

Development of a preoperative risk prediction model of muscle-invasive upper tract urothelial carcinoma using clinical and radiomic computed tomography features

Initial results from a multi-institutional, Canadian study

David-Dan Nguyen^{1,2}, Paola V. Nasute Fauerbach³, Jethro C.C. Kwong¹, Louise McLoughlin^{4,5}, Katherine Lajkosz⁶, Majed Al-Rumayyan⁷, Ammar Alam⁸, Ioana Fugaru⁹, Marie-Pier St-Laurent¹⁰, Paul Toren¹¹, Vincent Fradet¹¹, Ricardo A. Rendon¹², D. Robert Siemens¹³, Rodney H. Breau⁸, Wassim Kassouf⁸, Peter C. Black¹⁰, Girish S. Kulkarni^{1,2}, Masoom A. Haider³

¹Divisions of Urology and Surgical Oncology, Department of Surgery, University Health Network, University of Toronto, Toronto, ON, Canada; ²Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada; ³Joint Department of Medical Imaging, Sinai Health System, University of Toronto, Toronto, ON, Canada; ⁴Department of Urology, St James's Hospital, Dublin, Ireland; ⁵Department of Surgery, School of Medicine, Trinity College Dublin, Dublin, Ireland; ⁶Department of Biostatistics, University Health Network, University of Toronto, Toronto, ON, Canada; ⁷Department of Urology, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia; ⁸Division of Urology, University of Ottawa, Ottawa, ON, Canada; ⁹Division of Urology, McGill University Health Center, Montreal, QC, Canada; ¹⁰Department of Urologic Sciences, University of British Columbia, Vancouver, BC, Canada; ¹¹Division of Urology, Department of Surgery, Université Laval, Quebec City, QC, Canada; ¹²Department of Urology, Dalhousie University, Halifax, NS, Canada; ¹³Department of Urology, Queen's University, Kingston, ON, Canada

Cite as: Nguyen D-D, Nasute Fauerbach PV, Kwong JCC, et al. Development of a preoperative risk prediction model of muscle-invasive upper tract urothelial carcinoma using clinical and radiomic computed tomography features: Initial results from a multi-institutional, Canadian study. *Can Urol Assoc J* 2026;20(4):122-31 <http://dx.doi.org/10.5489/cuaj.9370>

Published online December 15, 2025

Appendix available at cuaj.ca

ABSTRACT

INTRODUCTION: Accurate pre-intervention staging of upper tract urothelial carcinoma (UTUC) remains a significant clinical challenge, particularly in identifying muscle-invasive disease ($\geq pT2$), where kidney-sparing surgery may not be appropriate. Current imaging and biopsy approaches are often inadequate. Radiomics, which extracts high-dimensional features from medical imaging, may improve non-invasive staging. This study assessed whether computed tomography (CT)-based radiomic features, alone or combined with clinical data, could predict $\geq pT2$ UTUC in a multicenter Canadian cohort.

METHODS: We retrospectively analyzed clinical, pathologic, and radiographic features of patients with UTUC who underwent extirpative surgery at five academic centers from January 2, 2001, to May 1, 2023. Radiomic features were extracted from machine-learning segmentations of the affected kidney using the excretory phase of CT. Predictive models were developed using clinical-only, radiomic-only, and combined data to predict stage $\geq pT2$. Feature selection included univariable logistic regression, correlation filtering, and LASSO. Model performance was assessed via five-fold cross-validation repeated 10 times, with area under the curve (AUC) as the primary metric.

RESULTS: Of 441 patients, 208 (47.2%) were included. Of the 208 patients, 97 (46.6%) had $\geq pT2$ disease. The clinical model (AUC 0.602) included age, hydronephrosis, and high-grade cytology. The radiomics model, based on two texture features, achieved an AUC of 0.653. The combined model achieved an AUC of 0.647. Radiomics and combined models significantly outperformed the clinical model ($p < 0.01$), but did not differ from each other. For 117 patients with renal pelvis cancers, the combined model's discrimination performance was statistically better than the clinical model (AUC 0.708 vs. AUC 0.607, $p < 0.001$). Likewise, the radiomics' AUC discrimination performance was statistically better than the clinical model (AUC 0.694 vs. AUC 0.607, $p = 0.004$). In contrast, we found no significant difference in model performance in the non-renal pelvis subgroup ($n = 91$).

CONCLUSIONS: Conventional radiomics improved the prediction of muscle-invasive UTUC compared to clinical models alone, but overall accuracy remained suboptimal for clinical use. Heterogeneity in CT protocols and challenges with tumor segmentation were the main limitations. Future work should develop more adaptable AI models trained on larger, more diverse datasets to better reflect real-world imaging conditions.

KEY MESSAGES

- In this multicenter, Canadian cohort of patients with UTUC, radiomic features from excretory-phase CT improved the prediction of muscle-invasive disease ($\geq pT2$) compared to clinical variables alone, although overall performance was modest.
- Differences in hydronephrosis rates and tumor size compared to other international cohorts suggest variations in Canadian referral patterns and highlight the importance of broader, more diverse datasets for generalizability.
- Technical challenges, including variability in CT protocols, tumor segmentation complexity, and the impact of pre-scan stents, limit current radiomic model performance and highlight areas for refinement.
- AI-based segmentation and inclusion of secondary features, such as ureteric obstruction, may enhance accuracy, but conventional radiomic methods remain constrained by contrast sensitivity and retrospective design.
- Future work from our group will assess advanced foundation models that integrate the entire urinary tract and accommodate heterogeneous imaging, with planned expansion to international cohorts and external validation.

INTRODUCTION

The standard surgical treatment for non-metastatic upper tract urothelial carcinoma (UTUC) is radical nephroureterectomy (RNU) with bladder cuff excision.^{1,2} Over the past decade, minimally invasive techniques, such as endoscopic laser ablation, have gained traction for the management of small, unifocal, low-grade tumors, even in patients with normal contralateral kidney function.^{3,4} Such kidney-sparing surgery (KSS) is both safe and effective, preserving oncologic outcomes in carefully selected patients.^{3,5}

In principle, KSS is intended for UTUC patients without muscle-invasive disease (i.e., stages lower than clinical T2); however, accurate preoperative diagnosis of muscle-invasive or otherwise high-risk disease remains

an unresolved challenge. First, while computed tomography (CT) urogram is widely regarded as the best available diagnostic tool for these carcinomas, CT imaging alone remains insufficient for accurate staging and risk stratification.^{6,7} Second, obtaining adequate tissue samples via ureteroscopic biopsy is itself an operative procedure and remains technically challenging, often yielding limited diagnostic material.^{8,9}

Given the limitations of current preoperative diagnostic methods, there is a need for innovative approaches to enhance disease staging and preoperative risk assessment. This is also pertinent as neoadjuvant chemotherapy is used in muscle-invasive UTUC.¹⁰ Reducing reliance on tissue biopsies and improving the diagnostic yield of imaging studies could enable more timely and accurate clinical decisions.

Radiomics, a field that extracts and analyzes quantitative imaging features using machine learning and deep learning, offers a promising solution.¹¹ These features, often imperceptible to the human eye, may serve as imaging biomarkers to improve cancer detection, staging, and prognosis by quantifying tumor intensity, shape, and internal heterogeneity, among other features.¹² By integrating radiomics, it may be possible to enhance non-invasive diagnostic accuracy and support more personalized treatment planning.

To explore this potential, we conducted a multicenter, Canadian feasibility study to develop a preoperative prediction model integrating clinical and CT radiomic data to identify muscle-invasive UTUC ($\geq pT2$). Here, we present the initial findings from this collaborative effort.

METHODS

Study design

We conducted a multicenter, retrospective cohort study of patients with UTUC who underwent extirpative surgery at five Canadian academic institutions (Dalhousie, University of British Columbia [UBC], University Health Network [UHN], McGill University Health Center, and University of Ottawa) between January 2, 2001, and May 1, 2023. As such, the data is representative of academic institutions across Canada.

All clinical, pathologic, and imaging assessments, as well as surgical resections, were conducted and collected locally at each participating institution. CT images were sent to the central study site for additional review by an independent radiologist blinded to the original interpretations and for radiomic data extraction.

Cohort inclusion and exclusion criteria

Patients were included if they 1) had a primary histology of urothelial carcinoma; 2) underwent extirpative surgery; 3) had a preoperative CT scan performed within 12 months before surgery; and 4) had pathology available for review. Patients were excluded if they 1) had metastatic disease at presentation; 2) received of neoadjuvant chemotherapy; 3) had a ureteral stent present at the time of CT imaging; 4) had no identifiable tumor on CT (clinically or on study review); 5) had a CT lacking an excretory phase; or 6) had a CT of insufficient quality, such as beam hardening artifacts through the tumor and patient motion. All imaging-related exclusion criteria were centrally reviewed by a radiologist (PVNF).

Outcome

Our outcome of interest was \geq pT2 disease on surgical specimen, as assessed by fellowship-trained pathologists at the included institutions.

Clinical and pathologic features

Relevant clinical and pathologic features were planned for data extraction based on a review of the literature and clinical expertise by trained abstractors.

Clinical data at the time of surgery were extracted from electronic medical records, including age, sex, body mass index (BMI), and smoking history. Smoking status was categorized as non-smoker, active smoker, quit within one year, and quit more than one year before UTUC diagnosis.

Pathologic data included presurgical urine cytology, categorized as positive for high-grade urothelial carcinoma or otherwise negative. Surgical histopathology served as the reference standard and included histologic subtype, pathologic stage, tumor size and location, tumor multifocality, and nodal stage (N stage). In the case of multiple tumors, the highest pathologic stage was used.

Imaging features

Conventional imaging features were planned for data extraction based on a review of the literature and clinical expertise. Imaging features were obtained from site-specific radiologic reports and further reviewed centrally by a genitourinary imaging radiology fellow.

Collected variables included tumor laterality, location, largest axial dimension, and the presence of ipsilateral hydronephrosis. Multifocality and abnormal regional lymph nodes were also documented. Clinical staging was then determined based on a combination of pre-surgical pathologic and imaging findings.

Additionally, CT images and features were reviewed at the central site. An independent, fellowship-trained radiologist (PVNF.), blinded to the original interpretations, re-evaluated all scans. Discrepancies were adjudicated by a senior fellowship-trained radiologist (MAH).

Radiomic features

Radiomic feature extraction was performed using artificial intelligence (AI)-based segmentation algorithms leveraging deep convolutional neural networks. The TotalSegmentator model was employed to segment the affected kidney.¹³ An iterative erosion and then dilation of the renal mask was used to fill in the renal hilum and proximal ureter. This final mask was subsequently used to define the region selected for radiomic analysis. A radiologist was not used to segment, thus there were no issues of interobserver variability in segmentation. An example of a segmentation is shown in Figure 1.

For mid and distal ureteral tumors, only secondary renal findings — such as hydronephrosis, delayed nephrogram, or parenchymal atrophy — contributed to radiomic signatures. Radiomic feature extraction was conducted using the Image Biomarker Standardization Initiative (IBSI)-compliant PyRadiomics package (version 3.1.0).¹⁴ The voxel size was normalized to 1 mm before feature extraction to reduce radiomics feature variability.



Figure 1. Example of a kidney segmentation using the TotalSegmentator. Preoperative CTU in the three planes (A: axial, B: sagittal, C: Coronal). A non-smoker, 81-year-old female patient with a 31 x 32 x 44 mm filling defect at the level of the right renal pelvis (blue arrow) in the excretory phase (note the contralateral kidney with a homogeneously contrast-enhanced pelvis). There was no evidence of hydronephrosis or other lesions. Ureter and bladder cytology did not show high-grade. She underwent radical nephroureterectomy and pathology revealed a unifocal urothelioma with a final stage pT3 pNX. With the TotalSegmentator model, the right kidney was segmented (yellow contour).

Only the 107 “original” Pyradiomics features were used without pre-filtering, given the available cohort size.

Missing data

To handle missing data, we employed a missing-indicator approach by creating explicit “missing” categories for variables with incomplete data. This method preserved the full sample size and allowed the model to account for potential informative missing data. This is particularly relevant for this initial report.

Feature selection and model development

Radiomic variables were normalized using the z-score transformation. Using the complete dataset, three predictive models were developed: 1) a model using clinical variables only; 2) a model using radiomic variables only; and 3) a combined model incorporating both clinical and radiomic variables. To select the radiomics variables, features with zero variance were first removed, and univariable logistic regression models predicting \geq pT2 status were fit with the remaining features. Features with p-values <0.15 were retained for further analysis. Pairwise Spearman’s correlation coefficients were estimated for the retained features, and features with a coefficient $>|0.8|$ were excluded to reduce redundancy.

The remaining features were incorporated into a least absolute shrinkage and selection operator (LASSO) model using five-fold cross-validation, and those that minimized the lambda parameter were selected for the final model. The random seed used for model development and resampling was set to ensure reproducibility. The selected features were incorporated into a multivariable logistic regression model with a probability of \geq pT2 status outputted.

The feature selection process for clinical characteristics followed a similar approach, except that variance and correlation were not assessed. The following clinical characteristics were considered: sex, age (continuous), presence of hydronephrosis, prior bladder cancer diagnosis, prior UTUC diagnosis, smoking status, and any prior high-grade cytology result. A complete case analysis was performed for model development; no imputation was used.

To develop a combined clinical and radiomics model, a radiomics score was calculated as the linear combination of the selected radiomics feature coefficients, which was then combined with the selected clinical characteristics into a single multivariable logistic regression model.

As a sensitivity analysis, feature selection and model fitting were repeated separately for the subgroup of patients with tumors confined to the renal pelvis and for those with tumors located exclusively outside the renal pelvis. These subgroup analyses were prespecified and exploratory, and model performance was compared across groups to assess robustness.

Model evaluation

The performance of all three models — radiomics, clinical, and combined — was assessed by estimating the area under the curve (AUC) using five-fold cross-validation repeated 10 times. The mean AUC with the 95% confidence interval (CI, estimated using 500 bootstrap replications) was reported and visualized using receiver operating characteristics (ROC) curves. Model calibration was assessed and visualized using calibration curves, applying 500 bootstrap resamples to estimate optimism-adjusted calibration curves. Clinical utility was evaluated using decision curve analysis (DCA).¹⁵ We believe that clinically meaningful risk thresholds, particularly for decisions like initiating neoadjuvant chemotherapy, generally range from 30–60%.

Statistical analysis

Cohort clinical characteristics stratified by pT stage were summarized using descriptive statistics. Differences in continuous and categorical characteristics were assessed using Mann-Whitney U and Chi-squared tests, respectively. The protocol of this initial report was not published a priori.

Statistical analyses were conducted using R version 4.4.2. All hypothesis tests were two-sided, with p-values <0.05 considered statistically significant.

Ethics and reporting statement

Ethical approval and waiver of informed consent were obtained at each participating institution before the conduct of the study.

Findings were reported according to APPRAISE-AI and TRIPOD-AI reporting guidelines.^{16,17}

RESULTS

Cohort characteristics

Of 441 patients, 208 (47.2%) met our inclusion criteria (Figure 2), with a similar percentage across all five sites (ranging from 38–61%). The distribution of patients was as follows: Dalhousie (n=14), UBC (n=61), UHN (n=58), McGill (n=27), and Ottawa (n=50). Patients were primarily excluded due to the lack of excretory

phase (n=113, 49%), followed by the presence of stents or nephrostomy tubes (n=59, 25%). Overall, the mean age was 72.1 (standard deviation [SD] 10.1), with a statistically significant difference in the mean age between the <pT2 cohort and those patients with a ≥ pT2 (71.0 vs. 74.3 years, p=0.04). Most patients were male (n=135, 65%).

Of all the patients, 97 (47%) had final pathologic stage tumors ≥pT2, while the remaining 111 (53%) were diagnosed with non-muscle-invasive UTUC (<pT2). Table 1, and Supplementary Tables 1, 2 (available at cuaj.ca) summarize the clinical and imaging features collected for eligible patients in the overall cohort, as well as the renal pelvis cancer and non-renal cancer subgroups, respectively. The main pathologic characteristics and surgical treatment for eligible patients in the overall cohort are also displayed in Table 1. Supplementary Table 3 (available at cuaj.ca) describes retained radiomic features.

Model comparisons for overall cohort

For the clinical model, age, hydronephrosis, and any high-grade cytology status were retained, with adjusted odds ratios (OR) of 1.03 (95% CI 1.00–1.06, p=0.048), 1.69 (95% CI 0.96–2.97, p=0.070), and 1.72 (95% CI 0.91–3.22, p=0.093), respectively (Table 2). The associated AUC was 0.602 (95% CI 0.526–0.670).

For the radiomics model, original Neighborhood Gray Tone Difference Matix (NGTDM) busyness and contrast were both selected, with adjusted OR of 1.59 (95% CI 1.07–2.36, p=0.022) and 1.44 (95% CI 1.04–2.00, p=0.027), respectively. The associated AUC was 0.653 (95% CI 0.562–0.700). The combined model has an AUC of 0.647 (95% CI 0.562–0.699). The combined model coefficients are summarized in Table 3.

Figure 3A presents the ROC curves for the overall cohort. When comparing the performance of

Table 1. Clinical and imaging characteristics

Variable	<pT2 (n=111, %)	≥pT2 (n=97, %)
Age, mean (SD)	71.4 (10.1)	74.3 (9.9)
Sex, female	39 (35.5%)	33 (34.0%)
Missing	1	0
BMI (kg/m ²), mean (SD)	27.7 (5.4)	26.9 (6.5)
Missing	16	15
Smoking history		
Non-smoker	39 (35.8%)	41 (42.7%)
Active smoker/quit <1 year	34 (31.2%)	18 (18.8%)
Quit ≥1 year	36 (33.0%)	37 (38.5%)
Missing	2	1
Laterality		
Right	67 (60.4%)	50 (51.5%)
Left	44 (39.6%)	47 (48.5%)
Clinical tumor location		
Renal pelvis	50 (59.5%)	28 (42.4%)
Ureter	27 (32.2%)	30 (45.5%)
Multifocal (ureter & renal pelvis)	7 (8.3%)	8 (12.1%)
Missing	27	31
CT tumor size (mm)		
Mean (SD)	22.4 (16.7)	28.2 (19.7)
Missing & not identified on CT	27	17
Hydronephrosis on CT	43 (38.7%)	49 (50.5%)
Clinical stage		
cT0	5 (6.2%)	0 (0.0%)
cT1	34 (42.5%)	20 (27.4%)
cT2	18 (22.5%)	13 (17.8%)
cT3	8 (10.0%)	34 (46.6%)
cTa	11 (13.8%)	5 (6.8%)
cTx	4 (5.0%)	1 (1.4%)
Positive urine cytology	24 (21.6%)	33 (34.0%)
Surgery		
Radical nephroureterectomy	101 (91.0%)	85 (87.6%)
Other	10 (9.0%)	9 (12.4%)

*Specimen high grade was determined as per WHO 2004 criteria. BMI: body mass index; CT: computed tomography; cTx: clinical stage unknown; SD: standard deviation.

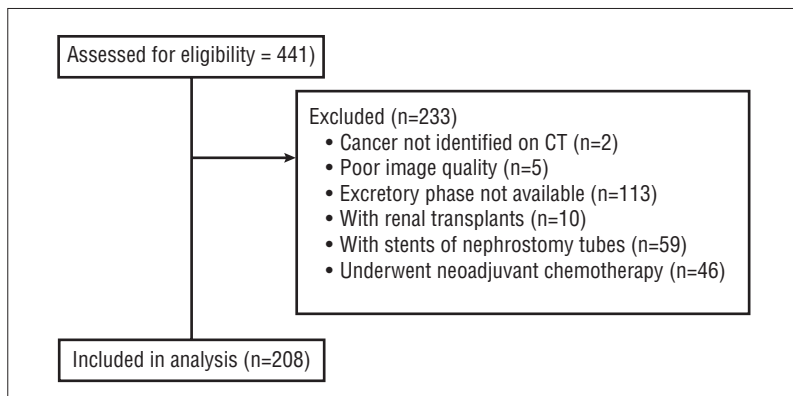


Figure 2. Patient flowchart. CT: computed tomography.

Table 1 (cont'd). Clinical and imaging characteristics

Variable	<pT2 (n=111, %)	≥pT2 (n=97, %)
pT Stage		
pT _a	67 (60.4%)	0 (0%)
pT _{is}	6 (5.4%)	0 (0%)
pT ₁	38 (34.2%)	0 (0%)
pT ₂	0 (0%)	25 (25.8%)
pT ₃	0 (0%)	64 (66.0%)
pT ₄	0 (0%)	8 (8.2%)
Specimen histology		
Urothelial	105 (94.6%)	76 (78.4%)
Urothelial with variant	6 (5.4%)	21 (21.6%)
Missing		
Specimen grade*		
Low-grade	54 (48.6%)	3 (3.1%)
High-grade	57 (52.3%)	94 (96.9%)
pN Stage		
pN ₀	29 (26.1%)	25 (25.8%)
pN ₊	0 (0%)	14 (14.4%)
pN _X	82 (73.9%)	58 (59.8%)

*Specimen high grade was determined as per WHO 2004 criteria. BMI: body mass index; CT: computed tomography; cTx: clinical stage unknown; SD: standard deviation.

each model using the AUCs, there was a statistical significance between the clinical and radiomics models ($p=0.006$) and between the clinical and combined models ($p=0.007$). In contrast, there was no significant difference between the performance of the radiologic and combined models ($p=0.48$).

Calibration plots showed that the combined model had good alignment between predicted and observed probabilities across the full range of risk, with minimal deviation after optimism adjustment (Supplementary Figure 1; available at [cuaj.ca](#)). In contrast, the clinical model underestimated risk at higher predicted probabilities, while the radiomics model exhibited slight overestimation in the mid-range. DCA showed that the combined model consistently provided the highest standardized net benefit across a broad range of threshold probabilities (approximately 0.2–0.8), outperforming both the clinical and radiomics models alone between clinically relevant thresholds (Supplementary Figure 2; available at [cuaj.ca](#)).

Model comparisons for renal pelvis cancer subgroup

This group had a tumor identified involving the renal pelvis ($n=117$). Cohort characteristics stratified by <pT2 vs. ≥pT2 are summarized in Supplementary Table 1 (available at [cuaj.ca](#)). In this subgroup, the features retained by the final clinical model were age and having a prior bladder cancer diagnosis, with OR of 1.03 (95% CI 0.99–1.07, $p=0.096$) and 0.36 (95% CI 0.12–1.09, $p=0.072$), respectively (Table 2). The clinical model's AUC was 0.607 (95% CI 0.522–0.700).

Four radiomic features were retained in the radiomics-only model, with an associated AUC of 0.694 (95% CI 0.572–0.774). The combined model had an AUC of 0.708 (95% CI 0.617–0.780). Figure 3B presents the ROC curves. The difference in AUCs was statistically significant between the clinical and radiomic models ($p=0.006$) and between the clinical and combined models ($p=0.008$) (Table 3). There was no significant difference between the radiomics and the combined models ($p=0.48$).

The combined model demonstrated superior calibration, with predicted probabilities closely matching observed outcomes (Supplementary Figure 1; available at [cuaj.ca](#)). The clinical model showed moderate underestimation at higher predicted probabilities, whereas the radiomics model tended to overpredict risk across most of the range. In DCA, the combined model showed modest gains in net benefit relative to the other models, particularly between clinically relevant threshold probabilities of 0.3 and 0.6 (Supplementary Figure 2; available at [cuaj.ca](#)). The radiomics model did not outperform the clinical model.

Model comparisons for non-renal pelvis subgroup

This group had a tumor identified not involving the renal pelvis ($n=91$). Cohort characteristics by <pT2 vs. ≥pT2 are summarized in Supplementary Table 2 (available at [cuaj.ca](#)). Only high-grade cytology was retained for the clinical model, with an OR 3.63 (95% CI 1.39–9.47, $p=0.008$) (Table 2). The AUCs for the clinical-only, radiomics-only, and combined models were 0.629 (95% CI 0.547–0.718), 0.608 (95% CI 0.478–0.706), and 0.677 (95% CI 0.599–0.770), respectively (Figure 3C). There were no statistically significant differences between clinical, radiomics, and combined models.

Calibration was best for the combined model, while both the clinical and radiomics models exhibited poor calibration at the extremes of predicted probability (Supplementary Figure 1; available at [cuaj.ca](#)).

Table 2. Multivariate logistic regression for selected clinical and CTU features

Cohort	Feature	Clinical only		Radiomics only		Radiomics + clinical	
		OR (95% CI)	p	OR (95% CI)	p	OR (95% CI)	p
Full cohort	Age	1.03 (1.00, 1.06)	0.048			1.03 (1.00, 1.06)	0.065
	High-grade cytology (ref=No)	1.72 (0.91, 3.22)	0.093			1.73 (0.91, 3.29)	0.097
	Hydronephrosis (ref=No)	1.69 (0.96, 2.97)	0.070			1.29 (0.71, 2.35)	0.45
	Original NGTDM busyness			1.59 (1.07, 2.36)	0.022		
	original NGTDM contrast			1.44 (1.04, 2.00)	0.027		
	Radiomics score					2.55 (1.37, 4.75)	0.003
	AUC (95% CI)	0.600 (0.527, 0.669)	0.653 (0.562, 0.700)	0.647 (0.562, 0.699)			
Renal pelvis subgroup	Age	1.03 (0.99, 1.07)	0.096			1.02 (0.98, 1.06)	0.39
	Prior BCa diagnosis (ref=No)	0.36 (0.12, 1.09)	0.072			0.28 (0.08, 0.97)	0.044
	Original shape flatness			1.62 (1.07, 2.47)	0.024		
	Original shape surface:volume ratio			1.80 (0.99, 3.24)	0.052		
	Original GLCM MCC			0.75 (0.48, 1.16)	0.20		
	Original NGTDM contrast			1.39 (0.81, 2.39)	0.24		
	Radiomics score					2.76 (1.46, 5.22)	0.002
	AUC (95% CI)	0.607 (0.522, 0.700)	0.694 (0.572, 0.774)	0.708 (0.617, 0.780)			
Non-renal subgroup	High-grade cytology (ref=No)	3.63 (1.39, 9.47)	0.008			3.29 (1.23, 8.82)	0.018
	Original first-order skewness			0.76 (0.47, 1.22)	0.25		
	Original GLSZM gray-level non-uniformity			1.50 (0.92, 2.44)	0.10		
	Original NGTDM contrast			1.34 (0.82, 2.18)	0.24		
	Radiomics score					2.54 (1.14, 5.65)	0.022
	AUC (95% CI)	0.629 (0.547, 0.718)		0.608 (0.478, 0.706)		0.677 (0.566, 0.770)	

AUC: area under the curve; BCa: bladder cancer; CI: confidence interval; OR: odds ratio.

The radiomics model tended to overpredict risk. DCA revealed a notable benefit of the combined model, with the highest standardized net benefit across nearly the entire range of threshold probabilities (especially 0.3–0.9) (Supplementary Figure 2; available at *cuaj.ca*).

DISCUSSION

In this Canadian, multicenter cohort of patients with UTUC who underwent extirpative surgery, we evaluated the feasibility and potential utility of radiomic features extracted from the excretory phase of CT scans to predict muscle-invasive disease (≥pT2). While the inclusion of radiomic features improved predictive accuracy compared to models based solely on clinical

variables, the overall performance remained modest. These findings demonstrate the potential of radiomics while underscoring key limitations and areas for refinement, providing important direction for future efforts to enhance predictive modeling in UTUC.

Although these findings represent initial results from our group, with additional sites and patients to be added, our current cohort suggests that patterns of care in Canada may differ from those reported elsewhere. For instance, prior studies have reported higher rates of hydronephrosis among patients with ≥pT2 disease and larger tumor sizes on preoperative CT.^{6,18} While our cohort is limited by its current sample size, these differences may reflect variations in referral

practices or patient selection at Canadian academic centers compared to other settings. This underscores the need for more diverse, multi-institutional cohorts to enhance generalizability. Nevertheless, our models still retained key clinical predictors established in the literature, such as hydronephrosis.¹⁸

Variability in CT imaging protocols and differences in the visibility of ureteral vs. renal pelvis tumors highlight the need for models that do not depend on a specific contrast phase. Many commonly used radiomic features are sensitive to the phase of CT enhancement. For example, the presence or absence of contrast in the collecting system can affect feature values regardless of tumor characteristics. Notably, recent single-center studies at academic institutions aiming to predict pre-surgical UTUC stage or differentiate UTUC from renal cell carcinoma using radiomics have reported excluding nearly 50% of patients due to the absence of high-quality, preoperative CT urography.¹⁹⁻²¹

In our cohort, such exclusions were primarily driven by the lack of excretory phase CT. This may be due to variations in CT protocol and patient history (such as renal function) influencing whether portal venous or non-contrast imaging is performed. Furthermore, many patients require stent placement before undergoing CT scans, which can affect the enhancement of the ureter walls, leading to potential misdiagnoses in staging.

Tumor segmentation in UTUC can be particularly challenging, especially when tumors are adjacent to complex anatomical structures. For example, distinguishing peripelvic fat invasion from renal parenchyma involvement in renal pelvis tumors is often difficult. In this study, we addressed these challenges by employing validated AI-based segmentation models to delineate the kidney

Table 3. Comparison of the performance of AUCs using the models in both UTUC cohorts				
Renal pelvis cancers	AUC (95% CI)	p-value of AUCs: clinical vs. radiomic	p-value of AUCs: clinical vs. combined	p-value of AUCs: radiomics vs. combined
Clinical model	0.607 (95% CI 0.522–0.700)	0.004	<0.001	0.25
Radiomics model	0.694 (95% CI 0.572–0.774)			
Combined model	0.708 (95% CI 0.617–0.78)			
Non-renal pelvis cancers	AUC (95% CI)	p-value of AUCs: clinical vs. radiomic	p-value of AUCs: clinical vs. combined	p-value of AUCs: radiomic vs. combined
Clinical model	0.629 (95% CI 0.547–0.718)	0.35	0.07	0.056
Radiomics model	0.608 (95% CI 0.478–0.706)			
Combined model	0.677 (95% CI 0.599–0.77)			

AUC: area under the curve; CI: confidence interval; UTUC: upper tract urothelial carcinoma.

and renal hilum, thereby minimizing observer variability. We also incorporated the renal parenchyma and secondary effects of ureteric obstruction — features that may enhance predictive accuracy for staging.

Radiomics represents a promising frontier in the management of UTUC, potentially offering non-invasive tools to improve diagnosis, risk stratification, and prognostication. Across multiple studies, radiomic features extracted from CT scans have consistently outperformed conventional clinical predictors, such as urine cytology, particularly in identifying high-grade or muscle-invasive disease.^{7,19-23} While current evidence is encouraging, most studies remain retrospective, single-center, and lack external validation. Notably, the largest study to date, conducted by Zheng et al, evaluated

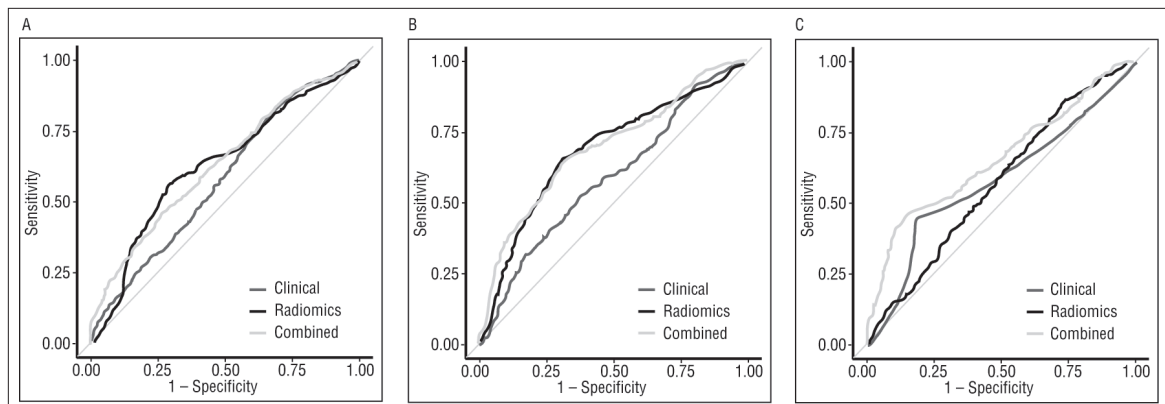


Figure 3. Receiving operating characteristic (ROC) curves for clinical, radiomics, and combined models for the overall cohort, renal pelvis subgroup, and non-renal subgroup. (A) The ROC curve for the overall cohort showed that the area under the curve (AUC) was higher in the radiomics and combined models compared to the clinical model (0.653, 0.647, 0.602, respectively). (B) The ROC curve for the renal pelvis cohort indicated a significant improvement in the AUC when using the radiomics model, both combined and alone, compared to the clinical model alone (0.708, 0.694, 0.607, respectively). (C) The ROC curve for the non-renal pelvis cohort showed a modest improvement in the AUC with the combined model compared to the clinical and radiomics models (0.677 vs. 0.629 and 0.608, respectively).

167 patients from a single institution and developed a nomogram integrating multiphase CT radiomics with hydronephrosis, a key clinical predictor.²¹ This model demonstrated strong performance, achieving an AUC of 0.84 and an accuracy of 79% in the validation set, with a 14% gain in diagnostic accuracy and a 33% improvement in sensitivity compared to urine cytology. These findings underscore both the promise of radiomics and the need for larger, multicenter studies to confirm its clinical utility and generalizability.

Limitations

Our study has several limitations. First, although our cohort is among the largest available for UTUC with corresponding CT imaging radiomic feature extraction,^{19-21,24} it remains relatively small compared to radiomic studies in other malignancies, which may limit model performance.

Second, while the dataset is multicentric, it is restricted to Canadian institutions. This geographic limitation may affect its generalizability to settings with differing healthcare systems, practice patterns, and resource availability. Considering these first two limitations, expanding the dataset to include international cohorts is an important next step.

Third, this study was retrospective in nature, which may introduce bias in the collection, availability, and interpretation of clinical variables.

Fourth, our analysis relied on traditional radiomic approaches. While this study provides valuable preliminary insights, it underscores the limitations of conventional radiomic-feature-extraction methods in addressing the diagnostic complexities of UTUC.

A subsequent study will evaluate the use of foundation models tailored to the unique challenges of UTUC diagnosis,²⁵ incorporating the entire urinary tract within the segmentation volume and including all scans regardless of the contrast excretion phase. The aim will be to assess whether these advanced models can more effectively extract features relevant to staging compared to traditional radiomics while accommodating heterogeneous CT protocols and avoiding the limitations of the PyRadiomics feature set. Additional patients from additional centers are planned, a subset of which will form an external validation, which is lacking in this study.

CONCLUSIONS

While our models leveraging conventional radiomic features demonstrated improved predictive performance over those based solely on clinical variables, their overall accuracy remains inadequate for clinical

implementation. This shortcoming is further exacerbated by variability in CT imaging protocols across institutions, which limits generalizability. Nonetheless, our findings are promising and emphasize the need for larger, more diverse patient cohorts to enhance model robustness and external validity. Future efforts by our group should focus on developing foundation models that integrate deep learning techniques to extract more robust and generalizable imaging features, ultimately improving staging accuracy across heterogeneous imaging environments.

FUNDING & COMPETING INTERESTS: A Princess Margaret Cancer Centre Catalyst Grant and a Bladder Cancer Canada Grant specifically supported this study. Dr. Louise McLoughlin received a Hold'em for Life Oncology Fellowship for the conduct of this study. Dr. David-Dan Nguyen is supported by a Canadian Institutes of Health Research (CIHR) Vanier Canada Graduate Scholarship (CGV-192647), the CMCC/Atrium Hold'em for Life Oncology Fellowship, the Ontario Ministry of Health Clinician-Investigator Program, and a Schwartz Reisman Fellowship from the Schwartz Reisman Institute for Technology and Society at the University of Toronto. Dr. St-Laurent has received honoraria from Bayer, EMD Serono and TerSera and been on advisory boards for Abbvie, EMD Serono, Ferring, Johnson & Johnson, Pfizer, and TerSera. Dr. Toren has been on advisory boards for Ferring, Janssen, and Knight; has received honoraria from AstraZeneca; and has participated in clinical trials supported by Janssen and Merck. Dr. Fradet has been on advisory boards for Accord, Astellas, Bayer, EMD Serono, Ferring, Janssen, and Knight; served as a speaker for Astellas, Bayer, Ferring, and Janssen; has received grants and/or honoraria from Abbvie, Accord, Astellas, Bayer, Ferring, Janssen, and TerSera; and has participated in clinical trials supported by Aragon, Astellas, AstraZeneca, Bayer, BMS, Cougar Biotech, Endocyte, Exelixis, GSK, Hoffman-La Roche, IMV, Janssen, MacroGenics, MedImmune, Medivation, Merck, Myovant, Novartis, Pfizer, Point, Progenics, Sanofi, Seattle Genetics, and Takeda. Dr. Rendon has been an advisory board and speakers' bureau member for and has received honoraria from Abbvie, Amgen, Astellas, AstraZeneca, Bayer, Ferring, Janssen, Pfizer, Roche, Sanofi, and Tolmar; has received honoraria/grants from Abbvie, Astellas, Bayer, Ferring, Janssen, Sanofi, Tersera, and Tolmar; holds investments in Myovant; and has participated in clinical trials supported by Abbvie, Astellas, Bavarian Nordic, Bayer, Ferring, Janssen, Myovant, and Sanofi. Dr. Siemens has participated in clinical trials supported by Astellas, Merck, Pfizer, and Verity. Dr. Breau has been on advisory boards for Abbvie, Astellas, CG Oncology, EMD Serono, Knight, Merck, and Tolmar. Dr. Kassouf has been on advisory boards for and received honoraria from Astellas, AstraZeneca, Bayer, BMS, CG Oncology, enGene, EMD Serono, Ferring, Janssen, Merck, Pfizer, and Photocure. Dr. Black has been on advisory boards for Abbvie, Astellas, AstraZeneca, Aura, Bayer, BMS, CG Oncology, Combat, EMD-Serono, enGene, Ferring, J&J, Merck, Nonagen, NanOlogy, Nanobot, Pfizer, Photocure, Prokarium, Sumitomo, TerSera, Tolmar, and Verity; been a speaker for Bayer, J&J, Pfizer, and TerSera; has received research grants from CIHR, CRS, CUASF, BCC, Hecht Foundation, Michael Smith Foundation, and TFR; and has participated in clinical trials supported by CG Oncology, enGene, J&J, Theralase, and Verity. Dr. Kulkarni has been on advisory boards for AAA/Novartis, Abbvie, Astellas, BMS, CG Oncology, EMD Serono, enGene, Ferring, J&J, Knight Therapeutics, Merck, Pfizer, Protara Therapeutics, Theralase, Tolmar, and Verity; has received grants and/or honoraria from AAA/Novartis, AstraZeneca, Ferring, TerSera, J&J, Merck, Pfizer, Photocure, and Verity; and has participated in clinical trials supported by CG Oncology, enGene, Ferring, J&J, Merck, Pfizer, Protara Therapeutics, Seagen, Theralase, and Verity. The remaining authors have no competing personal or financial interests to disclose.

This paper has been peer-reviewed.

REFERENCES

- Coleman JA, Clark PE, Bixler BR, et al. Diagnosis and management of non-metastatic upper tract urothelial carcinoma: AUA/SUO Guideline. *J Urol* 2023;209:1071-81. <https://doi.org/10.1097/JU.0000000000003480>
- Rouprêt M, Seisen T, Birtle AJ, et al. European association of urology guidelines on upper urinary tract urothelial carcinoma: 2023 Update. *Eur Urol* 2023;84:49-64. <https://doi.org/10.1016/j.eururo.2023.03.013>

3. Seisen T, Colin P, Rouprêt M. Risk-adapted strategy for the kidney-sparing management of upper tract tumours. *Nat Rev Urol* 2015;12:155-166. <https://doi.org/10.1038/nrurol.2015.24>
4. Pierorazio PM, Kleinmann N, Shabsigh A, et al. Long-term outcomes of primary chemoablation of low-grade upper tract urothelial carcinoma (LG UTUC) with UGN-101, a mitomycin reverse thermal gel. *J Urol* 2025;213:313-22. <https://doi.org/10.1097/JU.0000000000004331>
5. Paciotti M, Alkhatib KY, Nguyen DD, et al. Is Segmental ureterectomy associated with inferior survival for localized upper-tract urothelial carcinoma of the ureter compared to radical nephroureterectomy? *Cancers* 2023;15:1373. <https://doi.org/10.3390/cancers15051373>
6. Foerster B, Abufaraj M, Mari A, et al. The performance of tumor size as risk stratification parameter in upper tract urothelial carcinoma (UTUC). *Clin Genitourin Cancer* 2021;19:272.e1-272.e7. <https://doi.org/10.1016/j.clgc.2020.09.002>
7. Chung D, Ramjiawan R, Bal DS, et al. Radiographic predictors of muscle-invasive upper tract urothelial cancer: A Canadian cohort. *Can Urol Assoc J* 2024;18:413. <https://doi.org/10.5489/cuaj.8817>
8. Baard J, Cormio L, Dasgupta R, et al. Unveiling the challenges of UTUC biopsies and cytology: insights from a global real-world practice study. *World J Urol* 2024;42:177. <https://doi.org/10.1007/s00345-024-04866-w>
9. Baard J, Shariat SF, Roupret M, et al. Adherence to guideline recommendations in the management of upper tract urothelial carcinoma: an analysis of the CROES-UTUC registry. *World J Urol* 2022;40:2755-2763. <https://doi.org/10.1007/s00345-022-04168-z>
10. Coleman JA, Yip W, Wong NC, et al. Multicenter phase II clinical trial of gemcitabine and cisplatin as neoadjuvant chemotherapy for patients with high-grade upper tract urothelial carcinoma. *J Clin Oncol* 2023;41:1618-25. <https://doi.org/10.1200/JCO.22.00763>
11. Liu Z, Wang S, Dong D, et al. The applications of radiomics in precision diagnosis and treatment of oncology: Opportunities and challenges. *Theranostics* 2019;9:1303. <https://doi.org/10.7150/thno.30309>
12. Davnall F, Yip CSP, Ljungqvist G, et al. Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice? *Insights Imaging* 2012;3:573. <https://doi.org/10.1007/s13244-012-0196-6>
13. Wasserthal J, Breit HC, Meyer MT, et al. TotalSegmentator: Robust segmentation of 104 anatomic structures in CT images. *Radiol Artif Intell* 2023;5. <https://doi.org/10.1148/ryai.230024>
14. Van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res* 2017;77:e104. <https://doi.org/10.1158/0008-5472.CAN-17-0339>
15. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;26:565. <https://doi.org/10.1177/0272989X06295361>
16. Kwong JCC, Khondker A, Lajkosz K, et al. APPRAISE-AI tool for quantitative evaluation of AI studies for clinical decision support. *JAMA Netw Open* 2023;6:e2335377. <https://doi.org/10.1001/jamanetworkopen.2023.35377>
17. (17)Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* 2024;385. <https://doi.org/10.1136/bmj-2023-078378>
18. Messer JC, Terrell JD, Herman MP, et al. Multi-institutional validation of the ability of preoperative hydronephrosis to predict advanced pathologic tumor stage in upper-tract urothelial carcinoma. *Urol Oncol* 2013;31:904-8. <https://doi.org/10.1016/j.urolonc.2011.07.011>
19. Alqahtani A, Bhattacharjee S, Almopfi A, et al. Radiomics-based machine learning approach for the prediction of grade and stage in upper urinary tract urothelial carcinoma: a step towards virtual biopsy. *Int J Surg* 2024;110:3258-68. <https://doi.org/10.1097/JS9.0000000000001483>
20. Zheng Y, Shi H, Fu S, et al. A computed tomography urography-based machine learning model for predicting preoperative pathological grade of upper urinary tract urothelial carcinoma. *Cancer Med* 2024;13:6901. <https://doi.org/10.1002/cam4.6901>
21. Zheng Y, Shi H, Fu S, et al. Development and validation of a radiomics-based nomogram for predicting pathological grade of upper urinary tract urothelial carcinoma. *BMC Cancer* 2024 24:1. <https://doi.org/10.1186/s12885-024-13325-z>
22. Alqahtani A, Bhattacharjee S, Almopfi A, et al. Radiomics-based computed tomography urogram approach for the prediction of survival and recurrence in upper urinary tract urothelial carcinoma. *Cancers* 2024;16:3119. <https://doi.org/10.3390/cancers16183119>
23. Al Mopfi A, Alqahtani A, Alshehri AHD, et al. Evaluating the predictive capability of radiomics features of perirenal fat in enhanced CT images for staging and grading of UTUC tumours using machine learning. *Cancers* 2025;17:1220. <https://doi.org/10.3390/cancers17071220>
24. Laukhtina E, Muin D, Shariat SF. Imaging for upper tract urothelial carcinoma: Update of the evidence and a glimpse into the future. *Curr Opin Urol* 2025;35:103-7. <https://doi.org/10.1097/MOU.0000000000001241>
25. Pai S, Bontempi D, Hadzic I, et al. Foundation model for cancer imaging biomarkers. *Nat Mach Intell* 2024;6:354-67. <https://doi.org/10.1038/s42256-024-00807-9>

CORRESPONDENCE: Dr. Masoom A. Haider, Joint Department of Medical Imaging, Princess Margaret Hospital, Sinai Health System, University of Toronto, Toronto, ON, Canada; m.haider@utoronto.ca