Original Article

# Development of a machine learning-based predictive model for transitional cell carcinoma of the renal pelvis in White Americans: a SEER-based study

**Zhenyu Liu[1], Hang Ma[2], Ziqi Guo[3], Shuai Su[1], Xiangbiao He[1]**

[1]Department of Urology, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China; [2]Department of Rheumatology and Immunology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China; [3]Department of Urology, The First People Hospital of Lingbao, Lingbao, China

*Contributions:* (I) Conception and design: Z Liu, H Ma, S Su, X He; (II) Administrative support: Z Liu, S Su, X He; (III) Provision of study materials or patients: Z Liu, S Su, X He; (IV) Collection and assembly of data: Z Liu, H Ma, Z Guo, X He; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

*Correspondence to:* Shuai Su, PhD; Xiangbiao He, PhD. Department of Urology, The First Affiliated Hospital of Chongqing Medical University, No. 1 Youyi Road, Yuanjiagang, Yuzhong District, Chongqing 400016, China. Email: sushuai930809@163.com; hxblyy1989@sina.com.

**Background:** Transitional cell carcinoma (TCC) of the renal pelvis is a rare cancer within the urinary system. However, the prognosis is not entirely satisfactory. This study aims to develop a clinical model for predicting cancer-specific survival (CSS) at 1-, 3-, and 5-year for White Americans with renal pelvic TCC.

**Methods:** Data of all White American patients diagnosed with TCC of the renal pelvis from 2010 to 2015 were extracted and analyzed from the Surveillance, Epidemiology, and End Results (SEER) database in this retrospective study. Subsequently, after excluding the metastatic group, a subgroup analysis was performed on the data of 1,715 White Americans with non-metastatic renal pelvic TCC. Patients included in this study were randomly divided into the training and validation sets in a ratio of 7:3. In addition, the features in the training set were extracted by the Boruta algorithm. The importance of these features was visualized using the eXtreme Gradient Boosting (XGBoost)-based SHapley Additive exPlanation (SHAP) tool. To improve predictive accuracy, a nomogram model with these identified independent prognostic variables was developed.

**Results:** A total of 1,887 White American patients with renal pelvic TCC were included in this study. In the training set, the area under the curve (AUC) for CSS nomograms at 1-, 3-, and 5-year were 0.813 [95% confidence interval (CI): 0.774–0.852], 0.738 (95% CI: 0.702–0.774), and 0.733 (95% CI: 0.698–0.768), respectively. Correspondingly, the AUCs for CSS nomograms at the above time points were 0.781 (95% CI: 0.732–0.830), 0.785 (95% CI: 0.741–0.829), and 0.775 (95% CI: 0.729–0.820) in the validation set, respectively. The subgroup analysis results revealed that the AUCs for CSS nomograms at 1-, 3-, and 5-year were 0.788, 0.725, and 0.726 in the training set, respectively, while the AUCs for CSS nomograms at 1-, 3-, and 5-year were 0.831, 0.786, and 0.754 in the training set, respectively.

**Conclusions:** In this study, a nomogram that predicts CSS in White American patients diagnosed with renal pelvic TCC was efficiently constructed. The application of the nomogram may enhance patient care and assist clinicians in choosing the optimal treatment strategies.

**Keywords:** Surveillance, Epidemiology, and End Results (SEER); renal pelvic transitional cell carcinoma (renal pelvic TCC); cancer-specific survival (CSS); nomogram; SHapley Additive exPlanation (SHAP)

Submitted Aug 01, 2024. Accepted for publication Dec 03, 2024. Published online Dec 28, 2024.
doi: 10.21037/tau-24-385
**View this article at:** https://dx.doi.org/10.21037/tau-24-385

## Introduction

Upper urinary tract urothelial carcinoma (UTUC), affecting the renal pelvis and ureter, comprises roughly 5–10% of all urothelial cancers (1). The incidence of UTUC peaks in the population aged 70 to 90 years and is twice as high as in the rest of the population (2,3). In UTUC, common mutated genes include *FGFR3*, *KMT2D*, *KMT2A*, and *TP53*. Additionally, mutations in other oncogenes and tumor suppressor genes are also present, including *HRAS*, *NRAS*, *KRAS*, *ARID1A*, *PIK3CA*, and *CDKN2A* (4,5). Moreover, UTUC affects the renal pelvis more frequently than the ureter (6). An estimated 90% of all renal pelvic cancers and over 90% of ureteral cancers are attributed to transitional cell carcinoma (TCC) (7). Risk factors for TCC include smoking, occupational carcinogens, analgesic abuse, and Balkan endemic nephropathy (1). Despite its relatively low incidence, the prognosis of TCC is unfavorable.

The American Joint Committee on Cancer (AJCC) Staging Manual, updated every 5–7 years, is dedicated to providing valuable resources for cancer caregivers and physicians in the fight against cancer (8). The staging manual is constructed upon the principles of dynamic vision, international collaboration, and cancer classification. Moreover, it discusses the prognostic factors for various cancers in detail and views current clinically relevant cancer characteristics more comprehensively (8,9). Studies demonstrate that nomograms provide advantages over traditional AJCC staging for numerous cancer types (10,11). The nomogram is tailored to the patient's profile and can create a single numerical estimate for the likelihood of events, such as death or recurrence, using statistical prediction models (12). Despite the existence of current prognostic models for renal pelvic TCC (13), this study stands out in two key ways. Firstly, the feature selection was performed using the Boruta algorithm, and the significance of each feature was then visualized using the eXtreme Gradient Boosting (XGBoost)-based SHapley Additive explanation (SHAP) tool. Secondly, previous studies have overlooked the potential impact of racial differences on the prognosis of cancer patients. Therefore, this study constructed a nomogram for White American patients suffering from renal pelvic TCC.

Currently, competing risk analysis has been applied extensively in investigating various cancers (14-16). Nomogram is an effective and convenient tool for quantifying different biological and clinical variables (17,18). The objective of the current study was to determine the factors affecting the outcome of renal pelvic TCC among White Americans by competing risk analysis. Furthermore, this study sought to develop a concise model and construct a competing risk nomogram that individually assesses the risk of renal pelvic TCC. We present this article in accordance with the TRIPOD reporting checklist (available at https://tau.amegroups.com/article/view/10.21037/tau-24-385/rc).

## Methods

### *Data collection*

Patient data were collected using SEERStat software (version 8.4.3) from the database of Surveillance, Epidemiology, and End Results (SEER). From 2010 to 2015, White Americans diagnosed with TCC in the renal pelvis [International Classification of Diseases for Oncology, 3rd Edition (ICD-O-3): 8120, 8122, 8130, and 8131] were included in this study. The criteria for exclusion included: (I) missing data on grade, T staging, N staging, tumor size, surgery, lymph node dissection, survival time, race, and metastasis; (II) only lymph node biopsy was performed; (III) survival time was 0 month; (IV) The T stage and N stage could not be evaluated; (V) the first malignant primary indicator was not renal pelvic TCC; (VI) non-White American. This study encompassed a total of 1,887 participants, who were subsequently allocated randomly into two distinct groups (in a ratio of 7:3): a training set and a validation set. The prognostic factors that independently affect cancer-specific survival (CSS) were investigated. A prognostic nomogram was subsequently formulated and validated. Considering that metastatic TCC is typically regarded as a separate group. Hence, after excluding the metastatic group (n=172), 1,715 participants were included to repeat the aforementioned processes. Moreover, a nomogram for the non-metastatic group was established and validated. The detailed research process is illustrated in *Figure 1*.

For each patient, the following data were gathered: age, gender, race, year of diagnosis, T staging, N staging, tumor size, grade, histological type, surgery, lymph nodes, radiotherapy, chemotherapy, site of metastases, cause of death, and survival time. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

### *Feature selection*

Before conducting the multivariate analysis of the

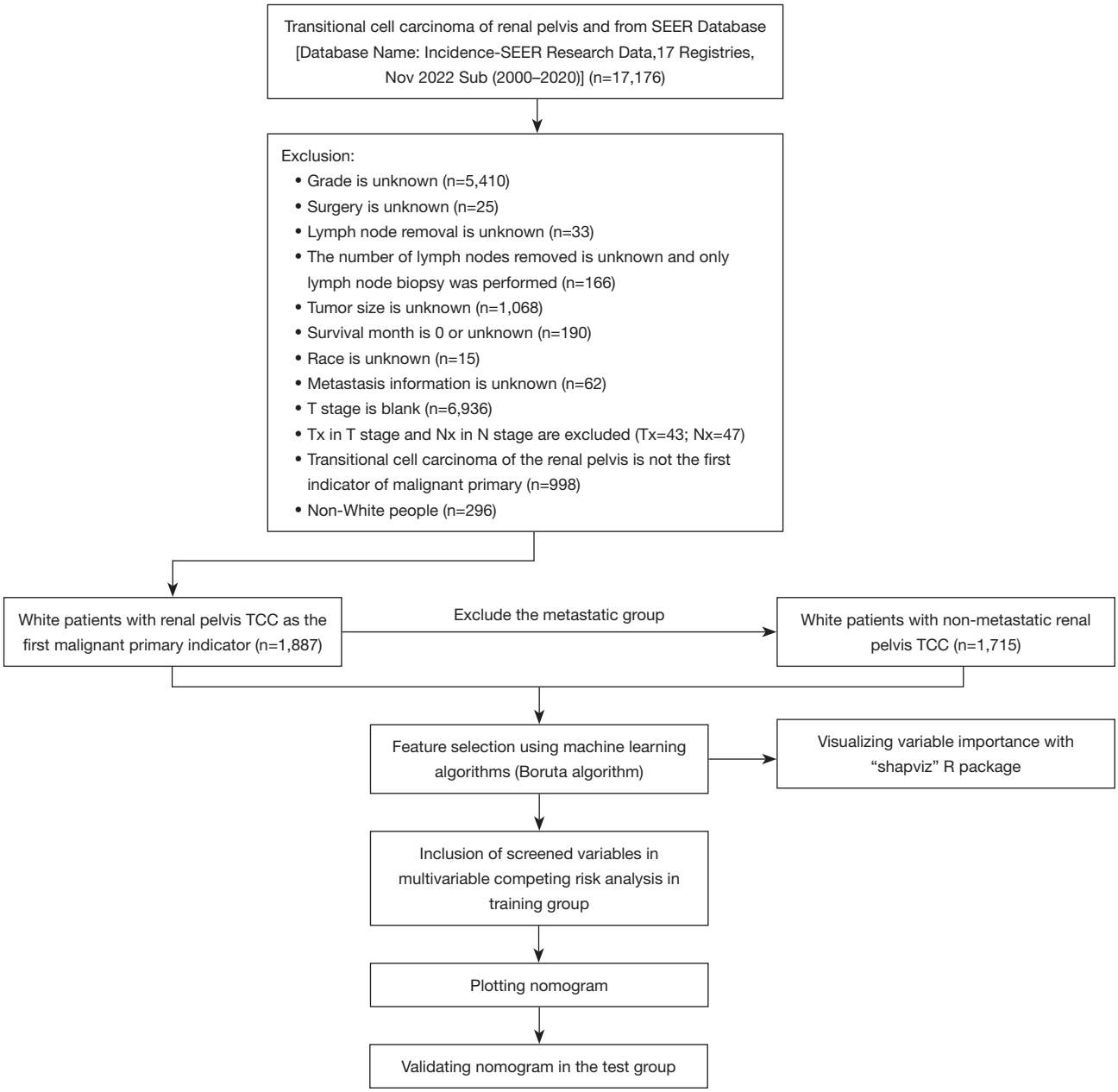


**Figure 1** The flowchart of this study. SEER, surveillance epidemiology, and end results; TCC, transitional cell carcinoma.

competing risks model, a feature selection using machine learning algorithms (mainly the Boruta algorithm) was performed to determine the importance of features within the prognostic model. The Boruta algorithm is a widely utilized feature selection technique based on the random forest classifier method. Moreover, Boruta replicated the features from the original dataset to compare the z-scores between the real feature and the shadow feature. Notably, the shadow features were produced during each iteration of the model development by the random forest classifier. A feature was considered essential and was kept if its z-score was higher than the highest z-score of the shadow features. Otherwise, the feature was eliminated (19). Subsequently, the features were further interpreted using the XGBoost-based SHAP tool to visualize the variable importance (20). XGBoost is an enhanced version of the gradient boosting

 

technique. Its foundation lies in a set of weak learners, and it is characterized by low variance and high bias. XGBoost approximates the value of the loss function by employing a second-order Taylor series. Besides, it can further reduce the likelihood of overfitting by regularization (21). SHAP was an explanatory framework proposed by Lundberg and Lee to explain the black-box model in 2017 (22). SHAP gained significant popularity in explaining models associated with medical and societal phenomena (23). SHAP, based on the computation of Shapley value, is an important coalition game theoretic approach for measuring how features affect the dependent variable. Briefly, this approach treats every feature as a contributor, computes their individual contribution values, and aggregates these values to get the final prediction of the model (24).

### *Statistical analysis*

Statistical analysis was performed using R (www.r-project.org, version 4.3.3). A two-sided P value less than 0.05 was considered statistically significant. The independent prognostic factors were identified by the Boruta algorithm and multivariate analysis. Moreover, the performance of the nomograms was assessed using receiver operating characteristic curves (ROCs). A calibration curve was subsequently established to compare the agreement between the nomogram predictions and the real results.

## Results

### *Baseline characteristics of patients*

*Table 1* presents the baseline characteristics of 1,887 White Americans with renal pelvic TCC (including the metastatic and non-metastatic groups) and 1,715 White Americans with non-metastatic renal pelvic TCC. Moreover, the mean age of 1,887 patients in the training set and validation set was 70.7±11.3 and 70.9±11.2 years, respectively. In the training and validation sets, the mean age of patients with non-metastatic renal pelvic TCC was 71.0±11.3 and 70.5±11.1 years, respectively. The majority of 1,887 patients had a grade IV tumor (training set: 58.5%; validation set: 61.9%). Lung was the most common site of metastasis for TCC in the renal pelvis both in the training and validation sets. Therapeutically, most patients underwent surgical intervention without regional lymph node dissection, radiotherapy, or chemotherapy.

### *Screening of prognostic factors*

Feature selection was conducted using the Boruta algorithm, and the results are visualized in *Figure 2*. It was observed that regional lymph node dissection, grade, radiotherapy, surgery, lung metastasis, tumor size, bone metastasis, liver metastasis, T staging, N staging, and age were associated with CSS of 1,887 patients (Table S1). It was also noticed that chemotherapy, grade, radiotherapy, regional lymph node dissection, tumor size, T staging, N staging, and age were factors affecting CSS in non-metastatic patients (Table S2). Subsequently, the XGBoost-based SHAP tool was utilized to visualize the importance of the variables (*Figure 3*), and the ROC curves of the two models are presented in Figures S1,S2. The area under the ROC curves (AUCs) were 0.841 [95% confidence interval (CI): 0.821–0.862] and 0.834 (95% CI: 0.812–0.856) respectively, proving that the models had high credibility. All CSS-related factors were included in a multivariate analysis of competing risk models. The results revealed that T staging, N staging, age, and tumor size were identified as independent CSS-related factors of 1,887 White Americans with renal pelvic TCC (*Table 2*). Additionally, the results also indicated that T staging, N staging, age, and chemotherapy were independent factors linked to CSS in White Americans with non-metastatic renal pelvic TCC (*Table 2*).

### *Construction and validation of CSS nomograms*

CSS prognostic nomograms were constructed by integrating the pertinent independent prognostic factors (*Figure 4*). The results from the data on metastatic and non-metastatic renal pelvic TCC in White Americans demonstrated that the 1-, 3-, and 5-year AUCs for the training set were 0.813, 0.738, and 0.733, respectively. Correspondingly, in the validation set, the AUCs were 0.781, 0.785, and 0.775, respectively (*Figure 5A,5B*). The training set yielded a C-index value of 0.747, whereas the validation set exhibited a C-index value of 0.805. The results from the data on non-metastatic renal pelvic TCC in White Americans demonstrated that 1-, 3-, and 5-year AUCs in the training set were 0.788, 0.725, and 0.726, respectively. In the validation set, the AUCs were 0.831, 0.786, and 0.754, respectively (*Figure 5C,5D*). Additionally, the calibration curves of the training and validation sets for both models exhibited a good consistency between the predicted and the actual results (Figures S3,S4).

**Table 1** Baseline information of the 1,887 White American patients with renal pelvic TCC (the metastatic and non-metastatic groups were included) and 1,715 White American patients with nonmetastatic renal pelvic TCC

| Characteristics | 1,887 White American patients with renal pelvic TCC | | | 1,715 White American patients with nonmetastatic renal pelvic TCC | | |
|---|---|---|---|---|---|---|
| | Training cohort (n=1,320) | Validation cohort (n=567) | P value | Training cohort (n=1,200) | Validation cohort (n=515) | P value |
| Sex | | | 0.99 | | | 0.98 |
| Female | 566 (42.9) | 244 (43.0) | | 510 (42.5) | 220 (42.7) | |
| Male | 754 (57.1) | 323 (57.0) | | 690 (57.5) | 295 (57.3) | |
| Age, years | 70.7±11.3 | 70.9±11.2 | 0.73 | 71.0±11.3 | 70.5±11.1 | 0.41 |
| Histology | | | 0.58 | | | 0.30 |
| 8120/3: TCC, NOS | 602 (45.6) | 273 (48.1) | | 545 (45.4) | 254 (49.3) | |
| 8122/3: TCC, spindle cell and 8131/3: TCC, micropapillary | 14 (1.06) | 5 (0.88) | | 12 (1.00) | 6 (1.17) | |
| 8130/3: papillary TCC | 704 (53.3) | 289 (51.0) | | 643 (53.6) | 255 (49.5) | |
| Grade | | | 0.31 | | | 0.12 |
| Well differentiated: grade I | 52 (3.94) | 22 (3.88) | | 50 (4.17) | 22 (4.27) | |
| Moderately differentiated: grade II | 166 (12.6) | 55 (9.70) | | 161 (13.4) | 56 (10.9) | |
| Poorly differentiated: grade III | 330 (25.0) | 139 (24.5) | | 265 (22.1) | 139 (27.0) | |
| Undifferentiated; anaplastic: grade IV | 772 (58.5) | 351 (61.9) | | 724 (60.3) | 298 (57.9) | |
| T stage | | | 0.27 | | | 0.54 |
| T1 | 418 (31.7) | 159 (28.0) | | 393 (32.8) | 159 (30.9) | |
| T2a/b | 141 (10.7) | 54 (9.52) | | 127 (10.6) | 61 (11.8) | |
| T3a/b/c/NOS | 590 (44.7) | 278 (49.0) | | 564 (47.0) | 236 (45.8) | |
| T4 | 171 (13.0) | 76 (13.4) | | 116 (9.67) | 59 (11.5) | |
| N stage | | | 0.24 | | | 0.03 |
| N0 | 1,103 (83.6) | 458 (80.8) | | 1,059 (88.2) | 444 (86.2) | |
| N1 | 130 (9.85) | 60 (10.6) | | 74 (6.17) | 49 (9.51) | |
| N2/N3 | 87 (6.59) | 49 (8.64) | | 67 (5.58) | 22 (4.27) | |
| Surgery | | | 0.31 | | | 0.84 |
| No | 75 (5.68) | 25 (4.41) | | 34 (2.83) | 13 (2.52) | |
| Yes | 1,245 (94.3) | 542 (95.6) | | 1,166 (97.2) | 502 (97.5) | |
| The number of regional lymph nodes removed | | | 0.66 | | | 0.59 |
| None | 949 (71.9) | 397 (70.0) | | 870 (72.5) | 369 (71.7) | |
| 1 to 3 regional lymph nodes removed | 214 (16.2) | 95 (16.8) | | 181 (15.1) | 87 (16.9) | |
| 4 or more regional lymph nodes removed | 157 (11.9) | 75 (13.2) | | 149 (12.4) | 59 (11.5) | |

**Table 1** (*continued*)

**Table 1** (*continued*)

| Characteristics | 1,887 White American patients with renal pelvic TCC | | | 1,715 White American patients with nonmetastatic renal pelvic TCC | | |
|---|---|---|---|---|---|---|
| | Training cohort (n=1,320) | Validation cohort (n=567) | P value | Training cohort (n=1,200) | Validation cohort (n=515) | P value |
| Radiation | | | 0.53 | | | 0.49 |
| No/unknown | 1,265 (95.8) | 539 (95.1) | | 1,165 (97.1) | 496 (96.3) | |
| Yes | 55 (4.17) | 28 (4.94) | | 35 (2.92) | 19 (3.69) | |
| Chemotherapy | | | >0.99 | | | >0.99 |
| No/unknown | 1,028 (77.9) | 441 (77.8) | | 975 (81.2) | 418 (81.2) | |
| Yes | 292 (22.1) | 126 (22.2) | | 225 (18.8) | 97 (18.8) | |
| Metastasis at bone | | | 0.92 | | | NA |
| No | 1,280 (97.0) | 551 (97.2) | | NA | NA | |
| Yes | 40 (3.03) | 16 (2.82) | | NA | NA | |
| Metastasis at brain | | | 0.30 | | | NA |
| No | 1,311 (99.3) | 566 (99.8) | | NA | NA | |
| Yes | 9 (0.68) | 1 (0.18) | | NA | NA | |
| Metastasis at liver | | | 0.81 | | | NA |
| No | 1,278 (96.8) | 547 (96.5) | | NA | NA | |
| Yes | 42 (3.18) | 20 (3.53) | | NA | NA | |
| Metastasis at lung | | | 0.78 | | | NA |
| No | 1,261 (95.5) | 544 (95.9) | | NA | NA | |
| Yes | 59 (4.47) | 23 (4.06) | | NA | NA | |
| Tumor size (mm) | 46.3±34.5 | 47.9±28.3 | 0.30 | 44.0±32.9 | 45.3±28.7 | 0.44 |
| Survival months | 52.9±38.4 | 50.7±38.6 | 0.27 | 56.0±37.5 | 56.6±38.3 | 0.75 |
| The cause of death | | | 0.45 | | | 0.78 |
| Alive | 519 (39.3) | 206 (36.3) | | 499 (41.6) | 218 (42.3) | |
| TCC | 272 (20.6) | 126 (22.2) | | 219 (18.2) | 99 (19.2) | |
| Other | 529 (40.1) | 235 (41.4) | | 482 (40.2) | 198 (38.4) | |

Data are presented as mean ± standard deviation or n (%). TCC, transitional cell carcinoma; NOS, not otherwise specified; NA, not applicable.

## Discussion

Urogenital cancer poses a significant threat to human health, bringing a heavy burden on global healthcare systems (25). Despite advancements in surgical, radiotherapy, and chemotherapy treatments, and emerging methods such as nano-drug therapy for urological cancers, the morbidity and mortality rates remain high in developed countries/regions (26,27). UTUC is a rare urological tumor and there are fewer studies on the genomic and transcriptomic features of UTUC. The current study suggests that *FGFR3* is the most frequently mutated and/or amplified gene in both high- and low-grade UTUCs (28). Additionally, mutations in genes such as *KMT2D*, *KMT2A*, and *TP53* are also present in UTUC (4,5). Fujii *et al.* classify the mutation subtypes of UTUC into five categories: hypermutated, *TP53/*

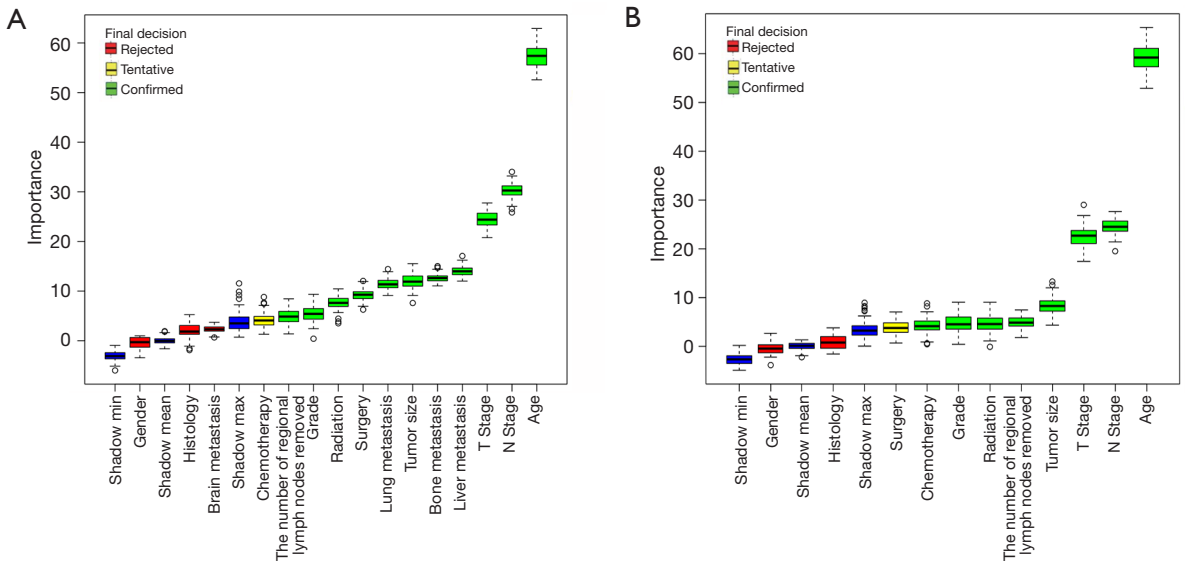


**Figure 2** Visualization of the results of the Boruta algorithm. (A) Visualization of the results of the Boruta algorithm based on 1,887 patients; (B) Visualization of the results of the Boruta algorithm based on 1,715 patients.

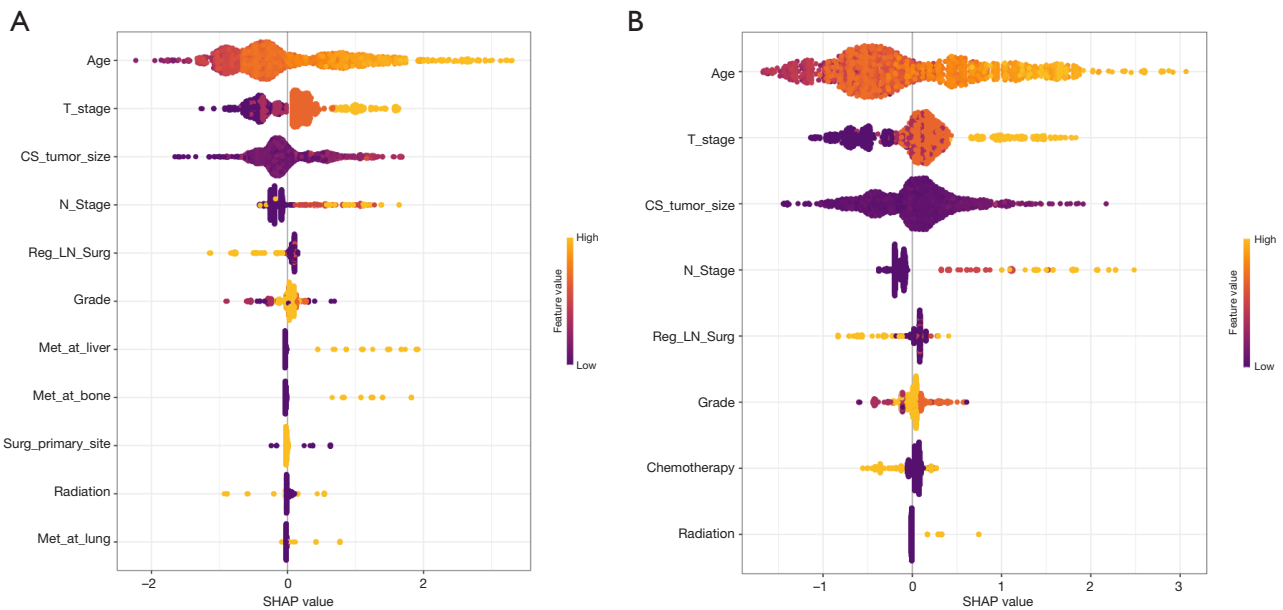


**Figure 3** The importance of the variables. (A) The importance of the variables based on 1,887 patients; (B) the importance of the variables based on 1,715 non-metastatic renal pelvic TCC patients. TCC, transitional cell carcinoma; CS tumor size, code the specific tumor size; Reg_LN_surg, regional lymph nodes surgery; Met, metastasis; surg_primary_site, surgery of primary site; SHAP, SHapley Additive exPlanation.

**Table 2** Multivariable competing risk analysis for renal pelvic TCC

| Variable | Metastatic group and non-metastatic group (n=1,887) | | Non-metastatic group (n=1,715) | |
|---|---|---|---|---|
| | SHR (95% CI) | P value | SHR (95% CI) | P value |
| Regional lymph node dissection | | | | |
| None | | | | |
| 1 to 3 regional lymph nodes removed | 1.079 (0.765, 1.521) | 0.67 | 0.829 (0.560, 1.229) | 0.35 |
| 4 or more regional lymph nodes removed | 0.903 (0.591, 1.379) | 0.64 | 0.663 (0.400, 1.100) | 0.11 |
| Grade | | | | |
| Well differentiated: grade I | | | | |
| Moderately differentiated: grade II | 1.376 (0.476, 3.977) | 0.56 | 1.069 (0.406, 2.817) | 0.89 |
| Poorly differentiated: grade III | 1.539 (0.578, 4.101) | 0.39 | 0.940 (0.370, 2.387) | 0.90 |
| Undifferentiated; anaplastic: grade IV | 1.645 (0.630, 4.295) | 0.31 | 1.066 (0.432, 2.631) | 0.89 |
| T stage | | | | |
| T1 | | | | |
| T2a/b | 1.673 (0.993, 2.819) | 0.053 | 1.733 (0.944, 3.184) | 0.08 |
| T3a/b/c/NOS | 2.205 (1.531, 3.176) | <0.001* | 2.811 (1.838, 4.297) | <0.001* |
| T4 | 3.514 (2.253, 5.480) | <0.001* | 5.545 (3.237, 9.497) | <0.001* |
| N stage | | | | |
| N0 | | | | |
| N1 | 2.407 (1.658, 3.493) | <0.001* | 3.293 (2.006, 5.405) | <0.001* |
| N2/N3 | 1.537 (0.927, 2.550) | 0.10 | 2.390 (1.311, 4.356) | 0.004* |
| Surgery | | | | |
| No | | | | |
| Yes | 0.579 (0.335, 1.003) | 0.051 | NA | NA |
| Radiation | | | | |
| No/Unknown | | | | |
| Yes | 1.113 (0.670, 1.849) | 0.68 | 1.181 (0.569, 2.451) | 0.65 |
| Metastasis at bone | | | | |
| No | | | | |
| Yes | 1.864 (0.993, 3.500) | 0.052 | NA | NA |
| Metastasis at liver | | | | |
| No | | | | |
| Yes | 0.888 (0.464, 1.698) | 0.72 | NA | NA |
| Metastasis at lung | | | | |
| No | | | | |
| Yes | 1.556 (0.911, 2.657) | 0.11 | NA | NA |
| Tumor size (mm) | 1.003 (1.001, 1.005) | 0.009* | 1.000 (0.997, 1.003) | 0.99 |
| Age (years) | 1.027 (1.015, 1.040) | <0.001* | 1.019 (1.005, 1.034) | 0.007* |
| Chemotherapy | | | | |
| No | | | | |
| Yes | NA | NA | 0.677 (0.459, 0.998) | 0.049* |

*, statistically significant (P<0.05). TCC, transitional cell carcinoma; SHR, subdistribution hazard ratio; CI, confidence interval; NOS, not otherwise specified; NA, not applicable.

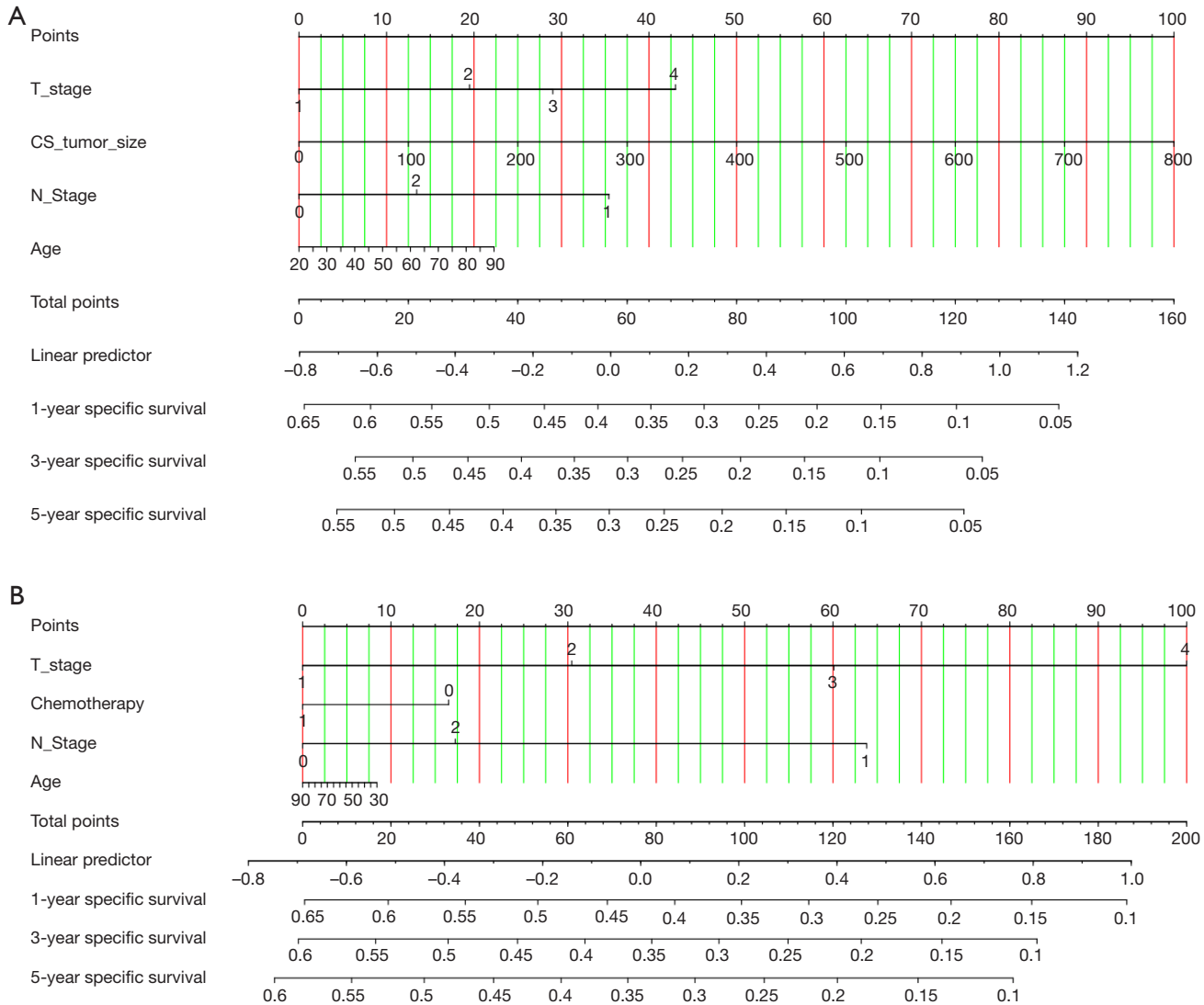


**Figure 4** Renal pelvic TCC of the CSS prognostic nomogram (A) Metastatic and non-metastatic renal pelvic TCC of the CSS prognostic nomogram; (B) non-metastatic renal pelvic TCC of the CSS prognostic nomogram. TCC, transitional cell carcinoma; CSS, cancer-specific survival; CS tumor size, code the specific tumor size.

*MDM2* mutations, *RAS* mutations, *FGFR3* mutations, and triple-negative. This classification aids in the diagnosis and management of UTUC (6). In addition to studies targeting the UTUC gene level, an effective predictive model is essential for steering cancer prevention and treatment. A previous study developed prognostic models to predict outcomes in patients with renal pelvic TCC (13). However, the previous study neglected to explore the race-specific prognosis of renal pelvic TCC. This study, based on machine learning, not only establishes and validates prognostic nomograms for all White Americans with renal pelvic TCC in the SEER database from 2010 to 2015, but also creates and validates prognostic nomograms for non-metastatic renal pelvic TCC during the same period. The two nomograms each include four variables: age, T staging, N staging, and tumor size for the renal pelvic TCC group, and age, T staging, N staging, and chemotherapy for the non-metastatic renal pelvic TCC group, demonstrating favorable predictive performance for survival outcomes in these two populations.

Multifactorial analysis indicated that age notably impacts the prognosis of White Americans diagnosed

 

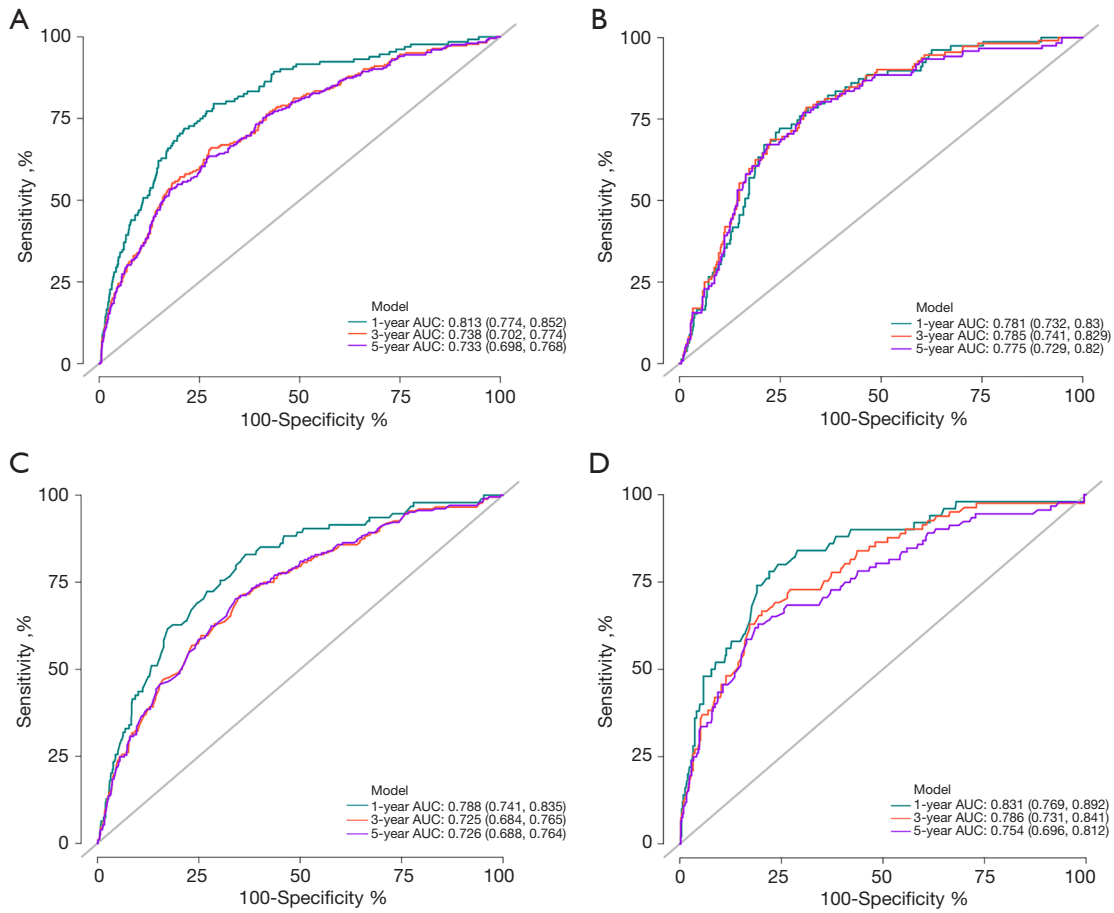


**Figure 5** ROC curves. (A) ROC curve of training set based on metastatic and non-metastatic renal pelvic TCC patients; (B) ROC curve of validation set based on metastatic and non-metastatic renal pelvic TCC patients; (C) ROC curve of training set based on non-metastatic renal pelvic TCC patients; (D) ROC curve of validation set based on non-metastatic renal pelvic TCC patients. ROC, receiver operating characteristic; TCC, transitional cell carcinoma; AUC, area under the curve.

with either renal pelvic TCC or non-metastatic renal pelvic TCC. Ferro *et al.* discovered that older patients (≥70 years old), who underwent radical nephroureterectomy for UTUC, had a worse prognosis compared to younger patients (29). The incidence of cancer is substantially evaluated with advancing age, and cancer stands as the primary cause of mortality among individuals aged 60–79 years. Moreover, age is strongly associated with the prognosis of patients diagnosed with cancer (30,31). Cancer and aging share a time-dependent accumulation of cellular damage, which may explain why older age is related to a higher cancer incidence. Additionally, aging may promote tumor progression and metastasis by affecting normal cells in the tumor microenvironment (30). Chemotherapy is one of the effective cancer treatments, and the current study found that patients receiving chemotherapy had a higher survival rate. Moreover, gemcitabine-platinum combination chemotherapy, administered within 90 days post-nephroureterectomy, has been found to markedly enhance the disease-free survival rates among individuals diagnosed with locally advanced UTUC (32). Cisplatin is the cornerstone of most urothelial cancer chemotherapy regimens (33). It is widely accepted that the primary anticancer mechanism of cisplatin involves binding to DNA through the creation of both intra-strand and inter-strand cross-links. This binding could trigger cell damage, prevent the cell cycle, and ultimately induce apoptosis in rapidly proliferating tumor cells (34,35). Multivariate analysis also reveals that in the population not excluding metastatic renal pelvic TCC, CSS decreases with increasing tumor

 

size. However, in the non-metastatic renal pelvic TCC population, tumor size is not linked to CSS. A previous retrospective study found that tumor size did not correlate with CSS (36). Another study found that tumor size was a key predictor of muscle-invasive or non-organ-localized urothelial carcinoma of the renal pelvis (37). Our subgroup analysis results suggest that distinguishing between metastatic and non-metastatic tumors may help explain the differences observed in investigating the relationship between tumor size and CSS in patients. Furthermore, tumor size is associated with a poor prognosis, possibly because it has been determined to be related to aggressive biological behavior such as tumor multifocality, tumor necrosis, and lymph node metastasis (38).

The feature selection results from the Boruta algorithm indicate that T staging, N staging, bone metastasis, liver metastasis, and lung metastasis are linked to CSS in all renal pelvic TCC patients included in this study. Multivariate analysis revealed that the T staging and N staging were strong predictors of TCC in the renal pelvis. Regarding distant metastatic disease, this study found that individuals with bone metastases exhibited a poor prognosis. Deuker *et al.* recommended a routine skeletal examination for UTUC patients, as bone metastases ranked as the second and fourth most prevalent metastatic sites in males and females, respectively (39). Moreover, bone metastases are most common in breast and prostate cancers (40). The "seed and soil" hypothesis may provide a compelling explanation for the occurrence of bone metastasis. It conceptualizes the bone microenvironment, comprising osteoclasts, osteoblasts, various bone marrow cells, the bone matrix, and the vascular network, as a fertile soil. When the bone matrix is disrupted, this environment releases cytokines that nourish the "seeds" (tumor cells) with abundant nutrients, promoting their proliferation and accelerating bone destruction (41,42). Certainly, we cannot ignore the impact of liver metastasis and lung metastasis on the prognosis of patients with TCC in the renal pelvis. The result of the Boruta algorithm analysis also demonstrated that liver and lung metastases were related to poor prognosis, and this result is consistent with previous studies (39,43).

This study introduced a pioneering approach by integrating machine learning with multivariate analysis to develop a nomogram specific to TCC of the renal pelvis in White Americans. Although the nomogram model showed high accuracy, there were some limitations in this study. First of all, this is a retrospective study, and a selection bias may be present during patient selection. Secondly, several valuable factors, such as chemotherapy cycles, lymphatic and vascular invasion, imaging data, and specific surgical information were unavailable in the SEER database. Finally, although the robustness of the internal validation results supported the reliability of the model, it is still necessary to validate the nomogram in an external cohort. Moreover, due to the limitations of the study conditions, no external validation was performed in this study. Therefore, further calibration of the nomogram is required.

## Conclusions

In summary, age, T staging, N staging, and tumor size are found to be closely linked to the survival of all White Americans diagnosed with renal pelvic TCC included in this study. In the subgroup analysis, age, T staging, N staging, and chemotherapy are also found to be associated with White Americans diagnosed with non-metastatic renal pelvic TCC. The data accessible to the public were utilized in this study to develop and verify a novel competing risks graph model. This model was highly accurate and offered distinct advantages over relying solely on prognostic factors. Future studies with expanded sample sizes and multicenter patients should be conducted to validate our findings and improve the accuracy and generalizability of the model.

## Acknowledgments

*Funding:* This work was supported by the doctoral program of the First Affiliated Hospital of Chongqing Medical University (No. CYYY-BSYJSCXXM-202332).

## Footnote

*Reporting Checklist:* The authors have completed the TRIPOD reporting checklist. Available at https://tau.amegroups.com/article/view/10.21037/tau-24-385/rc

*Peer Review File:* Available at https://tau.amegroups.com/article/view/10.21037/tau-24-385/prf

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at https://tau.amegroups.com/article/view/10.21037/tau-24-385/coif). The authors have no conflicts of interest to declare.

*Ethical Statement:* The authors are accountable for all aspects of the work in ensuring that questions related

 

to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).



## References

1. 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.
2. Shariat SF, Favaretto RL, Gupta A, et al. Gender differences in radical nephroureterectomy for upper tract urothelial carcinoma. World J Urol 2011;29:481-6.
3. Cinque A, Capasso A, Vago R, et al. MicroRNA Signatures in the Upper Urinary Tract Urothelial Carcinoma Scenario: Ready for the Game Changer? Int J Mol Sci 2022;23:2602.
4. Evmorfopoulos K, Mitrakas L, Karathanasis A, et al. Upper Tract Urothelial Carcinoma: A Rare Malignancy with Distinct Immuno-Genomic Features in the Era of Precision-Based Therapies. Biomedicines 2023;11:1775.
5. Hassler MR, Bray F, Catto JWF, et al. Molecular Characterization of Upper Tract Urothelial Carcinoma in the Era of Next-generation Sequencing: A Systematic Review of the Current Literature. Eur Urol 2020;78:209-20.
6. Fujii Y, Sato Y, Suzuki H, et al. Molecular classification and diagnostics of upper urinary tract urothelial carcinoma. Cancer Cell 2021;39:793-809.e8.
7. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin 2024;74:12-49.
8. Amin MB, Greene FL, Edge SB, et al. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. CA Cancer J Clin 2017;67:93-9.
9. Paner GP, Stadler WM, Hansel DE, et al. Updates in the Eighth Edition of the Tumor-Node-Metastasis Staging Classification for Urologic Cancers. Eur Urol 2018;73:560-9.
10. Sternberg CN. Are nomograms better than currently available stage groupings for bladder cancer? J Clin Oncol 2006;24:3819-20.
11. Adam MA, Thomas S, Roman SA, et al. Rethinking the Current American Joint Committee on Cancer TNM Staging System for Medullary Thyroid Cancer. JAMA Surg 2017;152:869-76.
12. Wang X, Zhao M, Zhang C, et al. Establishment and Clinical Application of the Nomogram Related to Risk or Prognosis of Hepatocellular Carcinoma: A Review. J Hepatocell Carcinoma 2023;10:1389-98.
13. Hu T, You S. Overall and Cancer-Specific Survival in Patients With Renal Pelvic Transitional Cell Carcinoma: A Population-Based Study. Front Med (Lausanne) 2022;8:719800.
14. Liu J, Lyu Y, He Y, et al. Competing risk nomogram and risk classification system for evaluating overall and cancer-specific survival in neuroendocrine carcinoma of the cervix: a population-based retrospective study. J Endocrinol Invest 2024;47:1545-57.
15. Wang T, Zhou J, Zheng Q, et al. A Competing Risk Model Nomogram to Predict the Long-Term Prognosis of Lung Carcinoid. Ann Surg Oncol 2023;30:5830-9.
16. Wang R, Dai W, Gong J, et al. Development of a novel combined nomogram model integrating deep learning-pathomics, radiomics and immunoscore to predict postoperative outcome of colorectal cancer lung metastasis patients. J Hematol Oncol 2022;15:11.
17. Cui Y, Zhang J, Li Z, et al. A CT-based deep learning radiomics nomogram for predicting the response to neoadjuvant chemotherapy in patients with locally advanced gastric cancer: A multicenter cohort study. EClinicalMedicine 2022;46:101348.
18. Zhou TH, Zhou XX, Ni J, et al. CT whole lung radiomic nomogram: a potential biomarker for lung function evaluation and identification of COPD. Mil Med Res 2024;11:14.
19. Degenhardt F, Seifert S, Szymczak S. Evaluation of variable selection methods for random forests and omics data sets. Brief Bioinform 2019;20:492-503.
20. Li X, Zhao Y, Zhang D, et al. Development of an interpretable machine learning model associated with heavy metals' exposure to identify coronary heart disease among US adults via SHAP: Findings of the US NHANES from 2003 to 2018. Chemosphere 2023;311:137039.
21. Wang K, Tian J, Zheng C, et al. Interpretable prediction

of 3-year all-cause mortality in patients with heart failure caused by coronary heart disease based on machine learning and SHAP. Comput Biol Med 2021;137:104813.
22. Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. Proceedings of the 31st International Conference on Neural Information Processing Systems; 2017: Association for computing Machinery.
23. Gonzalez-Estrada A, Park MA, Accarino JJO, et al. Predicting Penicillin Allergy: A United States Multicenter Retrospective Study. J Allergy Clin Immunol Pract 2024;12:1181-1191.e10.
24. Liu Z, Luo C, Chen X, et al. Noninvasive prediction of perineural invasion in intrahepatic cholangiocarcinoma by clinicoradiological features and computed tomography radiomics based on interpretable machine learning: a multicenter cohort study. Int J Surg 2024;110:1039-51.
25. Zhang ZH, Wang Y, Zhang Y, et al. The function and mechanisms of action of circular RNAs in Urologic Cancer. Mol Cancer 2023;22:61.
26. Li C, Zeng X, Qiu S, et al. Nanomedicine for urologic cancers: diagnosis and management. Semin Cancer Biol 2022;86:463-75.
27. Zi H, He SH, Leng XY, et al. Global, regional, and national burden of kidney, bladder, and prostate cancers and their attributable risk factors, 1990-2019. Mil Med Res 2021;8:60.
28. Robinson BD, Vlachostergios PJ, Bhinder B, et al. Upper tract urothelial carcinoma has a luminal-papillary T-cell depleted contexture and activated FGFR3 signaling. Nat Commun 2019;10:2977.
29. Ferro M, Chiujdea S, Vartolomei MD, et al. Advanced Age Impacts Survival After Radical Nephroureterectomy for Upper Tract Urothelial Carcinoma. Clin Genitourin Cancer 2024;22:27-37.
30. Fane M, Weeraratna AT. How the ageing microenvironment influences tumour progression. Nat Rev Cancer 2020;20:89-106.
31. Lin L, Li Z, Yan L, et al. Global, regional, and national cancer incidence and death for 29 cancer groups in 2019 and trends analysis of the global cancer burden, 1990-2019. J Hematol Oncol 2021;14:197.
32. Birtle A, Johnson M, Chester J, et al. Adjuvant chemotherapy in upper tract urothelial carcinoma (the POUT trial): a phase 3, open-label, randomised controlled trial. Lancet 2020;395:1268-77.
33. Shvero A, Hubosky SG. Management of Upper Tract Urothelial Carcinoma. Curr Oncol Rep 2022;24:611-9.
34. Yimit A, Adebali O, Sancar A, et al. Differential damage and repair of DNA-adducts induced by anti-cancer drug cisplatin across mouse organs. Nat Commun 2019;10:309.
35. Zhang C, Xu C, Gao X, et al. Platinum-based drugs for cancer therapy and anti-tumor strategies. Theranostics 2022;12:2115-32.
36. Milenkovic-Petronic D, Milojevic B, Djokic M, et al. The impact of tumor size on outcomes in patients with upper urinary tract urothelial carcinoma. Int Urol Nephrol 2014;46:563-9.
37. Collà Ruvolo C, Nocera L, Stolzenbach LF, et al. Tumor Size Predicts Muscle-invasive and Non-organ-confined Disease in Upper Tract Urothelial Carcinoma at Radical Nephroureterectomy. Eur Urol Focus 2022;8:498-505.
38. Ma R, Liu Z, Cheng Y, et al. Prognostic Value of Tumor Size in Patients with Upper Tract Urothelial Carcinoma: A Systematic Review and Meta-analysis. Eur Urol Open Sci 2022;42:19-29.
39. Deuker M, Rosiello G, Stolzenbach LF, et al. Sex- and Age-Related Differences in the Distribution of Metastases in Patients With Upper Urinary Tract Urothelial Carcinoma. J Natl Compr Canc Netw 2021;19:534-40.
40. Clézardin P, Coleman R, Puppo M, et al. Bone metastasis: mechanisms, therapies, and biomarkers. Physiol Rev 2021;101:797-855.
41. Satcher RL, Zhang XH. Evolving cancer-niche interactions and therapeutic targets during bone metastasis. Nat Rev Cancer 2022;22:85-101.
42. Coughlin TR, Romero-Moreno R, Mason DE, et al. Bone: A Fertile Soil for Cancer Metastasis. Curr Drug Targets 2017;18:1281-95.
43. Shinagare AB, Fennessy FM, Ramaiya NH, et al. Urothelial cancers of the upper urinary tract: metastatic pattern and its correlation with tumor histopathology and location. J Comput Assist Tomogr 2011;35:217-22.

**Cite this article as:** Liu Z, Ma H, Guo Z, Su S, He X. Development of a machine learning-based predictive model for transitional cell carcinoma of the renal pelvis in White Americans: a SEER-based study. Transl Androl Urol 2024;13(12):2681-2693. doi: 10.21037/tau-24-385

 