A Machine Learning Algorithm for Survival Prediction in Bladder Cancer Patients Undergoing Cystectomy

Francesco Andrea Causio[1,2], Vittorio De Vita[1,2], Andrea Nappi[2,3], Melissa Sawaya[4], Bernardo Rocco[5,6], Nazario Foschi[5,6], Giuseppe Maioriello[5,6], Pierluigi Russo[2,5,6,7]

[1] Section of Hygiene and Public Health, Department of Life Sciences and Public Health, Università Cattolica del Sacro Cuore, Rome, Italy.

[2] Italian Society for Artificial Intelligence in Medicine (SIIAM - Società Italiana Intelligenza Artificiale in Medicina), Rome, Italy.

[3] University of Twente, Twente, The Netherlands.

[4] Université Paris-Saclay, UVSQ, Inserm, Gustave Roussy, CESP, 94805, Villejuif, France.

[5] Department of Urology, Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Largo Francesco Vito 1, 00168 Rome, Italy.

[6] Department of Medicine and Translational Surgery, Università Cattolica Del Sacro Cuore, Rome, Italy.

[7] Department of Life Science, Health, and Health Professions, Università degli Studi LINK

Corresponding Author

Francesco Andrea Causio

francescoandrea.causio@siiam.it

Abstract

Background: Traditional statistical models often fail to capture the complex dynamics influencing survival outcomes in bladder cancer patients after radical cystectomy, a procedure where approximately 50% of patients develop metastases within two years. The integration of artificial intelligence (AI) offers a promising avenue for enhancing prognostic accuracy and personalizing treatment strategies.

Objectives: This study aimed to develop and evaluate a machine learning algorithm for predicting disease-free survival (DFS), overall survival (OS), and the cause of death in patients with bladder cancer undergoing cystectomy, utilizing a comprehensive dataset of clinical and pathological variables.

Methods: Retrospective data from 370 bladder cancer patients undergoing radical cystectomy at Fondazione Policlinico Gemelli, Rome, Italy, were collected. The dataset comprised 20 input variables, encompassing demographics, tumor characteristics, treatment, and inflammatory markers. The CatBoost algorithm was employed for regression tasks (DFS in 346 patients, OS in 347 patients) and a binary classification task (tumor-related death in 312 patients). Model performance was assessed using Mean Absolute Error (MAE) for

regression and F1-score for classification, prioritizing a minimum recall of 75% for tumor-related deaths. Five-fold cross-validation and SHAP (Shapley Additive exPlanations) values were employed to ensure robustness and interpretability.

Results: For DFS prediction, the CatBoost model achieved an MAE of 18.68 months, with clinical tumor stage (CTS) and pathological tumor classification (TC) identified as the most influential predictors. Overall survival prediction yielded an MAE of 17.2 months, which improved to 14.6 months after feature filtering, where TC and the Systemic Immune-Inflammation Index (SII) were most impactful. For tumor-related death classification, the model achieved a recall of 78.6% and an F1-score of 0.44 for the positive class (tumor deaths), correctly identifying 11 out of 14 cases. Bladder tumor position was the most influential feature for the cause of death.

Conclusions: The developed machine learning algorithm demonstrates promising accuracy in predicting survival and the cause of death in bladder cancer patients after cystectomy. Key predictors include clinical and pathological tumor staging, systemic inflammation (SII), and bladder tumor position. These findings highlight the potential of AI in providing clinicians with an objective, data-driven tool to improve personalized prognostic assessment and guide clinical decision-making.

Keywords: Urology, Urinary Bladder Cancer, Artificial Intelligence, Clinical Trials

Introduction

In the evolving landscape of healthcare, the integration of artificial intelligence (AI) into clinical decision-making has gained significant momentum, particularly in the realm of oncology. (1–3) With advancements in machine learning techniques, healthcare professionals are increasingly harnessing the power of AI to enhance diagnosis, prognosis, and treatment planning. The exponential growth of digital healthcare data, including electronic health records, medical imaging, genomic data, and real-time patient monitoring, fuels this rise in predictive algorithms. (3,4)

The field of urology is complex: cancerous conditions benefit from the leverage of additional data sources and decision-making algorithms that allow physicians to plan treatment while considering several complex factors. Urological cancers, including prostate, bladder, and renal cancers, pose a considerable burden on healthcare systems globally. (5) These malignancies often require complex management involving early diagnosis, accurate staging, and personalized treatment strategies to optimize patient outcomes. Traditional methods of assessing prognosis rely heavily on statistical models that may not capture the multifaceted nature of cancer behavior and patient responses to treatment. Conventional regression statistics often fail to provide the depth of analysis required to address the complexities of cancer management. In contrast, AI techniques, such as artificial neural networks (ANNs), Bayesian networks, and neuro-fuzzy modeling systems, offer innovative approaches to constructing data-driven models that can adapt to the heterogeneous nature of cancer. (6)

The potential of AI in predicting patient outcomes is particularly evident in its ability to analyze large datasets without the constraints of predetermined statistical distributions. By leveraging retrospective data, we can develop algorithms that not only identify patterns and correlations but also provide insights into individual patient behavior. This capability is crucial for clinicians who face the challenge of tailoring treatment plans to the unique characteristics of each patient. In the context of mortality and post-operative survival, the application of AI can provide critical insights that enhance our understanding of patient outcomes following surgical interventions. The ability to predict which patients are at higher risk of complications or recurrence can lead to more informed clinical decisions, ultimately improving the quality of care. (7) For instance, machine learning algorithms can analyze a multitude of variables, including clinical, pathological, and demographic factors, to generate individualized risk profiles that guide treatment strategies and follow-up plans. (8)

In this study, we will focus specifically on the training of an AI algorithm using retrospective data collected from patients diagnosed with bladder cancer and undergoing cystectomy. Patients with localized muscle-invasive or recurrent non-muscle invasive bladder cancer benefit most from radical cystectomy, preceded in some patients by neoadjuvant chemotherapy, in terms of local control. Even with sufficient local control achieved through cystectomy, approximately 50% of patients will develop metastases within two years and may ultimately die from the disease. This is likely due to the existence of regional or distant microscopic metastatic disease at the time of surgery. (9) The proposed methodology will involve the comprehensive examination of variables associated with patient demographics,

tumor characteristics, treatment modalities, and postoperative outcomes. By employing machine learning techniques, we aim to identify key predictors of mortality and post-operative survival, ultimately constructing a model that can inform clinical practice.

Methods

Study Design and Ethical Approval

We collected retrospective data on patients with high-risk and very high-risk non-muscle invasive bladder cancer (NMIBC) and muscle-invasive bladder cancer (MIBC) undergoing radical cystectomy (RC) from Fondazione Policlinico Gemelli in Rome, Italy. The dataset included 370 patients with various clinical and pathological variables. Ethical approval was obtained from the institutional ethical review board under protocol number 676-02.

Data Collection and Preprocessing

Clinical and pathological data were extracted from medical records, including demographic, lifestyle, tumor, treatment, and laboratory variables. The dataset was split into three outcome-specific subsets to maximize the usable sample for each task:

1. DFS dataset: Predicting disease-free survival (DFS) in months (346 Patients in total)
2. OS Dataset: Predicting overall survival (OS) in months (347 Patients in total)
3. Death Cause Dataset: For classification purposes, the classes “death from other causes” and “alive” were merged into a single negative class to create a binary variable. Therefore, the cause is defined as either "No" or "Yes", depending on the tumor relation (312 Patients in total).

The variables included in the dataset are detailed in Table 1. All Categorical variables were cast as strings to allow native handling by CatBoost.

A total of 20 input variables were selected for model development.

Table 1. Variables included in the study

Variable (English)	Description	Data Type	Input/Output
Patient Demographics and Lifestyle			
AGE	Age (years)	Numerical	Input
BMI	Body mass index (kg/m^2)	Numerical	Input
SEX	Biological Sex (0: man, 1: woman)	Categorical	Input
SMOKE	Smokes (0: no, 1: yes)	Categorical	Input
Patient Medical History			
DM	Patient is affected by diabetes mellitus (0: no, 1: yes)	Categorical	Input
PRIOR SURGERY	Patient had previous surgery in the abdominal area (0: no, 1: yes)	Categorical	Input

PRIOR RADIOTHERAPY	Patient had previous radiotherapy in the abdominal area (0: no, 1: yes)	Categorical	Input
PRIOR SYSTEMIC CHEMOTHERAPY	Patient was subject to previous systemic chemotherapy (0: no, 1: yes)	Categorical	Input
Tumor Characteristics			
BLADDER TUMOR POSITION	Identifier of the tumor's position (0: inter-trigonal zone, 1: Right peri-ostial, 2: Left peri-ostial, 3: Dome, 4: Posterior wall, 5: Right lateral wall, 6: Left lateral wall, 7: Prostatic urethra, 8: Anterior wall, 9: Entire bladder, 10: Bladder base)	Categorical	Input
TUMOR DIMENSION	Tumor dimension (cm)	Numerical	Input
PRE-HYDRONEPHROSIS	Hydronephrosis (0: no, 1: right hydronephrosis, 2: left hydronephrosis, 3: bilateral hydronephrosis)	Categorical	Input
H.E. TURV	Histological exam for TURB (0: Localized to mucosa +/- submucosa multirecurrent, 1: Muscle-invasive, 2: Squamous)	Categorical	Input
LVI	Lymphovascular invasion (0: Absent, 1: Present)	Categorical	Input
CTS	Clinical tumor stage (0: cTa, 1: cTis, 2: cT1, 3: cT2, 4: cT3, 5: cT4)	Categorical	Input
TC	Tumor classification (1: T0, 2: Ta, 3: Tis, 4: T1, 5: T2a, 6: T2b, 7: T3a, 8: T3b, 9: T4a, 10: T4b)	Categorical	Input
Inflammatory and Immune Markers			
SII	Systemic Immune-Inflammation index (decimals)	Numerical	Input
Treatment and Outcome			
UD	Urinary diversion type (0: Bricker/Ileal conduit, 1: Ureterocutaneostomy, 2: Vesicolleal Pouch)	Categorical	Input
RECURRENCE	Tumor recurrence (0: no, 1: yes)	Categorical	Input
DFS	Disease-free survival period after treatment (in months)	Numerical	Output
OST	Overall survival time from diagnosis/treatment start to death from any cause (in months)	Numerical	Output
DEATH CAUSE	Cause of death (X: Alive, 1: Other, 2: Cancer) Later merged (0: Alive + Other, 1: Cancer)	Categorical	Output

Machine Learning Models

To predict clinical outcomes, we utilized the CatBoost algorithm for both regression (disease-free survival and overall survival time) and classification (cause of death) tasks, as it is effective for small and structured datasets.

For disease-free survival and overall survival, we applied CatBoost regression models. For predicting the cause of death, we used the CatBoost classifier, with the binary outcome of death being tumor-related or not.

Model Evaluation

For the regression tasks (disease-free survival and overall survival time), we evaluated model performance using mean absolute error (MAE) to quantify the average prediction error in months.

For the classification task (cause of death), we evaluated performance using the F1 score. The F1 score is a single metric that balances precision and recall, particularly useful in cases of imbalanced classes where the positive class is of primary interest. For class 1 (tumor deaths), it is calculated as the harmonic mean of precision and recall:

$$F1 = 2 * (Precision * Recall) / (Precision + Recall)$$

Precision is the proportion of correctly identified positive predictions among all positive predictions, and recall is the proportion of correctly identified positive predictions among all actual positives.

$$Precision = TP / (TP + FP)$$

$$Recall = TP / (TP + FN)$$

Confusion matrices were used to examine prediction distributions, and probability thresholds were adjusted to optimize recall while limiting false positives. To account for class imbalance in the classification task, we applied custom class weights. We adjusted the decision threshold, aiming for a minimum recall of 75% to ensure that most tumor-related deaths were accurately identified and classified.

Cross-validation and Hyperparameter Tuning

All models were trained and evaluated using 5-fold cross-validation to ensure generalizability and reduce the risk of overfitting, especially given the relatively small dataset. In addition, we applied early stopping with a patience range of 30–50 rounds, allowing the model to terminate training once performance ceased to improve on the validation fold.

To enhance the interpretability and transparency of the developed machine learning models, we utilized violin plots and SHAP (Shapley Additive exPlanations) scatter plots to investigate the impact of variables on the prediction of the results. Violin plots show the effect of each variable on the results, both in terms of direction (favorable or unfavorable) and intensity. The SHAP scatter plot assigns an importance value to each feature for a particular prediction. For each patient, SHAP values revealed the specific features driving the predicted risk of tumor death. By aggregating SHAP values across the entire cohort, the overall impact and importance of each clinical and pathological variable on the model's outcome predictions

were determined. This enabled the identification of the most significant factors influencing bladder cancer patient outcomes after cystectomy.

This article presents only the most significant results of the analysis. The complete analysis can be accessed at https://figshare.com/s/ec2419a33ad7ad05f412.

TRIPOD AI Reporting

To enhance the transparency, interpretability, and reproducibility of our machine learning-based prediction models, this study adheres to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) Statement, specifically considering the extensions for Artificial Intelligence (TRIPOD AI). The TRIPOD AI guidelines provide a standardized framework for reporting studies that develop or validate prediction models, ensuring that sufficient detail is provided for critical appraisal and replication by other researchers. By following these guidelines, we aim to clearly articulate the study design, data characteristics, model development process, and performance evaluation, thereby contributing to the responsible and rigorous application of AI in medical research.

Table 2. Tripod+AI Checklist for Reporting AI Studies

TRIPOD AI Item	Description of Reporting in This Study
Title	A Machine Learning Algorithm for Survival Prediction in Bladder Cancer.
Abstract	The abstract summarizes the study's objectives, methods, key findings, and conclusions.
Introduction - Background	The introduction will establish the clinical context of bladder cancer, the prognostic challenges, and the rationale for using machine learning.
Methods - Participants	
4a. Eligibility criteria	Patients undergoing radical cystectomy for bladder cancer, with available data for selected variables.
4b. Settings and locations	Data collected retrospectively from a single institution: Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Rome, Italy.
4c. Source of data	Patient medical records.
5. How data was acquired	Retrospective data extraction into a spreadsheet.

Methods - Outcome	
6a. Definition of outcomes	DFS (Disease-Free Survival): Time in months from treatment to recurrence or death (event), or last follow-up (censored). OST (Overall Survival Time): Time in months from diagnosis/treatment start to death from any cause (event), or last follow-up (censored). DEATH CAUSE: Binary classification (0: Not died from the tumor, 1: Died from the tumor).
6b. How outcomes were measured	DFS and OST were calculated from documented dates. The cause of death was extracted from medical records and re-categorized for binary classification.
Methods - Predictors	
7a. Definition of all predictors	AGE: patient's age (years); BMI: Body mass index (kg/m^2) DM: Patient is affected by diabetes mellitus (0: no, 1: yes) PRIOR SURGERY: Patient had previous surgery in the abdominal area (0: no, 1: yes) PRIOR RADIOTHERAPY: Patient had previous radiotherapy in the abdominal area (0: no, 1: yes) PRIOR SYSTEMIC CHEMOTHERAPY: Patient was subject to previous systemic chemotherapy (0: no, 1: yes) BLADDER TUMOR POSITION: Identifier of the tumor's position (0: intertrigonal zone, 1: Right peri-ostial, 2: Left peri-ostial, 3: Dome, 4: Posterior wall, 5: Right lateral wall, 6: Left lateral wall, 7: Prostatic urethra, 8: Anterior wall, 9: Entire bladder, 10: Bladder base) TUMOR DIMENSION: Tumor dimension (cm) PRE-HYDRONEPRHOSIS: pre-treatment hydronephrosis (0: no, 1: right hydronephrosis, 2: left hydronephrosis, 3: bilateral hydronephrosis) SEX: Biological Sex (0: man, 1: woman) SMOKE: patient smokes (0: no, 1: yes) H.E. TURV: Histological exam for TURB (0: Localized to mucosa +/- submucosa multirecurrent, 1: Muscle-invasive, 2: Squamous) SII: Systemic Immune-Inflammation index (decimals) UD: Urinary diversion type (0: Bricker/Ileal conduit, 1: Ureterocutaneostomy, 2: VesicoIleal Pouch) LVI: Lymphovascular invasion (0: Absent, 1: Present) CTS: Clinical tumor stage (0: cTa, 1: cTis, 2: cT1, 3: cT2, 4: cT3, 5: cT4) TC: Tumor classification (1: T0, 2: Ta, 3: Tis, 4: T1, 5: T2a, 6: T2b, 7: T3a, 8: T3b, 9: T4a, 10: T4b)
7b. How predictors were measured	Predictors were measured clinically (e.g., age, BMI), from patient history (e.g., prior surgeries, smoking status), or derived from laboratory or pathology reports (e.g., SII, tumor dimension, LVI, CTS, TC).

Methods - Sample Size	
8. How sample size was determined	The available retrospective data determined the sample size. No formal power calculation was performed due to the exploratory nature of the study and the limitations of the data.
Methods - Data Handling	
9a. How missing data was handled	Rows with null values in specific critical variables ('TUMOR DIMENSION', 'LVI', 'TC', 'H.E. TURV', 'RECURRENCE', 'DFS', 'OST', 'DEATH CAUSE') were removed. No imputation was performed.
9b. Data transformations	Numeric variables were type-adjusted to int or float. Categorical variables were explicitly converted to category type. 'DEATH CAUSE' was recategorized into a binary format.
Methods - Model Development	
10a. Model type	CatBoost Regressor (for DFS, OST) and CatBoost Classifier (for DEATH CAUSE).
10b. Candidate predictors	All 17 selected independent variables were used as candidate predictors for each model, based on the relevant dataset (df1, df2, df3).
10c. How continuous predictors were handled	Continuous predictors (AGE, BMI, TUMOR DIMENSION, SII) were used directly by CatBoost, which handles them internally.
10c. How categorical predictors were handled	Categorical predictors were identified and explicitly converted to string type before training. CatBoost natively handles categorical features without explicit one-hot encoding.
10e. Details of model fitting	CatBoost Regressor (iterations=1000, learning_rate=0.05, depth=6, loss_function='RMSE', eval_metric='MAE', early_stopping_rounds=50, random_seed=42, verbose=0). Similar configurations for Classifier, with 'Logloss' or 'MultiClass' as loss function.
10f. How internal validation was done	Data was split into training (80%) and testing (20%) sets using train_test_split with random_state=42. 5-fold cross-validation (KFold, shuffle=True, random_state=42) was performed on the training set.

10g. Performance metrics	Regression (DFS, OST): Mean Absolute Error (MAE). Classification (DEATH CAUSE): F1-score for class 1 (tumor deaths), prioritizing recall.
11. How prediction performance was assessed	Performance was assessed on the independent test set. For classification, a confusion matrix was used.
12. Model interpretation methods	CatBoost's built-in feature importance was used. Shapley Additive exPlanations (SHAP) values were computed and visualized using violin and scatter plots to understand the individual contributions of each feature.

Results

The study included a final cohort of 370 patients. After excluding incomplete records, the analytical sample sizes were 346 for DFS, 347 for OS, and 312 for death cause prediction.

Table 3 presents the baseline clinical, pathological, and demographic characteristics of the study population, comprising 79.4% males and 20.6% females. A majority, 79.4% were active smokers, and 21.4% had a diagnosis of diabetes mellitus. Prior surgery was reported in 33.7% of patients, while previous radiotherapy and systemic chemotherapy were less common, 5.1% and 3.5%, respectively. Pre-operative hydronephrosis was present in approximately one-third of cases, most frequently unilateral.

Histologically, 57.0% of tumors were muscle-invasive, while 37.1% were localized to the mucosa or submucosa, and 5.9% exhibited squamous features. The most common urinary diversion method was Bricker ileal conduit (74.5%), followed by Vesico-Ileal pouch (15.8%) and ureterocutaneostomy (9.7%). Lymphovascular invasion was observed in 57.4% of patients.

In terms of staging, the most frequent clinical tumor stages (CTS) were cTa (36.1%) and cT1 (21.0%), while advanced stages (cT3 and cT4) were less common (11.0%). Tumor classification (TC) was heterogeneous, with T3a (21.9%) and Tis (16.4%) being most prevalent.

Regarding tumor location, the most frequent sites were the right lateral wall (19.5%), the posterior wall (15.9%), and the left lateral wall (15.9%). At the time of data collection, 56.5% of patients were alive, 24.0% had died due to cancer-related causes, and 19.6% had died from other causes.

Table 3. Characteristics of the patients in the dataset

	Mean	**STD**	
Continuous Variables			
Age	75.2	± 9.5 years	
Body Mass Index (BMI)	26.6	± 4.2	
Tumor Dimension	2.2 cm	IQR (1.1-2.8)	
SII	654.7	IQR (408-1047)	
DFS	23 Months	IQR (6-52.75)	
OS	29.05 Months	IQR (10.8-55.4)	
Categorical Variables	**Category**	**n**	**(%)**
Sex	Male	297	(79,4)
	Female	77	(20,6)
Smoking Status	No	77	(20,6)
	Yes	296	(79,4)
Diabetes Mellitus (DM)	No	294	(78,6)
	Yes	80	(21,4)

Prior Surgery	No	248 (66,3)
	Yes	126 (33,7)
Prior Radiotherapy	No	355 (94,9)
	Yes	19 (5,1)
Prior Chemotherapy	No	361 (96,5)
	Yes	13 (3,50)
Pre-Hydronephrosis	None	257 (68,7)
	Right	44 (11,8)
	Left	40 (10,7)
	Bilateral	33 (8,8)
Histological Exam (H.E. TURV)		
	Localized	138 (37,1)
	Muscle -invasive	212 (57,0)
	Squamous	22 (5,9)
Urinary Diversion Type		
	Bricker	278 (74,5)
	Ureterocutaneosto my	36 (9,7)
	Vesico-Ileal Pouch	59 (15,8)
Lymphovascular Invasion		
	Absent	158 (42,6)
	Present	213 (57,4)
Clinical Tumor Stage (CTS)		
	cTa	132 (36,1)
	cTis	43 (11,7)
	cT1	77 (21,0)
	cT2	74 (20,2)
	cT3	31 (8,5)
	cT4	9 (2,5)
Tumor Classification (TC)		
	T0	16 (4,4)
	Ta	18 (4,9)
	Tis	60 (16,4)
	T1	47 (12,8)
	T2a	54 (14,8)
	T2b	7 (1,9)
	T3a	80 (21,9)
	T3b	10 (2,7)
	T4a	42 (11,5)
	T4b	8 (2,2)
Bladder Tumor Position		
	Intertrigonal zone	44 (12,1)
	Right Peri-ostial	25 (6,8)
	Left peri-ostial	39 (10,7)
	Dome	22 (6,0)
	Posterior wall	58 (15,9)

	Right lateral wall	71 (19,5)
	Left lateral wall	58 (15,9)
	Prostatic urethra	8 (2,2)
	Anterior wall	26 (7,1)
	Entire bladder	12 (3,3)
	Bladder Base	2 (0,5)
Cause of Death		
	Alive	205 (56,5)
	Other	71 (19,6)
	Cancer	87 (24,0)

Disease-Free Survival (DFS) Prediction

The CatBoost Regressor model was trained to predict disease-free survival (DFS) in months. Input variables are referred to in Table 1. After 5-fold cross-validation and manual hyperparameter tuning, the model achieved a mean absolute error of 18.68 months, indicating that, on average, the model's predictions deviated by approximately 1.5 years from the observed DFS.

Figure 1. Global Feature Importance Ranking to Predict DFS. CTS: Clinical Tumor Stage; TC: Tumor Classification; SII: Systemic Immune-Inflammatory Index; PRE-HYDRONEPHROSIS: pre-treatment hydronephrosis; UD: Urinary Diversion type; H.E. TURV: Histological Exam for TURB; BMI: Body-Mass Index; LVI: lymphovascular invasion; DM: diabetes mellitus.

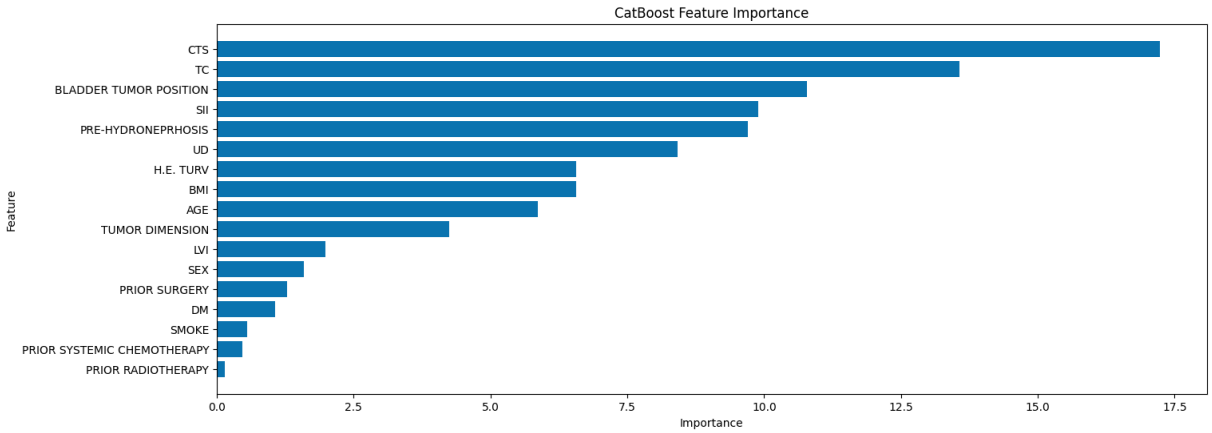


Figure 1 presents the global feature importance ranking from the CatBoost model trained to predict DFS. This ranking reflects the contribution of each variable to reducing the model's prediction across all patients. The most influential predictor was the clinical tumor stage (CTS), with an importance score of approximately 17, followed by the pathological tumor classification (TC), with an importance score of approximately 14, reflecting the role of tumor invasiveness, local extension, and accurate tumor staging in DFS prediction. The systemic immune-inflammation index (SII) ranked highly, with a score of approximately 9.5.

To a lesser extent, demographic and anatomical variables, such as Body Mass Index (BMI), age, and tumor dimension, also contributed to the model.

Figure 2 displays the SHAP summary plot for the DFS model, illustrating the distribution and direction of impact of each feature on the predicted survival time across all patients. CTS and TC exhibited the widest distribution of SHAP values, confirming their dominant influence, where they substantially increase or decrease predicted DFS based on their values. The SII displayed a more balanced distribution, with both positive and negative effects depending on the value. In contrast, features like prior treatment (surgery, radiotherapy, chemotherapy) and lifestyle factors (smoking status, diabetes) had SHAP distribution clustered near zero, indicating limited predictive power.

Figure 2. Violin Plot of Feature Influence on DFS Prediction from the SHAP Analysis. CTS: Clinical Tumor Stage; TC: Tumor Classification; PRE-HYDRONEPHROSIS: pre-treatment hydronephrosis; UD: Urinary Diversion type; SII: Systemic Immune-Inflammatory Index; BMI: Body-Mass Index; H.E. TURV: Histological Exam for TURB; LVI: lymphovascular invasion; DM: diabetes mellitus.

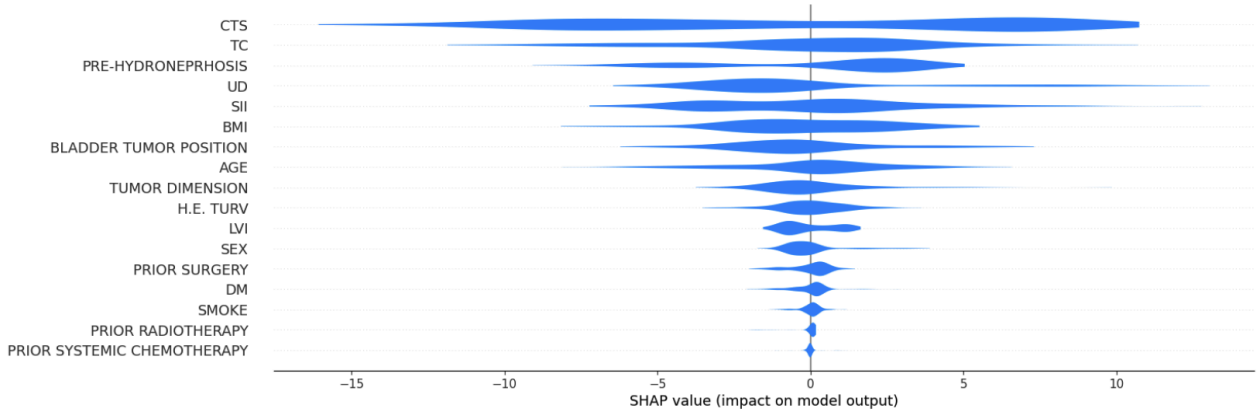
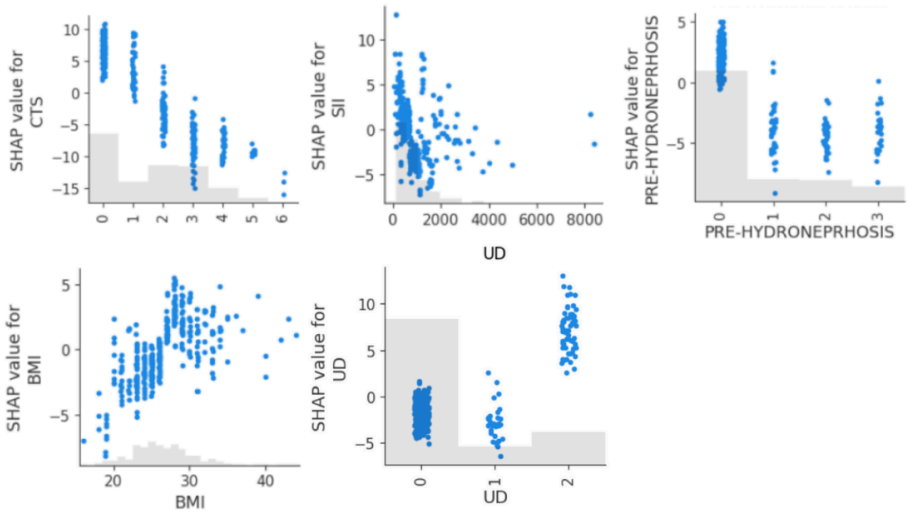


Figure 3 presents SHAP dependence plots for four of the most influential features affecting DFS predictions. The x-axis represents the feature value, and the y-axis shows the SHAP value (ie, the impact on the model's output). CTS showed a strong negative relationship with predicted DFS: as tumor stage increases, SHAP values shifted sharply downward, indicating a consistent reduction in predicted survival, aligning with the known prognostic role of tumor invasiveness in bladder cancer. SII demonstrated a non-linear relationship, showing that patients with lower SII values had better SHAP values, while those with elevated SII showed increasingly negative impacts on DFS. This suggests a threshold effect, where systemic inflammation beyond a certain level contributes to poorer prognosis. The presence of pre-treatment hydronephrosis had a negative impact on DFS prediction. Patients with low BMI had negative SHAP values, indicating reduced DFS, while those with moderate BMI experienced mildly negative predictions. At a BMI greater than 28, SHAP values become positive, suggesting a potential protective effect exerted by higher BMI. Urinary diversion (UD) showed positive SHAP values for the vesicoileal pouch, while other approaches had negative SHAP values.

Figure 3. SHAP scatter plots for the five most influential features influencing disease-free survival (DFS) predictions. CTS: Clinical Tumor Stage; SII: Systemic Immune-Inflammatory Index; pre-treatment hydronephrosis; BMI: Body-Mass Index; UD: Urinary Diversion type



Overall Survival Time (OST) Prediction

For overall survival prediction, the CatBoost model achieved a mean absolute error (MAE) of 17.2 months across the entire patient cohort. When restricted to the subgroup of patients who had died (n = 156), the prediction error improved to 15.8 months. After filtering features based on importance using a threshold of <0.5, the MAE further improved to 14.6, suggesting that a more compact feature set may improve predictive efficiency without compromising accuracy. This final model was selected for interpretation, as it maintained accuracy while reducing complexity.

Figure 4 Progressive improvement in model accuracy.

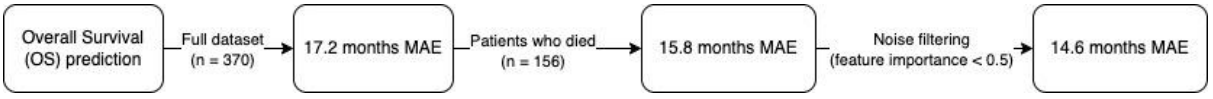


Figure 5 presents the CatBoost feature importance ranking for the best-performing overall survival time (OST) prediction model. TC emerged as the most influential predictor of overall survival in the final model with a value of approximately 17.5. SII followed closely with a value of approximately 15.5, highlighting the role of systemic inflammation in cancer progression and survival outcomes. The third feature was represented by histological findings (H.E. TURV). Other influencing factors include CTS, pre-treatment hydronephrosis, type of urinary diversion, BMI, and age.

Figure 5. CatBoost feature importance ranking for the best-performing overall survival time (OST) prediction model. TC: Tumor Classification; SII: Systemic Immune-Inflammatory Index; H.E. TURV: Histological Exam for TURB; CTS: Clinical Tumor Stage;

PRE-HYDRONEPHROSIS: pre-treatment hydronephrosis; UD: Urinary Diversion type; BMI: Body-Mass Index; LVI: lymphovascular invasion.

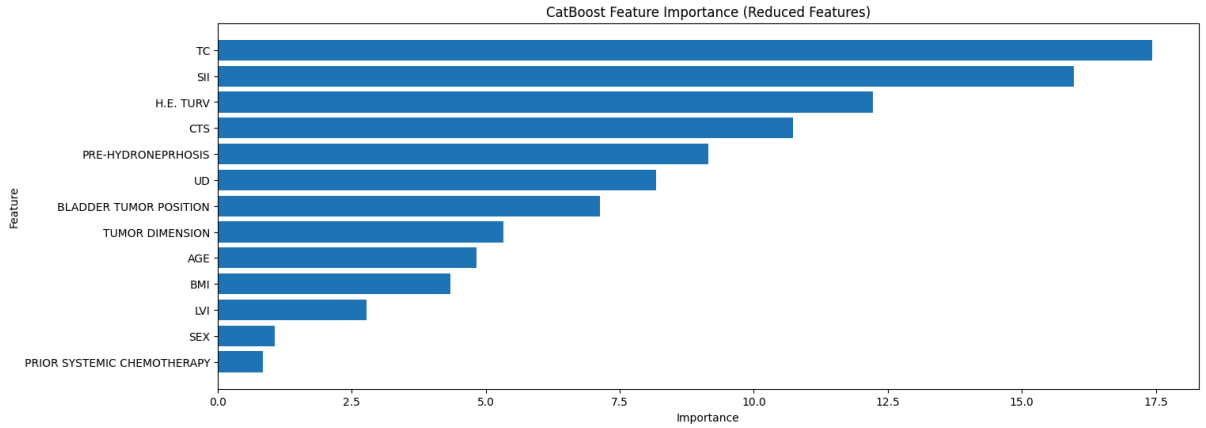


Figure 6 displays the SHAP summary plot for the final OST prediction model. As expected, the most impactful variable was TC, which showed a broad distribution. Pre-treatment hydronephrosis and SII exhibited wide SHAP distributions. CTS and histological findings (H.E. TURV) showed a similar overall effect on prognosis.

Figure 6. SHAP Violin Summary Plot for the Final Overall Survival (OST) Prediction Model. TC: Tumor Classification; PRE-HYDRONEPHROSIS: pre-treatment hydronephrosis; SII: Systemic Immune-Inflammatory Index; CTS: Clinical Tumor Stage; H.E. TURV: Histological Exam for TURB; UD: Urinary Diversion type; BMI: Body-Mass Index; LVI: lymphovascular invasion.

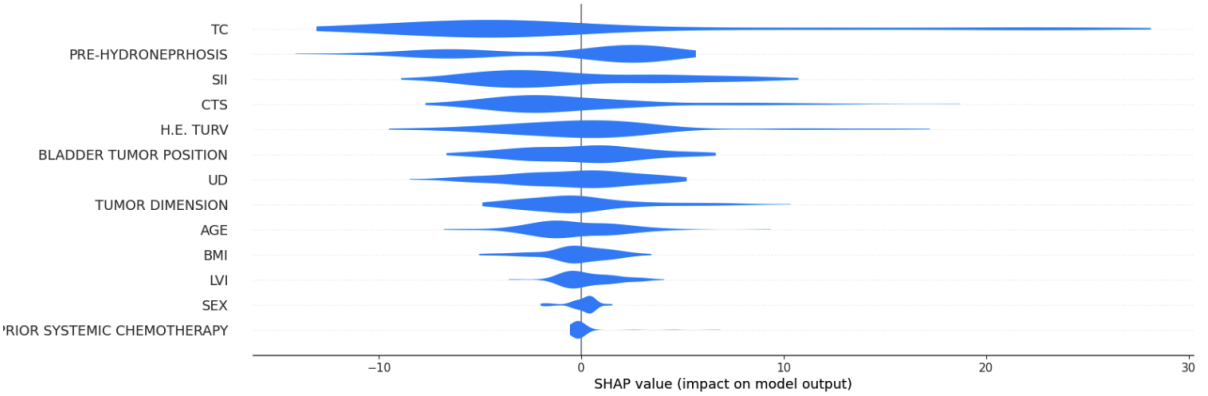
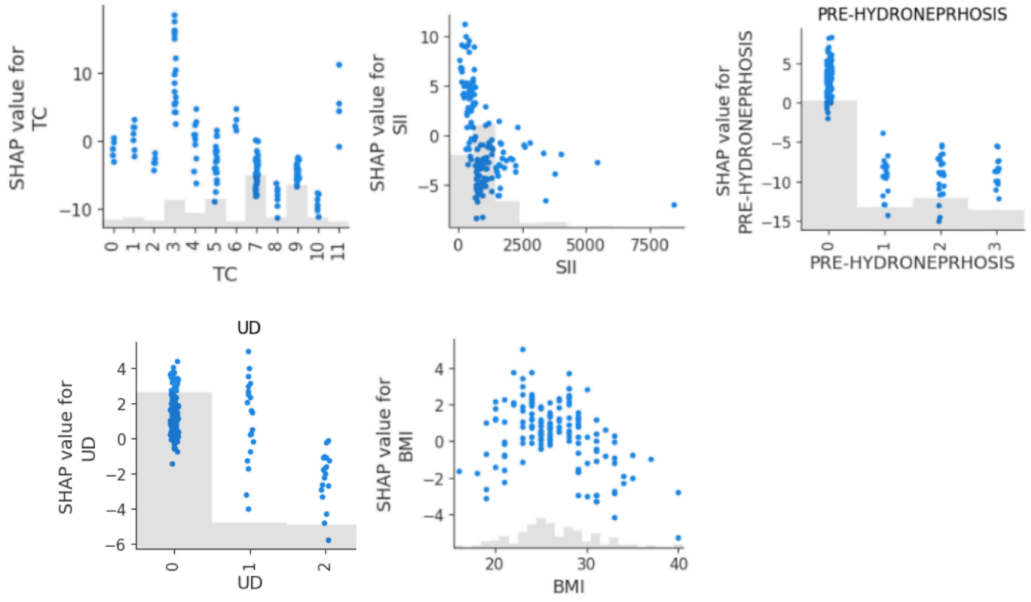


Figure 7 presents SHAP dependence plots for five key features influencing overall survival predictions. TC SHAP values indicate that patients with in situ cancers (value 3) have the best overall survival prediction, which gradually declines as the tumor stage increases. SII showed a threshold effect, where predictions remained relatively stable up to a value of approximately 1000, then fell sharply, indicating that elevated inflammation is associated with a poor overall outcome. Pre-treatment hydronephrosis was strongly linked to reduced predicted survival, where patients with this condition had uniformly negative SHAP values, not influenced by bilaterality. At the same time, BMI demonstrated a nonlinear pattern, where patients with very low BMIs had reduced survival predictions, moderate BMIs were associated with better outcomes, and at higher BMI values, SHAP values began to decline

again, suggesting that both underweight and obesity may be associated with increased mortality risk in this population. The type of urinary diversion showed a different impact than observed in DFS, with vesicoileal pouches being associated with a lower OST.

Figure 7. SHAP scatter plots for the four most influential features influence on overall survival time (OST) prediction. TC: Tumor Classification; SII: Systemic Immune-Inflammatory Index; PRE-HYDRONEPHROSIS: pre-treatment hydronephrosis; BMI: Body Mass Index; UD: Urinary Diversion type.



Cause of Death Classification

The CatBoost classifier was trained to predict whether a patient's death was tumor-related. Due to class imbalance, only 14 out of 78 deaths were cancer-related; custom class weights and a reduced decision threshold of 0.12 were applied to maximize recall and minimize false negatives. The final model achieved a recall of 78.6 % (Figure 8), correctly identifying 11 out of 14 tumor-related deaths. The overall F-1 score for the positive was 0.44, with a precision of 31%. The model prioritizes sensitivity over specificity.

Figure 8. Confusion Matrix for Cause of Death Classification.

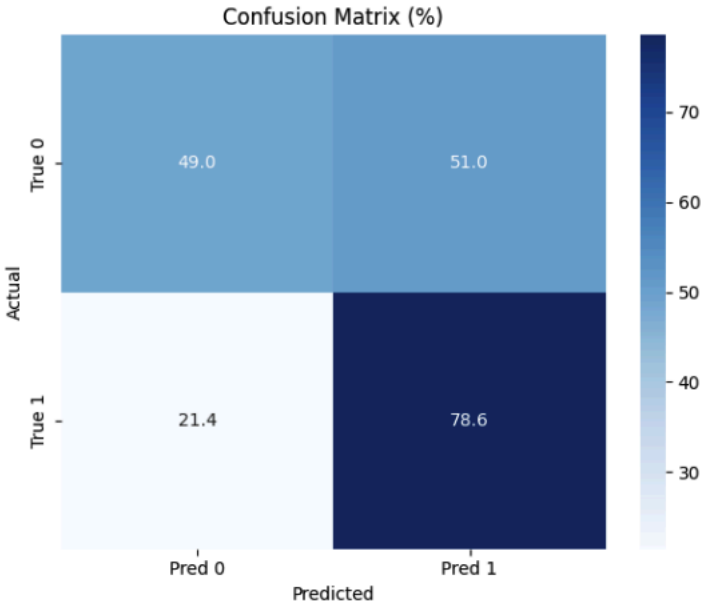


Figure 9 represents the CatBoost feature importance ranking for the death cause classification model. The most influential feature was the anatomical position of the bladder tumor, with a value of approximately 16.5. This was followed by tumor classification (TC) and pre-treatment hydronephrosis, both indicators of disease severity and progression.
Other key features included the clinical tumor stage (CTS) and systemic immune-inflammation index (SII), with values of approximately 11 and 11.5, respectively.

Figure 9. CatBoost feature importance ranking for the Cause of Death Classification. TC: Tumor Classification; PRE-HYDRONEPHROSIS: pre-treatment hydronephrosis; CTS: Clinical Tumor Stage; SII: Systemic Immune-Inflammatory Index; H.E. TURV: Histological Exam for TURB; UD: Urinary Diversion type; BMI: Body-Mass Index; DM: Diabetes Mellitus; LVI: lymphovascular invasion.

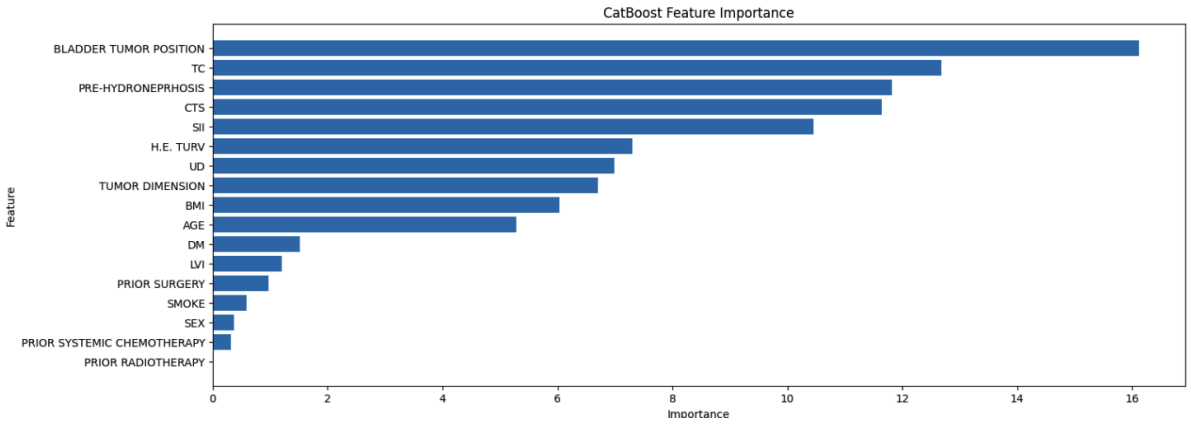


Figure 10 displays the SHAP summary plot for the tumor-related death classification model. The feature with the widest and most impactful distribution was bladder tumor position, which is addressed in detail in the discussion of Figure 11. The SII is also influential, with positive and negative SHAP values. Tumor classification (TC), pre-treatment hydronephrosis,

and clinical tumor stage (CTS) generally act to slightly decrease the predicted risk of tumor-related death for most patients, with little variability in their effect.

Figure 10. SHAP summary plot for the tumor-related death classification model. SII: Systemic Immune-Inflammatory Index; TC: Tumor Classification; BMI: Body-Mass Index; H.E. TURV: Histological Exam for TURB; CTS: Clinical Tumor Stage; PRE-HYDRONEPHROSIS: pre-treatment hydronephrosis; DM: Diabetes Mellitus; UD: Urinary Diversion type; LVI: lymphovascular invasion.

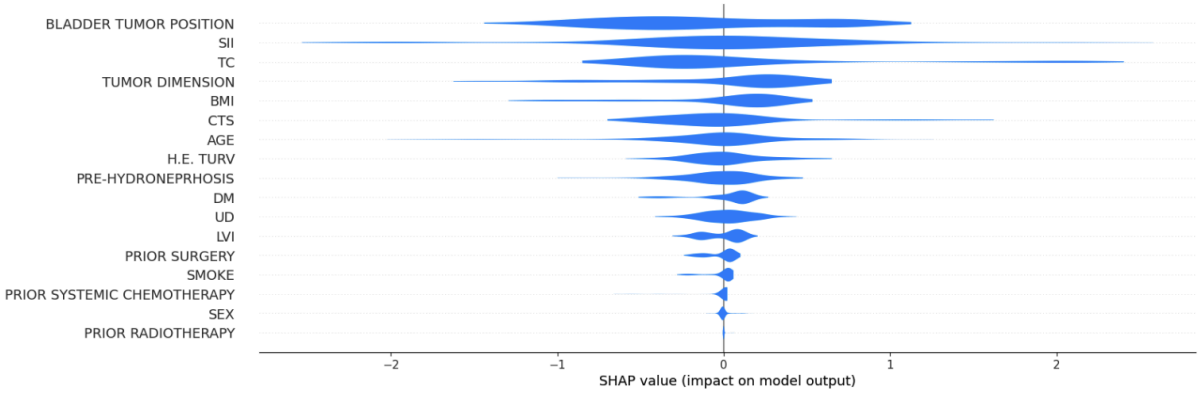
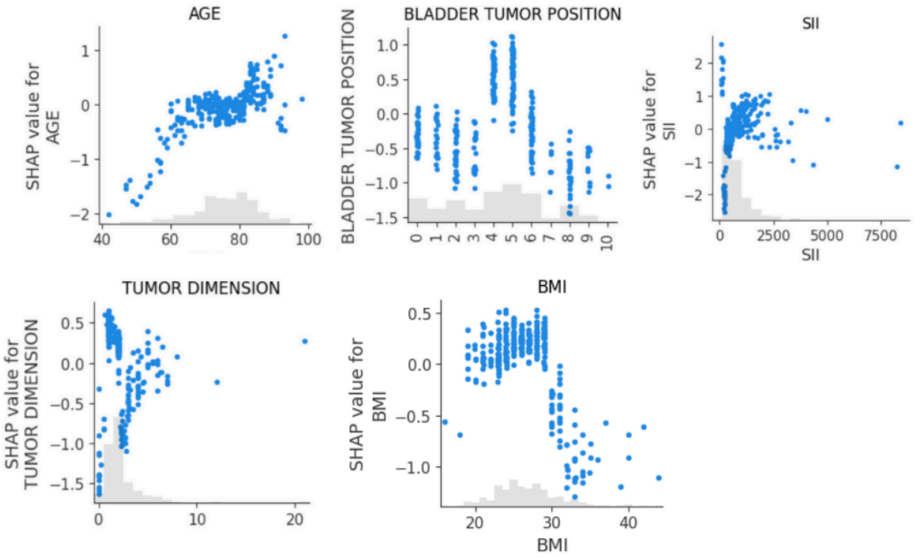


Figure 11 presents SHAP dependence plots for four key features influencing the tumor-related classification model: Age, Bladder tumor position, systemic immune-inflammation index (SII), and tumor dimension.

Figure 11. SHAP scatter plots for the five most influential features' influence on the tumor-related death classification model. SII: Systemic Immune-Inflammatory Index; BMI: Body Mass Index.



The scatter plot for Age shows a clear positive trend, showing that as patient age increases, the SHAP value for age also increases, indicating that older age consistently raises the

predicted risk of tumor-related death. Bladder tumor position is also another influential predictor in the tumor-related classification, highlighting that tumors located in the posterior wall and right lateral wall had a higher likelihood of being the patient's cause of death. In contrast, tumors located in the anterior wall, bladder base, or those spreading throughout the entire bladder had a lower likelihood. Increasing SII and tumor dimension also have a moderate predictive value, respectively increasing (SII) and decreasing (tumor dimension) the probability that the patient's cause of death is cancer. Finally, patients with a higher BMI show a higher likelihood.

Discussion

This study presents the development of a survival prediction model in bladder cancer patients using machine learning approaches. Our findings demonstrate that modern predictive algorithms show promising accuracy in forecasting disease-free survival, overall survival, and cause of death. The limited sample and the paucity of included categories for the analysis suggest that predictive algorithms trained with additional resources might significantly improve in the demonstrated accuracy.

The observation that age showed a positive correlation with survival outcomes is particularly intriguing and seemingly counterintuitive. This "age paradox" might be explained by several factors. First, older patients often receive more conservative treatment approaches, potentially leading to selection bias in surgical candidates. Additionally, younger patients with bladder cancer often present with more aggressive disease variants, which could account for their relatively poorer outcomes. (10)

Clinical tumor stage (CTS) was the strongest predictor, aligning with established prognostic factors in bladder cancer. (11) Additionally, inflammatory markers, particularly the Systemic Immune-Inflammation Index (SII), show a negative correlation with survival outcomes and support recent findings in other malignancies. (12) This relationship likely reflects the complex interplay between systemic inflammation and cancer progression, where elevated SII indicates a pro-tumoral inflammatory state. (13)

SHAP analysis revealed a clear, monotonic decline in predicted survival as CTS increased, reinforcing its primary prognostic role in both disease-free and overall survival. SII, by contrast, demonstrated a threshold effect where values beyond ~1000 were associated with a sharp drop in predicted survival time, suggesting a non-linear relationship between systemic inflammation and patient outcomes. The impact of urinary diversion type on survival outcomes represents an interesting finding. Previous studies had found a protective effect of orthotopic neobladder reconstruction against urethral recurrence in male patients undergoing radical cystectomy for bladder cancer. While that was not present in our patients, we found a positive association between vesico-ileal pouch construction and improved survival. This association may reflect both patient selection and the physiological advantages of this approach; however, this observation should be interpreted cautiously, as confounding factors such as surgical expertise and patient characteristics, which were not considered in the present study, could influence these results. Body Mass Index (BMI) shows a similar intriguing relationship: patients with an unhealthy BMI, either high or low, show poorer

outcomes; this can be related both to tumor characteristics and the surgical approach being limited in terms of radicality. This U-shaped association between BMI and survival was consistently observed across both DFS and OST outcomes, where moderate BMI ranges correlated with more favorable SHAP values. The findings support a metabolic vulnerability in both underweight and obese patients, which may influence recovery or treatment tolerance.

Our machine learning models achieved prediction accuracies comparable to those reported in previous studies. The accuracy in cause-of-death prediction, although modest, represents an encouraging level, given the limited resources and the paucity of categories considered for the analysis, when compared with studies published a few years ago that used samples significantly larger and achieved marginally higher accuracy in mortality and recurrence prediction. (14) A recently published systematic review investigating machine learning algorithms for bladder cancer cystectomy outcomes found that most of the algorithms would not exceed 70% accuracy, sometimes even 60%. (15) The integration of SII into predictive models represents an auspicious direction. As a low-cost, readily available biomarker, SII could enhance current prognostic tools without adding significant complexity or cost to patient evaluation. (16)

A notable limitation of our study is the relatively high MAE values in survival predictions. These MAE values render the algorithm unsuitable for precise individual patient counseling or treatment planning, where accurate timing is critical, such as in emergency settings or for patients exhibiting postoperative complications. (17) However, this level of accuracy remains acceptable for clinical trial patient stratification and allocation, particularly in trials where broad risk categories rather than precise survival estimates are needed for randomization, such as balancing treatment arms in clinical trials by identifying comparable risk groups or supporting enrollment decisions in competing risks studies where precise timing is less critical than overall risk assessment. (18)

Limitations and reproducibility

This study is subject to some limitations when interpreting the results. The relatively limited dataset size (n=370 initially, reduced to 312-347 for specific analyses) inherently constrains the generalizability and robustness of the developed models. While machine learning algorithms like CatBoost are robust on smaller datasets, their predictive power can be substantially enhanced with larger cohorts.

Secondly, the monocentric nature of the data collection, originating solely from Fondazione Policlinico Gemelli in Rome, Italy, introduces a potential for selection bias and limits external validity. Patient characteristics, treatment protocols, and population demographics can vary significantly across institutions and geographical regions. The findings from this study may not be directly transferable to other clinical settings without further validation on diverse, external datasets.

Thirdly, while rigorous data cleaning was performed, the inherent human factors associated with retrospective data extraction from medical records cannot be eliminated.

Conclusion

Our study demonstrates the potential utility of machine learning approaches in predicting bladder cancer outcomes following cystectomy. While the achieved accuracy levels are modest, they align with current literature benchmarks and provide a foundation for future development. The identification of clinical tumor stage as the primary predictor, along with the consistent negative correlation of SII with survival outcomes, validates these parameters as valuable prognostic indicators. In particular, SHAP analysis revealed a monotonic decline in predicted disease-free and overall survival as CTS increases, reaffirming its role in risk stratification. On the other hand, SII exhibited a threshold effect, where values above approximately 1000 were associated with a rapid drop in predicted survival, reinforcing the adverse prognostic impact of systemic inflammation. The current model's performance, though not suitable for precise individual prognostication, shows particular promise for clinical trial stratification and cohort allocation. Future studies with larger datasets and additional predictive variables may enhance the model's accuracy and broaden its clinical applications. Integrating readily available biomarkers, such as SII, represents a cost-effective approach to improving prognostic tools. These findings contribute to the growing body of evidence supporting the role of machine learning in oncological decision-making while acknowledging the need for continued refinement and validation in larger cohorts.

Authors contributions

FAC, BR, and PR elaborated on the first manuscript concept. AN and FAC performed the statistical analysis. FAC, VDV, and PR wrote the article. MS, NF, and GM reviewed and approved the final manuscript.

Open-Source Commitment

To ensure the reproducibility of our results and facilitate further research, all code used for this analysis is publicly available at https://figshare.com/s/ec2419a33ad7ad05f412.

Funding

This article received no funding.

References

1. Anagnostou T, Remzi M, Lykourinas M, Djavan B. Artificial Neural Networks for Decision-Making in Urologic Oncology. Eur Urol. 2003 Jun;43(6):596–603.
2. Schwarzer G, Schumacher M. Artificial neural networks for diagnosis and prognosis in prostate cancer. Semin Urol Oncol. 2002 May;20(2):asuro0200089.
3. Anagnostou T, Remzi M, Lykourinas M, Djavan B. Artificial Neural Networks for Decision-Making in Urologic Oncology. Eur Urol. 2003 Jun;43(6):596–603.
4. Daniel J. Sargent Ph.D. Comparison of artificial neural networks with other statistical approaches. 2001;
5. Zi H, Liu MY, Luo LS, Huang Q, Luo PC, Luan HH, et al. Global burden of benign prostatic hyperplasia, urinary tract infections, urolithiasis, bladder cancer, kidney cancer, and prostate cancer from 1990 to 2021. Mil Med Res. 2024 Sep 18;11(1):64.
6. Abbod MF, Catto JWF, Linkens DA, Hamdy FC. Application of Artificial Intelligence to the Management of Urological Cancer. Journal of Urology. 2007 Oct;178(4):1150–6.
7. Di Sarno L, Caroselli A, Tonin G, Graglia B, Pansini V, Causio FA, et al. Artificial Intelligence in Pediatric Emergency Medicine: Applications, Challenges, and Future Perspectives. Biomedicines. 2024 May 30;12(6):1220.
8. Mazzotta AD, Burti E, Causio FA, Orlandi A, Martinelli S, Longaroni M, et al. Machine Learning Approaches for the Prediction of Postoperative Major Complications in Patients Undergoing Surgery for Bowel Obstruction. J Pers Med. 2024 Oct 8;14(10):1043.
9. Yin M, Joshi M, Meijer RP, Glantz M, Holder S, Harvey HA, et al. Neoadjuvant Chemotherapy for Muscle-Invasive Bladder Cancer: A Systematic Review and Two-Step Meta-Analysis. Oncologist. 2016 Jun 1;21(6):708–15.
10. Withrow DR, Nicholson BD, Morris EJA, Wong ML, Pilleron S. Age‐related differences in cancer relative survival in the United States: A <scp>SEER</scp> ‐18 analysis. Int J Cancer. 2023 Jun 23;152(11):2283–91.
11. Kardoust Parizi M, Enikeev D, Glybochko P V., Seebacher V, Janisch F, Fajkovic H, et al. Prognostic value of T1 substaging on oncological outcomes in patients with non-muscle-invasive bladder urothelial carcinoma: a systematic literature review and meta-analysis. World J Urol. 2020 Jun 6;38(6):1437–49.
12. Kou J, Huang J, Li J, Wu Z, Ni L. Systemic immune-inflammation index predicts prognosis and responsiveness to immunotherapy in cancer patients: a systematic review and meta-analysis. Clin Exp Med. 2023 Mar 26;23(7):3895–905.
13. Wang Q, Zhu D. The prognostic value of systemic immune-inflammation index (SII) in patients after radical operation for carcinoma of stomach in gastric cancer. J Gastrointest Oncol. 2019 Oct;10(5):965–78.

14. Hasnain Z, Mason J, Gill K, Miranda G, Gill IS, Kuhn P, et al. Machine learning models for predicting post-cystectomy recurrence and survival in bladder cancer patients. PLoS One. 2019 Feb 20;14(2):e0210976.

15. Liu YS, Thaliffdeen R, Han S, Park C. Use of machine learning to predict bladder cancer survival outcomes: a systematic literature review. Expert Rev Pharmacoecon Outcomes Res. 2023 Aug 9;23(7):761–71.

16. Kou J, Huang J, Li J, Wu Z, Ni L. Systemic immune-inflammation index predicts prognosis and responsiveness to immunotherapy in cancer patients: a systematic review and meta-analysis. Clin Exp Med. 2023 Mar 26;23(7):3895–905.

17. Shi-Ang Qi NKMFWSLHKRRRHRG. An Effective Meaningful Way to Evaluate Survival Models. 2023;

18. Hassan N, Slight R, Morgan G, Bates DW, Gallier S, Sapey E, et al. Road map for clinicians to develop and evaluate AI predictive models to inform clinical decision-making. BMJ Health Care Inform. 2023 Aug 9;30(1):e100784.