A Machine Learning Approach for Long-Term Prognosis of Bladder Cancer based on Clinical and Molecular Features

Qingyuan Song[1], John D. Seigne[2], Alan R. Schned[3], Karl T. Kelsey[4], Margaret R. Karagas[5], Saeed Hassanpour[1,5,6]

[1]Department of Biomedical Data Science, Dartmouth College, Hanover, NH 03755, USA

[2]Department of Surgery, Division of Urology, Dartmouth-Hitchcock Medical Center, Lebanon, NH 03756, USA

[3]Department of Pathology and Laboratory Medicine, Dartmouth-Hitchcock Medical Center, Lebanon, NH 03756, USA

[4]Departments of Epidemiology and Pathology and Laboratory Medicine, Brown University, Providence, RI 02912, USA

[5]Department of Epidemiology, Dartmouth College, Hanover, NH 03755, USA

[6]Department of Computer Science Department, Dartmouth College, Hanover, NH 03755, USA

Corresponding author:

Saeed Hassanpour, PhD
Geisel School of Medicine at Dartmouth
One Medical Center Drive, HB 7261
Lebanon, NH 03756, USA
Telephone: +1 (603) 650-1983
Fax: +1 (603) 650-1966
Email: Saeed.Hassanpour@dartmouth.edu

PURPOSE: Improving the consistency and reproducibility of bladder cancer prognoses necessitates the development of accurate, predictive prognostic models. Current methods of determining the prognosis of bladder cancer patients relies on manual decision-making including factors with high intra- and inter-observer variability, such as tumor grade. To advance the long-term prediction of bladder cancer prognoses, we developed and tested a computational model to predict the 10-year overall survival for bladder cancer patients without considering tumor grade classification.

MATERIALS AND METHODS: We utilized a population-based dataset from the New Hampshire Cancer Registry with 1,225 bladder cancer patients diagnosed between 1994 and 2004. A weighted logistic regression model was trained using features including pre-treatment factors with high reproducibility including demographic characteristics, risk factors such as history of cigarette smoking, clinical information such as muscle invasiveness and tumor histology, and molecular features such as p53 immunohistochemical (IHC) positivity, while excluding less reliable measures such as tumor grade.

RESULT: Our model predictor of 10-year survival (F1 score = 0.78) was largely driven by age, muscle invasiveness and p53 IHC positivity and strongly related to patient survival in Cox models ($p = 0.0013$) even after adjustment for tumor grade and treatment. These results suggest that bladder cancer prognosis can be improved by machine learning methods and avoiding factors with high intra- and inter-observer variability.

CONCLUSION: Our study demonstrated a machine learning approach using a combination of clinical and molecular features could provide a better long-term prognosis for bladder cancer patients in comparison to tumor grade that suffers from low intra- and inter-observer variability.

If validated in clinical trials, this automated approach can guide personalized management and treatment for bladder cancer patients.

Introduction

Urothelial bladder cancer is one of the most common malignancies worldwide. It was estimated that in 2018, over 81,000 new cases of bladder cancer will be diagnosed in the United States, resulting in 17,000 deaths[1]. Although incidence rate of bladder cancer has been decreasing in recent years, disease frequently recurs while the mortality rate has remained stable[1]. Thus, the development of accurate, predictive prognostic markers is needed to better personalize healthcare management and improve patient survival, as patients with poorer prognoses may benefit from more intense follow-up, treatment, and healthcare planning.

While developing patient prognostic and treatment plans, clinicians follow complex guidelines that involve many factors, including classification of tumor grade, tumor stage, and the individual patient's general health status. Some of these factors depend upon subjective assessments lacking in consistency and reproducibility. The WHO 1973 and WHO/ISUP (World Health Organization/International Society of Urological Pathology) classification are the two most widely used methods for establishing tumor grade, but these have relatively high intra- and inter-observer reliability[2]. Robertson et al. reported that the inter-observer agreement among 11 pathologists using the WHO 1973 classification was slight to moderate ($\kappa = 0.19 - 0.44$)[3]. A study by Yorukoglu et al. compared the inter-observer agreement among 6 pathologists using both classifications and found a slightly better inter-observer agreement with the WHO/ISUP classification ($\kappa = 0.42 - 0.65$) than the WHO 1973 classification ($\kappa = 0.19 - 0.65$)[4]. Ooms et al. found a limited intra-observer variability (Spearman rank-order correlations coefficients of 0.50–0.67) using the WHO 1973 classification[5].

Automated computational approaches have demonstrated utility in providing unbiased and reliable guidance in clinical decision-making for bladder cancer patients by exploiting large datasets[7]. Furthermore, machine learning approaches require minimal time and resources. To demonstrate the potential utility of computational models, we had the opportunity to design and test a fully automated computational machine learning approach to predict 10-year bladder cancer survival based on factors with high reproducibility and low subjectivity, while excluding more variability histologic features such as tumor grade. Such models are of potentially clinical utility, particularly, because there is currently limited information about the predictors of long-term survival in the general population.

Methods

The overview of the pipeline for data collection, model training, and evaluation is shown in **Figure 1.**

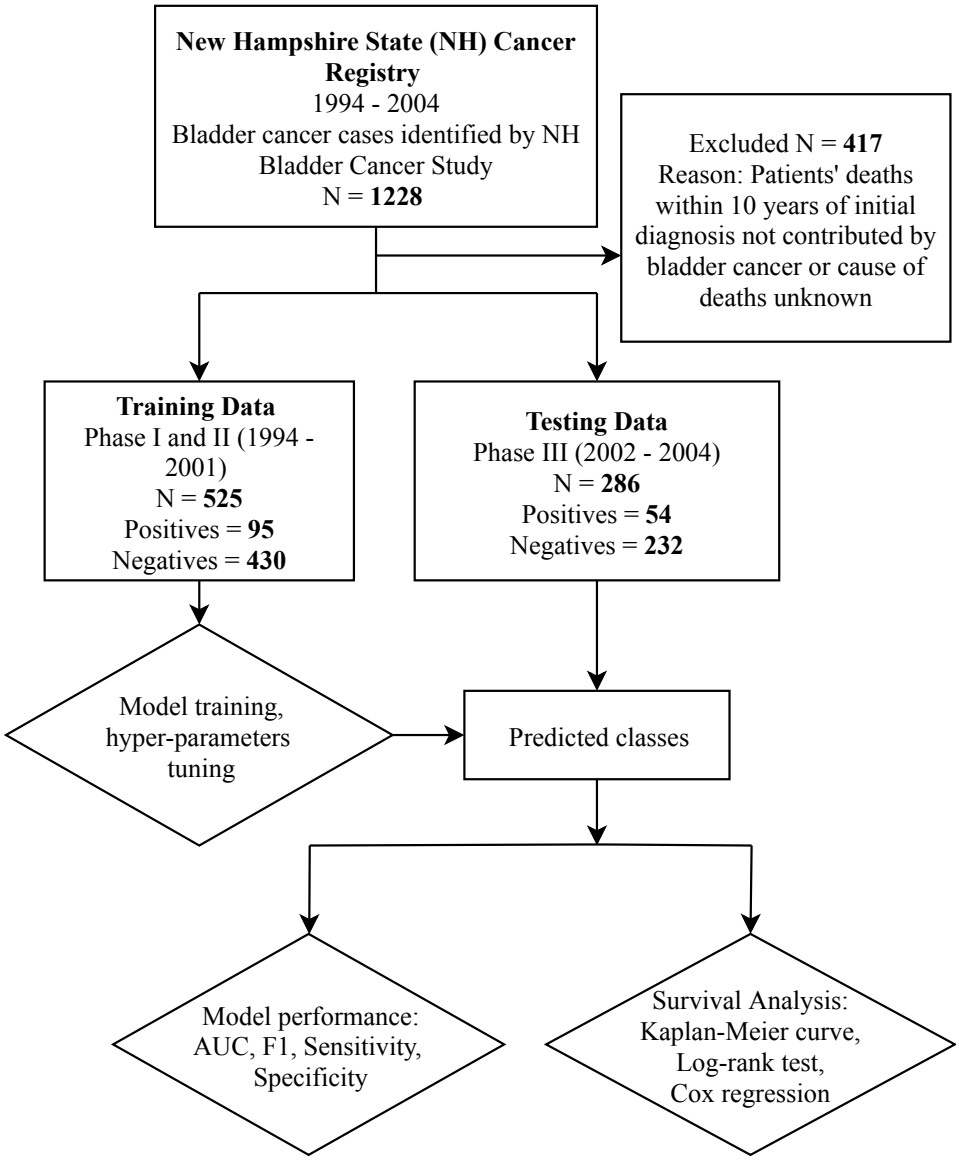

Figure 1. An overview of an automated machine learning pipeline for bladder cancer 10-year survival prediction. A total of 811 bladder cancer patients were split into training and testing datasets. Positive samples included patients with disease-specific deaths within 10 years from the initial diagnosis, and negative samples included patients with a survival of more than 10 years from the initial diagnosis. The model was trained on the training data then evaluated on the test dataset. The model's predictions were compared to histological classification through survival analysis.

Dataset

Our study utilized 1,228 previously collected histologically confirmed bladder cancer cases from the New Hampshire Bladder Cancer Study that were identified from the New Hampshire State

Cancer Registry[7]. This study and the use of human patient data in this project were approved by the Dartmouth institutional review board with a waiver of informed consent. The dataset consists of cases from three study phases: Phase I included 448 patients diagnosed from July 1st, 1994 to June 30th, 1998; Phase II included 385 patients diagnosed from July 1st, 1998 to December 31st, 2001; Phase III included 396 patients diagnosed from July 1st, 2002 to December 31st, 2004. Cases included patients who were ages 25 to 74 years at initial diagnosis in the first two study phases and 31 to 79 years at initial diagnosis in the third study phase. Demographic and risk factor information in this registry was obtained through personal interviews. Histologic classifications according to both WHO 1973[8] and WHO/ISUP[9] criteria were based on a standardized re-review of the original histopathology specimens. Tumor stage was reported using the TNM criteria of the American Joint Commission on Cancer[10]. *P53* positivity was based the presence of 50% or greater of *P53* immunohistochemical staining[11]. The distribution of our data is similar to that of a larger population-based dataset from the Surveillance, Epidemiology, and End Results (SEER) database[12], which suggests generalizability of our dataset.

Outcome Status

All patients who survived from phases I and II had follow-ups for at least 120 months. The survival status of patients was determined by examining the Social Security Administration Death Master File[13]. Survival time was calculated from the date of diagnosis to the date of death for patients who did not survive, or to the date when the Death Master File was queried for patients who survived. Patients were labeled as positives and negatives based on whether or not they survived at least 120 months after the initial diagnosis. Patients in phase III of the study were classified according to their death status at the last follow-up, which was approximately 10 years (129.0$\pm$8.90 months) for patients who survived. To avoid misclassification, we removed patients whose causes of death

were unknown or not contributed by bladder cancer. As a result, of the 1,228 bladder cancer patients, 417 were removed due to non-bladder-cancer-related deaths or unknown cause of deaths. The remaining patients in this study included 149 persons with disease-specific deaths within 10 years after initial diagnosis, labeled as positive class, and 662 persons who survived more than 10 years, labeled as negative class. To add to the generalizability of our model, we combined data from earlier phases (Phases I and II) into training data and used phase III as testing data. The ratios between the number of positives and negatives are consistent in both training and testing datasets. The baseline characteristics of patients included in this study are shown in **Table 1.**

Missing data imputation

We implemented a data imputation strategy for missing covariate data. Training data and testing data were imputed separately. Binary variables with missing values (i.e., *P53* mutation, *P53* intensity, PTCH LOH positivity, high-risk occupation, family history of cancer) were imputed with 0.5 (midpoint value). Body mass index (BMI) was mean imputed. Missing values for smoke pack-years were imputed with '0' for never smokers, and with the mean pack-years for others.

Predictive Model

We implemented a logistic regression model with demographic characteristics (age and sex), risk factors (history of cigarette smoking, high risk occupation, BMI, and family history of bladder cancer), clinical information (presence of muscle invasiveness and tumor histology being transitional cell carcinoma or other type) and molecular features (TP53 mutation, *P53* IHC positivity and *P53* IHC staining intensity; PTCH LOH) excluding tumor grade. The features included were putative or potential predictors of bladder cancer survival and known to be reliable or objective measures. Model implementation was done with Python 3.6 (Python Software

Foundation, Beaverton, OR) and Scikit Learn version 0.19.1[14]. Because the distribution of positive and negative classes in the dataset was highly imbalanced, we weighted each class reciprocal to its prevalence in the training dataset and maximized the weighted log-likelihood loss function to estimate the parameters of the model. The class weights were calculated using equation (1) as suggested by King[15] so that the minority class was weighted more and emphasized by the model:

$$W_{class} = \frac{\#\ of\ samples\ in\ the\ dataset}{\#\ of\ classes \times \#\ of\ samples\ in\ this\ class} \qquad (1)$$

Training and hyper-parameter tuning were done through 5-fold stratified cross-validation. Hyper-parameters included a confidence score threshold and an L2 regularization parameter. The confidence score threshold was selected to classify the samples into positive and negative classes. We used the optimal threshold to maximize the harmonic mean of sensitivity and specificity using equation (2) as suggested by Song et al.[16]:

$$\text{Harmonic mean} = 2 \times \frac{sensitivity \times specificity}{sensitivity + specificity} \qquad (2)$$

Also, the L2 regularization parameter was tuned by a log-spaced grid search in the cross-validation.

Evaluation

The trained model was evaluated on our independent, held-out test dataset using standard machine learning performance metrics of the area under the ROC curve, F1 score, sensitivity, and specificity. Survival analysis was implemented with package 'survival'[17] in R version 3.3.3 (R Foundation for Statistical Computing, Vienna, Austria). Kaplan-Meier curves and log-rank tests were employed to examine the survival difference between the patients from different prediction, tumor grade, and treatment groups. For this analysis, patients from WHO 1973 grades 3 and 4 were combined into a single group due to small sample sizes. For further comparison with tumor

grade, a multivariate Cox-proportional hazard model was also built with our model predictor and the WHO/ISUP classifications, while adjusting for the treatment information.

Results

Our fully automated machine learning approach utilized variables with high reproducibility such as patient's demographic characteristics (age and sex), risk factors (history of cigarette smoking, high risk occupation, BMI and family history of bladder cancer), clinical information (presence of muscle invasiveness and tumor histology being transitional cell carcinoma or other type) and molecular features (TP53 mutation, *p53* immunohistochemical (IHC) positivity and *p53* IHC staining intensity; PTCH LOH), without including less reliable measures like tumor grade to predict 10-year bladder cancer survival. In this evaluation we applied our model on an independent test dataset (N = 286 patients). The final model achieved an area under the Receiver Operating Characteristic curve (AUC-ROC) of 0.77 (**Figure 2**), and an overall F1 score of 0.78 on the test dataset. Figure 2 shows that our final model reached a sensitivity of 0.65 (95% confidence interval 0.60 – 0.71) and a specificity of 0.79 (95% confidence interval 0.74 – 0.84). The accuracy of the model prediction was 0.76 on the testing set (95% confidence interval 0.71 – 0.81).

To show the effect and utility of tumor grade in our long-term prognosis predictions, a logistic regression model including WHO/ISUP tumor grade as an independent variable was trained on the same training dataset, using the same class weighting and hyper-parameter tuning approaches, and tested for comparison with our final model, which does not include tumor grade features. The model including tumor grade achieved similar performances of an AUC-ROC of 0.77 and an overall F1 score of 0.78, with slightly lower sensitivity and specificity, indicating no significant

improvement upon our final model (Figure 2). The coefficients of predictive features in the final model, which indicate their directions and the magnitude of influence on the prediction results, are shown in **Figure 3**.

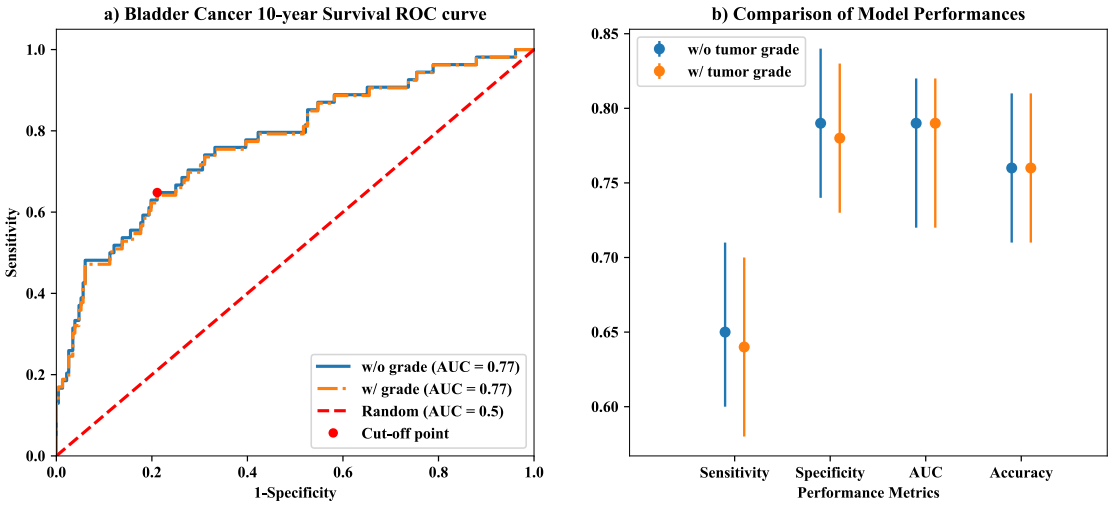

Figure 2. Model performance comparison of logistic regression model on the test dataset without and with WHO/ISUP tumor grade as a predictive feature**. a)** The receiver operating characteristic curve (ROC) of models' predictions on all patients from the test set. The blue line shows the ROC of our final model (without tumor grade), and the orange line shows the ROC of the logistic model with tumor grade. The red dotted line represents the performance of a model with random predictions. The red dot represents the confidence score cut-off point of our final model to predict high and low-risk groups. **b)** Comparison of the model performance measurements with their 95% confidence intervals.

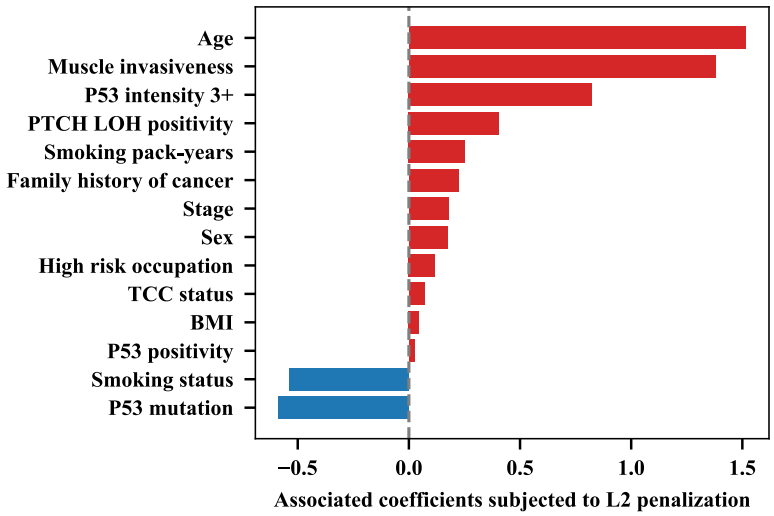

Figure 3. Coefficients of predictive features subjected to L2 penalization in final logistic regression model. The features with highest impacts on making positive predictions (i.e., survival prognosis of less than 10 years from the initial diagnosis due to bladder cancer), included age, muscle invasiveness, and *P53* intensity.

Figure 4 depicts the Kaplan-Meier survival curve of Phase III patients stratified by our predicted risk groups, the existing tumor grading schemes, and patients' treatment information. Our prediction model successfully distinguished the group with poorer prognoses from the rest, and the result was highly statistically significant (log-rank test $p = 4.06\times10^{-9}$). Of note, this result was more significant than both WHO 1973 (log-rank test $p = 4.61\times10^{-8}$) and WHO/ISUP grading schemes (log-rank test $p = 5.99\times10^{-7}$) for predicting long-term prognoses.

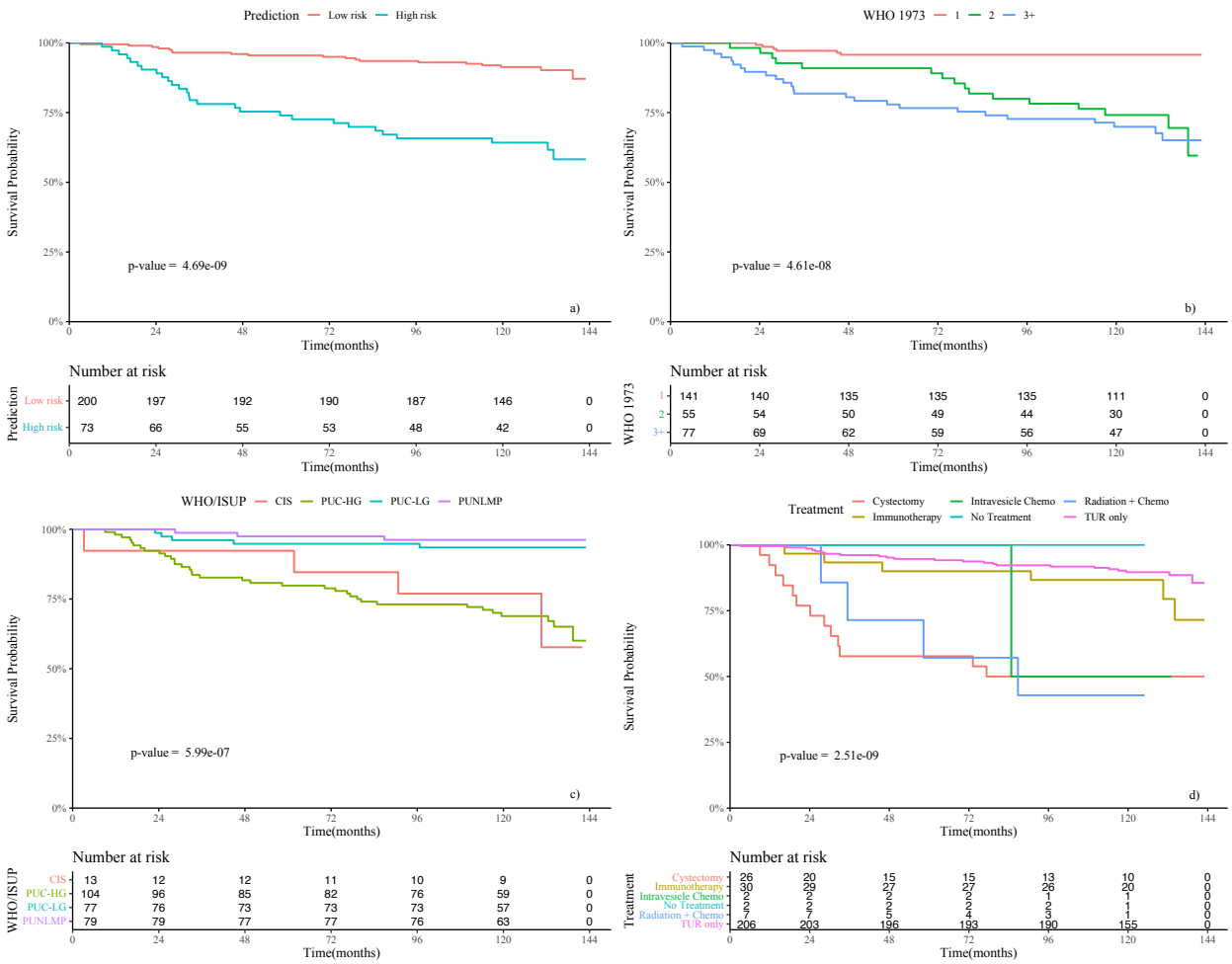

Figure 4. Survival analysis of bladder cancer patients from the test dataset stratified by **a)** our model's predictions, **b)** the WHO 1973 classification, **c)** the WHO/ISUP classification, and **d)** treatment groups.

A multivariate Cox-proportional hazard model was built with our model predictor, the WHO/ISUP classification, and the patients' first course of the treatment. The result indicates that our predictor provides a stronger estimate of patient survival ($p = 0.0013$, 95% confidence interval: 1.54 – 5.83) even after adjustment for tumor grade based on the WHO/ISUP criteria or first course of treatment (**Table 2**). Although the magnitude of the effect was smaller than our model predictor, the treatment effect was statistically significant with respect to the 10-year survival, while WHO/ISUP is not (**Table 2**).

Discussion

In this study, we developed a machine learning model that predicted 10-year survival of patients with bladder cancer. The few prior studies that have used machine learning approaches to predict bladder cancer prognoses have relied on small datasets from single hospitals, which could be biased toward specific sub-populations[6]. In order to avoid such bias and improve generalizability, we built a model using a population-based dataset. Unlike previous studies, we excluded both the WHO 1973 and the WHO/ISUP classifications from the predictive features, as these classifications are often inconsistent and add little value to our model predictions. Our model also excluded treatment information because its aim is to predict reliable prognoses before patients received their first course of treatments in order to serve as a constructive guide to aid treatment planning. The Cox regression analysis confirmed that our model's predictions were strongly associated with the patient survival, even after adjusting for treatment. As a result, our model achieved accurate performance on a held-out test dataset in predicting bladder cancer 10-year survival outcomes. Of note, we observed including the tumor grade as an independent variable in the model did not improve its performance. This result suggested the potential for our approach to create a reliable tool to support clinicians' decision-making and patient management in practice.

In making its predictions, our model prioritized factors that are critical to the bladder cancer prognosis. For example, age, muscle invasiveness of the tumors, and smoking pack-years were among the highest-ranked variables in our final model. It is widely accepted that age is the most significant risk factor for bladder cancer occurrence[18]. The elderly patients have higher probabilities of developing muscle-invasive bladder cancer, which is also a well-established risk

factor for a worse prognosis[19]. Muscle-invasive bladder cancer penetrates the central muscle layer of the bladder and is more likely to metastasize to other parts of the body[1]. More aggressive forms of treatment, such as radical cystectomy, are used for muscle-invasive bladder cancer and can pose additional risks, complications, and side effects for less tolerant elderly patients[19]. Elderly patients are also subjected to the accumulation of environmental exposure to carcinogens, such as those from cigarette smoking, which is another contributing factor to the incidence of bladder cancer[1,20].

An important advantage of our model is its ability to handle the imbalanced survival outcomes in the dataset. The estimated 10-year survival rate of bladder cancer patients is 70%, and the patients with lower stage disease have much higher survival rates[1]. Thus, more patients have survived the 10-year mark than who have not. A common weakness of using generic machine learning models on highly imbalanced data is the tendency for a bias toward the dominant class, which can result in low sensitivity. To minimize this bias, our model adopted a class weighting technique, placing greater emphasis on correctly predicting outcomes for the minor class. Additionally, we tuned the confidence score cutoff and chose a customized classification threshold that was well-suited for our imbalanced dataset. To further address this problem, we used the harmonic mean of sensitivity and specificity, and F1 score, rather than accuracy, as the primary evaluation metrics to account for the cost of both false positives and false negatives.

A limitation of our prediction model was its relatively low sensitivity, even after the utilization of several machine learning techniques to balance the decision boundary. While low sensitivity is likely inevitable in the presence of imbalanced data, incorporating more discriminating features in a prediction model may resolve this issue. Additionally, our model includes patients' data on

molecular features, such as *P53* alterations, which are not routinely available for most bladder cancer patients. *P53* alterations are highly prevalent among bladder tumors, and previous studies have implied the association between the mechanisms of *P53* alteration and the characteristics of bladder tumors[11,21]. Based on our results, the *P53* intensity ranked highly among all predictive features, suggesting its potential value in bladder cancer prognosis. In future work, we will pursue the extraction of additional clinicopathological features to be incorporated into our prediction model. These features may not be routinely recorded in medical records; but they can be extracted from immunohistochemistry slides through automatic image analysis techniques. We expect that integrating these clinicopathological features with the presented model will improve its predictive performance.

Acknowledgments

The authors would like to thank Lamar Moss, Naofumi Tomita, Sophie Montgomery, Alice Hsu for their editorial help on this manuscript.

Competing Interests

The authors declare no competing interests.

References

1. American Cancer Society. Cancer Facts & Figures 2018. 76 (2018).
2. Engers, R. Reproducibility and reliability of tumor grading in urological neoplasms. *World J. Urol.* **25**, 595–605 (2007).
3. Robertson, A. J. *et al.* Observer variability in histopathological reporting of transitional cell carcinoma and epithelial dysplasia in bladders. *J. Clin. Pathol.* **43**, 17–21 (1990).
4. Yorukoglu, K. *et al.* Reproducibility of the 1998 World Health Organization/International Society of Urologic Pathology classification of papillary urothelial neoplasms of the urinary bladder. *Virchows Arch. Int. J. Pathol.* **443**, 734–740 (2003).
5. Ooms, E. C., Anderson, W. A., Alons, C. L., Boon, M. E. & Veldhuizen, R. W. Analysis of the performance of pathologists in the grading of bladder tumors. *Hum. Pathol.* **14**, 140–143 (1983).
6. Kluth, L. A. *et al.* Prognostic and Prediction Tools in Bladder Cancer: A Comprehensive Review of the Literature. *Eur. Urol.* **68**, 238–253 (2015).
7. Karagas, M. R. *et al.* Design of an epidemiologic study of drinking water arsenic exposure and skin and bladder cancer risk in a U.S. population. *Environ. Health Perspect.* **106 Suppl 4**, 1047–1050 (1998).
8. Mostofi, F. K., Sobin, L. H., Torloni, H. & Organization, W. H. Histological typing of urinary bladder tumours. (1973).
9. The World Health Organization/International Society of Urological Pathology consensus classification of urothelial (transitional cell) neoplasms of... - PubMed - NCBI. Available at: https://www.ncbi.nlm.nih.gov/pubmed/9850170. (Accessed: 28th June 2018)

10. Wittekind, C., Compton, C. C., Greene, F. L. & Sobin, L. H. TNM residual tumor classification revisited. *Cancer* **94**, 2511–2516 (2002).
11. Kelsey, K. T. *et al.* A population-based study of immunohistochemical detection of p53 alteration in bladder cancer. *Br. J. Cancer* **90**, 1572–1576 (2004).
12. Niu, Q. *et al.* The effect of marital status on the survival of patients with bladder urothelial carcinoma: A SEER database analysis. *Medicine (Baltimore)* **97**, e11378 (2018).
13. Social Security Administration. Social Security Death Master File. Available at: https://www.ssdmf.com/. (Accessed: 3rd July 2018)
14. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Mach. Learn. PYTHON* 6
15. King, G. & Zeng, L. Logistic Regression in Rare Events Data. (2001).
16. Song, B., Zhang, G., Zhu, W. & Liang, Z. ROC operating point selection for classification of imbalanced data with application to computer-aided polyp detection in CT colonography. *Int. J. Comput. Assist. Radiol. Surg.* **9**, 79–89 (2014).
17. Therneau, T. M. *A Package for Survival Analysis in S*. (2015).
18. Shariat, S. F., Milowsky, M. & Droller, M. J. Bladder Cancer in the Elderly. *Urol. Oncol.* **27**, 653–667 (2009).
19. Slater, S. *et al.* Muscle invasive bladder cancer in the elderly: Treatment and prognosis in a prospective cohort study. *J. Clin. Oncol.* **31**, 275–275 (2013).
20. Skeldon, S. C. & Larry Goldenberg, S. Bladder cancer: A portal into men′s health. *Urol. Oncol. Semin. Orig. Investig.* **33**, 40–44 (2015).
21. Malats, N. *et al.* P53 as a prognostic marker for bladder cancer: a meta-analysis and review. *Lancet Oncol.* **6**, 678–686 (2005).

Tables

Table 1. Characteristics of 811 bladder cancer patients used in this study.

Variable	*Type*	*Overall*	*Training*	*Testing*
n		811	525	286
Age (SD)		61.17 (10.33)	59.82 (10.04)	63.67 (10.41)
Sex (%)	Men	586 (72.3)	385 (73.3)	201 (70.3)
	Women	225 (27.7)	140 (26.7)	85 (29.2)
Body mass index (SD)		27.80 (5.03)	28.17 (4.71)	27.69 (5.12)
Family history of bladder cancer (%)	Yes	40 (4.9)	25 (4.8)	15 (5.2)
	No	724 (89.3)	453 (86.3)	270 (94.8)
	Unknown	47 (5.8)	47 (9.0)	0 (0.0)
High risk occupation (%)	Yes	284 (35.0)	128 (24.4)	127 (44.6)
	No	258 (31.8)	130 (24.8)	156 (54.7)
	Not available	269 (33.2)	267 (50.9)	2 (0.7)
Smoke status (%)	Current	246 (30.3)	158 (30.1)	88 (30.8)
	Former	397 (49.0)	252 (48.0)	145 (50.7)
	Never	160 (19.7)	110 (21.0)	50 (17.5)
	Unknown	8 (1.0)	5 (1.0)	3 (1.0)
Smoke pack-years (SD)		40.01 (30.19)	41.24 (31.06)	37.88 (28.58)
*TCC status (%)**	Confirmed TCC	729 (89.9)	456 (86.9)	273 (95.5)
	Not TCC	17 (2.1)	12 (2.3)	5 (1.7)
	Not available	65 (8.0)	57 (10.9)	8 (2.8)
Muscle invasiveness (%)	Yes	100 (12.3)	58 (11.0)	42 (14.7)
	No	705 (86.9)	466 (88.8)	239 (83.6)
	Not available	6 (0.7)	1 (0.2)	5 (1.7)
WHO 1973 classification (%)	Grade 1	336 (41.4)	193 (36.8)	143 (50.0)
	Grade 2	166 (20.5)	110 (21.0)	56 (19.6)
	Grade 3	161 (19.9)	85 (16.2)	76 (26.6)
	Grade 4	65 (8.0)	61 (11.6)	4 (1.4)
	Pathology material not reviewed	17 (2.1)	17 (3.2)	0 (0.0)
	Not available	66 (8.1)	59 (11.2)	7 (2.4)
*WHO/ISUP classification (%)**	CIS	28 (3.5)	15 (2.9)	13 (4.5)
	Papilloma	2 (0.2)	2 (0.4)	0 (0.0)

	PUNLMP	201 (24.8)	122 (23.2)	79 (27.6)
	LG-PUC	225 (27.7)	148 (28.2)	77 (26.9)
	HG-PUC	196 (24.2)	109 (20.8)	87 (30.4)
	Non-PUCHG	53 (6.5)	36 (6.9)	17 (5.9)
	Other	42 (5.2)	35 (6.7)	7 (2.4)
	Not available	64 (7.9)	58 (11.1)	6 (2.1)
TNM Stage (%)	CIS	41 (5.1)	29 (5.5)	12 (4.2)
	0	541 (66.7)	347 (66.1)	194 (67.8)
	1	123 (15.2)	90 (17.1)	33 (11.5)
	2	38 (4.7)	23 (4.4)	15 (5.2)
	3	22 (2.7)	14 (2.7)	8 (2.8)
	4	40 (4.9)	21 (4.0)	19 (6.6)
	Not available	6 (0.7)	1 (0.2)	5 (1.7)
P53 mutation (%)	Yes	14 (1.7)	14 (2.7)	0 (0.0)
	No	186 (22.9)	186 (35.4)	0 (0.0)
	Not available	611 (75.3)	325 (61.9)	285 (100.0)
P53 intensity (%)	3+	490 (60.4)	298 (56.8)	192 (67.1)
	< 3	182 (22.4)	120 (22.9)	62 (21.7)
	Not available	139 (17.1)	107 (20.4)	32 (11.2)
P53 positivity (%)	50% +	323 (39.8)	232 (44.2)	91 (31.8)
	< 50%	322 (39.7)	159 (30.3)	163 (57.0)
	Not available	166 (20.5)	134 (25.5)	32 (11.2)
PTCH LOH positivity (%)	Yes	119 (14.7)	119 (22.7)	0 (0.0)
	No	45 (5.5)	45 (8.6)	0 (0.0)
	Not available	647 (79.8)	361 (68.8)	285 (100.0)
First Course of Therapy (%)	No treatment	39 (4.8)	36 (6.9)	3 (1.0)
	TUR only	577 (71.1)	368 (70.1)	209 (73.1)
	Immunotherapy	96 (11.8)	63 (12.0)	33 (11.5)
	Intravesical Chemotherapy	8 (1.0)	5 (1.0)	3 (1.0)
	Radiation + Chemotherapy	13 (1.6)	6 (1.1)	7 (2.4)
	Cystectomy	78 (9.6)	47 (9.0)	31 (10.9)
Survival time in months (SD)		143.94 (55.94)	159.89 (58.68)	114.65 (35.11)
Death status at 10-year mark(%)	Dead	149 (18.4)	95 (22.1)	54 (18.9)
	Alive	662 (81.6)	430 (77.9)	232 (81.4)

*TCC – transitional cell carcinoma, HG-PUC –high grade papillary urothelial carcinoma, LG-PUC – low grade papillary urothelial carcinoma, PUNLMP – papillary urothelial neoplasm of low malignant potential, CIS – carcinoma *in situ*

Table 2. Result of multivariate Cox-proportional hazard model with our model predictor and the WHO/ISUP classification, adjusted for treatments.

Variable	Hazard	95% CI	P-Value
Our model predictor	**2.99**	**(1.54 – 5.83)**	**0.0013**
WHO/ISUP	1.22	(0.90 – 1.65)	0.1922
Treatment	1.21	(1.07 – 1.37)	0.0028