIC; F:</li> </ul>	Muscle invasion status	CT	Pathology +CT	2	ITK (Insight Software Consortium)	327	120	Random sampling	68	LR

Studies	Country	Study type	Source of the patient	Age (years)	Gender (n)	Task	Radiomics source	Specimens source	Number of image researcher	ROI <sup>a</sup> segmentation software	The total number of cases (N)	Number of cases in the training set (n)	Generation method of validation set	Number of cases in validation set	Model type
				NMIBC: 66.2 (11.9); MIBC: 66.2 (11.7)	16; M: 44; MBIC: F: 8; M: 52										
				• Validation set, mean (SD): NMIBC: 65.4 (10.8); MIBC: 68.8 (9.3)	• Validation set, mean (SD): NMIBC: F: 7; M: 27; MBIC: F: 5; M: 29										
Zhang et al (2020) [49]	China	Case-control	Single center	• Training set, n (%): ≤60 years: 42 (38.9); >60 years: 66 (61.1)	• Training set: F: 28; M: 80 • Validation set: F: 8; M: 29	Histological grade	CT	Pathology +CT	2	Deepwise Research Platform (Deepwise Healthcare)	145	108	Random sampling	37	LR
				• Validation set, n (%): ≤60 years: 14 (37.8); >60 years: 23 (62.2)	• Training set: F: 17; M: 209 • Validation set: F: 11; M: 86	Muscle invasion status	CT	Pathology +CT	2	ITK-SNAP	375	226	Random sampling	internal validation: 97; external validation: 52	DL
Wei et al (2023) [50]	China	Case-control	Multicenter	• Training set, mean (SD): 65.86 (10.08)											
				• Validation set, mean (SD): 66.04 (9.56)	• Training set: F: 17; M: 113 • Validation set: F: 9; M: 60	Muscle invasion status	MRI	Pathology +MRI	2	3D Slicer version 4.9.0.	169	130	Grouping by different periods	69	LR
Zheng et al (2019) [51]	China	Case-control	Single center	• Training set, median (IQR): 64 (57-69)											
				• Validation set, median (IQR): 61 (54-70)	• Training set: F: 8; M: 56 • Validation set: 64.8 (10.6)	Muscle invasion status	MRI	Pathology +MRI	2	MATLAB 2016 (MathWorks)	106	64	External validation	42	LR
Wang et al (2019) [52]	China	Case-control	Multicenter	• Training set, mean (SD): 64.8 (10.6)											

Studies	Country	Study type	Source of the patient	Age (years)	Gender (n)	Task	Radiomics source	Specimens source	Number of image researcher	ROI <sup>a</sup> segmentation software	The total number of cases (N)	Number of cases in the training set (n)	Generation method of validation set	Number of cases in validation set	Model type
Zheng et al (2021) [53]	China	Case-control	Single center	<ul style="list-style-type: none"> <li>Validation set, mean (SD): 62.9 (11.0)</li> <li>Training set, n (%): &lt;65 years: 48 (37.2); ≥65 years: 81 (62.8)</li> </ul>	<ul style="list-style-type: none"> <li>Validation set: F: 7; M: 35</li> <li>Training set: F: 21; M: 108</li> <li>Validation set: F: 11; M: 45</li> </ul>	Histological grade	MRI	Pathology +MRI	2	ITK-SNAP	298	206	Random sampling	88	LASSO, RF, SVM
Wang et al (2022) [54]	China	Case-control	Single center	<ul style="list-style-type: none"> <li>Validation set, n (%): &lt;65 years: 22 (39.3); ≥65 years: 34 (60.7)</li> </ul>	<ul style="list-style-type: none"> <li>Total: F: 45; M: 146;</li> <li>Training set: F: 22; M: 63;</li> <li>Validation set: F: 8; M: 28;</li> <li>Test set: F: 15; M: 55</li> </ul>	Muscle invasion status	MRI	Pathology +MRI	2	ITK-SNAP	191	85	Random sampling	validation: 36; test: 70	LR
Yu et al (2023) [55]	China	Case-control	Single center	<ul style="list-style-type: none"> <li>Test set, median (IQR): 69 (62-76);</li> <li>Validation set, median (IQR): 70 (62-77);</li> <li>Training set (n): &gt;65 years: 99; ≤65: 57;</li> <li>Validation set (n): &gt;65 years: 19; ≤65 years: 20;</li> </ul>	<ul style="list-style-type: none"> <li>training set: F: 19; M: 137</li> <li>validation set: F: 9; M: 30;</li> <li>test set: F: 6; M: 37</li> </ul>	HER 2 status	MRI	Pathology +MRI	2	ITK-SNAP	195	156	Random sampling	39	SVM, RF, LR, NB, KNN, AdaBoost

Studies	Country	Study type	Source of the patient	Age (years)	Gender (n)	Task	Radiomics source	Specimens source	Number of image researcher	ROI <sup>a</sup> segmentation software	The total number of cases (N)	Number of cases in the training set (n)	Generation method of validation set	Number of cases in validation set	Model type
Wang et al (2023) [56]	China	Case-control	Single center	<ul style="list-style-type: none"> <li>• Test set (n): &gt;65 years: 25; ≤65 years: 28</li> <li>• Total, median (IQR): 70 (62-77)</li> <li>• Training set, median (IQR): 70 (64-77);</li> <li>• Validation set, median (IQR): 69 (61-79)</li> </ul>	<ul style="list-style-type: none"> <li>• Total: F: 55; M: 172</li> <li>• Training set: F: 35; M: 96</li> <li>• Validation set: F: 20; M: 76</li> </ul>	Histological grade	MRI	Pathology +MRI	2	ITK-SNAP	227	131	Random sampling	96	LR
Zhang et al (2022) [57]	China	Case-control	Single center	<ul style="list-style-type: none"> <li>• Training set, mean (SD): 66 (11)</li> <li>• Validation set, mean (SD): 67 (10)</li> <li>• Test set, mean (SD): 66 (11)</li> </ul>	<ul style="list-style-type: none"> <li>• Training set: F: 24; M: 215</li> <li>• Validation set: F: 9; M: 59;</li> <li>• Test set: F: 5; M: 35</li> </ul>	Muscle invasion status	MRI	Pathology +MRI	2	ITK-SNAP	342	239	Random sampling	Validation: 68; test: 35	LR
Zou et al (2022) [58]	China	Cohort study	Multicenter	<ul style="list-style-type: none"> <li>• Training set, median (range): 66 (26-95)</li> <li>• Validation set, median (range): 68 (11-91)</li> <li>• Retrospective, median (range): 64.5 (32-89)</li> <li>• Prospective, median</li> </ul>	<ul style="list-style-type: none"> <li>• Training set: F: 37; M: 253</li> <li>• Validation set: F: 6; M: 60</li> <li>• Retrospective: F: 5; M: 29</li> <li>• Prospective: F: 7; M: 32</li> <li>• Multicenter: F: 7; M: 32</li> </ul>	Muscle invasion status	MRI	Pathology +MRI	2	MBMIP model	468	290	Random sampling	178	DL

Studies	Country	Study type	Source of the patient	Age (years)	Gender (n)	Task	Radiomics source	Specimens source	Number of image researcher	ROI <sup>a</sup> segmentation software	The total number of cases (N)	Number of cases in the training set (n)	Generation method of validation set	Number of cases in validation set	Model type
Zhou et al (2022) [59]	China	Case-control	Multicenter	(range): 65 (47-91) • Multicenter, median (range): 68 (42-89)	—	Muscle invasion status	MRI	—	—	EvidentialNet	—	—	Random sampling	—	DL
Li et al (2023) [60]	China	Case-control	Multicenter	• Training set, median (IQR): 67 (57-70) • Test set, median (IQR): 63 (55-70)	• Training set, F: 13; M: 80 • Test set: F: 1; M: 27	Muscle invasion status	MRI	Pathology +MRI	1	ITK-SNAP	121	93	External validation	28	DL
Li et al (2024) [61]	China	Case-control	Single center	• Training set, median (range): 66 (27-90) • Validation set: 67 (27-90)	• Training set, F: 33; M: 145; • Validation set: F: 14; M: 63	Histological grade	MRI	Pathology +MRI	2	ITK-SNAP	255	178	Random sampling	77	LR
Chen et al (2022) [62]	China	Case-control	Single center	• Training set, mean (SD): 69.85 (11.57) • Validation set, mean (SD): 68.41 (11.17)	• Training set, F: 36; M: 81 • Validation set: F: 10; M: 41	Muscle invasion status	CT	Pathology +CT	4	ITK-SNAP	168	117	Random sampling	51	NB
Xu et al (2017) [63]	China	Case-control	Single center	—	—	Muscle invasion status	MRI	Pathology +MRI	2	MATLAB R2012b	68	—	Cross validation	—	SVM
Xu et al (2018) [64]	China	Case-control	Single center	—	—	Muscle invasion status	MRI	Pathology +MRI	2	MATLAB R2015b	54	—	Cross validation	—	SVM
Gao et al (2021) [65]	China	Case-control	Single center	—	—	Muscle invasion status	Ultrasound	Pathology +ultrasound	2	ITK-SNAP	157	110	Random sampling	47	NB

Studies	Country	Study type	Source of the patient	Age (years)	Gender (n)	Task	Radiomics source	Specimens source	Number of image researcher	ROI <sup>a</sup> segmentation software	The total number of cases (N)	Number of cases in the training set (n)	Generation method of validation set	Number of cases in validation set	Model type
Wang et al (2019) [66]	China	Case-control	Single center	<ul style="list-style-type: none"> <li>Total, mean (SD): 63.4 (10.4)</li> <li>Training set, mean (SD): 62.6 (11.0)</li> <li>Validation set, mean (SD): 65.2 (8.1)</li> </ul>	<ul style="list-style-type: none"> <li>Total: F: 14; M: 86</li> <li>Training set: F: 11; M: 59</li> <li>Validation set: F: 3; M: 27</li> </ul>	Histological grade	MRI	Pathology +MRI	2	MATLAB R2012b	100	70	The time of undergoing surgery varies.	30	LR
Zhang et al (2017) [67]	China	Case-control	Single center	—	—	Histological grade	MRI	Pathology +MRI	2	—	—	61	—	—	SVM
Gresser et al (2022) [68]	Germany	Case-control	Single center	<ul style="list-style-type: none"> <li>Training set, mean (SD): PNI-2: 68 (11); PNO: 68 (10)</li> <li>Validation set, mean (SD): PNI-2: 70 (9); PNO: 69 (11)</li> </ul>	<ul style="list-style-type: none"> <li>Training set: PNI-2: F: 24, M: 42; PNO: F: 57; M: 151</li> <li>Validation set: F: 9, M: 19; PNO: F: 16; M: 73</li> </ul>	Lymph node staging	CT	Pathology (a radical surgical specimen) +MRI	2	Medical Imaging Interaction Toolkit (MITK, DKFZ, Heidelberg; version 2018.04.2)	391	274	Random sampling	117	ANN
Tong et al (2018) [69]	United States	Case-control	Single center	<ul style="list-style-type: none"> <li>Mean (SD): 65.6 (10.5)</li> </ul>	<ul style="list-style-type: none"> <li>F: 18; M: 47</li> </ul>	Muscle invasion status	MRI	Pathology +MRI	2	—	65	—	Leave-one-out method	—	SVM
Tao et al (2024) [70]	China	Case-control	Single center	<ul style="list-style-type: none"> <li>Median (IQR): 65 (30-86)</li> </ul>	<ul style="list-style-type: none"> <li>F: 33; M: 105</li> </ul>	Histological grade	MRI	Pathology +MRI	—	3D Slicer (version 4.11.20210226)	138	110	Random sampling	28	DL
Liu et al (2022) [71]	China	Case-control	Single center	<ul style="list-style-type: none"> <li>Training set: 62.78 (11.65);</li> <li>Validation set: 64.41 (10.10)</li> </ul>	<ul style="list-style-type: none"> <li>Training set: F: 19; M: 146</li> <li>Validation set: F: 7; M: 34</li> </ul>	Muscle invasion status	MRI	Pathology +MRI	2	Manually depict	206	165	Cross validation	28	LASSO
Liu et al (2022) [72]	China	Case-control	Single center	—	—	Histological grade	CT	Pathology +CT	2	MSRN network (Multi-Scale Residual Network)	75	51	Random sampling	24	DL

Studies	Country	Study type	Source of the patient	Age (years)	Gender (n)	Task	Radiomics source	Specimens source	Number of image researcher	ROI <sup>a</sup> segmentation software	The total number of cases (N)	Number of cases in the training set (n)	Generation method of validation set	Number of cases in validation set	Model type
Razik et al (2021) [73]	India	Case-control	Single center	<ul style="list-style-type: none"> <li>Mean (SD): 57.6 (11.8)</li> </ul>	<ul style="list-style-type: none"> <li>F: 5; M: 35</li> </ul>	Histological grade	MRI	Pathology +MRI	2	TexRAD (Feedback Medical)	40	—	—	—	LR
Peng et al (2024) [74]	China	Case-control	Single center	<ul style="list-style-type: none"> <li>Training set, mean (SD): HER2<sup>x</sup> positive: 64.64 (11.56); HER2 negative: 64.24 (12.78)</li> <li>Validation set, mean (SD): HER2 positive: 59.71 (13.60); HER2 negative: 64.50 (7.89)</li> </ul>	—	HER2 status	CT	Pathology +CT	2	ITK-SNAP	124	100	Random sampling	24	SVM, RF, LR, MLP <sup>y</sup> , ExtraTrees, LightGBM <sup>z</sup> , XGBoost, KNN
Xiong et al (2024) [75]	China	Case-control	Single center	<ul style="list-style-type: none"> <li>Training set, median (IQR): 69.00 (63.00–75.00)</li> <li>Validation set, median (IQR): 68.50 (57.25–76.00)</li> </ul>	<ul style="list-style-type: none"> <li>Training set: F: 8; M: 65</li> <li>Validation set: F: 8; M: 24</li> </ul>	Muscle invasion status	CT	Pathology +CT	2	Darwin Research Platform	105	73	Random sampling	32	GBDT, K-NN, LR, RF, SVM, XGBoost
Feng et al (2024) [76]	China	Case-control	Single center	<ul style="list-style-type: none"> <li>Mean (SD): 66.16 (12.67)</li> </ul>	<ul style="list-style-type: none"> <li>F: 27; M: 108</li> </ul>	Ki-67 expression	CT	Pathology +CT	2	3D Slicer (version: 4.10.2)	135	94	Random sampling	41	LR
Wei et al (2024) [77]	China	Case-control	Multicenter	<ul style="list-style-type: none"> <li>Total, mean (SD): 67.52 (10.32);</li> </ul>	<ul style="list-style-type: none"> <li>Total: F: 22; M: 185</li> <li>Training set: F: 16; M: 138</li> </ul>	HER2 status	CT	Pathology +CT	2	ITK-SNAP	207	154	External validation	53	LR, SVM, KNN, RF, XGBoost

Studies	Country	Study type	Source of the patient	Age (years)	Gender (n)	Task	Radiomics source	Specimens source	Number of image researcher	ROI <sup>a</sup> segmentation software	The total number of cases (N)	Number of cases in the training set (n)	Generation method of validation set	Number of cases in validation set	Model type
Cai et al (2025) [78]	China	Case-control	Multicenter	<ul style="list-style-type: none"> <li>• Training set, mean (SD): 67.41 (9.80)</li> <li>• Validation set, mean (SD): 67.85 (11.80)</li> </ul>	<ul style="list-style-type: none"> <li>• Validation set: F: 6; M: 47</li> <li>• Training set: F: 37; M: 254</li> <li>• Validation set: F: 6; M: 60</li> <li>• Internal test set: F: 25; M: 139</li> <li>• External test set: F: 6; M: 32</li> </ul>	Muscle invasion status	MRI	Pathology +MRI	2	DL	559	291	External validation	Validation set: 66; internal test set: 164; external test set: 38	DL, VLRADS
Du et al (2025) [79]	China	Case-control	Single center	<ul style="list-style-type: none"> <li>• Training set, mean (SD): 66.47 (9.22); MBIC: 66.47</li> <li>• Validation set, mean (SD): 66.47 (9.22); MBIC: 67.27 (10.75); MBIC: 68.62 (8.45)</li> </ul>	<ul style="list-style-type: none"> <li>• Training set: NMBIC: F: 27; M: 70; MBIC: F: 12; M: 31</li> <li>• Validation set: NMBIC: F: 10; M: 34; MBIC: F: 6; M: 10</li> </ul>	Muscle invasion status	CT	Pathology +CT	2	ITK-SNAP (v.3.8.0)	200	140	Random sampling	60	DL
Du et al (2025) [80]	China	Case-control	Multicenter	<ul style="list-style-type: none"> <li>• Training set: total: 69.17 (9.92); NMBIC: 69.12 (9.51); MBIC: 69.42 (11.95)</li> </ul>	<ul style="list-style-type: none"> <li>• Training set: total: F: 51; M: 180; NMBIC: F: 47; M: 146; MBIC: F: 4; M: 34</li> </ul>	Muscle invasion status	CT	Pathology +CT	2	ITK-SNAP	325	231	External validation	94	Extra tree

Studies	Country	Study type	Source of the patient	Age (years)	Gender (n)	Task	Radiomics source	Specimens source	Number of image researcher	ROI <sup>a</sup> segmentation software	The total number of cases (N)	Number of cases in the training set (n)	Generation method of validation set	Number of cases in validation set	Model type
He et al (2025) [81]	China	Case-control	Multicenter	<ul style="list-style-type: none"> <li>Validation set: total: F: 68.03 (12.03); NMBIC: 67.90 (12.35); MBIC: 68.69 (10.66)</li> <li>Training set, (n): ≤60 years: 48; &gt;60 years: 70</li> <li>Test set (n): ≤60 years: 11; &gt;60 years: 19</li> <li>External test set: ≤60 years: 24; &gt;60 years: 39</li> </ul>	<ul style="list-style-type: none"> <li>Validation set: total: F: 22; M: 72; NMBIC: F: 20; M: 58; MBIC: F: 2; M: 14</li> <li>Training set: F: 29; M: 89</li> <li>Test set: F: 8; M: 22</li> <li>External test set: F: 21; M: 42</li> </ul>	Muscle invasion status	CT	Pathology +CT	1	Manual	211	118	External validation	Test set: 30; external validation: 63	LR, SVC (linear), SVC (poly), SVC (rbf)
Hu et al (2025) [82]	China	Case-control	Single center	<ul style="list-style-type: none"> <li>Training set, mean (SD): NMBIC: 65.92 (9.83); MBIC: 68.55 (9.18)</li> <li>Validation set, mean (SD): NMBIC: 67.40 (8.37); MBIC: 70.53 (7.67)</li> </ul>	<ul style="list-style-type: none"> <li>Training set: NMBIC: F: 16; M: 83; MBIC: F: 5; M: 37</li> <li>Validation set: NMBIC: F: 4; M: 38; MBIC: F: 7; M: 12</li> </ul>	Muscle invasion status	CT	Pathology +CT	2	ITK-SNAP (version 3.8.0)	202	141	Random sampling	61	DECT <sup>aa</sup> , Radiomics, Nomogram
Yu et al (2025) [83]	China	Case-control	Multicenter	—	—	Muscle invasion status	MRI	Pathology +MRI	4	Manual	401	313	External validation	Validation set: 26; internal test set: 34; external test set: 28	DL
Zhou et al (2025) [84]	China	Case-control	Single center	<ul style="list-style-type: none"> <li>L<sup>ab</sup>GUC<sup>ab</sup>, mean (SD): 61.98 (12.29);</li> </ul>	<ul style="list-style-type: none"> <li>L: F: 41; M: 141</li> <li>H: F: 30; M: 160</li> </ul>	Histological grade	Histological CT	Pathology +CT	2	3D Slicer image	372	259	Random sampling	LR, DT, SVM, AdaBoost	



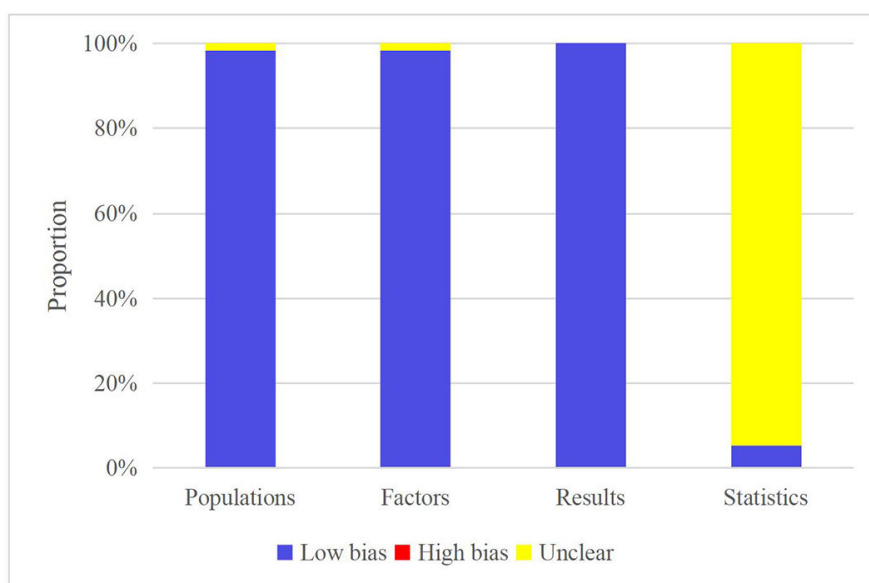
## Evaluation of ROB and Quality

### PROBAST-AI

Regarding the model development, all studies exhibited low ROB across all signaling questions in the “Participants” domain, with the sole exception of the question “Did the in- and exclusions of study participants result in a representative data set?” for which all studies were rated as “unclear.” For “Predictor” assessments, all included investigations indicated an uncertain risk pertaining to the question “Were predictor assessments made without knowledge of outcome data?” In the “Outcomes” domain, all studies demonstrated low ROB. Within the “Analysis” segment, uncertainty was

noted in 40 [29,30,32-34,36-38,40-45,47-50,55-57,59,60,62-65,67,69-73,77-79,81-84] investigations concerning “Was there evidence that the sample size was reasonable?” Two publications [32,35] also flagged uncertainty regarding “Were participants with missing or censored data handled appropriately in the analysis?” Furthermore, 21 [29,30,32,34,35,39,45,53,57-60,67,68,71-73,79-81,83] studies presented uncertainty on the item “If methods to address class imbalance were used, was the model or the model predictions recalibrated?” Forty-one [28-32,34-43,45-55,58-62,65-70,72,73,78,79] investigations indicated uncertainty for “Were methods used to address potential model overfitting?” (Figure 2).

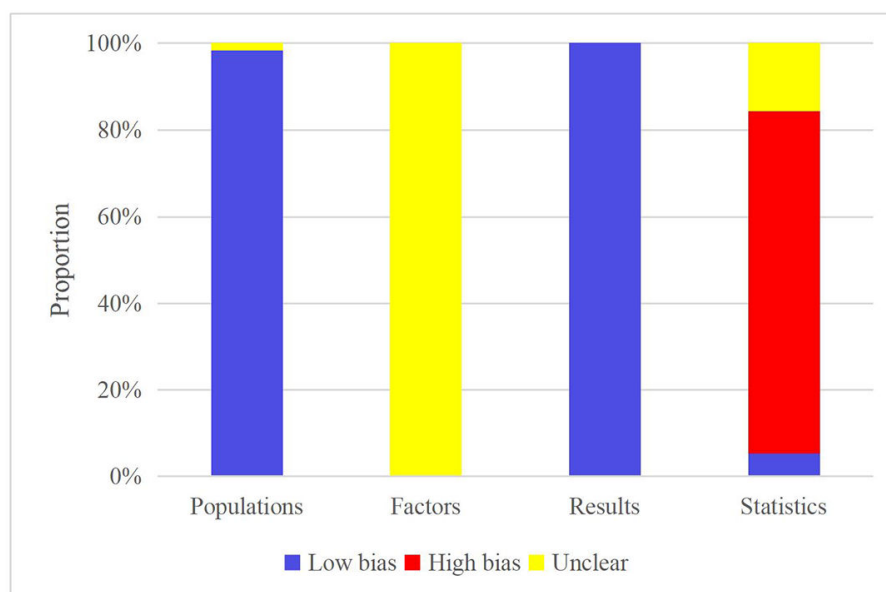
**Figure 2.** Risk of bias assessment results for model development.



During model evaluation, all but one [33] study demonstrated low ROB in the “Participants” domain. The exception was an unclear rating for the question “Did the in- and exclusions of study participants result in a representative data set?” For the “Predictors” domain, all studies were uniformly rated as unclear regarding “Were predictor assessments made without knowledge of outcome data?” The “Outcome” domain was consistently judged as low ROB across all studies. In the “Analysis” domain, 4 [33,38,67,73] studies were identified with a high ROB for “Was model evaluation based on only apparent performance avoided?” precluding further analysis. Forty-one [28,30-32,34-36,39,40,42-46,48-53,55,57,59-66,69,71,72,74-77,79,80,82,83] publications indicated a high ROB regarding “Was there evidence that the sample size

was reasonable?” Additionally, 37 [28-32,34-36,39-43,45-55,58-62,65,66,68-70,72,78,79] studies reported an unclear risk for “If resampling methods were used to evaluate model performance, were all model development steps replicated in the resampling process?” (Figure 3).

In summary, while the model development generally demonstrated a low ROB across participants and data sources, predictors, outcome, and analysis, the model evaluation phase presented a higher ROB specifically within the analysis dimension, particularly concerning the criteria “Was model evaluation based on only apparent performance avoided?” and “Was there evidence that the sample size was reasonable?”

**Figure 3.** Risk of bias assessment results for model evaluation.

## RQS

Two [32,38] studies did not provide complete imaging protocols. Multiple image segmentation was not performed in 7 [30-33,37,60,81] studies. All 57 studies failed to address scanner variability and vendor-dependent features, as well as imaging across multiple timepoints (ie, collecting individual images at different times). Three [32,37,69] studies did not use multiple testing correction or feature reduction to mitigate overfitting risk, nor did they use different data reduction methods (eg, principal component analysis and least absolute shrinkage and selection operator) for cross-analysis to reduce overfitting. Forty-two [29-31,33-38,40-50,55,56,58,60,62-65,67,69-73,75,78-84] studies did not perform multivariate analysis incorporating nonradiomic features. In all 57 [28-84] studies, biological correlations were neither detected nor discussed. The demonstration of phenotypic differences (potentially related to underlying gene-protein expression patterns) enhances the understanding of biology and radiomics. A cutoff value for disease presence or prediction risk was not provided in 51 [29,30,32-34,36-38,40-51,53-61,63-84] studies. Discrimination statistics were absent in 2 [32,69] studies. The reporting of these statistics (eg, ROC curve, C-statistic, and area under the curve [AUC]) and related statistical significance (eg, *P* values and CIs), or the application of resampling methods (eg, cross-validation and bootstrapping) was not performed. Calibration statistics were not reported in 47 [29-38,42-45,48-50,52-78,80,82,83] studies. Neither calibration statistics (eg, calibration plots, calibration in the large or slope) nor related statistical significance (eg, *P* values and CIs), or the application of resampling methods (eg, cross-validation and bootstrapping) were reported. Fifty-six [28-36,38-84] studies were not prospective studies registered in trial databases. Validation was lacking in 4 [33,38,67,73] studies. Comparisons with a “gold standard”—assessing the model’s agreement with the “gold standard” methods—were not performed in all 57 studies. Potential clinical utility—reporting potential and the applications of the model in a clinical setting (eg,

decision curve analysis)—was not addressed in 37 [30,32,33,36-39,42-45,48,49,54-73,78,80,82,83] studies. Cost-effectiveness analysis—reporting the cost-effectiveness of clinical application (eg, clinical impact curve)—was not performed in 56 [28-38,40-84] studies. Only 3 [28,32,35] studies did not adhere to open science and data principles by failing to release code and data. Regarding the RQS, a penalty of 3 points was applied for the absence of measures to control overfitting, and a deduction of 5 points was incurred for the lack of a validation set. Conversely, a bonus of 7 points was awarded to studies that were prospectively registered in a trial database. Among the 57 eligible investigations, 4 lacked any validation set. External validation was implemented in 16 [29-32,42,47,50,52,58-60,77,78,80,81,83] studies. Prospective registration was documented in only a single [37] study. The average quality score was 11.60 (range 1-17, IQR 3.5), representing an average percentage of 32.21% (Table S1 in [Multimedia Appendix 2](#)).

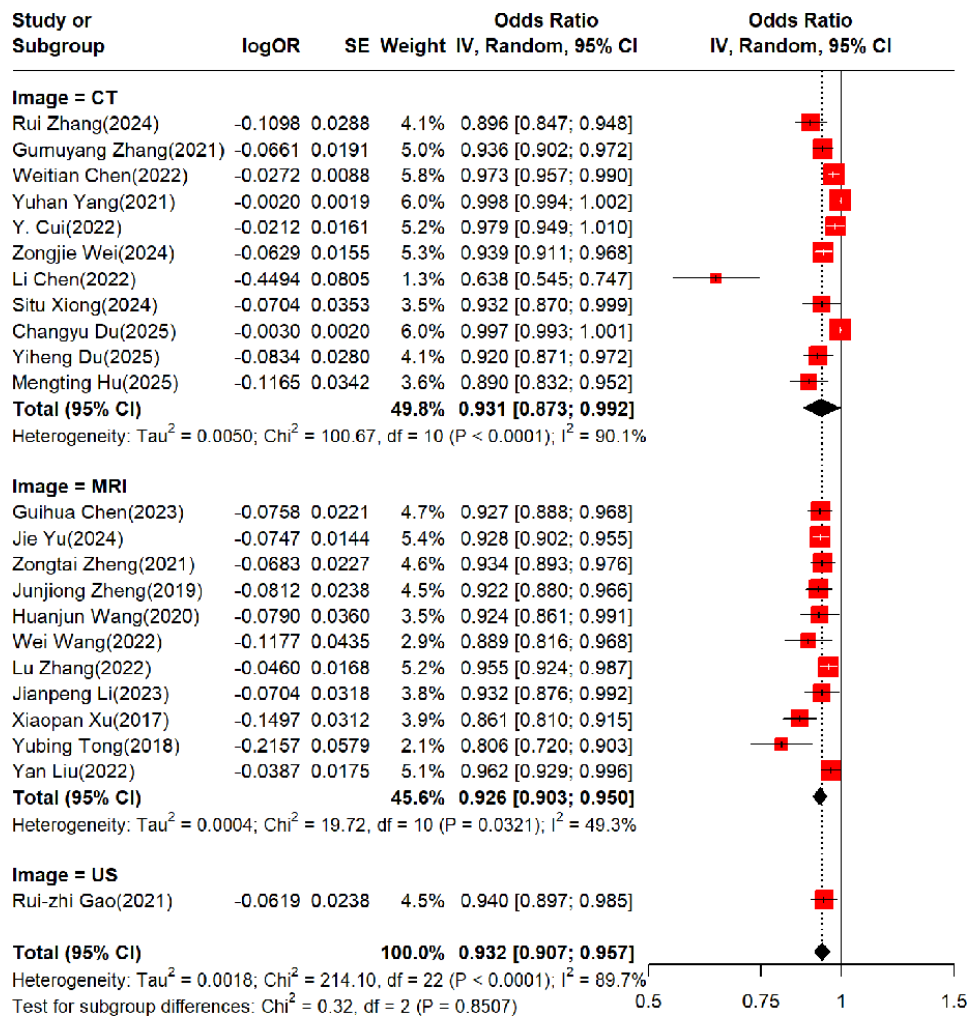
## Meta-Analysis

### Meta-Analysis for AUROC for Identifying Muscle Invasion

The meta-analysis of AUROC was conducted using a random-effects model. In the training sets, the pooled AUROC for radiomics-based ML was 0.932 (95% CI 0.907-0.957; grade: moderate [Figure 4](#) [28,29,34,41,42,45,46,48,50-52,54,57,60,62,63,65,69,71,75,79,80,82]).

The pooled AUROC was 0.932 (95% CI 0.853-1.000; grade: weak), 0.907 (95% CI 0.863-0.954; grade: weak), and 0.940 (95% CI 0.897-0.985; grade: weak) for CT-, MRI-, and ultrasound-based radiomics, respectively ([Table 2](#); [Figures S1 and S2 in Multimedia Appendix 3](#)). The pooled AUROC was 0.914 (95% CI 0.871-0.959; grade: weak) and 0.934 (95% CI 0.908-0.960; grade: moderate) for models integrating clinical features with CT- or MRI-based radiomics ([Table 2](#); [Figures S3 and S4 in Multimedia Appendix 3](#)).

**Figure 4.** Summary plot of radiomics for detecting muscle invasion in the training set [28,29,34,41,42,45,46,48,50-52,54,57,60,62,63,65,69,71,75,79,80,82]. For the 3 studies that included both radiomics-based and combined radiomics-clinical models, only the best-performing model in the validation set was retained to ensure independence across studies.



**Table 2.** Meta-analysis results for the identification of area under the receiver operating characteristic curve (AUROC) in muscle invasion.

Subgroup	Training set				Validation set							
	Number of models	AUROC <sup>a</sup> (95% CI)	tau <sup>2</sup>	tau	I <sup>2</sup>	Grade <sup>b</sup>	Number of models	AUROC (95% CI)	tau <sup>2</sup>	tau	I <sup>2</sup>	Grade
<b>Radiomics</b>												
<b>CT<sup>c</sup></b>												
Extra trees							1	0.825 (0.712-0.956)				⊕⊕ ⊕⊕
ANN <sup>d</sup>							1	0.950 (0.896-1.000)				⊕⊕ ⊕⊕
DL <sup>e</sup>	6	0.970 (0.940-1.000)	0.0007	0.0269	86	⊕⊕⊕⊕	5	0.903 (0.826-0.988)	0.0047	0.0687	88.1	⊕⊕ ⊕⊕
LR <sup>f</sup>	2	0.938 (0.513-1.000)	0.0038	0.0619	84.3	⊕⊕⊕⊕	2	0.891 (0.842-0.942)	0	0	0	⊕⊕ ⊕⊕
NB <sup>g</sup>	1	0.638 (0.545-0.747)				⊕⊕⊕⊕	1	0.665 (0.522-0.847)	--			⊕ ⊕⊕⊕
Overall	9	0.932 (0.853-1.000)	0.0083	0.0914	89.8	⊕⊕⊕⊕	10	0.893 (0.840-0.948)	0.004	0.0634	85.4	⊕⊕ ⊕⊕
<b>MRI<sup>h</sup></b>												
DL	2	0.929 (0.910-0.948)	0	0	0	⊕⊕⊕⊕	5	0.908 (0.892-0.924)	0	0	0	⊕⊕ ⊕⊕
LASSO <sup>i</sup>	2	0.951 (0.793-1.000)	<0.0001	0.005	5.8	⊕⊕⊕⊕	2	0.906 (0.900-0.913)	0	0	0	⊕⊕⊕ ⊕
LR	1	0.933 (0.895-0.972)				⊕⊕⊕⊕	1	0.931 (0.847-1.000)				⊕⊕ ⊕⊕
RF <sup>j</sup>							1	0.907 (0.856-0.961)				⊕⊕ ⊕⊕
SVM <sup>k</sup>	3	0.840 (0.772-0.914)	0	0	0	⊕⊕⊕⊕	2	0.913 (0.299-1.000)	0.0132	0.1151	85.4	⊕ ⊕⊕⊕
Overall	8	0.907 (0.863-0.954)	0.0022	0.047	71	⊕⊕⊕⊕	11	0.916 (0.891-0.942)	0.0001	0.01	0	⊕⊕ ⊕⊕
Ultrasonography	1	0.940 (0.897--0.985)				⊕⊕⊕⊕	1	0.840 (0.737-0.958)				⊕⊕ ⊕⊕
<b>Radiomics+clinical features</b>												
CT	3	0.914 (0.871-0.959)	0	0	0	⊕⊕⊕⊕	4	0.874 (0.852-0.896)	0	0	0	⊕⊕ ⊕⊕

Subgroup	Training set				Validation set							
	Number of models	AUROC <sup>a</sup> (95% CI)	tau <sup>2</sup>	tau	<i>I</i> <sup>2</sup>	Grade <sup>b</sup>	Number of models	AUROC (95% CI)	tau <sup>2</sup>	tau	<i>I</i> <sup>2</sup>	Grade
MRI	5	0.934 (0.908-0.960)	<0.0001	0.0077	0	⊕⊕⊕⊕	5	0.921 (0.867-0.979)	0.0005	0.0228	19.8	⊕⊕ ⊕⊕

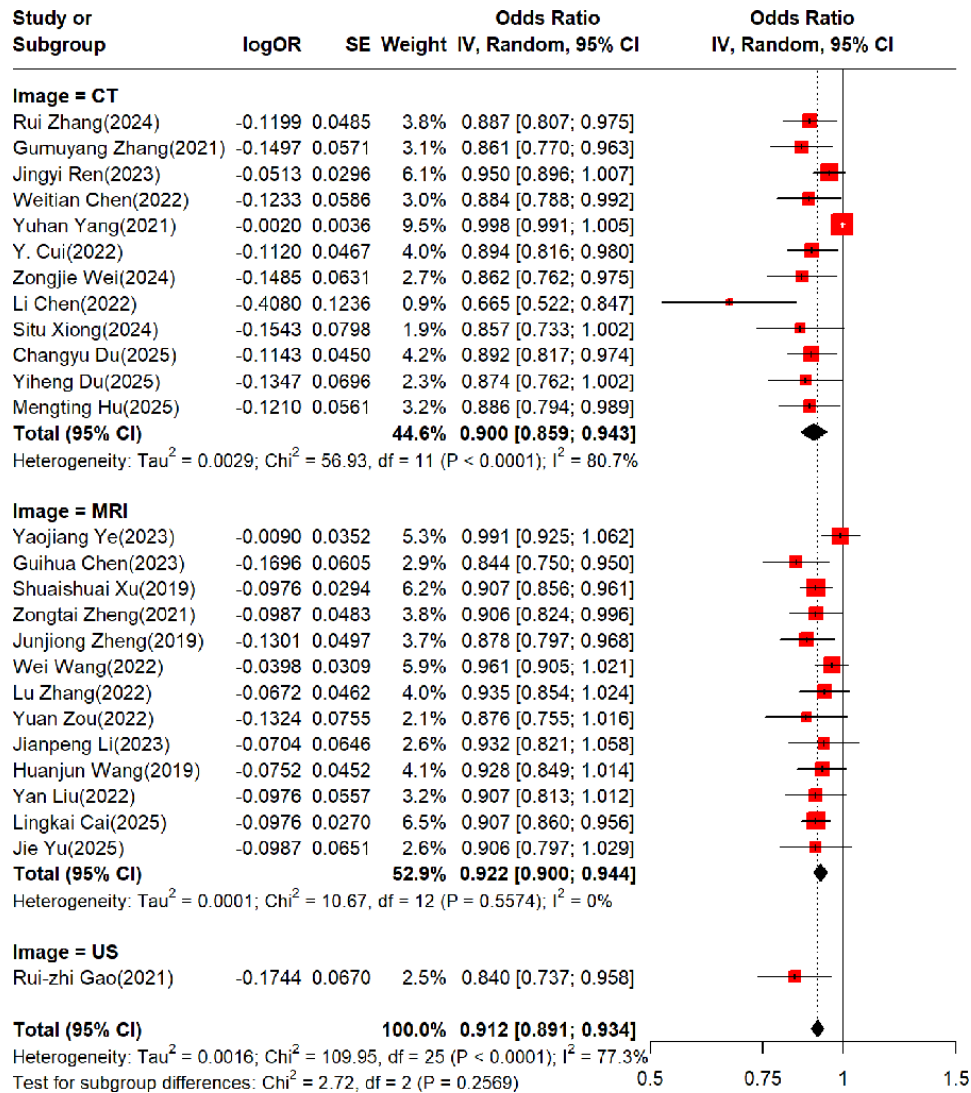
<sup>a</sup>AUROC: area under the receiver operating characteristic curve.  
<sup>b</sup>In the GRADE assessment section, ⊕⊕⊕⊕ represents high quality of evidence, ⊕⊕⊕ represents moderate quality of evidence, ⊕⊕ represents low quality of evidence, and ⊕⊕⊕⊕ represents very low quality of evidence.  
<sup>c</sup>CT: computed tomography.  
<sup>d</sup>ANN: artificial neural network.  
<sup>e</sup>DL: deep learning.  
<sup>f</sup>LR: logistic regression.  
<sup>g</sup>NB: naive Bayes.  
<sup>h</sup>MRI: magnetic resonance imaging.  
<sup>i</sup>LASSO: least absolute shrinkage and selection operator.  
<sup>j</sup>RF: random forest.  
<sup>k</sup>SVM: support vector machine.

In the validation sets, the pooled AUROC for radiomics-based ML was 0.912 (95% CI 0.891-0.934; grade: weak) [28-31,34,41,43,45,46,48,50,51,54,57,58,60,62,65,66,69,75,78-80,82,83].

The pooled AUROC was 0.893 (95% CI 0.840-0.948; grade: weak), 0.916 (95% CI 0.891-0.942; grade: weak), and

0.840 (95% CI 0.737-0.958; grade: weak) for CT-, MRI-, and ultrasound-based radiomics. The pooled AUROC was 0.874 (95% CI 0.852-0.896; grade: weak) and 0.921 (95% CI 0.867-0.979; grade: weak) for models integrating clinical features with CT- or MRI-based radiomics (Table 2 and Figures S5-S8 in Multimedia Appendix 3).

**Figure 5.** Summary plot of radiomics for detecting muscle invasion in the validation set [28-31,34,41,43,45,46,48,50,51,54,57,58,60,62,65,66,69,75,78-80,82,83].



### Meta-Analysis for Sensitivity and Specificity in Muscle Invasion Identification

The meta-analysis of 2x2 diagnostic contingency tables was conducted via a bivariate mixed effects model. In the training sets, the pooled sensitivity and specificity were 0.92 (95% CI 0.84-0.96) and 0.88 (95% CI 0.83-0.92), 0.80 (95% CI 0.76-0.84) and 0.90 (95% CI 0.86-0.92), and 0.81 and 0.91 for CT-, MRI-, and ultrasound-based radiomics. The pooled sensitivity and specificity were 0.79-0.93, 0.58-0.92, and 0.86 (95% CI 0.79-0.91) and 0.90 (95% CI 0.86-0.92) for models integrating clinical features with CT- or MRI-based radiomics (Table 3).

In the validation sets, the pooled sensitivity and specificity were 0.87 (95% CI 0.80-0.91) and 0.83 (95% CI 0.76-0.88), and 0.87 (95% CI 0.79-0.92) and 0.89 (95% CI 0.84-0.92) for CT- and MRI-based radiomics. The pooled sensitivity and specificity were 0.86 (95% CI 0.77-0.92) and 0.75 (95% CI 0.64-0.84) for models integrating clinical features with CT-based radiomics, and 0.88 (95% CI 0.81-0.93) and 0.88 (95% CI 0.76-0.94) for models integrating clinical features with MRI-based radiomics (Table 3).

**Table 3.** Meta-analysis results for sensitivity and specificity in the identification of muscle invasion.

Subgroup	Training set			Validation set		
	Number of models	Sensitivity (95% CI)		Number of models	Specificity (95% CI)	
		Sensitivity (95% CI)	Specificity (95% CI)		Sensitivity (95% CI)	Specificity (95% CI)
<b>Radiomics</b>						
<b>CT<sup>a</sup></b>						
Extra trees	1	0.82	0.81	1	0.81	0.8
ANN <sup>b</sup>				1	0.9	0.93
DL <sup>c</sup>	4	0.95 (0.85-0.99)	0.89 (0.82-0.93)	5	0.88 (0.77-0.94)	0.83 (0.71-0.91)
LR <sup>d</sup>	2	0.76-0.95	0.86-0.95	3	0.74-0.95	0.72-0.88
Overall	7	0.92 (0.84-0.96)	0.88 (0.83-0.92)	10	0.87 (0.80-0.91)	0.83 (0.76-0.88)
<b>MRI<sup>e</sup></b>						
DL	2	0.75-0.90	0.87-0.93	4	0.87 (0.78-0.92)	0.93 (0.88-0.96)
LASSO <sup>f</sup>	2	0.80-0.92	0.93-0.90	2	0.92-0.94	0.78-0.87
LR	2	0.82-0.83	0.84-0.90	1	0.81	0.83
<b>RF<sup>g</sup></b>						
SVM <sup>h</sup>	2	0.74-0.80	0.75-0.82	3	0.67-1	0.88-0.96
Overall	8	0.80 (0.76-0.84)	0.90 (0.86-0.92)	11	0.87 (0.79-0.92)	0.89 (0.84-0.92)
Ultrasound	1	0.81	0.91	— <sup>i</sup>	—	—
<b>Radiomics+ Clinical features</b>						
<b>CT</b>						
Overall	3	0.79-0.93	0.58-0.92	4	0.86 (0.77-0.92)	0.75 (0.64-0.84)
<b>MRI</b>						
LR	4	0.86 (0.79-0.91)	0.90 (0.86-0.92)	5	0.88 (0.78-0.94)	0.91 (0.79-0.96)
Overall	4	0.86 (0.79-0.91)	0.90 (0.86-0.92)	6	0.88 (0.81-0.93)	0.88 (0.76-0.94)

<sup>a</sup>CT: computed tomography.

<sup>b</sup>ANN: artificial neural network.

<sup>c</sup>DL: deep learning.

<sup>d</sup>LR: logistic regression.

<sup>e</sup>MRI: magnetic resonance imaging.

<sup>f</sup>LASSO: least absolute shrinkage and selection operator.

<sup>g</sup>RF: random forest.

<sup>h</sup>SVM: support vector machine.

<sup>i</sup>Not available.

### Subgroup Analysis

Subgroup analyses of AUROC for the diagnosis of muscle invasion were performed based on different imaging modalities and model types. Among CT-based radiomics models, deep learning (DL) was the predominant approach. In the validation sets, the pooled AUROC of DL was 0.903 (95% CI 0.826-0.988; grade: weak), with a sensitivity of 0.88 (95% CI 0.77-0.94) and a specificity of 0.83 (95% CI 0.71-0.91). Among MRI-based radiomics models, DL was the most common modeling approach. In the validation sets, the pooled AUROC of DL was 0.908 (95% CI 0.892-0.924; grade: weak), with a sensitivity of 0.87 (95% CI 0.78-0.92) and specificity of 0.93 (95% CI 0.88-0.96). In the validation sets, the model integrating CT-based radiomics with clinical characteristics yielded an AUROC of 0.874 (95% CI

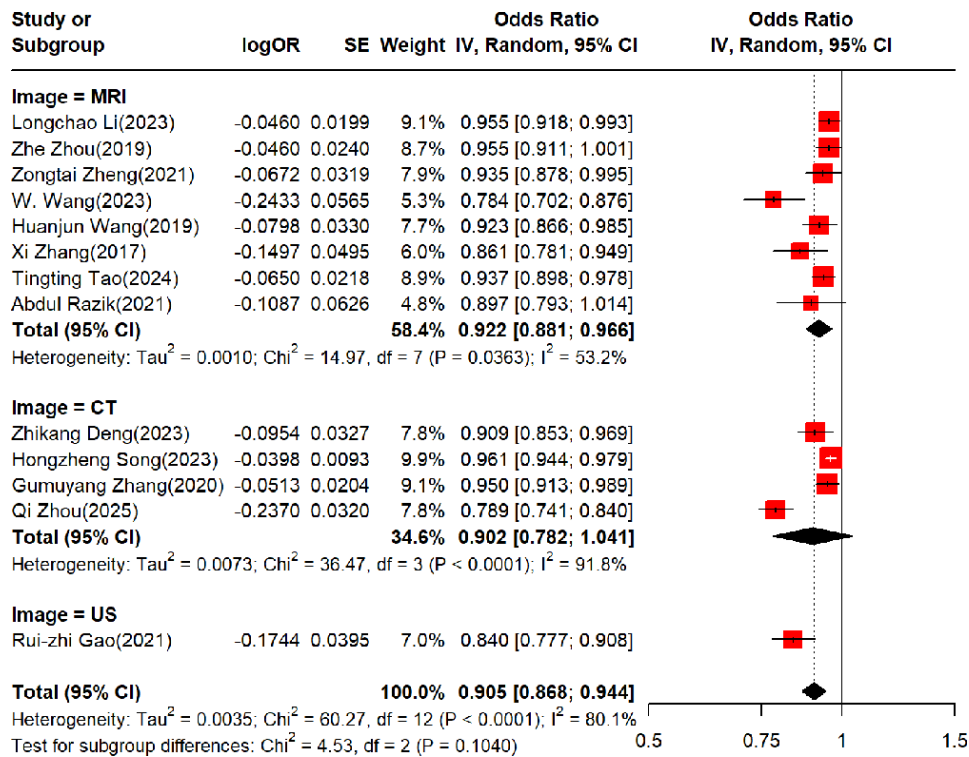
0.852-0.896; grade: weak). The corresponding sensitivity was 0.86 (95% CI 0.77-0.92), and the specificity was 0.75 (95% CI 0.64-0.84). Conversely, for MRI-based models combined with clinical variables, the pooled AUROC was 0.921 (95% CI 0.867-0.979; grade: weak). The corresponding sensitivity and specificity were 0.88 (95% CI 0.81-0.93) and 0.88 (95% CI 0.76-0.94), respectively (Tables 2-3).

### Diagnosis of High-Grade Tumors

#### Pooled Results

A meta-analysis of AUROC was conducted using a random-effects model. In the training sets, the pooled AUROC for radiomics-based ML was 0.905 (95% CI 0.868-0.944; grade: weak; Figure 6 [39,40,44,47,49,53,56,65-67,70,73,84]).

**Figure 6.** Summary plot of radiomics for identifying high-grade tumors in the training set [39,40,44,47,49,53,56,65-67,70,73,84].

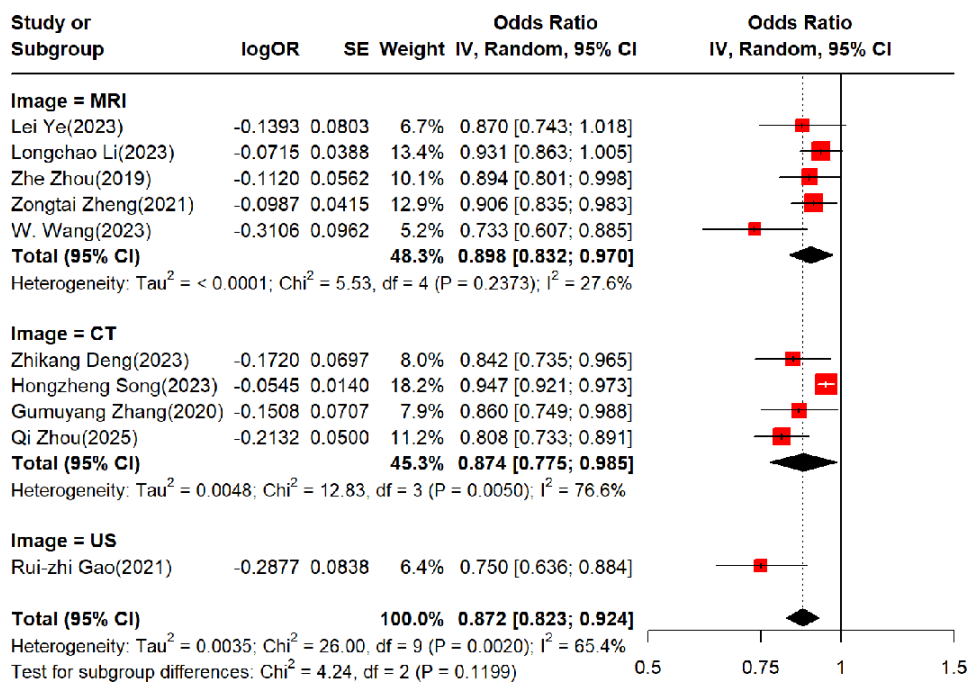


The pooled AUROC was 0.902 (95% CI 0.782-1.000; grade: weak) for CT-based radiomics, 0.895 (95% CI 0.814-0.985; grade: weak) for MRI-based radiomics, and 0.840 (95% CI 0.777-0.908; grade: weak) for ultrasound-based radiomics (Figures S9 and S10 in Multimedia Appendix 3 and Table 3). The AUROC for MRI-based radiomics combined with

clinical features was 0.944 (95% CI 0.904-0.985; grade: moderate; Figure S11 in Multimedia Appendix 3).

In the validation sets, the pooled AUROC for radiomics-based ML was 0.872 (95% CI 0.823-0.924; grade: weak; Figure 7 [36,39,40,44,47,49,53,56,65,84]).

**Figure 7.** Summary plot of radiomics for identifying high-grade tumors in the validation set [36,39,40,44,47,49,53,56,65,84].



The pooled AUROC was 0.874 (95% CI 0.775-0.985; grade: weak) for CT-based radiomics, 0.846 (95% CI 0.663-1.000; grade: very low) for MRI-based radiomics, and 0.750 (95% CI 0.636-0.884; grade: very low) for ultrasound-based

radiomics. (Figures S12-S13 in Multimedia Appendix 3). The AUROC for MRI-based radiomics combined with clinical features was 0.919 (95% CI 0.774-1.000; grade: weak; Table 4 and Figure S14 in Multimedia Appendix 3).

**Table 4.** Meta-analysis results for area under the receiver operating characteristic curve (AUROC) in the identification of high-grade tumors.

Subgroup	Training set					Validation set						
	No. of models	AUROC (95% CI)	tau <sup>2</sup>	tau	I <sup>2</sup>	Grade <sup>a</sup>	No. of models	AUROC (95% CI)	tau <sup>2</sup>	tau	I <sup>2</sup>	Grade
Radiomics												
CT <sup>b</sup>												
DL <sup>c</sup>	1	0.961 (0.944-0.979)				⊕⊕⊕ ⊖	1	0.947 (0.921-0.973)				⊕⊕⊕⊖
LR <sup>d</sup>	1	0.950 (0.913-0.989)				⊕⊕⊕ ⊖	1	0.860 (0.749-0.988)				⊕⊕⊖⊖
SVM <sup>e</sup>	1	0.909 (0.853-0.969)				⊕⊕ ⊖⊖	2	0.819 (0.639-1.000)	0	0	0	⊕⊖⊖⊖
AdaBoos t	1	0.789 (0.7410-0.840)				⊕⊕ ⊖⊖						
Overall	4	0.902 (0.782-1.000)	0.0073	0.0854	91.8	⊕⊕ ⊖⊖	4	0.874 (0.7750-0.985)	0.0048	0.0689	76.6	⊕⊕⊖⊖
MRI <sup>f</sup>												
DL	1	0.937 (0.898-0.978)				⊕⊕ ⊖⊖						
LR	3	0.882 (0.685-1.000)	0.0084	0.0915	81.1	⊕ ⊖⊖⊖	3	0.846 (0.663-1.000)	0.0031	0.0555	38.3	⊕⊖⊖⊖
SVM	1	0.861 (0.781-0.949)				⊕⊕ ⊖⊖						
Overall	5	0.895 (0.814-0.985)	0.004	0.0633	69.3	⊕⊕ ⊖⊖	3	0.846 (0.663-1.000)	0.0031	0.0555	38.3	⊕⊖⊖⊖
Ultrasound	1	0.840 (0.777-0.908)				⊕⊕ ⊖⊖	1	0.750 (0.636-0.884)			NA	⊕⊖⊖⊖
Radiomics+ clinical features												
CT												
MRI	3	0.944 (0.904-0.985)	0	0	0	⊕⊕⊕ ⊖	2	0.919 (0.774-1.000)	0	0	0	⊕⊕⊖⊖

<sup>a</sup>In the GRADE assessment section, ⊕⊕⊕⊕ represents high quality of evidence, ⊕⊕⊕⊖ represents moderate quality of evidence, ⊕⊕⊖⊖ represents low quality of evidence, and ⊕⊖⊖⊖ represents very low quality of evidence.

<sup>b</sup>CT: computed tomography.

<sup>c</sup>DL: deep learning.

<sup>d</sup>LR: logistic regression.

<sup>e</sup>SVM: support vector machine.

<sup>f</sup>MRI: magnetic resonance imaging.

A meta-analysis of 2x2 diagnostic contingency tables was performed via a bivariate mixed-effects model. In the training sets, the pooled sensitivity and specificity were 0.91 (95% CI 0.84-0.95) and 0.74 (95% CI 0.67-0.79) for CT-based radiomics, and 0.86 (95% CI 0.57-0.96) and 0.89 (95% CI 0.83-0.93) for MRI-based radiomics. The sensitivity and specificity were 0.78 (95% CI 0.65-0.87) and 0.94 (95% CI 0.79-0.99) for MRI-based radiomics combined with clinical features.

In the validation sets, the pooled sensitivity and specificity were 0.65-0.95 and 0.69-0.73 for CT-based radiomics, and 0.81~0.89 and 0.70~0.91 for MRI-based radiomics. The sensitivity and specificity were 0.54-0.77 and 0.88-0.98 for MRI-based radiomics combined with clinical features (Table 5).

**Table 5.** Meta-analysis results for sensitivity and specificity in identification of high-grade tumors.

Subgroup	Training set				Validation set		
	Number of models	Sensitivity (95% CI)	Specificity (95% CI)		Number		
					Sensitivity (95% CI)	Specificity (95% CI)	
Radiomics							

Subgroup	Training set			Validation set		
	Number of models	Sensitivity (95% CI)	Specificity (95% CI)	Number of models	Sensitivity (95% CI)	Specificity (95% CI)
<b>CT<sup>a</sup></b>						
LR <sup>b</sup>	1	0.96	0.71	1	0.89	0.73
NB <sup>c</sup>	1	0.89	0.77			
SVM <sup>d</sup>	1	0.93	0.77	2	0.65-0.95	0.69-0.72
AdaBoost	1	0.85	0.73			
Overall	4	0.91 (0.84-0.95)	0.74 (0.67-0.79)	3	0.65-0.95	0.69-0.73
<b>MRI<sup>e</sup></b>						
DL <sup>f</sup>	1	0.99	0.9			
LR	2	0.60-0.74	0.82-1	3	0.81-0.89	0.70-0.91
SVM	1	0.78	0.87			
<b>XGBoost</b>						
Overall	4	0.86 (0.57-0.96)	0.89 (0.83-0.93)	3	0.81-0.89	0.70-0.91
Ultrasound	1	0.8	0.76			
<b>Radiomics+ Clinical features</b>						
<b>MRI</b>						
Overall	4	0.78 (0.65-0.87)	0.94 (0.79-0.99)	2	0.54-0.77	0.88-0.98

<sup>a</sup>CT: computed tomography.

<sup>b</sup>LR: logistic regression.

<sup>c</sup>NB: naive Bayes.

<sup>d</sup>SVM: support vector machine.

<sup>e</sup>MRI: magnetic resonance imaging.

<sup>f</sup>DL: deep learning.

## Subgroup Analysis

Subgroup analyses of AUROC for the diagnosis of high-grade tumors were performed based on different imaging modalities and model types. In CT-based radiomics, the support vector machine (SVM) model was the predominant type. The pooled AUROC of SVM was 0.819 (95% CI 0.639-1.000; grade: very low) in the validation sets, with corresponding sensitivity of 0.65-0.95 and specificity of 0.69-0.72. Likewise, the LR model was the predominant type in MRI-based radiomics. The pooled AUROC of LR was 0.846 (95% CI 0.663-1.000; grade: very low) in the validation sets, with corresponding sensitivity of 0.81-0.89 and specificity of 0.70-0.91. In the validation sets, the model integrating MRI-based radiomics with clinical characteristics achieved an AUROC of 0.919 (95% CI 0.774-1.000; grade: weak). The corresponding sensitivity was 0.54-0.77, and the specificity was 0.88-0.98. Given the paucity of available studies using CT-based radiomics combined with clinical feature models, subgroup analyses could not be performed (Tables 4 and 5).

## Other Risk Stratification Outcomes

HER2 is encoded by the ERBB2 gene. It plays a critical role in the development of various malignancies, including breast, gastric, bladder, ovarian, and lung cancers. HER2-targeted therapies have become first-line treatments for patients with advanced cancers exhibiting HER2 overexpression. Three studies reported on radiomics-based approaches to assess HER2 expression in bladder cancer. Yu et al [55] used MRI radiomics-based ML for noninvasive assessment of HER2.

In their test set, an SVM model demonstrated an AUC of 0.712 (95% CI 0.535-0.889), with a sensitivity of 0.857 and a specificity of 0.533. They highlighted the potential value of this approach in cases where patients cannot undergo invasive procedures such as biopsies or diagnostic TUR. Based on contrast-enhanced computed tomography (CE-CT), Peng et al [74] assessed a clinical-radiomics model in evaluating HER2 status in urothelial bladder cancer. This CE-CT-based model exhibited the highest effectiveness in forecasting HER2 status, with an AUC of 0.814 (95% CI 0.642-0.986) in the test set. Wei et al [77] conducted a multicenter study using explainable ML based on CT radiomics to preoperatively diagnose HER2 status in bladder cancer. In their test set, AUCs were 0.803, 0.709, 0.679, 0.794, and 0.815 for LR, SVM, KNN, XGBoost, and RF models, respectively. These studies suggest that radiomics-based ML has promising potential for detecting HER2 expression in bladder cancer.

Ki-67 expression is associated with a poor prognosis and advanced clinicopathological features in cancer. Two single-center studies [35,76] explored radiomics-based prediction of Ki-67 expression in bladder cancer, with the potential to improve prognostic assessment and clinical decision-making. Zheng et al [35] reported on an MRI-based radiomics study. Their SMOTE-LASSO model achieved an AUC of 0.819 (0.658-0.980) in the validation set, with a sensitivity of 0.795 and a specificity of 0.867%. Feng et al [76] developed a radiomics nomogram based on CE-CT, which demonstrated AUCs of 0.836 and 0.887 in their validation set. These findings indicate that radiomics-based

ML holds promise for detecting Ki-67 expression in bladder cancer.

Additionally, 2 [37,68] studies reported on the use of radiomics to predict LN staging. Gresser et al [68] used CT radiomics to evaluate LN staging in bladder cancer. Their combined model incorporated manually segmented radiomic features and radiologist assessment. The model achieved an AUC of 0.81 (0.71-0.92), a sensitivity of 0.73 (0.55-0.88), and a specificity of 0.84 (0.76-0.92), respectively, in the test set. However, Starmans et al [37] conducted a multicenter study using preoperative CT radiomics. They found that in patients with cT2-T4aN0-N1M0 muscle-invasive bladder cancer (MIBC), radiomics was not helpful in differentiating pN+ and pN0 disease. This suggests that further research is needed to explore the effectiveness of radiomics-based ML in detecting LN staging in bladder cancer.

### Small-Study Effect Analysis

Funnel plots were generated to evaluate small-study effects in radiomics-based ML models for detecting muscle invasion and high-grade tumors. Significant small-study effects were observed in both the training and validation sets for assessing muscle invasion ( $P<.05$ ; Figures S15-S16 in [Multimedia Appendix 3](#)), as well as for identifying high-grade tumors ( $P<.05$ ; Figures S17-S18 in [Multimedia Appendix 3](#)).

### Meta-Regression Analysis

For the tasks of detecting muscle invasion, meta-regression was performed. Independent variables included case number, model type (LR [Reference] vs other ML), and modeling variables (Radiomics [Reference] vs. Radiomics +clinical). Separate analyses were conducted for models developed on CT or MRI in the training set and validated on CT or MRI in the validation set. The results indicated a significant association between case number and the AUC only in the validation set of MRI-based models for detecting muscle invasion ( $P<.05$ ). No other variables demonstrated a significant influence ( $P>.05$ ; Table S2 in [Multimedia Appendix 4](#) and Table S3 in [Multimedia Appendix 5](#)). Meta-regression was not conducted for pathological grade due to an insufficient number of studies.

## Discussion

### Summary of the Main Findings

This systematic review incorporated 57 studies [28-84] to evaluate radiomics-based ML for preoperative risk stratification in bladder cancer. The findings indicated a high AUROC for detecting both muscle invasion and high-grade tumors. Additionally, radiomics showed potential for identifying HER2 and Ki-67 expression. However, evaluation via the RQS revealed an overall low methodological quality among the eligible studies. PROBAST-AI assessment revealed that the primary source of bias was in the model evaluation phase, primarily due to small validation set sizes or an absence of external validation. Consequently, the findings of this review should be interpreted with caution. Based on the present findings, radiomics is considered to have application

potential. Nevertheless, the current evidence faces significant challenges, including methodological shortcomings and a high ROB, which currently preclude its readiness for clinical translation.

### Comparison With Previous Reviews

Previous studies have examined the application of radiomics in bladder cancer. Kozikowski et al [10] reviewed the prediction ability of radiomics in muscle-invasive cancer, ultimately including eight articles. The pooled estimated sensitivity and specificity were 82% (95% CI 77%-86%) and 81% (95% CI 76%-85%). However, the study did not classify or analyze CT and MRI image sources separately nor attempt to combine different ML approaches with clinical radiomics models. Boca et al [14] reviewed MRI-based radiomics studies in bladder cancer, ultimately including 26 articles with 2991 participants. Radiomics in these studies was primarily used for preoperative prediction of tumor stage or molecular correlations ( $n=9$ ), preoperative tumor grading ( $n=13$ ), and prediction of prognosis or response to neoadjuvant therapy ( $n=4$ ). Most radiomics models incorporated second-order features from filtered images, with quality scores ranging from 8.33% to 52.77% [14]. However, the study only discussed MRI-based radiomics and failed to consider other imaging modalities. The study also did not analyze sensitivity, specificity, or dataset differences, nor did it incorporate a discussion of radiomics combined with clinical features. Building on these prior efforts, the current study conducted a more systematic and comprehensive review of the current state of radiomics-based ML in bladder cancer. The current study also considered different image features, accounted for various image sources, analyzed different modeling approaches, and examined the detection performance across diverse datasets. This provides more comprehensive evidence for future developments.

In the clinical management of bladder cancer, CT often struggles to accurately diagnose flat lesions and prostate-adjacent bladder base lesions, especially in patients with benign prostatic hyperplasia. This difficulty arises from the challenge of distinguishing tumor recurrence from inflammatory wall thickening following intravesical chemotherapy, as well as from scar tissue after TURBT [85]. MRI assessment is a laborious, slice-by-slice process, with its effectiveness depending on the radiologist's experience [86]. The integration of radiomics and ML with bladder MRI holds promise for improving staging and treatment response assessment. Bladder cancer management guidelines would be enhanced by the integration of MRI into their staging strategies [87]. However, radiomics may experience similar slow progress as molecular biology-based diagnostic and therapeutic techniques. This can be attributed to several factors, such as technical complexities, a lack of validation standards, and inadequate study design (particularly conflating hypothesis generation with hypothesis testing) [11]. Other contributing factors include data overfitting, incomplete reporting of results, and unidentified confounding variables in the datasets used (especially in retrospective datasets). Therefore, as with all biomarker studies, retrospective radiomics studies require

validation on independent datasets, ideally from another institution [11].

In clinical practice, MRI has become the most accurate imaging modality for evaluating local invasiveness in bladder cancer. It can be used to assess regional LN involvement and tumor spread to pelvic bones and the upper urinary tract [21, 88]. While bladder cancer appears as a soft tissue lesion on CT scans, it is more easily identified as a filling defect during CT urography. Ultrasound is a dynamic imaging modality. It can distinguish bladder cancer from other conditions that appear similar. Its ability to detect blood flow aids in distinguishing the solid tissue of the tumor from blood clots or debris [89]. This meta-analysis reveals that in studies examining muscle invasion, the primary image sources are CT, MRI, and, to a lesser extent, ultrasound. Radiomics combined with clinical features has not consistently demonstrated superior detection performance, potentially due to limited available data. The diagnostic effectiveness of radiomics warrants further investigation. In studies investigating histological grading, the primary image sources are CT and MRI. Only a few studies have explored combining radiomics with clinical features, resulting in limited current evidence.

Traditional ML-based radiomics requires substantial upfront effort, including texture extraction, manual image segmentation, and model construction. This process carries the risk of image data loss and imposes a significant workload. Manual segmentation, in particular, can be influenced by clinical experience, habits, and prior research [86,90]. Consequently, some researchers have explored DL, leveraging its capabilities for staging, grading, automated tumor detection, intelligent segmentation, bladder wall segmentation, and prediction of recurrence, response to chemotherapy, and overall survival. The goal is to improve disease management [58,91-94]. While external validation of intelligent segmentation accuracy is lacking, the potential of DL to reduce workload and automatically interpret images in an intelligent manner is encouraging the development of better, smarter tools. Of the studies included in this review, only a few DL models have demonstrated good performance. Thus, future research should explore DL in addition to traditional methods.

While the models in this study demonstrated relatively high AUROC values, the widespread absence of calibration metrics represents a major limitation. This gap substantially undermines the reliability of these models for practical application [95,96]. Dependence on discriminatory metrics such as AUC, sensitivity, and specificity only reflects the ability of a model to differentiate between positive and negative events. These metrics do not, however, inform on the accuracy of the predicted probabilities—a critical shortfall that poses challenges for safe integration into clinical decision support [97,98]. Consequently, future studies must report comprehensive calibration metrics, including calibration curves, calibration slope, intercept, and the Brier score. Furthermore, decision curve analysis should be performed to evaluate the clinical net benefit across different decision thresholds, a crucial step for assessing clinical utility.

Before initiating this meta-analysis, a prospective registration was completed. However, several adjustments were made during the actual research process. First, to mitigate the risk of missing newly published literature and ensure the completeness of evidence, supplementary database searches were conducted on October 17, 2025. Second, to better reflect the ROB in the included studies, PROBAST-AI was adopted in addition to RQS for the assessment of bias. Third, the GRADE approach was applied to evaluate the certainty of evidence. Fourth, subgroup analyses were carried out based on the type of dataset (training vs validation), imaging source, and type of model. Finally, analyses of small-study effects and meta-regression were also performed. The registered protocol will be updated accordingly.

### **Strengths and Limitations**

This meta-analysis is the first comprehensive and systematic review of preoperative risk stratification for bladder cancer, summarizing the available evidence. However, it is subject to the following limitations. First, the present analysis incorporated only 16 studies [29-32,42,47,50,52,58-60,77,78, 80,81,83] that used external validation. Among these, 13 investigations focused on muscle invasion, one examined histological grade, one assessed HER2 expression, and one addressed preoperative LN staging. The analytical strategy required subgrouping based on the imaging modality used for model development and the integration of clinical features. Incorporating the data source as a subgroup variable was precluded due to the limited number of multicenter datasets, which would have yielded statistically unreliable comparisons. Furthermore, the validation sets in this study were predominantly created via random sampling. The lack of independent external validation limits the interpretation of the results to some extent. Second, different mathematical models exhibit varying performance in processing images. Although subgroup analyses were conducted to account for these differences, the small number of models within each subgroup may limit the interpretation of our findings. Therefore, the comparative performance of models within these subgroups should be interpreted with caution. Third, evidence for the preoperative identification of LN metastasis and Ki-67 expression is extremely limited. Fourth, while subgroup analyses were performed based on different model types, definitive conclusions cannot be drawn. However, given the paucity of available studies and the small number of articles included in specific model categories, definitive conclusions cannot be drawn. Fifth, the RQS and PROBAST-AI instruments incorporate elements requiring subjective judgment, which may introduce variability into the overall interpretation of the quality assessments. Sixth, a notable limitation of the current evidence base is the infrequent reporting of calibration metrics, such as the calibration-in-the-large, calibration slope, and Brier score. This omission considerably diminishes the clinical applicability of the model performance estimates. Seventh, while all eligible studies confirmed diagnoses pathologically, the transparency regarding the origin of pathological specimens was often insufficient. This lack of detail concerning specimen sourcing

is a potential source of bias. Together, these limitations collectively curtail the generalizability of the study findings. Eighth, a notable small-study effect was observed across the included studies. Thus, the findings should be interpreted with caution.

### **Challenges and Future Directions**

In summary, significant challenges remain prior to the clinical deployment of these models. First, current investigations have not rigorously examined the impact of imaging modality. Variations in acquisition protocols can influence image quality; however, the potential effect of such heterogeneity on modeling outcomes has not been adequately addressed in the existing literature. Second, independent external validation is essential for establishing the generalizability of models. The predominant reliance on internal validation in the eligible studies implies that these findings propose a potential methodological approach rather than one ready for clinical application. Substantial additional evidence is required to confirm clinical utility and generalizability. Third, sample size presents a major constraint. Only a few studies provided a sufficient number of cases, and most encompassed fewer than 200 subjects. Consequently, discussions regarding model robustness necessitate larger, more powerful datasets. Fourth, the image segmentation process depends heavily on manual delineation. This approach is time-consuming and susceptible to variability introduced by the operator's prior knowledge. Future work should prioritize developing

and implementing DL-based automated or semiautomated segmentation techniques.

To facilitate successful clinical translation, future investigations should prioritize the following areas: conducting robust pilot studies, performing reproducibility analyses on different imaging protocols and segmentation methods, ensuring adequate sample sizes, implementing multicenter external validation, comparing diverse ML architectures, and establishing protocols for continuous model updating and refinement.

### **Conclusions**

This is the first systematic review to comprehensively assess radiomics for preoperative risk stratification in bladder cancer. The findings provide evidence to support the development and refinement of future ML-based tools for image analysis. However, due to limitations in the current evidence, such as methodological flaws and a high ROB, low GRADE level, clinical translation is not yet warranted. Future research should standardize radiomics workflows, incorporate multicenter images from diverse geographical regions, and minimize the impact of imaging protocols and pre-processing steps. These measures are essential to advance radiomics toward successful clinical implementation. Such efforts are essential for fully elucidating and validating the potential of radiomics in the noninvasive diagnosis of bladder cancer.

---

### **Acknowledgments**

This study was conducted without the assistance of artificial intelligence technologies in its writing process. The authors assume full responsibility and accountability for ensuring the precision, novelty, and scholarly integrity of every element within the manuscript, encompassing all referenced sources and citation practices.

---

### **Funding**

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

---

### **Data Availability**

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

---

### **Authors' Contributions**

Conceptualization: ZH, YL

Data curation: ZH, QJ, BX

Formal analysis: ZH, YL, QJ, FH

Investigation: ZH, YL, QJ, FH

Methodology: ZH, QJ, BX

Project administration: FH, BX

Resources: ZH, BX

Software: ZH, YL

Supervision: FH, BX

Validation: ZH, YL, BX

Visualization: ZH, YL

Writing – original draft: ZH, YL

Writing – review & editing: ZH, BX

All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

---

### **Conflicts of Interest**

None declared.

---

### **Multimedia Appendix 1**

Literature search strategy.

[\[DOCX File \(Microsoft Word File\), 17 KB-Multimedia Appendix 1\]](#)

## Multimedia Appendix 2

Quality assessment of included studies using the radiomics quality score (RQS) tool.

[\[DOCX File \(Microsoft Word File\), 45 KB-Multimedia Appendix 2\]](#)

## Multimedia Appendix 3

Forest and funnel plots for preoperative risk stratification using radiomics-based machine learning.

[\[DOCX File \(Microsoft Word File\), 15727 KB-Multimedia Appendix 3\]](#)

## Multimedia Appendix 4

Meta-regression of area under the curve (AUC) for machine learning models based on computed tomography (CT) and magnetic resonance imaging (MRI) radiomics in detecting muscle invasion (training set).

[\[DOCX File \(Microsoft Word File\), 15 KB-Multimedia Appendix 4\]](#)

## Multimedia Appendix 5

Meta-regression of area under the curve (AUC) for machine learning models based on computed tomography (CT) and magnetic resonance imaging (MRI) radiomics in detecting muscle invasion (validation set).

[\[DOCX File \(Microsoft Word File\), 15 KB-Multimedia Appendix 5\]](#)

## Checklist 1

PRISMA 2020 checklist.

[\[PDF File \(Adobe File\), 606 KB-Checklist 1\]](#)

## References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. May 2021;71(3):209-249. [doi: [10.3322/caac.21660](https://doi.org/10.3322/caac.21660)] [Medline: [33538338](https://pubmed.ncbi.nlm.nih.gov/33538338/)]
2. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. *CA Cancer J Clin*. 2024;74(1):12-49. [doi: [10.3322/caac.21820](https://doi.org/10.3322/caac.21820)] [Medline: [38230766](https://pubmed.ncbi.nlm.nih.gov/38230766/)]
3. 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*. Jan 2022;81(1):75-94. [doi: [10.1016/j.eururo.2021.08.010](https://doi.org/10.1016/j.eururo.2021.08.010)] [Medline: [34511303](https://pubmed.ncbi.nlm.nih.gov/34511303/)]
4. Mariappan P, Zachou A, Grigor KM, Edinburgh Uro-Oncology Group. Detrusor muscle in the first, apparently complete transurethral resection of bladder tumour specimen is a surrogate marker of resection quality, predicts risk of early recurrence, and is dependent on operator experience. *Eur Urol*. May 2010;57(5):843-849. [doi: [10.1016/j.eururo.2009.05.047](https://doi.org/10.1016/j.eururo.2009.05.047)] [Medline: [19524354](https://pubmed.ncbi.nlm.nih.gov/19524354/)]
5. Poletajew S, Krajewski W, Kaczmarek K, et al. The learning curve for transurethral resection of bladder tumour: how many is enough to be independent, safe and effective surgeon? *J Surg Educ*. 2020;77(4):978-985. [doi: [10.1016/j.jsurg.2020.02.010](https://doi.org/10.1016/j.jsurg.2020.02.010)] [Medline: [32147466](https://pubmed.ncbi.nlm.nih.gov/32147466/)]
6. Svatek RS, Shariat SF, Novara G, et al. Discrepancy between clinical and pathological stage: external validation of the impact on prognosis in an international radical cystectomy cohort. *BJU Int*. Mar 2011;107(6):898-904. [doi: [10.1111/j.1464-410X.2010.09628.x](https://doi.org/10.1111/j.1464-410X.2010.09628.x)] [Medline: [21244604](https://pubmed.ncbi.nlm.nih.gov/21244604/)]
7. Yanagisawa T, Kawada T, von Deimling M, et al. Repeat transurethral resection for non-muscle-invasive bladder cancer: an updated systematic review and meta-analysis in the contemporary era. *Eur Urol Focus*. Jan 2024;10(1):41-56. [doi: [10.1016/j.euf.2023.07.002](https://doi.org/10.1016/j.euf.2023.07.002)] [Medline: [37495458](https://pubmed.ncbi.nlm.nih.gov/37495458/)]
8. Noh MR, Ezz Eldin O. Precise vesical wall staging of bladder cancer in the era of precision medicine: has it been fulfilled? *Abdom Radiol (NY)*. Jul 2025;50(7):3084-3091. [doi: [10.1007/s00261-024-04786-8](https://doi.org/10.1007/s00261-024-04786-8)] [Medline: [39725735](https://pubmed.ncbi.nlm.nih.gov/39725735/)]
9. Gulddammer CS, Vásquez JL, Kristensen VM, et al. Cystoscopy accuracy in detecting bladder tumors: a prospective video-confirmed study. *Cancers (Basel)*. Dec 28, 2023;16(1):160. [doi: [10.3390/cancers16010160](https://doi.org/10.3390/cancers16010160)] [Medline: [38201586](https://pubmed.ncbi.nlm.nih.gov/38201586/)]
10. Kozikowski M, Suarez-Ibarrola R, Osiecki R, et al. Role of radiomics in the prediction of muscle-invasive bladder cancer: a systematic review and meta-analysis. *Eur Urol Focus*. May 2022;8(3):728-738. [doi: [10.1016/j.euf.2021.05.005](https://doi.org/10.1016/j.euf.2021.05.005)] [Medline: [34099417](https://pubmed.ncbi.nlm.nih.gov/34099417/)]
11. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. Feb 2016;278(2):563-577. [doi: [10.1148/radiol.2015151169](https://doi.org/10.1148/radiol.2015151169)] [Medline: [26579733](https://pubmed.ncbi.nlm.nih.gov/26579733/)]

12. Kwong JCC, Wu J, Malik S, et al. Predicting non-muscle invasive bladder cancer outcomes using artificial intelligence: a systematic review using APPRAISE-AI. *NPJ Digit Med*. Apr 18, 2024;7(1):98. [doi: [10.1038/s41746-024-01088-7](https://doi.org/10.1038/s41746-024-01088-7)] [Medline: [38637674](https://pubmed.ncbi.nlm.nih.gov/38637674/)]
13. Gelikman DG, Rais-Bahrami S, Pinto PA, Turkbey B. AI-powered radiomics: revolutionizing detection of urologic malignancies. *Curr Opin Urol*. Jan 1, 2024;34(1):1-7. [doi: [10.1097/MOU.0000000000001144](https://doi.org/10.1097/MOU.0000000000001144)] [Medline: [37909882](https://pubmed.ncbi.nlm.nih.gov/37909882/)]
14. Boca B, Caraianni C, Telecan T, et al. MRI-based radiomics in bladder cancer: a systematic review and radiomics quality score assessment. *Diagnostics (Basel)*. Jul 6, 2023;13(13):2300. [doi: [10.3390/diagnostics13132300](https://doi.org/10.3390/diagnostics13132300)] [Medline: [37443692](https://pubmed.ncbi.nlm.nih.gov/37443692/)]
15. Marvaso G, Isaksson LJ, Zaffaroni M, et al. Can we predict pathology without surgery? Weighing the added value of multiparametric MRI and whole prostate radiomics in integrative machine learning models. *Eur Radiol*. Oct 2024;34(10):6241-6253. [doi: [10.1007/s00330-024-10699-3](https://doi.org/10.1007/s00330-024-10699-3)] [Medline: [38507053](https://pubmed.ncbi.nlm.nih.gov/38507053/)]
16. Kulkarni S, Seneviratne N, Baig MS, Khan AHA. Artificial intelligence in medicine: where are we now? *Acad Radiol*. Jan 2020;27(1):62-70. [doi: [10.1016/j.acra.2019.10.001](https://doi.org/10.1016/j.acra.2019.10.001)] [Medline: [31636002](https://pubmed.ncbi.nlm.nih.gov/31636002/)]
17. Pak S, Park SG, Park J, Cho ST, Lee YG, Ahn H. Applications of artificial intelligence in urologic oncology. *Investig Clin Urol*. May 2024;65(3):202-216. [doi: [10.4111/icu.20230435](https://doi.org/10.4111/icu.20230435)] [Medline: [38714511](https://pubmed.ncbi.nlm.nih.gov/38714511/)]
18. Wang Z, Shi H, Wang Q, et al. AI-driven and traditional radiomic model for predicting muscle invasion in bladder cancer via multi-parametric imaging: a systematic review and meta-analysis. *Acad Radiol*. Dec 2025;32(12):7215-7243. [doi: [10.1016/j.acra.2025.08.035](https://doi.org/10.1016/j.acra.2025.08.035)] [Medline: [40914722](https://pubmed.ncbi.nlm.nih.gov/40914722/)]
19. He C, Xu H, Yuan E, et al. The accuracy and quality of image-based artificial intelligence for muscle-invasive bladder cancer prediction. *Insights Imaging*. Aug 1, 2024;15(1):185. [doi: [10.1186/s13244-024-01780-y](https://doi.org/10.1186/s13244-024-01780-y)] [Medline: [39090234](https://pubmed.ncbi.nlm.nih.gov/39090234/)]
20. Mastroleo F, Marvaso G, Jereczek-Fossa BA. Artificial intelligence in muscle-invasive bladder cancer: opportunities, challenges, and clinical impact. *Curr Opin Urol*. Sep 1, 2025;35(5):543-548. [doi: [10.1097/MOU.0000000000001309](https://doi.org/10.1097/MOU.0000000000001309)] [Medline: [40765429](https://pubmed.ncbi.nlm.nih.gov/40765429/)]
21. Pastorino GL, Mercinelli C, Necchi A. The role of MRI in muscle-invasive bladder cancer: an update from the last two years. *Curr Opin Urol*. Mar 1, 2025;35(2):165-170. [doi: [10.1097/MOU.0000000000001249](https://doi.org/10.1097/MOU.0000000000001249)] [Medline: [39529480](https://pubmed.ncbi.nlm.nih.gov/39529480/)]
22. Esteva A, Feng J, van der Wal D, et al. Prostate cancer therapy personalization via multi-modal deep learning on randomized phase III clinical trials. *NPJ Digit Med*. Jun 8, 2022;5(1):71. [doi: [10.1038/s41746-022-00613-w](https://doi.org/10.1038/s41746-022-00613-w)] [Medline: [35676445](https://pubmed.ncbi.nlm.nih.gov/35676445/)]
23. Collins GS, Dhiman P, Andaur Navarro CL, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ Open*. Jul 9, 2021;11(7):e048008. [doi: [10.1136/bmjopen-2020-048008](https://doi.org/10.1136/bmjopen-2020-048008)] [Medline: [34244270](https://pubmed.ncbi.nlm.nih.gov/34244270/)]
24. Moons KGM, Damen JAA, Kaul T, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *BMJ*. Mar 24, 2025;388:e082505. [doi: [10.1136/bmj-2024-082505](https://doi.org/10.1136/bmj-2024-082505)] [Medline: [40127903](https://pubmed.ncbi.nlm.nih.gov/40127903/)]
25. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. Dec 2017;14(12):749-762. [doi: [10.1038/nrclinonc.2017.141](https://doi.org/10.1038/nrclinonc.2017.141)] [Medline: [28975929](https://pubmed.ncbi.nlm.nih.gov/28975929/)]
26. Debray TP, Damen JA, Riley RD, et al. A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. *Stat Methods Med Res*. Sep 2019;28(9):2768-2786. [doi: [10.1177/0962280218785504](https://doi.org/10.1177/0962280218785504)] [Medline: [30032705](https://pubmed.ncbi.nlm.nih.gov/30032705/)]
27. IntHout J, Ioannidis JPA, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Med Res Methodol*. Feb 18, 2014;14:25. [doi: [10.1186/1471-2288-14-25](https://doi.org/10.1186/1471-2288-14-25)] [Medline: [24548571](https://pubmed.ncbi.nlm.nih.gov/24548571/)]
28. Zhang R, Jia S, Zhai L, Wu F, Zhang S, Li F. Predicting preoperative muscle invasion status for bladder cancer using computed tomography-based radiomics nomogram. *BMC Med Imaging*. Apr 27, 2024;24(1):98. [doi: [10.1186/s12880-024-01276-7](https://doi.org/10.1186/s12880-024-01276-7)] [Medline: [38678222](https://pubmed.ncbi.nlm.nih.gov/38678222/)]
29. Zhang G, Wu Z, Xu L, et al. Deep learning on enhanced CT images can predict the muscular invasiveness of bladder cancer. *Front Oncol*. 2021;11:654685. [doi: [10.3389/fonc.2021.654685](https://doi.org/10.3389/fonc.2021.654685)] [Medline: [34178641](https://pubmed.ncbi.nlm.nih.gov/34178641/)]
30. Ye Y, Luo Z, Qiu Z, et al. Radiomics prediction of muscle invasion in bladder cancer using semi-automatic lesion segmentation of MRI compared with manual segmentation. *Bioengineering (Basel)*. Nov 25, 2023;10(12):1355. [doi: [10.3390/bioengineering10121355](https://doi.org/10.3390/bioengineering10121355)] [Medline: [38135946](https://pubmed.ncbi.nlm.nih.gov/38135946/)]
31. Ren J, Gu H, Zhang N, Chen W. Preoperative CT-based radiomics for diagnosing muscle invasion of bladder cancer. *Egypt J Radiol Nucl Med*. 2023;54(1):131. [doi: [10.1186/s43055-023-01044-7](https://doi.org/10.1186/s43055-023-01044-7)]
32. Qureshi TA, Chen X, Xie Y, et al. MRI/RNA-seq-based radiogenomics and artificial intelligence for more accurate staging of muscle-invasive bladder cancer. *Int J Mol Sci*. Dec 20, 2023;25(1):88. [doi: [10.3390/ijms25010088](https://doi.org/10.3390/ijms25010088)] [Medline: [38203254](https://pubmed.ncbi.nlm.nih.gov/38203254/)]

33. Özdemir H, Azamat S, Sam Özdemir M. Can only the shape feature in radiomics help machine learning show that bladder cancer has invaded muscles? *Cureus*. Sep 2023;15(9):e45488. [doi: [10.7759/cureus.45488](https://doi.org/10.7759/cureus.45488)] [Medline: [37859896](https://pubmed.ncbi.nlm.nih.gov/37859896/)]
34. Chen W, Gong M, Zhou D, et al. CT-based deep learning radiomics signature for the preoperative prediction of the muscle-invasive status of bladder cancer. *Front Oncol*. 2022;12:1019749. [doi: [10.3389/fonc.2022.1019749](https://doi.org/10.3389/fonc.2022.1019749)] [Medline: [36544709](https://pubmed.ncbi.nlm.nih.gov/36544709/)]
35. Zheng Z, Gu Z, Xu F, et al. Magnetic resonance imaging-based radiomics signature for preoperative prediction of Ki67 expression in bladder cancer. *Cancer Imaging*. Dec 4, 2021;21(1):65. [doi: [10.1186/s40644-021-00433-3](https://doi.org/10.1186/s40644-021-00433-3)] [Medline: [34863282](https://pubmed.ncbi.nlm.nih.gov/34863282/)]
36. Ye L, Wang Y, Xiang W, Yao J, Liu J, Song B. Radiomic analysis of quantitative T2 mapping and conventional MRI in predicting histologic grade of bladder cancer. *J Clin Med*. Sep 11, 2023;12(18):5900. [doi: [10.3390/jcm12185900](https://doi.org/10.3390/jcm12185900)] [Medline: [37762841](https://pubmed.ncbi.nlm.nih.gov/37762841/)]
37. Starman MPA, Ho LS, Smits F, et al. Optimization of preoperative lymph node staging in patients with muscle-invasive bladder cancer using radiomics on computed tomography. *J Pers Med*. Apr 30, 2022;12(5):726. [doi: [10.3390/jpm12050726](https://doi.org/10.3390/jpm12050726)] [Medline: [35629148](https://pubmed.ncbi.nlm.nih.gov/35629148/)]
38. Sarkar S, Min K, Ikram W, et al. Performing automatic identification and staging of urothelial carcinoma in bladder cancer patients using a hybrid deep-machine learning approach. *Cancers (Basel)*. Mar 8, 2023;15(6):1673. [doi: [10.3390/cancers15061673](https://doi.org/10.3390/cancers15061673)] [Medline: [36980557](https://pubmed.ncbi.nlm.nih.gov/36980557/)]
39. Li L, Zhang J, Zhe X, et al. An MRI-based radiomics nomogram in predicting histologic grade of non-muscle-invasive bladder cancer. *Front Oncol*. 2023;13:1025972. [doi: [10.3389/fonc.2023.1025972](https://doi.org/10.3389/fonc.2023.1025972)] [Medline: [37007156](https://pubmed.ncbi.nlm.nih.gov/37007156/)]
40. Deng Z, Dong W, Xiong S, et al. Machine learning models combining computed tomography semantic features and selected clinical variables for accurate prediction of the pathological grade of bladder cancer. *Front Oncol*. 2023;13:1166245. [doi: [10.3389/fonc.2023.1166245](https://doi.org/10.3389/fonc.2023.1166245)] [Medline: [37223680](https://pubmed.ncbi.nlm.nih.gov/37223680/)]
41. Chen G, Fan X, Wang T, et al. A machine learning model based on MRI for the preoperative prediction of bladder cancer invasion depth. *Eur Radiol*. Dec 2023;33(12):8821-8832. [doi: [10.1007/s00330-023-09960-y](https://doi.org/10.1007/s00330-023-09960-y)] [Medline: [37470826](https://pubmed.ncbi.nlm.nih.gov/37470826/)]
42. Yu J, Cai L, Chen C, et al. A novel predict method for muscular invasion of bladder cancer based on 3D mp-MRI feature fusion. *Phys Med Biol*. Feb 22, 2024;69(5). [doi: [10.1088/1361-6560/ad25c7](https://doi.org/10.1088/1361-6560/ad25c7)] [Medline: [38306973](https://pubmed.ncbi.nlm.nih.gov/38306973/)]
43. Xu S, Yao Q, Liu G, et al. Combining DWI radiomics features with transurethral resection promotes the differentiation between muscle-invasive bladder cancer and non-muscle-invasive bladder cancer. *Eur Radiol*. Mar 2020;30(3):1804-1812. [doi: [10.1007/s00330-019-06484-2](https://doi.org/10.1007/s00330-019-06484-2)] [Medline: [31773297](https://pubmed.ncbi.nlm.nih.gov/31773297/)]
44. Zhou Z, Liu L, Xue K, Ma Y, Liu J, Zhang M. Assessment of pathological grading of bladder cancer using texture features from MRI. Presented at: 2019 IEEE International Conference on Mechatronics and Automation (ICMA); Aug 4-7, 2019:1333-1337; Tianjin, China. 2019.URL: <https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=8801712> [doi: [10.1109/ICMA.2019.8816242](https://doi.org/10.1109/ICMA.2019.8816242)]
45. Yang Y, Zou X, Wang Y, Ma X. Application of deep learning as a noninvasive tool to differentiate muscle-invasive bladder cancer and non-muscle-invasive bladder cancer with CT. *Eur J Radiol*. Jun 2021;139:109666. [doi: [10.1016/j.ejrad.2021.109666](https://doi.org/10.1016/j.ejrad.2021.109666)] [Medline: [33798819](https://pubmed.ncbi.nlm.nih.gov/33798819/)]
46. Zheng Z, Xu F, Gu Z, et al. Combining Multiparametric MRI Radiomics Signature With the Vesical Imaging-Reporting and Data System (VI-RADS) Score to Preoperatively Differentiate Muscle Invasion of Bladder Cancer. *Front Oncol*. 2021;11:619893. [doi: [10.3389/fonc.2021.619893](https://doi.org/10.3389/fonc.2021.619893)] [Medline: [34055600](https://pubmed.ncbi.nlm.nih.gov/34055600/)]
47. Song H, Yang S, Yu B, et al. CT-based deep learning radiomics nomogram for the prediction of pathological grade in bladder cancer: a multicenter study. *Cancer Imaging*. Sep 18, 2023;23(1):89. [doi: [10.1186/s40644-023-00609-z](https://doi.org/10.1186/s40644-023-00609-z)] [Medline: [37723572](https://pubmed.ncbi.nlm.nih.gov/37723572/)]
48. Cui Y, Sun Z, Liu X, Zhang X, Wang X. CT-based radiomics for the preoperative prediction of the muscle-invasive status of bladder cancer and comparison to radiologists' assessment. *Clin Radiol*. Jun 2022;77(6):e473-e482. [doi: [10.1016/j.crad.2022.02.019](https://doi.org/10.1016/j.crad.2022.02.019)] [Medline: [35367051](https://pubmed.ncbi.nlm.nih.gov/35367051/)]
49. Zhang G, Xu L, Zhao L, et al. CT-based radiomics to predict the pathological grade of bladder cancer. *Eur Radiol*. Dec 2020;30(12):6749-6756. [doi: [10.1007/s00330-020-06893-8](https://doi.org/10.1007/s00330-020-06893-8)] [Medline: [32601949](https://pubmed.ncbi.nlm.nih.gov/32601949/)]
50. Wei Z, Liu H, Xv Y, et al. Development and validation of a CT-based deep learning radiomics nomogram to predict muscle invasion in bladder cancer. *Heliyon*. Jan 30, 2024;10(2):e24878. [doi: [10.1016/j.heliyon.2024.e24878](https://doi.org/10.1016/j.heliyon.2024.e24878)] [Medline: [38304824](https://pubmed.ncbi.nlm.nih.gov/38304824/)]
51. Zheng J, Kong J, Wu S, et al. Development of a noninvasive tool to preoperatively evaluate the muscular invasiveness of bladder cancer using a radiomics approach. *Cancer*. Dec 15, 2019;125(24):4388-4398. [doi: [10.1002/cncr.32490](https://doi.org/10.1002/cncr.32490)] [Medline: [31469418](https://pubmed.ncbi.nlm.nih.gov/31469418/)]

52. 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*. Sep 2020;30(9):4816-4827. [doi: [10.1007/s00330-020-06796-8](https://doi.org/10.1007/s00330-020-06796-8)] [Medline: [32318846](https://pubmed.ncbi.nlm.nih.gov/32318846/)]
53. Zheng Z, Xu F, Gu Z, et al. Integrating multiparametric MRI radiomics features and the Vesical Imaging-Reporting and Data System (VI-RADS) for bladder cancer grading. *Abdom Radiol (NY)*. Sep 2021;46(9):4311-4323. [doi: [10.1007/s00261-021-03108-6](https://doi.org/10.1007/s00261-021-03108-6)] [Medline: [33978825](https://pubmed.ncbi.nlm.nih.gov/33978825/)]
54. Wang W, Li W, Wang K, et al. Integrating radiomics with the vesical imaging-reporting and data system to predict muscle invasion of bladder cancer. *Urol Oncol*. Jun 2023;41(6):294. [doi: [10.1016/j.urolonc.2022.10.024](https://doi.org/10.1016/j.urolonc.2022.10.024)] [Medline: [36526525](https://pubmed.ncbi.nlm.nih.gov/36526525/)]
55. Yu R, Cai L, Gong Y, et al. MRI-based machine learning radiomics for preoperative assessment of human epidermal growth factor receptor 2 status in urothelial bladder carcinoma. *J Magn Reson Imaging*. Dec 2024;60(6):2694-2704. [doi: [10.1002/jmri.29342](https://doi.org/10.1002/jmri.29342)] [Medline: [38456745](https://pubmed.ncbi.nlm.nih.gov/38456745/)]
56. Wang W, Wang K, Qiu J, et al. MRI-based radiomics analysis of bladder cancer: prediction of pathological grade and histological variant. *Clin Radiol*. Nov 2023;78(11):e889-e897. [doi: [10.1016/j.crad.2023.07.020](https://doi.org/10.1016/j.crad.2023.07.020)] [Medline: [37633748](https://pubmed.ncbi.nlm.nih.gov/37633748/)]
57. Zhang L, Li X, Yang L, et al. Multi-sequence and multi-regional MRI -based radiomics nomogram for the preoperative assessment of muscle invasion in bladder cancer. *J Magn Reson Imaging*. Jul 2023;58(1):258-269. URL: <https://onlinelibrary.wiley.com/toc/15222586/58/1> [doi: [10.1002/jmri.28498](https://doi.org/10.1002/jmri.28498)] [Medline: [36300676](https://pubmed.ncbi.nlm.nih.gov/36300676/)]
58. Zou Y, Cai L, Chen C, et al. Multi-task deep learning based on T2-weighted images for predicting muscular-invasive bladder cancer. *Comput Biol Med*. Dec 2022;151(Pt A):106219. [doi: [10.1016/j.compbiomed.2022.106219](https://doi.org/10.1016/j.compbiomed.2022.106219)] [Medline: [36343408](https://pubmed.ncbi.nlm.nih.gov/36343408/)]
59. Zhou X, Yue X, Xu Z, Denoeux T, Chen Y. PENet: Prior evidence deep neural network for bladder cancer staging. *Methods*. Nov 2022;207:20-28. [doi: [10.1016/j.ymeth.2022.08.010](https://doi.org/10.1016/j.ymeth.2022.08.010)] [Medline: [36031139](https://pubmed.ncbi.nlm.nih.gov/36031139/)]
60. Li J, Qiu Z, Cao K, et al. Predicting muscle invasion in bladder cancer based on MRI: a comparison of radiomics, and single-task and multi-task deep learning. *Comput Methods Programs Biomed*. May 2023;233:107466. [doi: [10.1016/j.cmpb.2023.107466](https://doi.org/10.1016/j.cmpb.2023.107466)] [Medline: [36907040](https://pubmed.ncbi.nlm.nih.gov/36907040/)]
61. Li L, Zhang J, Zhe X, et al. Prediction of histopathologic grades of bladder cancer with radiomics based on MRI: comparison with traditional MRI. *Urol Oncol*. Jun 2024;42(6):176. [doi: [10.1016/j.urolonc.2024.02.008](https://doi.org/10.1016/j.urolonc.2024.02.008)] [Medline: [38556403](https://pubmed.ncbi.nlm.nih.gov/38556403/)]
62. Chen L, Zhang G, Xu L, et al. Preoperative CT features to predict risk stratification of non-muscle invasive bladder cancer. *Abdom Radiol (NY)*. Feb 2023;48(2):659-668. [doi: [10.1007/s00261-022-03730-y](https://doi.org/10.1007/s00261-022-03730-y)] [Medline: [36454277](https://pubmed.ncbi.nlm.nih.gov/36454277/)]
63. Xu X, Liu Y, Zhang X, et al. Preoperative prediction of muscular invasiveness of bladder cancer with radiomic features on conventional MRI and its high-order derivative maps. *Abdom Radiol (NY)*. Jul 2017;42(7):1896-1905. [doi: [10.1007/s00261-017-1079-6](https://doi.org/10.1007/s00261-017-1079-6)] [Medline: [28217825](https://pubmed.ncbi.nlm.nih.gov/28217825/)]
64. Xu X, Zhang X, Tian Q, et al. Quantitative identification of nonmuscle-invasive and muscle-invasive bladder carcinomas: a multiparametric MRI radiomics analysis. *J Magn Reson Imaging*. May 2019;49(5):1489-1498. URL: <https://onlinelibrary.wiley.com/toc/15222586/49/5> [Accessed 2026-04-26] [doi: [10.1002/jmri.26327](https://doi.org/10.1002/jmri.26327)] [Medline: [30252978](https://pubmed.ncbi.nlm.nih.gov/30252978/)]
65. Gao RZ, Wen R, Wen DY, et al. Radiomics analysis based on ultrasound images to distinguish the tumor stage and pathological grade of bladder cancer. *J Ultrasound Med*. Dec 2021;40(12):2685-2697. [doi: [10.1002/jum.15659](https://doi.org/10.1002/jum.15659)] [Medline: [33615528](https://pubmed.ncbi.nlm.nih.gov/33615528/)]
66. Wang H, Hu D, Yao H, et al. Radiomics analysis of multiparametric MRI for the preoperative evaluation of pathological grade in bladder cancer tumors. *Eur Radiol*. Nov 2019;29(11):6182-6190. [doi: [10.1007/s00330-019-06222-8](https://doi.org/10.1007/s00330-019-06222-8)] [Medline: [31016445](https://pubmed.ncbi.nlm.nih.gov/31016445/)]
67. Zhang X, Xu X, Tian Q, et al. Radiomics assessment of bladder cancer grade using texture features from diffusion-weighted imaging. *J Magn Reson Imaging*. Nov 2017;46(5):1281-1288. URL: <https://onlinelibrary.wiley.com/toc/15222586/46/5> [Accessed 2026-04-26] [doi: [10.1002/jmri.25669](https://doi.org/10.1002/jmri.25669)] [Medline: [28199039](https://pubmed.ncbi.nlm.nih.gov/28199039/)]
68. Gresser E, Woźnicki P, Messmer K, et al. Radiomics signature using manual versus automated segmentation for lymph node staging of bladder cancer. *Eur Urol Focus*. Jan 2023;9(1):145-153. [doi: [10.1016/j.euf.2022.08.015](https://doi.org/10.1016/j.euf.2022.08.015)] [Medline: [36115774](https://pubmed.ncbi.nlm.nih.gov/36115774/)]
69. Tong Y, Udupa JK, Wang C, et al. Radiomics-guided therapy for bladder cancer: using an optimal biomarker approach to determine extent of bladder cancer invasion from t2-weighted magnetic resonance images. *Adv Radiat Oncol*. 2018;3(3):331-338. [doi: [10.1016/j.adro.2018.04.011](https://doi.org/10.1016/j.adro.2018.04.011)] [Medline: [30202802](https://pubmed.ncbi.nlm.nih.gov/30202802/)]
70. Tao T, Chen Y, Shang Y, He J, Hao J. SMMF: a self-attention-based multi-parametric MRI feature fusion framework for the diagnosis of bladder cancer grading. *Front Oncol*. 2024;14:1337186. [doi: [10.3389/fonc.2024.1337186](https://doi.org/10.3389/fonc.2024.1337186)] [Medline: [38515574](https://pubmed.ncbi.nlm.nih.gov/38515574/)]

71. Liu Y, Xu X, Wang H, et al. The additional value of tri-parametric MRI in identifying muscle-invasive status in bladder cancer. *Acad Radiol*. Jan 2023;30(1):64-76. [doi: [10.1016/j.acra.2022.04.014](https://doi.org/10.1016/j.acra.2022.04.014)] [Medline: [35676179](https://pubmed.ncbi.nlm.nih.gov/35676179/)]
72. Liu D, Wang S, Wang J. The effect of CT high-resolution imaging diagnosis based on deep residual network on the pathology of bladder cancer classification and staging. *Comput Methods Programs Biomed*. Mar 2022;215:106635. [doi: [10.1016/j.cmpb.2022.106635](https://doi.org/10.1016/j.cmpb.2022.106635)] [Medline: [35063711](https://pubmed.ncbi.nlm.nih.gov/35063711/)]
73. Razik A, Das CJ, Sharma R, et al. Utility of first order MRI-Texture analysis parameters in the prediction of histologic grade and muscle invasion in urinary bladder cancer: a preliminary study. *Br J Radiol*. Jun 1, 2021;94(1122):20201114. [doi: [10.1259/bjr.20201114](https://doi.org/10.1259/bjr.20201114)] [Medline: [33882245](https://pubmed.ncbi.nlm.nih.gov/33882245/)]
74. Peng J, Tang Z, Li T, Pan X, Feng L, Long L. Contrast-enhanced computed tomography-based radiomics nomogram for predicting HER2 status in urothelial bladder carcinoma. *Front Oncol*. 2024;14:1427122. [doi: [10.3389/fonc.2024.1427122](https://doi.org/10.3389/fonc.2024.1427122)] [Medline: [39206159](https://pubmed.ncbi.nlm.nih.gov/39206159/)]
75. Xiong S, Fu Z, Deng Z, et al. Machine learning-based CT radiomics enhances bladder cancer staging predictions: a comparative study of clinical, radiomics, and combined models. *Med Phys*. Sep 2024;51(9):5965-5977. URL: <https://aapm.onlinelibrary.wiley.com/toc/24734209/51/9> [doi: [10.1002/mp.17288](https://doi.org/10.1002/mp.17288)] [Medline: [38977273](https://pubmed.ncbi.nlm.nih.gov/38977273/)]
76. Feng S, Zhou D, Li Y, et al. Prediction of Ki-67 expression in bladder cancer based on CT radiomics nomogram. *Front Oncol*. 2024;14:1276526. [doi: [10.3389/fonc.2024.1276526](https://doi.org/10.3389/fonc.2024.1276526)] [Medline: [38482209](https://pubmed.ncbi.nlm.nih.gov/38482209/)]
77. Wei Z, Bai X, Xv Y, et al. A radiomics-based interpretable machine learning model to predict the HER2 status in bladder cancer: a multicenter study. *Insights Imaging*. Oct 28, 2024;15(1):262. [doi: [10.1186/s13244-024-01840-3](https://doi.org/10.1186/s13244-024-01840-3)] [Medline: [39466475](https://pubmed.ncbi.nlm.nih.gov/39466475/)]
78. Cai L, Yang X, Yu J, et al. Deep learning on T2WI to predict the muscle-invasive bladder cancer: a multi-center clinical study. *Sci Rep*. Mar 22, 2025;15(1):9942. [doi: [10.1038/s41598-024-82909-3](https://doi.org/10.1038/s41598-024-82909-3)] [Medline: [40121216](https://pubmed.ncbi.nlm.nih.gov/40121216/)]
79. Du C, Wei W, Hu M, et al. Multi-DECT image-based interpretable model incorporating habitat radiomics and vision transformer deep learning for preoperative prediction of muscle invasion in bladder cancer. *Acad Radiol*. Dec 2025;32(12):7204-7214. [doi: [10.1016/j.acra.2025.08.018](https://doi.org/10.1016/j.acra.2025.08.018)] [Medline: [40887351](https://pubmed.ncbi.nlm.nih.gov/40887351/)]
80. Du Y, Li H, Sui Y, et al. Habitat-based radiomic model for predicting muscle invasion in bladder cancer: a multi-center study using enhanced-CT and machine learning. *Med Phys*. Aug 2025;52(8):e18021. [doi: [10.1002/mp.18021](https://doi.org/10.1002/mp.18021)] [Medline: [40781767](https://pubmed.ncbi.nlm.nih.gov/40781767/)]
81. He X, Chen Y, Zhou S, et al. A machine learning model based on multi-phase contrast-enhanced CT for the preoperative prediction of the muscle-invasive status of bladder cancer. *Curr Med Imaging*. 2025;21:e15734056377754. [doi: [10.2174/0115734056377754250304040058](https://doi.org/10.2174/0115734056377754250304040058)] [Medline: [40103464](https://pubmed.ncbi.nlm.nih.gov/40103464/)]
82. Hu M, Zhang J, Cheng Q, et al. Multi-DECT image-based intratumoral and peritumoral radiomics for preoperative prediction of muscle invasion in bladder cancer. *Acad Radiol*. Jan 2025;32(1):287-297. [doi: [10.1016/j.acra.2024.08.010](https://doi.org/10.1016/j.acra.2024.08.010)] [Medline: [39168722](https://pubmed.ncbi.nlm.nih.gov/39168722/)]
83. Yu J, Cai L, Chen C, et al. Multi-path neural network based on mp-MRI for predicting muscle-invasive bladder cancer. *Intell Data Anal*. Nov 2025;29(6):1568-1581. [doi: [10.1177/1088467X241313324](https://doi.org/10.1177/1088467X241313324)]
84. Zhou Q, Ma L, Yu Y, et al. Development of a radiomics and clinical feature-based nomogram for preoperative prediction of pathological grade in bladder cancer. *Front Oncol*. 2025;15:1661979. [doi: [10.3389/fonc.2025.1661979](https://doi.org/10.3389/fonc.2025.1661979)] [Medline: [40936708](https://pubmed.ncbi.nlm.nih.gov/40936708/)]
85. Verma S, Rajesh A, Prasad SR, et al. Urinary bladder cancer: role of MR imaging. *Radiographics*. 2012;32(2):371-387. [doi: [10.1148/rg.322115125](https://doi.org/10.1148/rg.322115125)] [Medline: [22411938](https://pubmed.ncbi.nlm.nih.gov/22411938/)]
86. Bleker J, Kwee TC, Rouw D, et al. A deep learning masked segmentation alternative to manual segmentation in biparametric MRI prostate cancer radiomics. *Eur Radiol*. Sep 2022;32(9):6526-6535. [doi: [10.1007/s00330-022-08712-8](https://doi.org/10.1007/s00330-022-08712-8)] [Medline: [35420303](https://pubmed.ncbi.nlm.nih.gov/35420303/)]
87. van der Pol CB, Chung A, Lim C, et al. Update on multiparametric MRI of urinary bladder cancer. *J Magn Reson Imaging*. Oct 2018;48(4):882-896. [doi: [10.1002/jmri.26294](https://doi.org/10.1002/jmri.26294)] [Medline: [30221801](https://pubmed.ncbi.nlm.nih.gov/30221801/)]
88. Caglic I, Panebianco V, Vargas HA, et al. MRI of bladder cancer: local and nodal staging. *J Magn Reson Imaging*. Sep 2020;52(3):649-667. [doi: [10.1002/jmri.27090](https://doi.org/10.1002/jmri.27090)] [Medline: [32112505](https://pubmed.ncbi.nlm.nih.gov/32112505/)]
89. Wentland AL, Dessler TS, Troxell ML, Kamaya A. Bladder cancer and its mimics: a sonographic pictorial review with CT/MR and histologic correlation. *Abdom Radiol (NY)*. Dec 2019;44(12):3827-3842. [doi: [10.1007/s00261-019-02276-w](https://doi.org/10.1007/s00261-019-02276-w)] [Medline: [31676920](https://pubmed.ncbi.nlm.nih.gov/31676920/)]
90. Akin O, Lema-Dopico A, Paudyal R, et al. Multiparametric MRI in era of artificial intelligence for bladder cancer therapies. *Cancers (Basel)*. Nov 18, 2023;15(22):5468. [doi: [10.3390/cancers15225468](https://doi.org/10.3390/cancers15225468)] [Medline: [38001728](https://pubmed.ncbi.nlm.nih.gov/38001728/)]
91. Cai L, Bai R, Cao Q, et al. A non-invasive MRI-based multimodal fusion deep learning model (MF-DLM) for predicting overall survival in bladder cancer: a multicentre retrospective study. *EClinicalMedicine*. Dec 2025;90:103640. [doi: [10.1016/j.eclinm.2025.103640](https://doi.org/10.1016/j.eclinm.2025.103640)] [Medline: [41324013](https://pubmed.ncbi.nlm.nih.gov/41324013/)]

92. Wei Z, Xu 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*. May 1, 2024;110(5):2922-2932. [doi: [10.1097/JS9.0000000000001194](https://doi.org/10.1097/JS9.0000000000001194)] [Medline: [38349205](https://pubmed.ncbi.nlm.nih.gov/38349205/)]
93. Jiang F, Hong G, Zeng H, et al. Deep learning-based model for prediction of early recurrence and therapy response on whole slide images in non-muscle-invasive bladder cancer: a retrospective, multicentre study. *EClinicalMedicine*. Mar 2025;81:103125. [doi: [10.1016/j.eclinm.2025.103125](https://doi.org/10.1016/j.eclinm.2025.103125)] [Medline: [40093987](https://pubmed.ncbi.nlm.nih.gov/40093987/)]
94. Lucas M, Jansen I, van Leeuwen TG, Oddens JR, de Bruin DM, Marquering HA. Deep learning-based recurrence prediction in patients with non-muscle-invasive bladder cancer. *Eur Urol Focus*. Jan 2022;8(1):165-172. [doi: [10.1016/j.euf.2020.12.008](https://doi.org/10.1016/j.euf.2020.12.008)] [Medline: [33358370](https://pubmed.ncbi.nlm.nih.gov/33358370/)]
95. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. *BMC Med*. Dec 16, 2019;17(1):230. [doi: [10.1186/s12916-019-1466-7](https://doi.org/10.1186/s12916-019-1466-7)] [Medline: [31842878](https://pubmed.ncbi.nlm.nih.gov/31842878/)]
96. Zhang T, Nikouline A, Lightfoot D, Nolan B. Machine learning in the prediction of trauma outcomes: a systematic review. *Ann Emerg Med*. Nov 2022;80(5):440-455. [doi: [10.1016/j.annemergmed.2022.05.011](https://doi.org/10.1016/j.annemergmed.2022.05.011)] [Medline: [35842343](https://pubmed.ncbi.nlm.nih.gov/35842343/)]
97. Dong J, Jiang Z, Pan D, et al. A survey on confidence calibration of deep learning-based classification models under class imbalance data. *IEEE Trans Neural Netw Learn Syst*. Sep 2025;36(9):15664-15684. [doi: [10.1109/TNNLS.2025.3565159](https://doi.org/10.1109/TNNLS.2025.3565159)] [Medline: [40531642](https://pubmed.ncbi.nlm.nih.gov/40531642/)]
98. Ehrmann DE, Joshi S, Goodfellow SD, Mazwi ML, Eytan D. Making machine learning matter to clinicians: model actionability in medical decision-making. *NPJ Digit Med*. Jan 24, 2023;6(1):7. [doi: [10.1038/s41746-023-00753-7](https://doi.org/10.1038/s41746-023-00753-7)] [Medline: [36690689](https://pubmed.ncbi.nlm.nih.gov/36690689/)]

## Abbreviations

**AI:** artificial intelligence  
**AUC:** area under the curve  
**AUROC:** area under the receiver operating characteristic curve  
**CE-CT:** contrast-enhanced computed tomography  
**CT:** computed tomography  
**DL:** deep learning  
**GRADE:** Grading of Recommendations, Assessment, Development, and Evaluation  
**HER2:** human epidermal growth factor receptor 2  
**LN:** lymph node  
**LR:** logistic regression  
**MIBC:** muscle-invasive bladder cancer  
**ML:** machine learning  
**MRI:** magnetic resonance imaging  
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses  
**PROBAST-AI:** Prediction Model Risk of Bias Assessment Tool for artificial intelligence  
**PROSPERO:** International Prospective Register of Systematic Reviews  
**ROB:** risk of bias  
**ROI:** region of interest  
**RQS:** Radiomics Quality Scoring  
**SVM:** support vector machine  
**TURBT:** Transurethral resection of bladder tumor

*Edited by Stefano Brini; peer-reviewed by Huasheng Lv, Long Zhou, Sara Zamani; submitted 22.Jul.2025; final revised version received 03.Apr.2026; accepted 06.Apr.2026; published 12.Jun.2026*

*Please cite as:*

*He Z, Liu Y, Jiang Q, Hou F, Xu B*

*Application Value of Radiomics-Based Machine Learning for Preoperative Risk Stratification of Bladder Cancer: Systematic Review and Meta-Analysis*

*J Med Internet Res 2026;28:e81084*

*URL: <https://www.jmir.org/2026/1/e81084>*

*doi: [10.2196/81084](https://doi.org/10.2196/81084)*

© Zirong He, Yinghua Liu, Qin Jiang, Fang Hou, Bing Xu. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 12.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.