

Development and use of machine learning models for prediction of male sling success: A proof-of-concept institutional evaluation

Jin K. Kim^{*1,2}, Kurt A. McCammon^{*3}, Kellie J. Kim⁴, Mandy Rickard², Armando J. Lorenzo^{1,2}, Michael E. Chua^{2,5}

¹Division of Urology, Department of Surgery, University of Toronto, Toronto, ON, Canada; ²Division of Urology, Department of Surgery, The Hospital for Sick Children, Toronto, ON, Canada; ³Department of Urology, Eastern Virginia Medical School, Norfolk, VA; Devine-Jordan Center for Reconstructive Surgery and Pelvic Health, Urology of Virginia, Virginia Beach, VA, United States; ⁴Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada; ⁵Institute of Urology, St. Luke's Medical Center, Quezon City, Philippines

**Equal contributors*

Cite as: Kim JK, McCammon KA, Kim KJ, et al. Development and utilization of machine learning models for prediction of male sling success: A proof-of-concept institutional evaluation. *Can Urol Assoc J* 2023 July 21; Epub ahead of print. <http://dx.doi.org/10.5489/cuaj.8265>

Published online July 21, 2023

Corresponding author: Dr. Jin K. Kim, Division of Urology, Hospital for Sick Children, Toronto, ON, Canada; jjk.kim@mail.utoronto.ca

ABSTRACT

Introduction: For mild to moderate male stress urinary incontinence (SUI), transobturator male slings remain an effective option for management. We aim to use a machine learning (ML)-based model to predict those who will have a long-term success in managing SUI with male sling.

Methods: All transobturator male sling cases from August 2006 to June 2012 by a single surgeon were reviewed. Outcome of interest was defined as 'cure': complete dryness with 0 pads used, without the need for additional procedures. Clinical variables included in ML models were: number of pads used daily, age, height, weight, race, incontinence type, etiology of incontinence, history of radiation, smoking, bladder neck contracture, and prostatectomy. Model performance was assessed using AUROC, AUPRC, and F1-score.

Results: A total of 181 patients were included in the model. The mean followup was 56.4 months (standard deviation [SD] 41.6). Slightly more than half (53.6%, 97/181) of patients had procedural success. Logistic regression, K-nearest neighbor (KNN), naive Bayes, decision tree, and random forest models were developed using ML. KNN model had the best performance,

with AUROC of 0.759, AUPRC of 0.916, and F1-score of 0.833. Following ensemble learning with bagging and calibration, KNN model was further improved, with AUROC of 0.821, AUPRC of 0.921, and F-1 score of 0.848.

Conclusions: ML-based prediction of long-term transobturator male sling is feasible. The low numbers of patients used to develop the model prompt further validation and development of the model but may serve as a decision-making aid for practitioners in the future.

INTRODUCTION

Stress urinary incontinence in males is a bothersome condition that usually results from sphincter deficiency secondary to prostate surgeries or insults to the pelvic floor or sphincter.¹ While conservative management including pelvic floor therapy can be helpful, those who have refractory symptoms are considered for surgical management, usually in the form of artificial urinary sphincter (AUS) and more recently, transobturator male slings.²⁻⁴ AUS remains a preferred option for moderate to severe stress urinary incontinence (SUI) following radical prostatectomy, male slings can be considered for mild to moderate SUI.

The success rates of transobturator male slings have been reported as >70%. Clinical characteristics that were identified as predictors of success based on traditional regression included concomitant urge incontinence symptoms and preoperative SUI severity.⁵ Machine learning (ML) has been utilized in clinical medicine, especially in area of personalized medicine, to use data to predict and classify patients based on diagnosis or likelihood of treatment success.⁶ Despite its potential benefits, the use of ML has not been applied in creating a model to predict male SUI patients who may benefit from transobturator male slings. Hence, this investigation aims to develop an ML algorithm to predict male SUI patients who will have success following transobturator male sling insertion.

METHODS

Following approval by the institutional research ethics board (REB# 18-10-WC-0236), a retrospective assessment of prospectively maintained database on all male patients who underwent transobtruator male sling (AdVance Boston Scientific, Minnetonka, MN) insertion between August 2006 and June 2012 by a single surgeon was assessed. All redo cases were excluded from the analysis. Patients with missing data were excluded from the analysis. The operative technique has been described previously.⁷ While patients with severe stress urinary incontinence are counseled to pursue AUS, patients who prefer to pursue male sling despite this are still offered the procedure.⁵ The collected data was internally validated by a random counter-verification of 15% of total extracted data.

The machine learning classifier models assessed included logistic regression, K-nearest neighbor (KNN), Naïve Bayes, Decision Tree, and Random Forest. The pre-operative clinical variables included in the database were: age, smoking status, diagnosis of diabetes, race, height,

weight, prior prostatectomy, prior pelvic radiation, prior stress urinary incontinence management, bladder neck contractures, type of incontinence, severity of urgency incontinence (if present), potential need for concomitant procedures, number of pre-operative pads, and etiology of incontinence. Plot densities of these variables were created to determine whether each variable affected outcomes (Supplementary Figure 1). The closer the overlapping lines on these graphs, the less discrimination for ‘cure’ based on the assessed variable. Based on these, diagnosis of diabetes and concomitant procedures were less likely to contribute significantly to the model and was removed for feature selection.

Python 3.8.13 (Python Software Foundation, <http://python.org>) was used for model development. Following confirmation of lack of significant outliers, the continuous variables were standardized and scaled to each feature, avoiding biases due to variables being measured at different scales and contributing unequally to models. The models were built using an 85:15 train-test split (85% of data used for model training, 15% used for model performance evaluation, split at random). Grid search was performed to optimize and tune the hyperparameters of KNN and Random Forest models. Based on the highest performing algorithm, further ensemble learning method was applied. For this study, bagging method was chosen as it allows a random sample data in a training set to be selected with replacement to allow individual data to be utilized more than once. These new samples are trained independently and parallel to each other. In the end, based on these independently trained models, the majority prediction is taken and used to produce a more accurate outcome.

The model performance was assessed using sensitivity, specificity, area under receiver operating characteristic curve (AUROC), area under precision-recall curve (AUPRC), and F1-score. Validation curve of the ensemble model was assessed using validation curve. The model was subsequently calibrated using sigmoid method with 3 cross-validations to improve validation. We further interpreted the explainability of the KNN model by identifying the most important features in the final calibrated ensemble model using permutation importance.

RESULTS

A total of 181 patients were included in our analysis. Following train-test split at random, 153 patients were included in the training set and 28 patients were included in the testing set. Overall, the mean follow up was 56.4 months (SD 41.6), with at least 24-month follow-up data available for all patients. 53.6% (97/181) of patients had procedural success. The baseline characteristics of patients in each training and testing set are summarized in Table 1.

The five classifier models were developed using our data. Grid search showed that the best hyperparameters for Random Forest classifier was n-estimator of 57 and best hyperparameters for KNN was n-neighbor of 23 with uniform weight. The performances for each model are summarized in Table 2.

The AUROC and AUPRC curves were developed for all models (Figure 1). KNN model had the highest balanced performance among the five models with AUROC of 0.759, AUPRC of 0.916, and F-1 score of 0.833. The ensemble KNN model was developed using the bagging

method. The bagging KNN model had AUROC of 0.791, AUPRC of 0.919, and F-1 score of 0.812 (Table 2).

A validation curve was built for the bagging CNN model to assess for model reliability. The initial curve had a poor predictability when the predicted probability of cure was low. It was also more likely to over-forecast when predicted and true probability were <0.5 , and under-forecast when predicted and true probability were >0.5 . The bagged CNN model was then calibrated using sigmoid method. Following calibration, validation curve showed a curve closer to perfect calibration. It was much less likely to over-forecast with lower predictabilities. While it is still likely to under-forecast the likelihood of cure, it may serve as a more conservative tool to prevent false positives (Figure 2). The calibrated model had the best performance amongst all models, with AUROC of 0.821, AUPRC of 0.921, and F-1 score of 0.848 (Figure 3). The calibrated model hyperparameters are detailed in Appendix A.

Using permutation importance, the top features contributing to the model were identified. Pre-operative number of pads used was the most important feature in predicting success of male sling. In addition, weight and height (likely interacting to behave like body-mass index), as well as severity of incontinence and type of incontinence had relatively greater importance compared to other features (Supplementary Figure 2).

DISCUSSION

Artificial intelligence and machine learning allows creation of models that predict outcomes by performing challenging tasks as humans would do.⁹ Experts in their field often utilizes personalized medicine with plethora of learned experience and prior knowledge. However, for those with less experience, being able to have a model that will provide predictions based on large patient dataset will allow validation of one's own prediction and aid clinical decision making for healthcare practitioners and informed decision making for patients. With increasing computing power available, there is growing interest in utilizing machine learning to aid medical decision making and predictions. In urology, there has been efforts in incorporating machine learning to predict surgical outcomes in benign prostatic hyperplasia, as well as oncologic outcomes such as biochemical recurrence following radical prostatectomy.^{9,10} Machine learning has also been utilized for endourologic procedures and urolithiasis.¹¹ Despite numerous ongoing efforts to utilize machine learning in urologic practice, there has yet been a study that assessed its utility in prediction of male urinary incontinence surgeries. Therefore, this study assessed the potential clinical utility of machine learning in prediction of surgical outcome following transobturator male sling surgery. Using our institutional single surgeon database, we were able to train a machine learning model to achieve high predictive potential with AUROC and AUPRC of 0.821 and 0.921, respectively.

While personalized medicine is becoming increasingly popular in fields such as oncology where there is heterogeneity across disease and genes that may be strongly associated with outcomes, it can be possible across all fields of urology, including male functional urology.^{12,13} Mourmouris et al. (2021) suggested that this is possible as they describe prediction of clinical

outcomes of urinary flow following benign prostatic hyperplasia surgery using random forest classifier model.⁹ Chua et al. (2019) reported using the same database used in this study that long-term outcomes of the preoperative moderate to severe stress urinary incontinence was the only independent predictor for failure to achieve cure in long-term follow up.⁵ To supplement this knowledge, our machine learning based model using the same patient data can allow patients to understand their individualized risk of failing to achieve cure based on additional clinical characteristics that are incorporated into the machine learning algorithm training.

There are several limitations to this study. The first is the sample size – there were limited number of patients involved in developing this model, which may make our model noisy with high degree of variance. Nonetheless, as KNN performed the best among the five models that were initially assessed, we attempted to decrease the amount of noise and variance that would be affecting the prediction and were able to do so with ensemble learning. This is likely due to the nature of KNN, which performs relatively well in small datasets. It is simple and accurate for small datasets and as the dataset grows, KNN may become financially and computationally inefficient as it requires more memory and data storage compared to other models. It is true that there are concerns with overfitting (model too closely resembles the training dataset and may not perform well to external data) with KNN. However, we attempted to minimize this effect by using hyperparameter tuning to attain a k of 23, meaning prediction is based on 23 patients with most similar characteristics as the one being assessed (i.e. if ≥ 12 , the majority of the neighbors, had achieved ‘cure’ among the 23 neighbors, the model will predict ‘cure’ for the assessed patient; Supplementary Figure 3). Generally, the higher number of k leads to less overfitting as it averages the value over a greater neighbourhood.¹⁴ Moreover, using plot densities, we have attempted to reduce the number of features to reduce the noise by removing variables that did not seem to significantly contribute to the model.

We also constructed validation curves to assess for model reliability and was able to further calibrate our bagged KNN model to create the model with best prediction. The validation curves allow us to assess how well calibrated the model is in predicting its outcome. Our bagging KNN with calibration showed that while it may have a slightly higher likelihood of overpredicting success for some, there is a good concordance with a theoretical perfect calibrated model (Figure 2). Moreover, the more significantly contributing variables in the model (height, weight, severity of incontinence and type of incontinence) appear to be variables that have been shown in the past to be predictive of male sling success or stress urinary incontinence outcomes.^{15, 16, 17} While the high predictive potential from our current model suggests there may be a role for utilization of machine learning models in clinical practice of male stress urinary incontinence, its immediate utilization and generalization is limited with both small numbers and single surgeon series of patients. As the model was developed using a retrospectively collected data, there may also be predisposed to selection bias. Moreover, the outcome measures did not utilize validated questions such as ICIQ-SF to assess patient-reported outcomes. However, as our

outcome measure was complete dryness or zero pads used, there may be less subjectivity in reporting by patients.

Despite the limitations, this is the first study utilizing machine learning models to predict outcomes in male stress urinary incontinence. Our model shows promise for utilization in clinical practice. With increasing the amount of patient data by prospectively maintaining our database and involving other institutions, we hope to continue to evolve this model to improve its predictive performance and generalizability.

CONCLUSIONS

ML-based prediction of long-term transobturator male sling is feasible. The low numbers of patients used to develop the model prompt further validation and development of the model but may serve as a decision-making aid for practitioners in the future.

DRAFT

REFERENCES

1. Chung E, Katz DJ, Love C. Adult male stress and urge urinary incontinence - A review of pathophysiology and treatment strategies for voiding dysfunction in men. *Aust Fam Physician*. 2017 Sep;46(9):661-666. PMID: 28892597.
2. Rehder P, Haab F, Cornu JN, Gozzi C, Bauer RM. Treatment of postprostatectomy male urinary incontinence with the transobturator retroluminal repositioning sling suspension: 3-year follow-up. *Eur Urol*. 2012 Jul;62(1):140-5. doi: 10.1016/j.eururo.2012.02.038. Epub 2012 Feb 25. PMID: 22386196.
3. Kowalik CG, DeLong JM, Mourtzinou AP. The advance transobturator male sling for post-prostatectomy incontinence: subjective and objective outcomes with 3 years follow up. *Neurourol Urodyn*. 2015 Mar;34(3):251-4. doi: 10.1002/nau.22539. Epub 2013 Dec 7. PMID: 24318744.
4. Bauer RM, Grabbert MT, Klehr B, Gebhartl P, Gozzi C, Homberg R, May F, Rehder P, Stief CG, Kretschmer A. 36-month data for the AdVance XP® male sling: results of a prospective multicentre study. *BJU Int*. 2017 Apr;119(4):626-630. doi: 10.1111/bju.13704. Epub 2016 Dec 2. PMID: 27862836.
5. Chua ME, Zuckerman J, Mason JB, DeLong J, Virasoro R, Tonkin J, McCammon KA. Long-term Success Durability of Transobturator Male Sling. *Urology*. 2019 Nov;133:222-228. doi: 10.1016/j.urology.2019.07.032. Epub 2019 Aug 8. PMID: 31401219.
6. Schork NJ. Artificial Intelligence and Personalized Medicine. *Cancer Treat Res*. 2019;178:265-283. doi:10.1007/978-3-030-16391-4_11
7. Zuckerman JM, Edwards B, Henderson K, Beydoun HA, McCammon KA. Extended outcomes in the treatment of male stress urinary incontinence with a transobturator sling. *Urology*. 2014 Apr;83(4):939-45. doi: 10.1016/j.urology.2013.10.065. Epub 2014 Feb 5. PMID: 24507897.
8. Maadi M, Akbarzadeh Khorshidi H, Aickelin U. A Review on Human-AI Interaction in Machine Learning and Insights for Medical Applications. *Int J Environ Res Public Health*. 2021 Feb 22;18(4):2121. doi: 10.3390/ijerph18042121. PMID: 33671609; PMCID: PMC7926732.
9. Mourmouris P, Tzelves L, Feretzakis G, Kalles D, Manolitsis I, Berdempes M, Varkarakis I, Skolarikos A. The use and applicability of machine learning algorithms in predicting the surgical outcome for patients with benign prostatic enlargement. Which model to use? *Arch Ital Urol Androl*. 2021 Dec 20;93(4):418-424. doi: 10.4081/aiua.2021.4.418. PMID: 34933537.
10. Ekşi M, Evren İ, Akkaş F, Arıkan Y, Özdemir O, Özlü DN, Ayten A, Sahin S, Tuğcu V, Taşçı AI. Machine learning algorithms can more efficiently predict biochemical recurrence after robot-assisted radical prostatectomy. *Prostate*. 2021 Sep;81(12):913-920. doi: 10.1002/pros.24188. Epub 2021 Jul 5. PMID: 34224165.
11. Suarez-Ibarrola R, Hein S, Reis G, Gratzke C, Miernik A. Current and future applications of machine and deep learning in urology: a review of the literature on urolithiasis, renal cell carcinoma, and bladder and prostate cancer. *World J Urol*. 2020 Oct;38(10):2329-2347. doi: 10.1007/s00345-019-03000-5. Epub 2019 Nov 5. PMID: 31691082.

12. Zhang S, Bamakan SMH, Qu Q, Li S. Learning for Personalized Medicine: A Comprehensive Review From a Deep Learning Perspective. *IEEE Rev Biomed Eng.* 2019;12:194-208. doi: 10.1109/RBME.2018.2864254. Epub 2018 Aug 7. PMID: 30106692.
13. MacEachern SJ, Forkert ND. Machine learning for precision medicine. *Genome.* 2021 Apr;64(4):416-425. doi: 10.1139/gen-2020-0131. Epub 2020 Oct 22. PMID: 33091314.
14. Sun K, Du W, Shi N. A survey of kNN algorithm. *Inf Eng Appl Comput.* 2018 May 10;1(1).
15. Monn MF, Jarvis HV, Gardner TA, Mellon MJ. Impact of obesity on male urethral sling outcomes. *Ther Adv in Urol.* 2020;12. doi:10.1177/1756287220927997. Epub 2020 Jun 09. PMID: 32565915.
16. Collado SA, Resel FL, Domínguez-Escrig JL, Gómez-Ferrer A, Rubio-Briones J, Solsona NE. AdVance/AdVance XP transobturator male slings: Preoperative degree of incontinence as predictor of surgical outcome. *Urology.* 2013 May;81(5):1034-9. doi: 10.1016/j.urology.2013.01.007. Epub 2013 Mar 7. PMID: 23465151.
17. Cornu JN, Sèbe P, Ciofu C, Peyrat L, Cussenot O, Haab F. Mid-term evaluation of the transobturator male sling for post-prostatectomy incontinence: Focus on prognostic factors. *BJU Int.* 2011 Jul;108(2):236-40. doi: 10.1111/j.1464-410X.2010.09765.x. Epub 2010 Oct 18. PMID: 20955265.

DRAFT

FIGURES AND TABLES

Figure 1. (A) Receiver operating curves for evaluated models. (B) Precision recall curves for evaluated models.

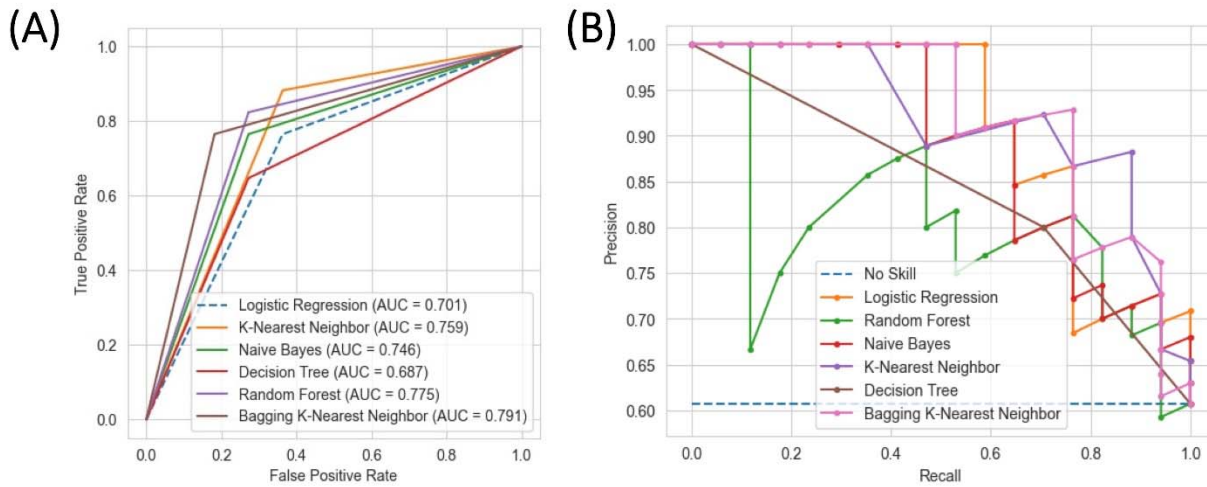


Figure 2. Validation curves for bagging K-nearest neighbor (KNN) algorithm pre- and post-calibration.

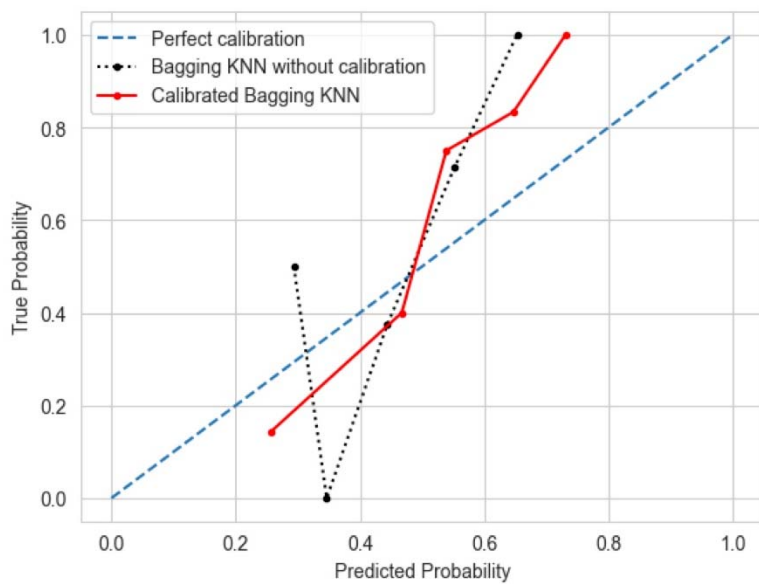
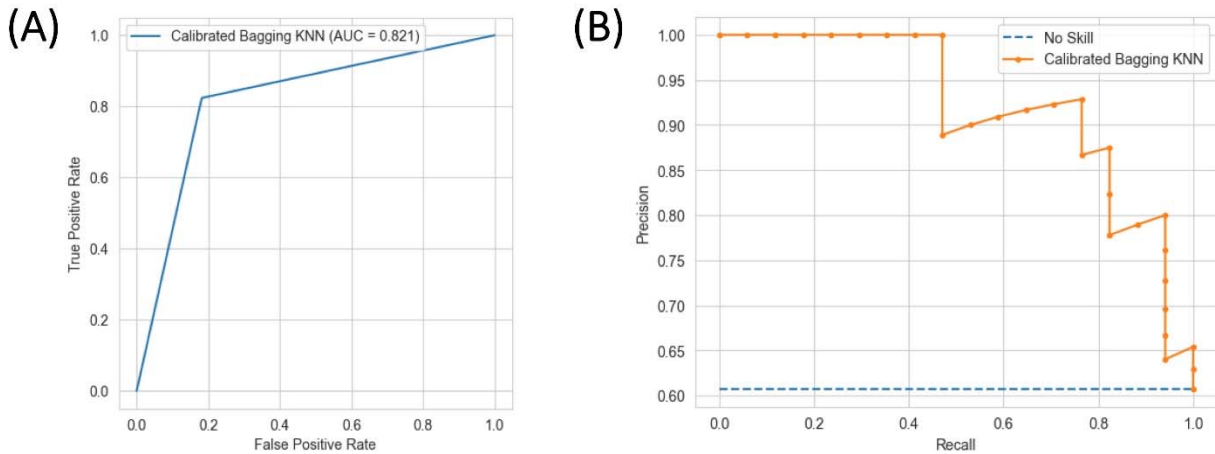


Figure 3. (A) Received operating characteristics curve for calibrated bagged K-nearest neighbor (KNN). (B) Precisions recall curve for calibrated bagged KNN.**Table 1. Summary of baseline characteristics for training and testing groups**

		Training (n=153)		Testing (n=28)	
		Median, n	IQR, %	Median, n	IQR, %
Age at sling surgery (years)		67.5	63.6-72.0	67.6	60.8-74.4
Height at sling surgery (cm)		177.8	172.7-182.9	177.8	172.9-182.9
weight at sling surgery (kg)		89.6	81.4-98.9	90.2	84.0-94.2
Pre-op number of pads		3.5	2.0-5	3.5	2.0-4.6
Race	White	105	68.6%	19	67.9%
	Black	39	25.5%	7	25.0%
	Other	9	5.9%	2	7.1%
Incontinence type	Stress urinary incontinence	109	71.2%	20	71.4%
	Mixed urinary incontinence	44	28.8%	8	28.6%
Incontinence etiology	Prostatectomy	140	91.5%	25	89.3%
	Other prostate therapy	7	4.6%	1	3.6%
	Radiation	3	2.0%	2	7.1%
	Neurogenic bladder/spinal cord injury	3	2.0%	0	0.0%
No		136	88.9%	24	85.7%

Concomitant procedure required	Yes	17	11.1%	4	14.3%
Pelvic radiation	No	121	79.1%	21	75.0%
	Yes	32	20.9%	7	25.0%
Prostatectomy	No	15	9.8%	4	14.3%
	Yes	138	90.2%	24	85.7%
Smoking history	No	72	47.1%	12	42.9%
	Prior smoker	69	45.1%	14	50.0%
	Current smoker	12	7.8%	2	7.1%
Bladder neck contracture	No	118	77.1%	24	85.7%
	Yes	35	22.9%	4	14.3%

IQR: interquartile range.

Model	Sensitivity	Specificity	AUROC	AUPRC	F-1 score
Logistic regression	0.764	0.636	0.701	0.765	0.765
K-nearest neighbor	0.882	0.636	0.759	0.916	0.833
Naïve Bayes	0.764	0.727	0.746	0.891	0.788
Decision tree	0.647	0.727	0.687	0.842	0.750
Random forest	0.823	0.727	0.775	0.801	0.750
Bagging K-nearest neighbor	0.764	0.818	0.791	0.919	0.812
Calibrated bagging K-nearest neighbor	0.823	0.818	0.821	0.921	0.848