ACCSCIENCE PUBLISHING

Eurasian Journal of Medicine and Oncology

ORIGINAL RESEARCH ARTICLE

A comparative analysis of U-Net-based architectures for robust segmentation of bladder cancer lesions in magnetic resonance imaging

Ishak Pacal[1,2]* and Yigitcan Cakmak[1,2]

[1]Department of Computer Engineering, Faculty of Engineering, Igdir University, Igdir, Türkiye

[2]Department of Electronics and Information Technologies, Faculty of Architecture and Engineering, Nakhchivan State University, Nakhchivan, Azerbaijan

(This article belongs to the *Special Issue: New Developments in Bladder Cancer Treatment and Management*)

Abstract

Introduction: Bladder cancer (BCa) represents a significant uro-oncological challenge due to its aggressive nature and high recurrence rates. Although magnetic resonance imaging (MRI) is a cornerstone modality in BCa management, the manual segmentation of lesions is time-consuming and suffers from low reproducibility due to inter- and intra-observer variability, morphological heterogeneity, and MRI artifacts.

Objective: This study aims to address these limitations by conducting a rigorous comparative evaluation of four distinct U-Net-based deep learning architectures.

Methods: The models were evaluated using the publicly available, multi-center FedBCa dataset, comprising 275 T2-weighted MRI scans from 228 patients. Using a standardized training protocol, performance was rigorously assessed with a suite of quantitative metrics, including the Dice coefficient, intersection over union (IoU), and Hausdorff distance, supplemented by qualitative visual comparison.

Results: Cross-scale mixer U-Net (CMUNet) achieved the best overall performance, yielding the highest Dice coefficient (0.7937), IoU (0.7033), and boundary delineation accuracy (Hausdorff distance: 8.4550 mm. Architectural trade-offs were evident: CMUNeXt was the most computationally efficient and offered the highest lesion sensitivity (0.9656), whereas Attention U-Net recorded the highest precision (0.8380).

Conclusion: CMUNet provides the most balanced and accurate performance for BCa segmentation. However, the optimal architecture choice is application-dependent; high-sensitivity models such as CMUNeXt are ideal for screening, while high-precision models like Attention U-Net are better suited for treatment planning. Deep learning models serve as powerful assistive tools to improve efficiency and objectivity in clinical workflows, though expert oversight remains essential. The top model's accuracy approached, but did not surpass, the inter-rater reliability of human experts (Dice: 0.870).

Keywords: Bladder cancer; Deep learning; U-Net; Image segmentation; Magnetic resonance imaging; Clinical decision support systems

***Corresponding author:**
Ishak Pacal
(ishak.pacal@igdir.edu.tr)

Citation: Pacal I, Cakmak Y. A comparative analysis of U-Net-based architectures for robust segmentation of bladder cancer lesions in magnetic resonance imaging. *Eurasian J Med Oncol*. 2025;9(4):268-283.
doi: 10.36922/EJMO025260276

Received: June 26, 2025

Revised: September 14, 2025

Accepted: September 22, 2025

Published online: October 27, 2025

1. Introduction

Bladder cancer (BCa) is a serious global health problem, ranking among the most prevalent cancers of the urinary tract.[1,2] It is defined as cancer that arises from the urothelial cells lining the inside of the bladder. When diagnosed early, BCa typically has a high treatment success rate; however, advanced BCa can lead to significant morbidity and mortality.[3] Major risk factors of BCa include smoking, age, male sex, and occupational exposure to certain chemicals.[4-6] According to estimates from the American Cancer Society for 2024, the projected number of new BCa cases in the United States was 83,190, with 63,070 in men and 20,120 in women. Estimates also suggest that in 2024, there were 16,840 deaths due to BCa (12,290 men and 4,550 women), further emphasizing the disease's significance and the need for effective diagnostic and therapeutic methodologies.[7]

Medical imaging methods play a pivotal role in the diagnosis and staging of BCa.[8] Cystoscopy, considered the cornerstone of the diagnostic process, allows for direct endoscopic observation of the bladder's interior and enables biopsies of suspicious lesions.[9] Nevertheless, cross-sectional imaging techniques such as computed tomography (CT) and magnetic resonance imaging (MRI) are indispensable for obtaining critical information regarding the depth of tumor invasion into the bladder wall, its relationship with surrounding tissues, and potential lymph node or distant organ metastasis.[10,11] These methods guide accurate disease staging and inform the selection of appropriate treatment strategies (e.g., surgery, chemotherapy, and radiotherapy). However, the interpretation of these images, even by experienced radiologists, can be time-consuming and is susceptible to inter-observer variability.[12,13]

With rapid advances in artificial intelligence, including machine learning and deep learning, high-capacity models now power perception and decision-making systems across numerous domains.[14-16] In medicine, these techniques are increasingly tailored for medical image analysis.[17] They enable detection, segmentation, and classification across a wide range of cancers and related conditions, including brain, lung,[18] breast,[19,20] prostate, bladder, liver, pancreatic, colorectal, head and neck, gynecologic, musculoskeletal, and hematologic malignancies, as well as dental, renal, and other urological pathologies.[21-24] Together, these developments illustrate a general-purpose capability for objective, scalable, and reproducible visual interpretation tailored to clinical tasks.[25] One of the most important applications in this field is segmentation, the process of precisely delineating the boundaries of specific anatomical structures or pathologies within an image.[26,27] Accurate segmentation of cancerous tissue is highly relevant for objective tumor volume measurement, treatment response monitoring, and surgical planning.[28,29] Deep learning-based algorithms can automate the segmentation of urological tumors, including BCas, on CT or MRI scans. Ongoing studies continue to improve these methods to enhance the speed and accuracy of diagnosis and treatment.[30-32]

To address the limitations of manual interpretation in medical imaging, deep learning, a subfield of artificial intelligence, has revolutionized medical image analysis. Deep learning models, particularly convolutional neural networks (CNNs), have an exceptional capacity to automatically learn hierarchical feature representations directly from image data. This eliminates the need for manual feature engineering and allows the models to capture a broad spectrum of patterns, from simple edges and textures to complex contextual and semantic information. For the task of semantic segmentation, the U-Net architecture has established itself as a gold standard in the biomedical domain. Its elegant design consists of a contracting path (encoder) that captures image context and a symmetric expanding path (decoder) that enables precise localization. The critical innovation of U-Net lies in its "skip connections," which fuse high-resolution feature maps from the encoder with upsampled maps in the decoder. This mechanism allows the network to combine high-level semantic information with low-level spatial details, which is essential for delineating the fine and often ambiguous boundaries of pathological tissues.

The success of this foundational framework has inspired a plethora of U-Net variants aimed at further improving performance. However, despite the proliferation of these advanced architectures, a rigorous, head-to-head comparison on a large, contemporary, and clinically relevant BCa dataset is conspicuously lacking in the literature. This gap makes it difficult for the research community to objectively assess the true performance gains of newer models and to select the optimal architecture for this specific clinical problem.

This study addresses this critical gap by providing a comprehensive and systematic benchmark of leading U-Net-based models. Our primary contribution is the rigorous evaluation of four influential architectures: the foundational U-Net, the widely adopted Attention U-Net, and the cutting-edge cross-scale mixer U-Net (CMUNet) and CMUNeXt models. We performed this comparative analysis on a novel and extensive BCa dataset, which represents one of the largest and most current collections available for this purpose. By meticulously assessing the performance of these models under standardized conditions, we aim to identify the most robust and accurate architecture for BCa segmentation. The findings of this work are intended to serve as a crucial guideline

for future research and to accelerate the development of reliable, automated tools that can be integrated into clinical workflows for managing BCa.

1.1. Related works

The application of deep learning for delineating organs-at-risk in radiotherapy planning has garnered significant attention. Researchers have demonstrated the utility of U-Net-based architectures for the automated segmentation of the bladder, a critical organ-at-risk in pelvic radiotherapy. For example, do Nascimento Silva and Idrobo Pizo[33] developed a U-Net model for bladder delineation in cone beam CT images, achieving a high Dice similarity coefficient (DSC) of 81% through the use of data augmentation. Similarly, addressing the specific needs of gynecological cancer treatment, Das *et al.*[34] proposed an enhanced U-Net with both long and short skip connections for bladder segmentation in T2-weighted (T2W) magnetic resonance images for high-dose-rate brachytherapy. Their approach yielded superior intersection over union (IoU) metrics compared to the standard U-Net, underscoring the potential of customized deep learning models to provide consistent and accurate segmentations across different imaging modalities, which are essential for precise dose delivery.

Beyond radiotherapy, deep learning models are being developed to enhance the diagnostic accuracy of endoscopic procedures. Challenges in this domain, such as blurred tumor boundaries and high shape variability in cystoscopy images, demand more sophisticated network architectures. To address this, Zhao *et al.*[35] introduced NAFF-Net, a framework featuring a nested attentional feature fusion mechanism and a weighted pyramid pooling module. This design effectively refined edge representations, leading to state-of-the-art performance in bladder tumor segmentation with a mean IoU of 84.05%. In a complementary study, Ye *et al.*[36] utilized the High-Resolution Network Version 2 (HRNetv2) model for detecting bladder lesions in white-light cystoscopy videos. Their work highlighted the model's high sensitivity (91.6%) and precision (91.3%) while also revealing the significant impact of image resolution on performance, pointing to the necessity of high-quality imaging for optimal automated lesion detection.

The adaptability of deep learning in urological imaging extends to adjacent and intricate structures beyond the bladder itself. Takagi *et al.*[37] showcased this by developing 2D and 3D U-Net models to accurately localize the prostatic urethra on planning CT scans, a historically difficult task. By achieving an average centerline distance error of approximately 2 mm, their work demonstrated that these models can reliably identify small structures when trained on large datasets. In another advanced application, Mei *et al.*[38] employed the nnU-Net v2 architecture to segment both the gross tumor volume and neurovascular bundles from prostate biparametric MRI scans. This detailed segmentation was subsequently used to guide surgical decision-making in robot-assisted radical prostatectomy, directly linking automated image analysis to improved functional outcomes for patients.

A further evolution in the use of image segmentation involves leveraging these delineations for predictive analytics through radiomics. Moving beyond anatomical outlining, Wang *et al.*[39] pioneered an unsupervised clustering-based radiomics model to noninvasively predict clinical outcomes from preoperative CT images. By segmenting the tumor and then subdividing it into distinct intertumoral subregions, they extracted a rich set of features. A fusion model integrating these subregional features demonstrated remarkable performance, achieving area under the receiver operating curve scores of 0.884 for predicting muscle invasion and 0.832 for American Joint Committee on Cancer staging. This approach of analyzing intertumoral heterogeneity presents a powerful paradigm for preoperative risk stratification and the personalization of treatment strategies in BCa.

The scope of deep learning in BCa imaging has also expanded from anatomical delineation to sophisticated predictive modeling. Researchers are now developing models that integrate various data types to forecast critical clinical outcomes noninvasively. For example, Sun *et al.*[40] and Wei *et al.*[41] engineered nomograms that combine deep learning features with handcrafted radiomics from CT scans. These models demonstrated excellent performance in preoperatively predicting lymph node metastasis and muscle invasion, respectively, providing valuable tools for clinical staging. The exploration of novel architectures is also a key research avenue, with Kurata *et al.*[42] showing that a vision transformer-based model significantly outperformed traditional CNNs in diagnosing muscle-invasive BCa from MRI data. Beyond clinical outcomes, deep learning is being used to predict molecular markers directly from imaging. Han *et al.*[43] successfully constructed an interpretable model to predict programmed death-ligand 1 expression from CT scans, even employing Shapley Additive Explanations technology to make the model's decision-making process transparent. Further refinement of the segmentation task continues, with studies like Gumus *et al.*[44] comparing various model backbones, such as MAnet and PSPnet, to show that specific combinations of architectures and loss functions can yield superior tumor delineation on multiparametric MRI.

In a parallel and highly significant research direction, deep learning is being applied to the domain of digital pathology, analyzing whole-slide images to extract

prognostic and molecular information directly from tissue histology. This field, often termed "pathomics," has shown great promise for personalizing postoperative management. For example, both Jiang *et al.*[45] and Wang *et al.*[46] developed powerful models capable of predicting the probability of early recurrence in patients with non-muscle-invasive BCa by analyzing routine hematoxylin and eosin (H&E)-stained slides. Jiang *et al.*[45] further enhanced their predictive model for therapy response by uniquely integrating immunohistochemistry-stained slides alongside H&E data. A particularly groundbreaking application is the prediction of molecular status from standard, low-cost H&E images, thus bypassing the need for specialized and more expensive staining techniques. Jiao *et al.*[47] pioneered this by creating a model to predict human epidermal growth factor receptor 2 expression status with high accuracy. Similarly, Zheng *et al.*[48] developed a model that identifies basal and luminal molecular subtypes from H&E slides, outperforming even senior pathologists in a comparative evaluation. These advancements are creating a powerful link between histology and molecular profiling, offering rapid and economical solutions for guiding targeted treatment decisions.

2. Materials and methods

2.1. Segmentation models

2.1.1. U-Net

The U-Net architecture is a popular starting point for those aiming to achieve high performance in biomedical image segmentation, especially when working with a small number of annotated images. U-Net is based on an encoder–decoder network structure and is symmetrical in design, similar to SegNet, but with fewer convolutional layers. Its architecture is defined by skip connections, which combine the feature maps from the encoder path with the input feature maps of each corresponding layer in the decoder path. This merging is accomplished by copying the output feature map from an encoder layer and concatenating it with the input feature map of the corresponding decoder layer. The intent is to merge high-resolution, local spatial information within a single layer with high-level semantic information from the global object, which is important for biomedical image segmentation.[49]

During model training, a composite loss function was selected, consisting of a combination of binary cross-entropy (BCE) and Dice coefficient loss, which are commonly employed in U-Net and its variants. This combined function has been shown to enhance segmentation quality while optimizing pixel-level accuracy. The U-Net architecture, whose schematic structure is illustrated in Figure 1, was used as the baseline model in our study and trained using the composite loss function expressed in Equation I.

$$L = -\frac{1}{N}\sum_{k=1}^{N}\left[\left(y_k \log(t_k) + (1-y_k)\log(1-t_k)\right) + \left(\frac{2y_k t_k}{y_k + t_k}\right)\right] \quad \text{(I)}$$

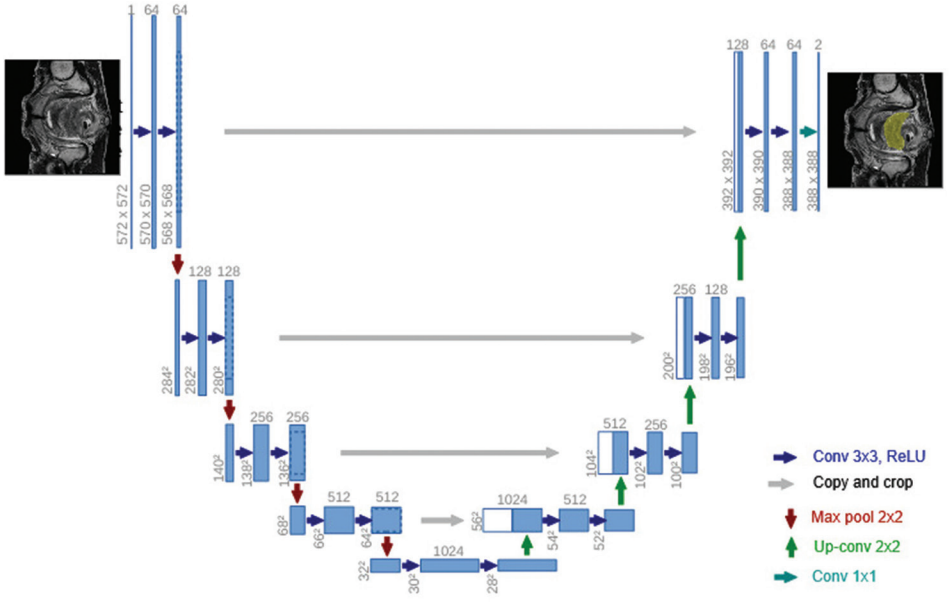

Figure 1. Schematic diagram of the U-Net architecture
Abbreviations: Conv: Convolution; Max: Maximum; ReLU: Rectified linear unit.

2.1.2. CMUNet

The CMUNet model was developed to build upon the success of the U-Net and its derivatives in medical image segmentation by addressing two of their fundamental limitations. First, the inherently local receptive field of standard convolution operations in U-Net architectures impedes the effective extraction of global context information, which is critical for accurately localizing lesions. Second, the simple skip connections between the encoder and decoder paths constrain the model's performance by indiscriminately transferring both relevant and irrelevant features. CMUNet introduces two main contributions to address these issues. First, at the bottom level of the encoder path (the bottleneck), a ConvMixer module replaces the standard block, as depicted in Equations II and III.

$$f_1^{'} = BN(\sigma_1\{DepthwiseConv(f_1 - 1)\}) + f_1 - 1 \quad \text{(II)}$$

$$f_1 = BN(\sigma_1\{PointwiseConv(f_1^{'})\}) \quad \text{(III)}$$

The ConvMixer layer performs depth-wise convolution to process the spatial information of each feature map, followed by pointwise convolution to mix cross-channel information. This design enables the network to achieve a global receptive field by integrating features from spatially distant locations. The second contribution lies in the use of a multi-scale attention gate (AG) within the skip connections. This gate processes the encoder's feature map using three parallel convolutional layers with different receptive fields (pointwise, ordinary, and dilated). The outputs of these three scales are combined to form an attention map, which measures and filters the most valuable information for the decoder. In this way, task-relevant features are enhanced and irrelevant information, such as background noise, is filtered, resulting in a more sophisticated and efficient information transfer. The complete architecture of the CMUNet model is illustrated in Figure 2.[50]

2.1.3. Attention U-Net

The Attention U-Net is designed to improve the performance of medical image segmentation by enabling the model to automatically learn and focus on target structures of varying shapes and sizes. This architecture was introduced to address a key limitation of the standard U-Net and other fully convolutional networks, which often require multi-stage cascaded models to first localize a region of interest before performing detailed segmentation. Such cascaded frameworks are computationally expensive and involve redundant feature extraction. The Attention U-Net eliminates this necessity by integrating novel AGs directly into the standard U-Net architecture. These AGs are placed within the skip connections, where they filter the features passed from the encoder path to the decoder path. The core mechanism involves using a gating signal (g), taken from a coarser scale (i.e., a deeper layer of the network), to provide contextual information. This gating signal modulates the feature map from the corresponding encoder layer (x^l), generating attention coefficients (α) that identify salient image regions and prune feature responses, preserving only the activations relevant to the specific segmentation task. The model employs an additive attention mechanism to compute these gating coefficients, which has been shown to be more accurate than multiplicative attention. The fundamental steps of this mechanism, detailing the computation of the attention query and final attention coefficient, are presented in Equations IV and V, respectively.[51]

$$q_{att}^l = \psi\left(\sigma_1(W_\chi^T \chi_i^l + W_g^T g_i + b_g)\right) + b_\psi \quad \text{(IV)}$$

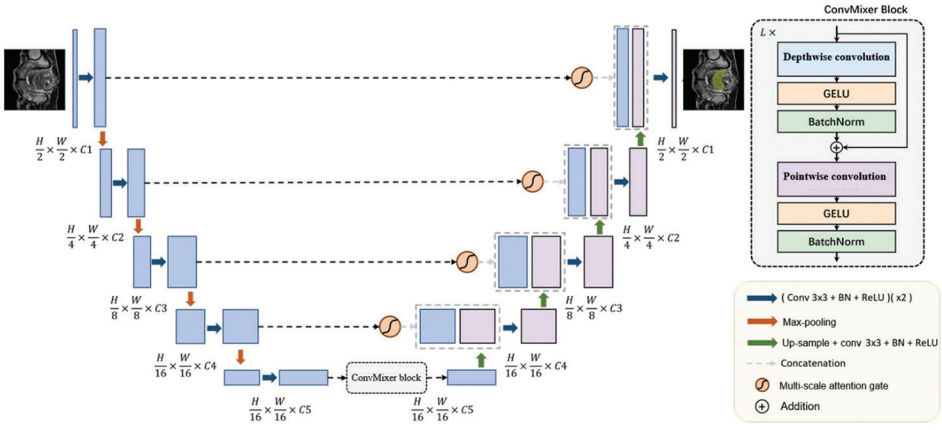

Figure 2. Schematic diagram of the cross-scale mixer U-Net architecture
Abbreviations: BN: BatchNorm; Conv: Convolution; GELU: Gaussian error linear unit; Max: Maximum; ReLU: Rectified linear unit.

$$\alpha_i^l = \sigma_2\left(q_{att}^l(\chi_i^l, g_i; \Theta_{att})\right) \quad \text{(V)}$$

Here, the attention coefficient α is derived from linear transformations (W_x, W_g, ψ) of the encoder features and the gating signal, followed by sigmoid (σ_2) and rectified linear unit (σ_1) activation functions. The output of the AG is the element-wise multiplication of the input feature map and these attention coefficients. By progressively suppressing feature responses in irrelevant background regions, this mechanism allows the model to learn "where to look" without external supervision, thereby increasing model sensitivity and prediction accuracy with minimal computational overhead. The Attention U-Net architecture, which integrates AGs into its skip connections, is detailed in Figure 3.

2.1.4. CMUNeXt

The structure of the CMUNeXt architecture, a lightweight and fully convolutional network designed to balance segmentation performance with computational efficiency, is shown in Figure 4. This architecture addresses a key challenge in medical image segmentation: capturing the global context needed for precise lesion localization, without incurring the high computational costs associated with large transformer-based networks. CMUNeXt introduces two fundamental innovations. The first is the CMUNeXt block, used in the encoder layers, which employs a large-kernel depth-wise convolution to integrate distant spatial information. This is followed by two pointwise convolutions incorporating an inverted bottleneck design, as illustrated in Equations VI–VIII.

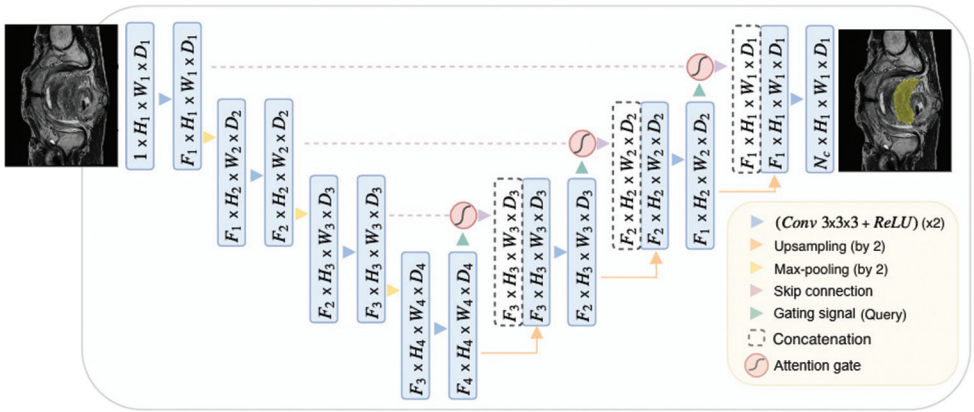

Figure 3. Schematic structure of the attention U-Net architecture
Abbreviations: Conv: Convolution; Max: Maximum; ReLU: Rectified linear unit.

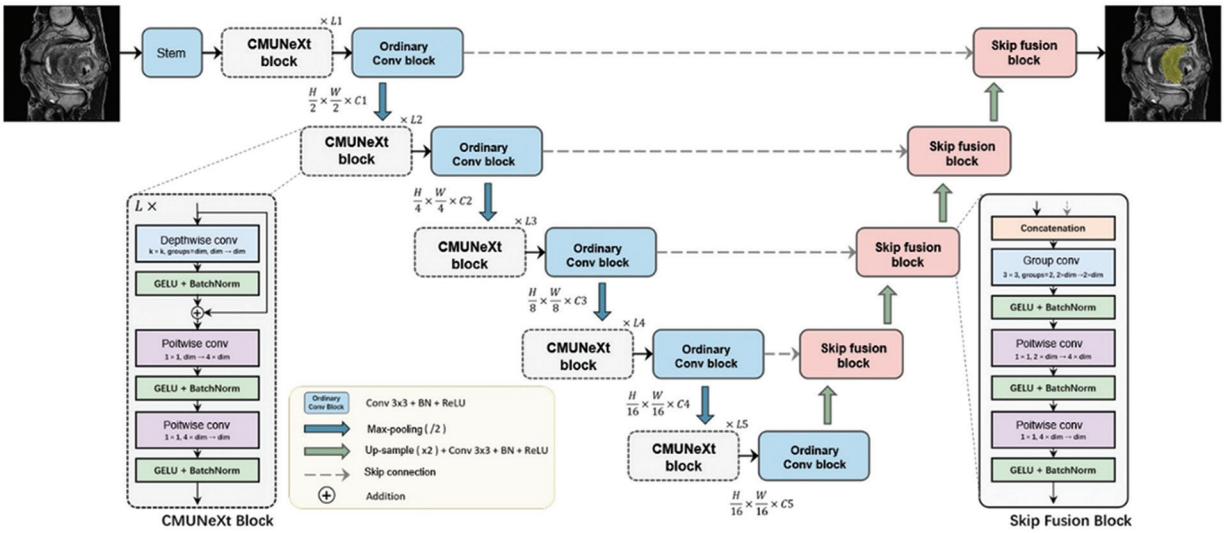

Figure 4. CMUNeXt: Architecture of an efficient segmentation network based on large kernels and skip-fusion
Abbreviations: Conv: Convolution; GELU: Gaussian error linear unit; Max: Maximum; ReLU: Rectified linear unit.

$$f_1^{'} = BN(\sigma_1\{DepthwiseConv(f_1 - 1)\}) + f_1 - 1 \quad \text{(VI)}$$

$$f_1^{''} = BN(\sigma_1\{PointwiseConv(f_1^{'})\}) \quad \text{(VII)}$$

$$f_1 = BN(\sigma_1\{PointwiseConv(f_1^{''})\}) \quad \text{(VIII)}$$

The second innovation is the skip-fusion block, developed to enable more seamless feature fusion in the skip connections. Rather than a simple concatenation, this block first applies group convolution to features from the encoder and decoder paths, followed by pointwise convolutions to perform dense and efficient fusion.[52]

2.2. Dataset

This study utilized the FedBCa dataset, a publicly available, multi-center MRI collection curated specifically for BCa research.[53,54] The dataset was compiled retrospectively under a waiver of informed consent granted by the ethics committees of four participating medical centers in China. It comprises 275 3D T2W MRI scans sourced from a final cohort of 228 patients (198 male, 30 females; mean age: 65.4 ± 11.2 years). Imaging was performed using scanners from multiple vendors (including Siemens, GE, and Philips) at 1.5T and 3.0T field strengths, reflecting real-world clinical heterogeneity. The dataset includes both non-muscle-invasive and muscle-invasive BCa cases, providing a representative distribution for clinical evaluation.

To ensure the highest annotation quality, all tumor labels were initially created on axial T2W sequences by a radiologist with 14 years of experience. These annotations were subsequently reviewed and, if necessary, refined by a second radiologist of equal experience. A consensus protocol was employed to resolve discrepancies through discussion. The final inter-rater reliability was quantitatively validated with a DSC of 0.870 and an intraclass correlation coefficient of 0.988, indicating a high degree of agreement between annotators. For ease of use and to protect patient privacy, all images were fully anonymized and converted from their original DICOM format to the Nifti (.nii.gz) standard. The data were meticulously organized by center, preserving the distinct data distributions of each institution, thereby providing a realistic testbed for federated learning, domain adaptation, and generalization studies. Figure 5 displays representative T2W imaging examples from the four centers included in the dataset.

2.3. Experimental setup

To ensure a fair and reproducible comparison, all experiments were conducted on a single, consistent computational platform. This approach eliminates hardware and software-related variables, ensuring that performance differences can be attributed solely to architectural design. The technical specifications were as follows: Ubuntu 24.04 operating system, an NVIDIA RTX 4090 graphics processing unit (GPU), an Intel Core i9-14900K central processing unit, and 64 GB of DDR5 memory. All models were developed and evaluated using the PyTorch framework with CUDA 12.8 for GPU acceleration. To support reproducibility, the code used for training, evaluation, along with the trained model weights, is available from the corresponding author upon reasonable request.

2.4. Performance metrics

To quantitatively evaluate the performance of the segmentation models, a pixel-by-pixel comparison was performed between the model's predicted segmentation map and the ground truth map. This comparison yielded four fundamental outcomes: a true positive (TP) refers to a pixel correctly identified by the model as part of the lesion; a false positive (FP) denotes a pixel erroneously identified as part of the lesion but actually belonging to the background; a true negative is a pixel correctly classified as background; and a false negative (FN) represents a pixel that belongs to the lesion but was missed by the model and classified as background. Based on these outcomes, the performance metrics were calculated.

IoU, also known as the Jaccard index, is one of the most widely used evaluation metrics in segmentation tasks. It measures the ratio of overlap between the predicted segmentation and the ground truth. Mathematically, it is calculated by dividing the area of intersection between the prediction and the ground truth by the area of their union. Its value ranges from 0 to 1, with values closer to 1 indicating more accurate predictions. IoU provides a balanced measure of segmentation accuracy, as both FP and FN errors increase the union area while leaving the intersection area unchanged. The mathematical formulation of this metric, based on TP, FP, and FN, is presented in Equation IX.

$$IoU = \frac{TP}{TP + FP + FN} \quad \text{(IX)}$$

The Dice coefficient, closely related to the F1-score, is another prevalent similarity metric, particularly in medical image segmentation. Like IoU, it quantifies the overlap between the predicted and ground truth masks. It is calculated as twice the area of intersection divided by the sum of pixels in both masks. The Dice coefficient also ranges from 0 to 1, with values approaching 1 indicating high similarity and successful segmentation. Although mathematically related to IoU, Dice scores are typically

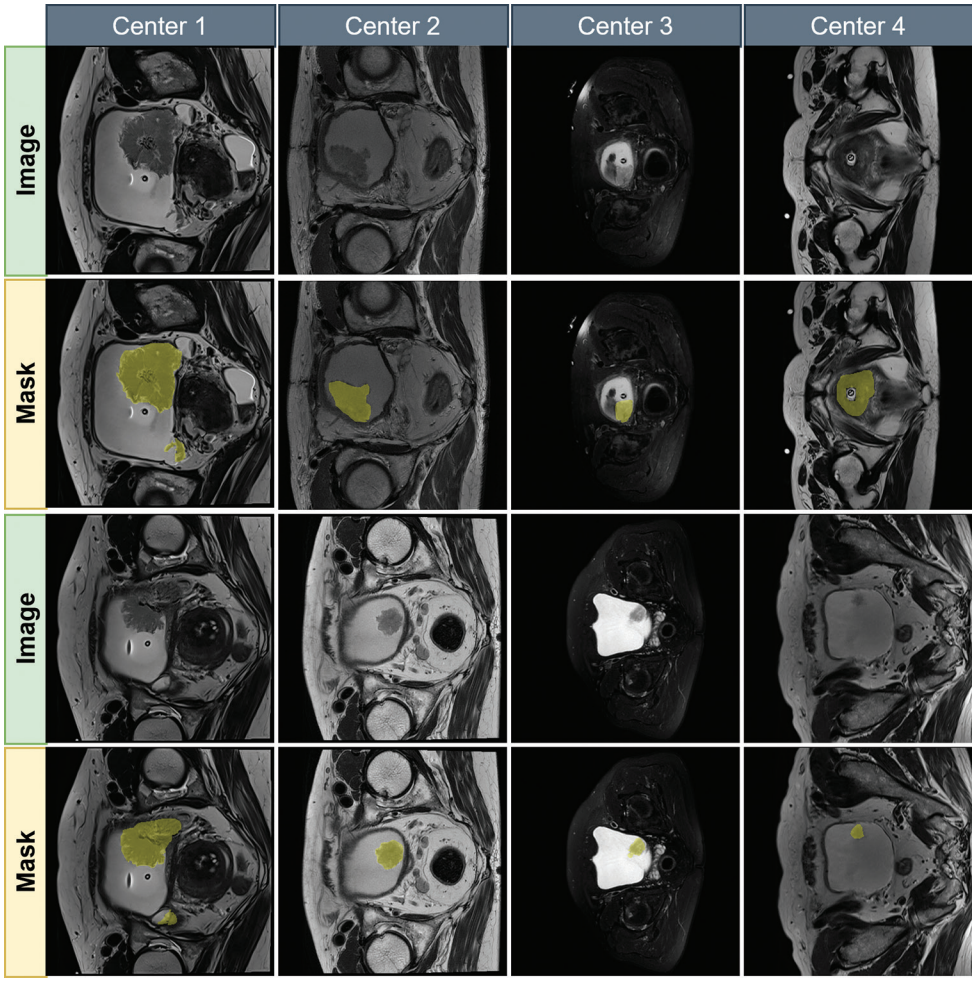

Figure 5. Representative magnetic resonance images and corresponding tumor masks from the multi-center bladder cancer dataset

slightly higher for the same prediction. The Dice coefficient, which measures the similarity between the predicted and ground truth masks, is calculated as shown in Equation X.

$$Dice = \frac{2 \times TP}{2 \times TP + FP + FN} \quad \text{(X)}$$

Sensitivity, alternatively referred to as recall or TP rate, quantifies the model's ability to correctly identify actual positive cases (i.e., pixels belonging to a lesion). It represents the proportion of true lesion pixels correctly labeled by the model. Sensitivity is particularly important in medical diagnosis, where missing a lesion (i.e., a high FN rate) could have critical consequences. High sensitivity indicates a low probability of missing existing lesions, which is vital for early diagnosis and treatment. The formula for sensitivity is provided in Equation XI.

$$Recall\ (sensitivity) = \frac{TP}{TP + FN} \quad \text{(XI)}$$

Precision, also known as positive predictive value, quantifies the proportion of positive predictions made by the model that are actually correct. In other words, it represents the fraction of pixels identified as "lesion" that truly belong to the lesion class. A high precision score indicates a low FP rate, which is especially important in clinical applications to prevent misclassifying healthy tissue as abnormal, thereby reducing unnecessary biopsies or treatments. Precision is calculated using the formula in Equation XII.

$$Precision = \frac{TP}{TP + FP} \quad \text{(XII)}$$

2.5. Data preprocessing, partitioning, and augmentation

All 3D T2W MRI volumes were processed on a 2D slice-by-slice basis. Before training, several preprocessing steps were applied. First, image intensities were normalized

using Z-score normalization (subtracting the mean and dividing by the standard deviation of non-zero voxels) to standardize intensity ranges across scans. Subsequently, all slices and their corresponding masks were resampled to a uniform resolution of 256×256 pixels, ensuring consistency in model input dimensions.

To prevent data leakage and ensure robust model evaluation, the dataset was split at the patient level. The 228 patients were randomly divided into training (70%; 160 patients), validation (15%; 34 patients), and testing (15%; 34 patients) sets. Patient-level partitioning ensures that no slices from the same patient appear in more than one set. A fixed random seed was used during the split to enable reproducibility.

To enhance model robustness and generalizability, a comprehensive online data augmentation strategy was applied exclusively to the training set. This strategy incorporated diverse transformations, including random rotations (±15°), elastic deformations, horizontal flipping, random scaling (0.8–1.2×), random cropping (to 224×224 pixels), Gaussian blur (σ: 0–1.5), Gaussian noise (mean: 0, standard deviation: 0.02), and random adjustments to brightness, contrast, and gamma.

2.6. Training protocol and model selection

All network architectures were trained under a standardized protocol to ensure a fair and reproducible comparative analysis. For the optimization process, the AdamW algorithm was selected for its proven efficacy. Training hyperparameters were determined based on established best practices and empirical testing: a learning rate of 1×10^{-5}, a weight decay factor of 2×10^{-5}, and a momentum of 0.9 were consistently applied across all models.

To address the inherent class imbalance in segmentation and optimize for spatial overlap, a composite loss function was employed, defined as the sum of BCE loss and Dice loss (L = L_BCE + L_Dice). This hybrid loss function provides a balanced optimization target, penalizing both pixel-level classification errors and inaccuracies in boundary delineation.

Models were trained for a maximum of 300 epochs. To prevent overfitting and ensure the selection of the best-performing model state, an early stopping mechanism was implemented. The training process was automatically halted if the DSC on the validation set did not improve for 20 consecutive epochs. For the final evaluation on the independent test set, the model weights corresponding to the epoch that achieved the highest validation DSC were selected. This rigorous procedure ensures that each model is evaluated at its optimal performance point.

3. Experimental results

This section presents a comprehensive quantitative evaluation of four distinct deep learning architectures—U-Net, CMUNet, AttU_Net, and CMUNeXt—for the task of BCa segmentation. Model performance was assessed using a wide array of metrics to evaluate segmentation accuracy, lesion detection capability, boundary delineation precision, and computational efficiency. The consolidated results are shown in Table 1.

The results summarized in Table 1 reveal that the CMUNet architecture delivers superior performance in segmentation accuracy for BCa. It achieved the highest scores in key overlap-based metrics, with an IoU of 0.7033 and a Dice coefficient of 0.7937. This accuracy extends to the more clinically relevant task of lesion delineation, where CMUNet achieved an outstanding lesion F1-Score of 0.9521, indicating exceptional reliability in both identifying and precisely outlining tumor regions. Furthermore, its leading performance in distance-based metrics, such as Hausdorff distance (8.4550 mm) and average symmetric surface distance (2.4066 mm), confirms that the segmentation boundaries it produces are significantly more congruent with the ground truth, which is critical for clinical applications, including treatment planning and volumetric assessment.

Table 1. Comparative performance evaluation of different models for bladder cancer segmentation

Parameter	U_Net	CMUNet	AttU_Net	CMUNeXt
IoU	0.6708	0.7033	0.6603	0.6554
Dice	0.7637	0.7937	0.7482	0.7547
Sensitivity	0.7776	0.8130	0.7488	0.8044
Precision	0.8175	0.8250	0.8380	0.7768
F1-score	0.7637	0.7937	0.7482	0.7547
Accuracy	0.9967	0.9972	0.9968	0.9963
Specificity	0.9985	0.9986	0.9987	0.9981
Lesion sensitivity	0.9411	0.9579	0.9358	0.9656
Lesion F1-score	0.8839	0.9521	0.9031	0.9192
Hausdorff distance (mm)	12.1977	8.4550	11.5668	12.5334
HD95	9.0538	6.4048	8.7746	9.3209
ASSD (mm)	3.2723	2.4066	3.2021	3.3388
Total inference time (s)	0.91	1.35	1.00	0.50
Avg. time per image (ms)	3.76	5.54	4.12	2.07
FPS	265.85	180.47	242.99	482.55

Abbreviations: ASSD: Average symmetric surface distance; AttU_Net: Attention U-Net; Avg.: Average; CMUNet: Cross-scale Mixer U-Net; FPS: Frames per second; HD95: Hausdorff 95% percentile index; IoU: Intersection over union.

A nuanced comparison across the models highlighted distinct performance profiles. While CMUNet demonstrated the best overall balance, other models showed specific strengths. For example, Attention U-Net recorded the highest precision (0.8380), although at the cost of lower sensitivity, while CMUNeXt achieved the maximum lesion sensitivity (0.9656). However, CMUNeXt's lower lesion F1-score suggests that its high detection rate may be accompanied by less precise delineations or a higher number of frames per second (FPs). It is also noteworthy that while all models displayed exceptionally high pixel-wise accuracy and specificity, these metrics are less discriminative in this context due to the inherent class imbalance in medical images.

A clear trade-off between segmentation accuracy and computational efficiency was also observed. CMUNeXt distinguished itself as the most efficient architecture, capable of processing 482.55 FPS, making it a strong candidate for real-time or high-throughput applications. In contrast, CMUNet, the most accurate model, was also the most computationally intensive, operating at 180.47 FPS. Therefore, the optimal model choice is application-dependent: for diagnostic scenarios where maximal precision and boundary fidelity are paramount, CMUNet is the superior option. Conversely, where processing speed is the primary constraint, CMUNeXt offers a compelling alternative.

To qualitatively evaluate the segmentation performance of the models in addition to the quantitative results, the predictions of all architectures on a representative test image were visually compared. All models were evaluated with the same test image to ensure a consistent and fair comparison. The resulting prediction masks (green) are presented side-by-side with the ground truth mask (red). As shown in Figure 6, CMUNet produced the closest match to the expert mask, particularly along the lumen-facing border and the irregular basal interface, yielding a wide overlap region while preserving fine concavities. Attention U-Net remained conservative, tending to prune the ambiguous penumbral signal, which reduced small FPs but slightly under-segments the superior edge. CMUNeXt broadly recovered the lesion extent and was sensitive to faint tumor signal, yet it occasionally extended into adjacent mucosal folds, indicating a higher propensity for small FPs near textural transitions. The baseline U-Net tracked the main tumor core but smoothed thin protrusions and small satellite foci. These visual patterns reflect the quantitative results, where CMUNet leads in overlap and boundary quality, Attention U-Net shows stronger precision, and CMUNeXt demonstrates high lesion sensitivity.

To further assess model robustness and identify potential failure modes, a particularly challenging test case was selected for comparative visual analysis. This case presents a significant challenge for automated segmentation algorithms due to the low tissue contrast between the tumor and surrounding tissue, as well as the lesion's poorly defined boundaries. As presented in Figure 7, even the top-performing model, CMUNet, exhibited certain inaccuracies in its prediction, struggling to perfectly delineate the tumor's full extent. A side-by-side comparison revealed that all models faced difficulties with this specific example, each displaying varying degrees of under-segmentation or boundary inaccuracies. This qualitative analysis of a challenging case is crucial as it highlights the limitations of current models and the need for future improvements in handling images with low signal-to-noise ratios and ambiguous anatomical features.

As illustrated in Figure 7, this example exhibited low tissue contrast and poorly defined margins, with subtle intensity gradients between the tumor and the surrounding bladder wall. Even in this challenging context, CMUNet preserved the bulk of the lesion and captured heterogeneous interior signal, though it under-represented thin spicules at the periphery. Attention U-Net produced a compact and clean contour by suppressing equivocal regions, which limited over-segmentation, but omitted shallow extensions at the posterior wall. CMUNeXt detected more of the faint peripheral signal and highlighted potential submucosal spread, yet this came with spillover into edema-like areas and partial-volume zones. The baseline U-Net emphasized contiguous high-contrast areas and missed fragmented edges. Together, these failure modes illustrate known trade-offs among sensitivity, precision, and boundary fidelity in low-signal, low-contrast conditions, aligning with the model-specific strengths reported in the quantitative results.

4. Discussion

The quantitative results provide deep insights into the architectural strengths and weaknesses of the evaluated models. The clear superiority of CMUNet in key accuracy and boundary metrics (IoU, Dice, Hausdorff distance, and average symmetric surface distance) suggests that its architecture achieves a more effective synthesis of local and global features essential for BCa segmentation. Its ability to maintain a robust balance between high sensitivity and high precision establishes it as a reliable, general-purpose model for this clinical task. While surpassed by CMUNet, the foundational U-Net provided a solid performance baseline, validating that the architectural enhancements in models like CMUNet offer tangible improvements.

This study's most critical finding, however, is the clear demonstration that the optimal model choice is not

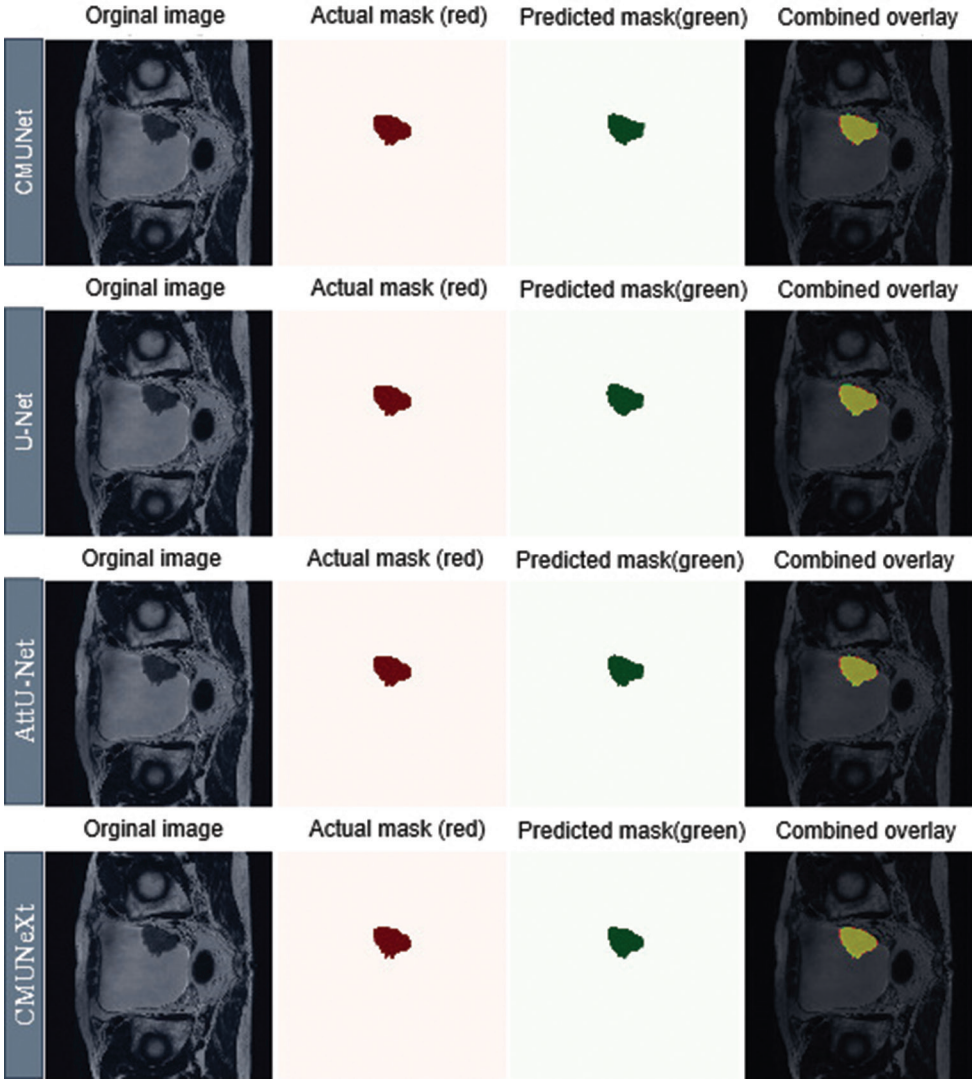

Figure 6. Qualitative comparison of U-Net-based segmentation models on randomly selected bladder-cancer lesions
Abbreviations: AttU_Net: Attention U-Net; CMUNet: Cross-scale mixer U-Net.

universal, but rather dictated by specific clinical utility and the evidentiary thresholds for a given task. This principle mirrors the established framework in biomarker development, where clinical value is determined not by a single performance metric but by the alignment of sensitivity–specificity profiles with specific diagnostic needs and the rigorous validation required for guideline adoption. For example, the CMUNeXt model, with its specialist high-sensitivity profile, is analogous to a screening biomarker, making it ideally suited for scenarios where minimizing FNs is the primary imperative, such as in pre-cystoscopy triage and surveillance imaging. Conversely, the Attention U-Net model functions as a high-precision specialist, akin to a confirmatory test. Its conservative yet accurate delineations are invaluable in workflows where avoiding FPs is paramount, such as in radiotherapy planning to minimize radiation toxicity and in surgical planning to ensure complete tumor capture. This strategic alignment of model performance with distinct clinical priorities represents a paradigm shift from seeking a single "best" model to developing a validated toolkit of specialized, application-specific solutions ready for prospective clinical integration.

An essential aspect of assessing the clinical viability of automated models is benchmarking their performance against human experts' consistency. The inter-rater reliability for manual segmentations within our dataset was established at a high standard, with a DSC of 0.870. In comparison, our top-performing architecture, CMUNet,

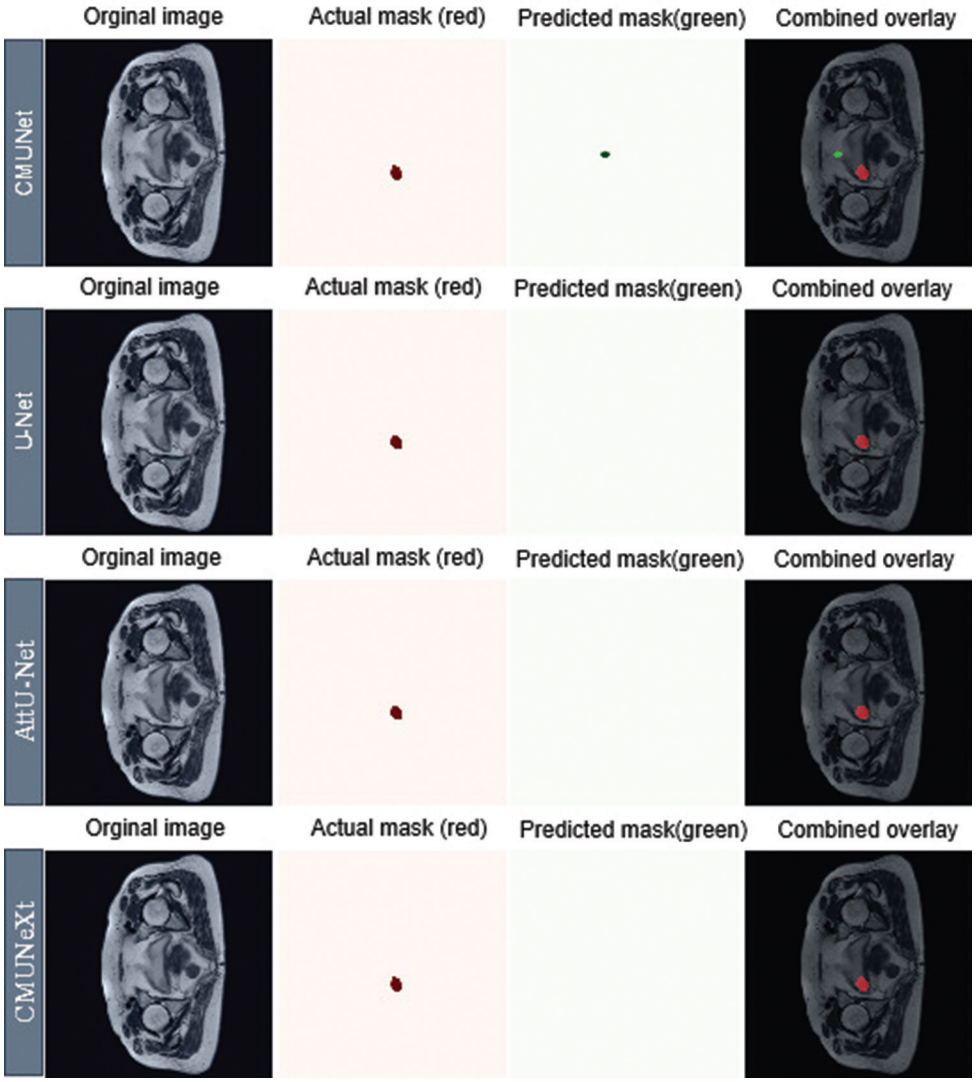

Figure 7. Visual evaluation of UNet-based segmentation models on the most challenging bladder-cancer lesions
Abbreviations: AttU_Net: Attention U-Net; CMUNet: Cross-scale mixer U-Net.

achieved a DSC of 0.7937. This result indicates that while the automated model demonstrates robust and highly accurate performance, it does not yet surpass the level of agreement observed between two experienced radiologists. This finding positions the model as a powerful assistive tool, capable of generating reliable initial segmentations to accelerate clinical workflows and reduce inter-observer variability. However, it also underscores the continued importance of expert review and correction, particularly for challenging cases with ambiguous tumor boundaries, to ensure the highest diagnostic accuracy.

4.1. Generalizability and robustness

We assessed model generalizability and robustness using stratified internal validation within the available dataset. Training, model selection, and performance estimation were performed without sharing patients across splits, thereby reducing the risk of leakage. The resulting estimates, therefore, reflect internal validity only (test data) and should not be interpreted as evidence of performance across different institutions, vendors, or acquisition protocols.

Generalizability can be affected by distribution shifts originating from scanner hardware, pulse sequences, pulse reconstruction settings, local clinical workflows, and population characteristics. Given that an independent external cohort was not accessible in the present study, we were unable to test the model using data acquired at other sites or using different vendor platforms. Likewise, we did not run formal stress tests for common perturbations, such as motion, susceptibility artifacts, post-Bacillus Calmette–Guérin (BCG) vaccine inflammation, hematuria, or partial

volume effects. As a result, performance under these conditions remains unknown.

From a deployment perspective, decision thresholds, calibration, and failure modes often change with a distribution shift. In the current study, operating points were selected using internal validation data. These operating points may require re-tuning when the model is applied in new environments. Therefore, we view the current results as a necessary first step that establishes internal reproducibility while recognizing that clinical utility depends on performance under external and potentially shifted conditions.

To strengthen generalizability in future work, we plan a staged validation program. First, we will perform external testing on a multi-center cohort that spans distinct institutions, vendors, and protocols. Second, we will conduct leave-one-site-out experiments to quantify site-level sensitivity. Third, we will include robustness evaluations that approximate clinical perturbations through controlled experiments, for example, simulated motion, susceptibility-related distortions, post-BCG inflammatory changes, hematuria-related signal, and partial volume scenarios. When full external data are not immediately available, we will complement these efforts with informative internal surrogates, including domain shift experiments, test-time augmentation, and uncertainty estimation using techniques such as Monte Carlo dropout. Together, these steps will provide a clearer picture of the model's behavior under realistic clinical variation.

4.2. Limitations and future directions

This study has several limitations that define the scope of our conclusions. First, we did not perform external validation on an independent cohort, nor did we conduct leave-one-site-out analyses; therefore, generalizability to other institutions, vendor platforms, and protocols remains unknown. Second, we did not carry out robustness and stress tests against realistic perturbations, including motion, susceptibility, post-BCG inflammation, hematuria, and partial volume effects, which limits our ability to characterize failure modes and threshold stability. Third, we reported point estimates only and did not compute confidence intervals, run paired significance tests across models on the same cases, or apply corrections for multiple comparisons. Finally, operating points and calibration were derived from internal data and may require transfer calibration; in addition, variability in annotations and local workflows could introduce residual bias.

Future work will address these limitations along three axes. For validation, we will assemble external cohorts that diversify sites, vendors, and protocols, and we will include leave-one-site-out validation to quantify sensitivity to site-level differences. For robustness, we will design targeted stress tests that emulate clinically relevant perturbations, and include internal surrogates when external data are not immediately available—such as controlled domain shift experiments, test time augmentation, and uncertainty estimation using techniques like Monte Carlo dropout. For statistical rigor and decision analysis, we will report patient-level, stratified bootstrap 95% confidence intervals for all metrics; conduct paired non-parametric tests on per-case metrics with Holm or Benjamin–Hochberg adjustments for multiple comparisons; and include decision-analytic summaries (e.g., cost-sensitive analyses and decision curve analysis) that map sensitivity and precision to plausible clinical utilities for application-specific thresholds.

Furthermore, this study was confined to image-based analysis. The clinical management of BCa increasingly adopts a multifaceted approach that integrates diverse biomarkers. Urinary biomarkers provide a noninvasive avenue for detection and surveillance and serve as a critical complement to radiological imaging.[55] Future research should therefore develop multimodal frameworks that integrate features extracted from deep learning-based image analysis with urinary biomarker data, pathology reports, and other clinical information. Such holistic models could yield more powerful and comprehensive decision support, improving both diagnosis and patient stratification.

5. Conclusion

This study conducted a descriptive comparative analysis of four U-Net architectures for the automated segmentation of BCa lesions on MRI images. Within this dataset, CMUNet achieved the highest point estimates for IoU (0.7033) and Dice (0.7937), suggesting competitive delineation quality relative to the alternatives. These findings support the view that deep learning can help address limitations of manual segmentation by providing more objective and efficient outputs, though we note that our evidence reflects internal validation only. A key insight is that the optimal architecture choice appears to be application-dependent rather than universal. In our experiments, CMUNeXt displayed a high-sensitivity profile that may be advantageous for triage and screening settings where minimizing FNs is critical. In contrast, Attention U-Net exhibited higher precision, a property better suited to treatment planning and response assessment, where boundary accuracy and the avoidance of FPs are paramount. Future work will prioritize external validation across diverse sites, vendors, and protocols, coupled with decision-oriented analyses that relate sensitivity and precision to plausible clinical utilities. These steps aim to translate the models into robust decision support systems that can enhance the management of BCa.

Acknowledgments

This work was supported by the Grant provided by TÜSEB under the "2023-C1-YZ" call and Project No: "33934." We would like to thank TÜSEB for their financial support and scientific contributions. Experimental computations were carried out on the computing units at Igdir University's Artificial Intelligence and Big Data Application and Research Center.

Funding

None.

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: Ishak Pacal
Formal analysis: All authors
Investigation: Ishak Pacal
Methodology: All authors
Writing–original draft: All authors
Writing–review & editing: All authors

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

The dataset used in this study is publicly available on Zenodo: https://zenodo.org/records/13622759.

References

1. Bashkami A, Nasayreh A, Makhadmeh SN, *et al*. A review of artificial intelligence methods in bladder cancer: Segmentation, classification, and detection. *Artif Intell Rev*. 2024;57:339.

   doi: 10.1007/s10462-024-10953-6

2. Li L, Jiang L, Yang K, Luo B, Wang X. A novel artificial intelligence segmentation model for early diagnosis of bladder tumors. *Abdom Radiol*. 2024;50(7):3092-3099.

   doi: 10.1007/S00261-024-04715-9

3. Wang S, Wei Z, Shu H, *et al*. Early diagnosis and prognostic potential of RAC3 in bladder tumor. *Int Urol Nephrol*. 2024;56(2):475-482.

   doi: 10.1007/S11255-023-03781-0

4. Gercek O, Ulusoy K, Yazar VM, Topal K. Effects of delayed diagnosis on tumor size, stage and grade in bladder cancer. *Int Urol Nephrol*. 2024;56(3):935-940.

   doi: 10.1007/S11255-023-03829-1

5. Nasrallah AA, Williams SB. Epidemiology of bladder cancer. In: *Biology of Bladder Cancer*. Berlin: Springer; 2024. p. 3-22.

   doi: 10.1007/978-3-031-68505-7_1

6. Wang Y, Zhou H, Tang Z, *et al*. Smoking and bladder cancer: Insights into pathogenesis and public health implications from a bibliometric analysis (1999-2023). *Subst Abuse Treat Prev Policy*. 2025;20(1):25.

   doi: 10.1186/S13011-025-00658-6

7. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. *CA Cancer J Clin*. 2024;74:12-49.

   doi: 10.3322/caac.21820

8. 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*. 2024;15(1):185.

   doi: 10.1186/S13244-024-01780-Y

9. Gnyawali D, Sigdel PR, Rai BDK, Chapagain S, Gyawali PR. Storz professional image enhancement system (SPECTRA A) enhancing detection of carcinoma urinary bladder by white light cystoscopy. *Afr J Urol*. 2024;30(1):27.

   doi: 10.1186/S12301-024-00429-Y

10. Badawy MA, Zaky SM, Korkor MS, Fattah SA. Role of intravoxel incoherent motion MRI in diagnosis and assessment of invasiveness of urothelial bladder carcinoma. *Egyptian Journal of Radiology and Nuclear Medicine*. 2024;55(1):40.

    doi: 10.1186/S43055-024-01212-3

11. 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*. 2024;24(1):98.

    doi: 10.1186/S12880-024-01276-7

12. Arita Y, Kwee TC, Akin O, *et al*. Multiparametric MRI and artificial intelligence in predicting and monitoring treatment response in bladder cancer. *Insights Imaging*. 2025;16(1):7.

    doi: 10.1186/S13244-024-01884-5

13. Donswijk ML, Ettema RH, Meijer D, *et al*. The accuracy and intra- and interobserver variability of PSMA PET/CT for the local staging of primary prostate cancer. *Eur J Nucl Med Mol Imaging*. 2024;51(6):1741-1752.

    doi: 10.1007/S00259-024-06594-0

14. Ozdemir B, Sermet F, Pacal I. Attention-enhanced ConvNeXt for accurate, efficient, and interpretable crack detection. *Expert Syst Appl*. 2026;296:129165.

    doi: 10.1016/J.ESWA.2025.129165

15. Pacal I, Kilicarslan S, Ozdemir B, Deveci M, Kadry S. Efficient

and autonomous detection of olive leaf diseases using AI-enhanced MetaFormer. *Artif Intell Rev.* 2025;58:303.

doi: 10.1007/S10462-025-11131-Y

16. Cakmak Y. Machine learning approaches for enhanced diagnosis of hematological disorders. *Comput Syst Artif Intell*. 2025;1(1):8-14.

doi: 10.69882/ADBA.CSAI.2025072

17. Pacal I. Investigating deep learning approaches for cervical cancer diagnosis: A focus on modern image-based models. *Eur J Gynaecol Oncol*. 2025;46(1):125-141.

doi: 10.22514/ejgo.2025.012

18. Cakmak Y, Maman A. Deep learning for early diagnosis of lung cancer. *Comput Syst Artif Intell*. 2025;1(1):20-25.

doi: 10.69882/ADBA.CSAI.2025074

19. Cakmak Y, Pacal N. Deep learning for automated breast cancer detection in ultrasound: A comparative study of four CNN architectures. *Artif Intell Appl Sci*. 2025;1(1):13-19.

doi: 10.69882/ADBA.AI.2025073

20. Cakmak Y, Zeynalov J. A comparative analysis of convolutional neural network architectures for breast cancer classification from mammograms. *Artif Intell Appl Sci*. 2025;1(1):28-34.

doi: 10.69882/ADBA.AI.2025075

21. Pacal I, Attallah O. Hybrid deep learning model for automated colorectal cancer detection using local and global feature extraction. *Knowl Based Syst*. 2025;319:113625.

doi: 10.1016/J.KNOSYS.2025.113625

22. Pacal I, Attallah O. InceptionNeXt-Transformer: A novel multi-scale deep feature learning architecture for multimodal breast cancer diagnosis. *Biomed Signal Process Control*. 2025;110:108116.

doi: 10.1016/J.BSPC.2025.108116

23. Pacal I, Akhan O, Deveci RT, Deveci M. NeXtBrain: Combining local and global feature learning for brain tumor classification. *Brain Res*. 2025;1863:149762.

doi: 10.1016/J.BRAINRES.2025.149762

24. Aruk I, Pacal I, Toprak AN. A novel hybrid ConvNeXt-based approach for enhanced skin lesion classification. *Expert Syst Appl*. 2025;283:127721.

doi: 10.1016/J.ESWA.2025.127721

25. Ince S, Kunduracioglu I, Algarni A, Bayram B, Pacal I. Deep learning for cerebral vascular occlusion segmentation: A novel ConvNeXtV2 and GRN-integrated U-Net framework for diffusion-weighted imaging. *Neuroscience*. 2025;574:42-53.

doi: 10.1016/J.NEUROSCIENCE.2025.04.010

26. Dahbi H, Lazaar M, Hssayni EH. Deep learning based approach for medical image segmentation: A systematic literature review. In: *Lecture Notes in Networks and Systems*. Berlin: Springer; 2024. p. 1-16.

doi: 10.1007/978-3-031-74491-4_1

27. Kumar RR, Priyadarshi R. Denoising and segmentation in medical image analysis: A comprehensive review on machine learning and deep learning approaches. *Multimed Tools Appl*. 2024;84(12):10817-10875.

doi: 10.1007/S11042-024-19313-6

28. Adeel Abid M, Munir K. A systematic review on deep learning implementation in brain tumor segmentation, classification and prediction. In: *Multimedia Tools and Applications*. 2025. Berlin: Springer; 2025. p. 1-40.

doi: 10.1007/S11042-025-20706-4

29. Sherifa CA, Kanisha B. *A Review on Medical Image Segmentation using Deep Learning*. Berlin: Springer; 2024. p. 65-77.

doi: 10.1007/978-981-97-5791-6_6

30. Jain S, Pal A, Andrew J, Maheshwari S. Adaptive treatment planning via multi-class segmentation of GI tract tumours. In: *Lecture Notes in Networks and Systems*. Vol. 1122. Berlin: Springer; 2025. p. 477-492.

doi: 10.1007/978-981-97-7426-5_36

31. Rasool N, Bhat JI. A critical review on segmentation of glioma brain tumor and prediction of overall survival. *Arch Comput Methods Eng*. 2024;32(3):1525-1569.

doi: 10.1007/S11831-024-10188-2

32. Seeram E, Kanade V. Image processing and analysis. In: *Artificial Intelligence in Medical Imaging Technology*. Berlin: Springer; 2024. p. 83-103.

doi: 10.1007/978-3-031-64049-0_5

33. Do Nascimento Silva JH, Idrobo Pizo GA. Unsupervised bladder segmentation in cone beam-CT imaging via U-net model. *In: 1st IFMBE Latin American Conference on Digital Health*. Berlin: Springer; 2025. p. 28-46.

doi: 10.1007/978-3-031-88064-3_3

34. Das S, Sanki P, Mondal S, Bhattacharya S, Ghosh S. Bladder segmentation in MRI for high dose rate brachytherapy using deep network. In: *Machine Learning for Social Transformation*. Berlin: Springer; 2024. p. 55-67.

doi: 10.1007/978-981-97-7532-3_5

35. Zhao X, Lai L, Li Y, *et al*. A lightweight bladder tumor segmentation method based on attention mechanism. *Med Biol Eng Comput*. 2024;62(5):1519-1534.

doi: 10.1007/S11517-024-03018-X

36. Ye Z, Li Y, Sun Y, He C, He G, Ji Z. Leveraging deep learning in real-time intelligent bladder tumor detection during cystoscopy: A diagnostic study. *Ann Surg Oncol*. 2025;32(5):3220-3226.

doi: 10.1245/S10434-025-17015-3

37. Takagi H, Takeda K, Kadoya N, *et al*. Development of deep learning-based novel auto-segmentation for the prostatic urethra on planning CT images for prostate cancer radiotherapy. *Radiol Phys Technol*. 2024;17(4):819-826.

doi: 10.1007/S12194-024-00832-8

38. Mei H, Yang R, Huang J, *et al*. Artificial intelligence-assisted segmentation of prostate tumors and neurovascular bundles: Applications in precision surgery for prostate cancer. *Ann Surg Oncol*. 2025;18:1-9.

doi: 10.1245/S10434-025-17659-1

39. Wang Y, Wang H, Li N, *et al*. Unsupervised learning-based quantitative analysis of CT intratumoral subregions predicts risk stratification of bladder cancer patients. *BMC Med*. 2025;23(1):328.

doi: 10.1186/S12916-025-04163-2

40. Sun R, Zhang M, Yang L, *et al*. Preoperative CT-based deep learning radiomics model to predict lymph node metastasis and patient prognosis in bladder cancer: A two-center study. *Insights Imaging*. 2024;15(1):21.

doi: 10.1186/s13244-023-01569-5

41. 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*. 2024;10(2):e24878.

doi: 10.1016/j.heliyon.2024.e24878

42. Kurata Y, Nishio M, Moribata Y, *et al*. Development of deep learning model for diagnosing muscle-invasive bladder cancer on MRI with vision transformer. *Heliyon*. 2024;10(16):e36144.

doi: 10.1016/j.heliyon.2024.e36144

43. Han X, Guan J, Guo L, *et al*. A CT-based interpretable deep learning signature for predicting PD-L1 expression in bladder cancer: A two-center study. *Cancer Imaging*. 2025;25(1):27.

doi: 10.1186/s40644-025-00849-1

44. Gumus KZ, Nicolas J, Gopireddy DR, Dolz J, Jazayeri SB, Bandyk M. Deep learning algorithms for bladder cancer segmentation on multi-parametric MRI. *Cancers (Basel)*. 2024;16(13):2348.

doi: 10.3390/cancers16132348

45. 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*. 2025;81:103125.

doi: 10.1016/j.eclinm.2025.103125

46. Wang GY, Zhu JF, Wang QC, *et al*. Prediction of non-muscle invasive bladder cancer recurrence using deep learning of pathology image. *Sci Rep*. 2024;14(1):18931.

doi: 10.1038/s41598-024-66870-9

47. Jiao P, Zheng Q, Yang R, *et al*. Prediction of HER2 status based on deep learning in H&E-stained histopathology images of bladder cancer. *Biomedicines*. 2024;12(7):1583.

doi: 10.3390/biomedicines12071583

48. Zheng Z, Dai F, Liu J, *et al*. Pathology-based deep learning features for predicting basal and luminal subtypes in bladder cancer. *BMC Cancer*. 2025;25(1):310.

doi: 10.1186/s12885-025-13688-x

49. Weng W, Zhu X. INet: Convolutional networks for biomedical image segmentation. *IEEE Access*. 2021;9:16591-16603.

doi: 10.1109/ACCESS.2021.3053408

50. Tang F, Wang L, Ning C, Xian M, Ding J. *CMU-Net: A Strong Convmixer-Based Medical Ultrasound Image Segmentation Network*. United States: Cornell University; 2022.

doi: 10.1109/ISBI53787.2023.10230609

51. Oktay O, Schlemper J, Le Folgoc L, *et al*. *Attention U-Net: Learning Where to Look for the Pancreas*. United States: Cornell University; 2018.

52. Tang F, Ding J, Wang L, Ning C, Zhou SK. *CMUNeXt: An Efficient Medical Image Segmentation Network Based on Large Kernel and Skip Fusion*. Available from: https://www.github.com/fenghetan9/cmunext [Last accessed on 2025 Jun 23].

53. Cao K, Zou Y, Zhang C, *et al*. A multicenter bladder cancer MRI dataset and baseline evaluation of federated learning in clinical application. *Sci Data*. 2024;11(1):1147.

doi: 10.1038/s41597-024-03971-0

54. Cao K, Zou Y, Zhang C, *et al*. *A Multi-Center MRI Dataset for Bladder Cancer and Baseline Evaluations of Federated Learning in its Clinical Application*. Switzerland: Zenodo; 2024.

doi: 10.5281/ZENODO.10409145

55. Matuszczak M, Kiljańczyk A, Salagierski M. A liquid biopsy in bladder cancer-the current landscape in urinary biomarkers. *Int J Mol Sci*. 2022;23(15):8597.

doi: 10.3390/IJMS23158597