Predicting intraoperative and postoperative consequential events using machine learning techniques in patients undergoing robotic partial nephrectomy (RPN): Vattikuti Collective Quality Initiative (VCQI) database study

Authors:

Mahendra Bhandari[1], Anubhav Reddy Nallabasannagari[2], Madhu Reddiboina[2], James R. Porter[3], Wooju Jeong[1], Alexandre Mottrie[4], Prokar Dasgupta[5], Ben Challacombe[6], Ronney Abaza[7], Koon Ho Rha[8], Dipen J. Parekh[9], Rajesh Ahlawat[10], Umberto Capitanio[11], Thyavihally B. Yuvaraja[12], Sudhir Rawal[13], Daniel A. Moon[14], Nicolò M Buffi's[15], Ananthakrishnan Sivaraman[16], Kris K. Maes[17], Francesco Porpiglia[18], Gagan Gautam[19], Levent Turkeri[20], Kohul Raj Meyyazhgan[2], Preethi Patil[21], Mani Menon[1], Craig Rogers[1]

[1] Vattikuti Urology Institute, Henry Ford Hospital, Detroit, USA
[2] Rediminds Research Inc., Southfield, MI, USA
[3] Swedish Medical Center, Seattle, WA, USA
[4] OLV Vattikuti Institute, Aalst, Belgium
[5] MRC Centre of Transplantation, King's College London, London, United Kingdom
[6] Guy's & St Thomas' Hospitals, London, United Kingdom
[7] Ohio Health Dublin Methodist Hospital, Dublin, OH, USA
[8] Yonsei University, Seoul, Korea
[9] University of Miami Miller School of Medicine and Sylvester Comprehensive Cancer Center, Miami, FL, USA
[10] Medanta Vattikuti Institute, Medanta – The Medicity, Gurugram, Haryana, India
[11] Urology Clinic, San Raffaele Hospital, Milan, Italy
[12] Kokilaben Dhirubhai Ambani Hospital, Mumbai, India
[13] Rajiv Gandhi Cancer Institute and Research Centre, New Delhi, India
[14] Peter MacCallum Cancer Center, Melbourne, Victoria, Australia
[15] Humanitas Research Hospital, Milan, Italy
[16] Apollo Hospitals, Chennai, India
[17] Center for Robotic and Minimally Invasive Surgery, Hospital Da Luz, Luz Sáude, Portugal
[18] San Luigi Gonzaga Hospital, University of Turin, Orbassano, Italy
[19] Max Institute of Cancer Care, Saket, India
[20] Acıbadem University School of Medicine, Istanbul, Turkey
[21] Vattikuti Foundation

Keywords: Deep Learning, Intraoperative Complications, Machine Learning, Postoperative Complications, Postoperative Morbidity, Robotic Partial Nephrectomy

Corresponding Author Details

Name: Mahendra Bhandari

Address: Suit 700, 1000 Town Center, Southfield MI 48075

Telephone Number: +1 248-205-0605

E-mail: mahendra@vattikutifoundation.com

PROF. MAHENDRA BHANDARI (Orcid ID : 0000-0003-1789-5931)

DR. JAMES ROSCOE PORTER (Orcid ID : 0000-0001-9275-4552)

PROF. KOON HO RHA (Orcid ID : 0000-0001-8588-7584)

DR. UMBERTO CAPITANIO (Orcid ID : 0000-0001-9225-1795)

MR. DANIEL A MOON (Orcid ID : 0000-0002-1557-2584)

DR. FRANCESCO PORPIGLIA (Orcid ID : 0000-0001-8567-6795)

DR. GAGAN GAUTAM (Orcid ID : 0000-0002-3106-706X)

DR. CRAIG G. ROGERS (Orcid ID : 0000-0002-7781-6644)

Article type : Original Article

Article category: Urological Oncology

Abstract

Objective: To predict intraoperative events (IOE) and postoperative events (POE) consequential to the derailment of the ideal clinical course of patient recovery.

Material and Methods: Vattikuti Collective Quality Initiative (VCQI), a multi-institutional dataset of patients who underwent Robotic Partial Nephrectomy for kidney tumors. Machine Learning (ML) models were constructed to predict IOE, and POE using Logistic Regression, Random Forest, and Neural Networks. The models to predict IOE used patient demographics and preoperative data. In addition to the above, intraoperative data was used to predict POE. Performance on the test dataset was assessed using Area Under Receiver Operating Curve (AUC-ROC) and Area Under Precision-Recall Curve (PR-AUC).

Results: The rate of IOE and POE was 5.62% and 20.98%, respectively. Models for predicting IOE were constructed using data from 1690 patients and 38 variables; the best model had AUC-ROC of 0.858 (95% CI, 0.762, 0.936), and PR-AUC of 0.590 (95% CI, 0.400, 0.759). Models for predicting POE were trained using data from 1406 patients and 59 variables; the best model had AUC-ROC of 0.875 (95% CI, 0.834, 0.913), and PR-AUC 0.706 (95% CI, 0.610, 0.790).

Conclusions: The performance of the ML models in this study is encouraging. Further validation in a multi-institutional clinical setting with larger datasets would be necessary to establish their clinical value. ML models can be used to predict significant events during and after surgery with good accuracy, paving the way for application in clinical practice to predict and intervene at an opportune time to avert complications and improve patient outcomes.

1. Introduction

Current guidelines recommend partial nephrectomy (PN) for T1a masses, as and when indicated. The robotic surgery has gained popularity among surgeons[1] but with an overall complication rate of up to 30% and major complications (Clavien-Grade≥3) rate of 3%-6%[2]. The morbidity of the PN is attributed to the combination of tumor complexity, patient-related comorbidities[3], tumor surroundings[4], and surgeon experience[5,6]. Significant efforts have been made to develop tools to individualize patient risk profiles for better surgical planning. Instruments such as RENAL[7], PADUA[8], and MAP[9] scores are mainly based on preoperative imaging. However, their utility in clinical practice remains questionable[10,11]. Surgical risk calculators were made with suboptimal predictability and virtually no clinical utility[12].

Artificial Intelligence (AI), a subfield of Machine Learning (ML) has been leveraged to improve clinical diagnosis and decision-making in the many areas of healthcare: Radiology, Dermatology, Ophthalmology, Pathology, Genome interpretation, Biomarker discovery, clinical outcome prediction, patient monitoring, inferring health through wearable devices, and Autonomous robotic surgery[13]. The application of AI techniques to the management of urological cancer has also been reported[14]. Access to large datasets in combination with the rapid progress of modern machine learning, data mining, data engineering techniques offers a promising opportunity to building models that could translate to valuable clinical practice tools for the personalized care of a patient.

The objective of this retrospective study was to apply ML/AI models to predict the probability of a patient having Intraoperative events (IOE) and postoperative events (POE), which are likely to adversely impact the clinical course for the patient undergoing robotic partial nephrectomy (RPN). We further propose a design for the potential deployment of these models for clinical validation in a prospective clinical setting.

2. Material and Methods

2.1 Dataset

VCQI, a multi-institutional dataset of patients who underwent Robotic Partial Nephrectomy for kidney tumors, was leveraged for this study. Eighteen centers from around the world contributed to the dataset. Ethics committee/IRB clearance for data collection was obtained by each center contributing to VCQI database.

2.1.1 Intraoperative Events (IOE)

Gross violation of tumor bed, major bleeding from the tumor bed, injury to major vessels, injury to abdominal organs, conversion to open, and intraoperative blood transfusion >= 1 unit were defined as significant intraoperative events that are likely to delay the recovery of a patient (Table 1). After accounting for missing values and outliers, the resultant dataset for IOE comprised 1690 patients and 38 predictors. The data processing steps are presented in subsection 2.2.1. The median age of patients was 58.0 [Q1-Q3, 48.0-66.0] years, and the median BMI was 27.3 [Q1-Q3, 24.3-30.1]. The rate of POE was 5.62%. Over 90% of the patients had a clinical staging of T1 tumors. Supplementary tables S1 and S2 illustrate the summary statistics for demographic variables and preoperative variables (Table 2).

2.1.2 Postoperative Events (POE)

Clavien-Dindo grade ≥ 3, and length of hospital stay greater than 75th percentile (>4 days in our database) within 30 days of surgery were considered as POE (Table 1). After accounting for missing values and outliers, the resultant dataset comprised 1406 patients and 59 predictors. The data processing steps are presented in subsection 2.2.1. The median age of patients was 57.0 [Q1-Q3, 48.0-66.0] years, and the median BMI was 27.2 [Q1-Q3, 24.3-30.0]. The male to female ratio was 66:34. The POE rate was 20.98%. Over 90% of the patients had a clinical staging of T1 tumors. Supplementary tables S3, S4, and S5 illustrate summary stats for demographic, preoperative, and intraoperative variables (Table 2).

2.2 Model Development

We trained classification models using Logistic Regression (LR), Random Forests (RF), and Neural Networks (NN). LR has been traditionally used due to its ease of result interpretation and analysis. RF and NN are non-parametric models capable of modeling complex non-linear relationships.

2.2.1 Data Processing

Missing values for numeric variables were imputed with mean value, and missing values for categorical variables were encoded as a separate category for each variable. Categorical variables were one-hot encoded, and numeric variables were standardized for LR and NN. Multicollinearity of the predictor variables was tested, and it was observed that none of the predictors were highly correlated with other predictors. The models were trained using balanced weights (inverse ratio of class samples multiplied by the number of classes to total samples) to account for imbalance in the dataset. All transformations were performed on the training dataset and then applied to validation and test datasets. Stratified random sampling was performed to assign 30% of the dataset as the test dataset and remaining data as the training dataset. The test dataset was used for final evaluation only, to prevent overfitting.

2.2.2 Logistic Regression

Logistic Regression is an extension of Linear Regression where the output of the model is restricted between 0 and 1 using the logistic function. It models the probabilities for classification problems with a binary outcome. L2 regularization was used to prevent overfitting. The regularization strength, and the solver for optimizing cost function were tuned as part of hyperparameter tuning.

2.2.3 Random Forest

Random Forest is an ensemble method for classification and regression tasks. Many uncorrelated decision trees are created at training time using a subset of features and bootstrapped samples (sampling with replacement) of training data. Once trained, the predictions from individual trees are aggregated and provided as the output of the model. The process of generating aggregated results from uncorrelated tress makes Random Forest less susceptible to overfitting. The number of trees, the maximum number of features used at each split, and the minimum sample size per leaf were identified using hyperparameter tuning.

2.2.4 Neural Network Architecture

Neural Networks are composed of units of calculation called neuron and are capable of modelling complex-patterns present in the data. Neural networks can contain many types of layers that perform calculations on input received from the previous layer. Our NN models (Supplementary Figure S1) comprised of the following layers: an input layer, dense layers, dropout layers, and an output layer[15,16]. The input layer was provided with data in the form of a 2-d array. After the input layer, a 2-layered dense network with a dropout layer after each dense layer was implemented. Each dense layer comprised of neurons and used rectified linear unit (ReLU) for non-linear activation. Output layer comprised of a single neuron with a sigmoid activation function. The dropout layers were used to prevent overfitting of the network to training data[17]. The number of neurons in dense layers, the rate of dropout for each hidden layer, and the learning rate were tuned as part of the hyperparameter tuning for each model. All models were trained to minimize the loss of function, i.e., binary cross-entropy using Adam optimizer.

2.2.3 Hyperparameter Tuning

Evaluation of a hyperparameter setting for each model was performed using 'Grid-search' with 10-fold cross-validation. In 10-fold cross-validation, the training dataset was split into 10 stratified smaller sets. For each of 10 folds, a model with a specific set of hyperparameters was trained on 9 sets and evaluated on the remaining 1 set. Model with the best average performance over 10-folds was selected as the final model. We used Python (Python Software Foundation) with Scikit-learn[18] package and TensorFlow[19] for analysis.

2.3 Model Evaluation and Statistical Analysis

We assessed the comparative performance of our models using Area Under Receiver Operating Curve (AUC-ROC) and Area Under Precision-Recall curve (PR-AUC). In highly skewed datasets, PR-AUC is shown to be more informative in the evaluation of model performance[20,21]. We used bootstrapping to generate confidence intervals for the scores[22,23] and performed a permutation test to assess if the observed difference between the models was significant. We resampled the test data 10,000 times with replacement to generate a 95% confidence interval (CI). We performed the permutation test by simulating a bootstrapped population and checked the likelihood of getting the observed difference in AUC-ROC and PR-AUC.

The permutation feature importance technique was used to assess the importance of the features[24]. This is a model agnostic procedure where the values of a feature are randomly shuffled to break the relationship between the feature and the target, and the drop in the model score is observed. The drop in the model score indicates the importance of the feature. Each feature was permuted 100 times, and the drop in the model score was assessed, and a 95% confidence interval was constructed.

3. Results

Three models each were trained for predicting IOE and POE, resulting in a total of six models. The models were trained using LR, RFC, and NN algorithms. Each algorithm was trained using balanced weights. The performance of the models constructed for each outcome was assessed using the same test data.

3.1 Intraoperative Events

The models developed for predicting IOE denoted as LR_IOE, RFC_IOE, and NN_IOE for Logistic Regression, Random Forest Classifier, and Neural Network, respectively. The performance of the models on the test dataset is shown in figure 1. The sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F-1 score are reported in table 3.

The AUC-ROC and PR-AUC for model RFC_IOE was 0.858 (95% CI 0.762, 0.936) and 0.590 (95% CI 0.400, 0.759), respectively. The AUC-ROC for model LR_IOE and NN_IOE was 0.826 (95% CI 0.731, 0.905) and 0.856 (95% CI 0.779, 0.923), respectively. The PR-AUC for model LR_IOE and NN_IOE was 0.372 (95% CI 0.189, 0.552) and 0.398 (95% CI 0.212, 0.605), respectively (Table 4). The observed PR-AUC difference between RFC_IOE and LR_IOE was 0.221 (p-value=0.035), and between RFC_IOE and NN_IOE was 0.208 (p-value=0.067). RFC_IOE outperformed LR_IOE and NN_IOE.

3.2 Postoperative Events

The models constructed for predicting POE are denoted as LR_POE, RFC_POE, and NN_POE for Logistic Regression, Random Forest Classifier, and Neural Network, respectively. The performance of the models on the test dataset is shown in figure 1. The sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F-1 score are reported in table 3.

The AUC-ROC and PR-AUC for model RFC_POE was 0.875 (95% CI 0.834, 0.913) and 0.706 (95% CI 0.610, 0.790), respectively. The AUC-ROC for model LR_POE and NN_POE was 0.837 (95% CI 0.786, 0.882) and 0.837 (95% CI 0.786, 0.883), respectively. The PR-AUC for model LR_POE and NN_POE was 0.591 (95% CI 0.477, 0.701) and 0.649 (95% CI 0.549, 0.739), respectively (Table 4). The observed PR-AUC difference between RFC_IOE and LR_IOE was 0.121 (p-value = 0.068), and between RFC_IOE and NN_IOE was 0.057 (p-value=0.21). RFC_POE outperformed LR_POE and NN_POE.

3.3 Feature Importance

The permutation feature importance procedure was performed for the best model for each outcome, i.e., RF_IOE for IOE and RF_POE for POE. The feature importance of features used for constructing models RF_IOE and RF_POE are reported in supplementary table S6 and supplementary table S7, respectively. *Multifocality, Clinical staging – regional lymph nodes, Center code, and Solitary kidney* were observed to be the most important features for RF_IOE (Supplementary table S6). *Center code, Haemostatic agents, Center volume, Ischemia time(min), and Race* were observed to be top five important features for RF_POE (Supplementary table S7). The most important feature for IOE was Multifocality with a value of 0.192 [95% CI 0.136, 0.217] and for POE was Center Code with a value 0.037 [95% CI 0.018, 0.054]. The feature importance value is the degradation in the model score when the values of a feature are randomly shuffled.

4. Discussion

Surgeons always endeavor to bring predictability to their actions and decisions taken to manage patients under their care. With the successful application of machine learning, it is possible to predict many intraoperative and postoperative events, which could be consequential to the smooth clinical course of the patient. Predictive models, thus created, can enable surgeons to identify high-risk cohorts, preempt consequential events, and plan therapeutic strategies to improve patient outcomes. The Robotic partial nephrectomy for small T1 renal tumors is establishing itself as a safe and day-case surgery procedure. For a strict case selection, the surgeon needs accurate predictive tools to preempt consequential events as a precursor to emergency readmission of the patient. Currently, different risk profile tools are in practice to identify high-risk cohorts.

The existing surgical risk predictors such as PADUA, MAP, and RENAL played a significant role in assisting clinicians in the risk assessment of a patient, but it is time their application needs to be enhanced to the development of tools prompting an intervention to prevent complications and improve overall patient outcomes. ACS NSQIP risk calculator was proposed as an alternative tool but did not establish clinical value for predicting post-surgical complications for patients who underwent RPN[12]. AI-driven predictive models hold promise to fill in the gap in the field, open an opportunity to assess the potential utility of machine learning, and the continued use of these models to develop intervention protocols to manage predicted complications. In a recent study on the treatment of sepsis, the treatment selected by the AI model was, on average, reliably more objective and effective than that chosen by humans[25]. Manaktala et al. applied algorithms to detect sepsis and deliver highly sensitive specific decision support tools to the point of care using a mobile application. In their practice, the sepsis mortality decreased by 53%, and the 30 days readmission rate dropped from 19.08% to 13.21%[26]. Recently, in urology, machine learning was used to predict urinary continence recovery[27] and early biochemical recurrence after robot-assisted prostatectomy[28], and to detect low and high-grade clear cell renal cell carcinomas (cc-RCCs)[29].

Conversion from minimally invasive surgery to open surgery is generally considered as a part of the procedure and not a complication. However, the prediction of such an event preoperatively would be of high value to the surgeon. Conversion is known to increase the hospital stay and adds to significant post-surgical 30-day events. Shumate et al., in their report, discuss that conversion to open surgery was a significant reason for a prolonged hospital stay, which increases 30 days morbidity[30]. Accurate prediction of prolonged length of stay could improve case selection for RPN as an outpatient procedure and to consul the patient and the family with certainty. While deciding on the cutoff limit of 4 days as the (>75th percentile) as a postoperative event, we agree with Sperling et al., which included hospital stay greater than 75th percentile in their dataset as a secondary outcome in the study[31]. Shumate et al. reported that 80% of patients discharged within the 3rd day of the surgery[30]. Center volume was calculated as the number of surgeries performed by each center before operating on a given patient, and it was based on the information available in our database.

In this study, AI/ML algorithms were used to construct models to predict consequential IOE and POE. Three models were built for each outcome, and the best model was selected based on the performance on an unseen test dataset. The variables for these models were selected based on the availability of data points in VCQI, taking a cue from the published research on the subject[12,32–34]. We used AUC-ROC and PR-AUC to assess model performance. In highly skewed datasets, PR-AUC is shown to be more informative in the evaluation of model performance[20,21]. AUC-ROC Curve takes into account True Positive Rate (TPR) and False Positive Rate (FPR). In the case of highly imbalanced datasets, where the bulk of the target variable is composed of negative class, true negatives can greatly influence change in FPR. This can result in an overly optimistic representation when the classifier is ineffective at predicting positive class but predicts negative class accurately. In contrast, PR-AUC depends on RECALL, i.e., TPR and Precision (ratio of number true positives divided by the sum of true positives and false negatives). Precision indicates how many of the positive predictions were positive labels. Hence, PR-AUC is a better measure for evaluating the performance of binary classifiers on highly imbalanced datasets.

The permutation feature importance was calculated to identify the features which contribute most to the overall performance of the model. The feature importance score does not reflect the intrinsic predictive value of a feature but indicates its contribution to the performance of the model on the test dataset. A feature found to be unimportant for a model under inspection could be an important feature for a model with higher performance. *Multifocality, Clinical staging, Center experience, and Solitary kidney* were observed to be the most important features for a model constructed for predicting intraoperative vents using random forest, i.e., RF_IOE. *Center experience, Haemostatic agents, Center volume, Ischemia time(min),* and *Race* were observed to be the most important features for a model constructed to predict postoperative events using random forest, i.e., RF_POE. In recent studies, surgeons with high volumes were reported to have lower perioperative outcomes[5] and significant variability was observed between surgeon for outcomes post partial Nephrectomy[6]. The *center code* was used as a surrogate for *center experience*. In another study, ischemia times were found to be significantly associated with longer length of stay and postoperative complications[32,35,36].

Hospital volume is known to be significantly associated with postoperative complications and prolonged length of stay[37–39]. Racial disparities were reported to exist for in the use of PN across hospitals and postoperative complications after partial nephrectomy[40,41].

This study has some limitations. First, the variables in the dataset do not account for the temporal shift in patient characteristics during the admission. Second, intraoperative outcomes are known to vary significantly between surgeons. Unfortunately, surgeon information was not available for many patients; therefore, it could not be leveraged for constructing the models in this study. Third, patient data, such as imaging data and caregiver notes, were not available. Fourth, predicting individual intraoperative events or postoperative events (clavien-grade>=3) was not possible due to very few occurrences of these individual events. Fifth, the cohort of patients was small, i.e., 1690 patients for intraoperative events and 1406 patients for postoperative events. Further studies will be required with larger cohorts to validate the findings of this study.

We propose deploying the models at participating centers contributing to the database. The deployment process will include the following steps: (1) Re-validate models with new data in the VCQI database; (2) Deploy models in co-ordination with respective centers; (3) Monitor performance and retrain models when performance falls below a predetermined threshold (Figure 2). This would enable the centers to leverage the models for identifying potential patients likely to have a complication during or after surgery and innovate intervention protocols and provide feedback on model performance. This would also enable us to monitor the performance of the models in a clinical setting.

5. Conclusion

The clinician's efforts to profile risk before partial nephrectomy with the objective to tailor the surgical plan accordingly is a standard of care. In this study, ML/AI models performed well in predicting intraoperative events and 30-day postoperative events. We hope to turn these models into regularly applied clinical tools. However, the true value of effort could only be known in the years to come.

6. Acknowledgements

We gratefully acknowledge discussions and comments on the manuscript by our colleagues Trevor Zeffiro. This are grateful to Vattikuti Foundation for granting access to the VCQI database and RediMinds for funding this work. This publication only reflects the authors views. Funding agency is not liable for any use that may be made of the information contained herein.

Conflicts of interest

None disclosed

7. References:

1. Bertolo R, Autorino R, Simone G, Derweesh I, Garisto JD, Minervini A, et al. Outcomes of robot-assisted partial nephrectomy for clinical T2 renal tumors: a multicenter analysis (ROSULA Collaborative Group). Eur Urol. 2018;74(2):226–32.

2. Tanagho YS, Kaouk JH, Allaf ME, Rogers CG, Stifelman MD, Kaczmarek BF, et al. Perioperative complications of robot-assisted partial nephrectomy: analysis of 886 patients at 5 United States centers. Urology. 2013;81(3):573–80.

3. Ficarra V, Bhayani S, Porter J, Buffi N, Lee R, Cestari A, et al. Predictors of warm ischemia time and perioperative complications in a multicenter, international series of robot-assisted partial nephrectomy. Eur Urol. 2012;61(2):395–402.

4. Khene Z-E, Peyronnet B, Kocher NJ, Robyak H, Robert C, Pradere B, et al. Predicting morbidity after robotic partial nephrectomy: the effect of tumor, environment, and patient-related factors. In: Urologic Oncology: Seminars and Original Investigations. 2018. p. 338--e19.

5. Khandwala YS, Jeong IG, Kim JH, Han DH, Li S, Wang Y, et al. The impact of surgeon volume on perioperative outcomes and cost for patients receiving robotic partial

nephrectomy. J Endourol. 2017;31(9):851–7.

6. Dagenais J, Bertolo R, Garisto J, Maurice MJ, Mouracade P, Kara O, et al. Variability in Partial Nephrectomy Outcomes: Does Your Surgeon Matter? Eur Urol. 2019;75(4):628–34.

7. Kutikov A, Uzzo RG. The RENAL nephrometry score: a comprehensive standardized system for quantitating renal tumor size, location and depth. J Urol. 2009;182(3):844–53.

8. Ficarra V, Novara G, Secco S, Macchi V, Porzionato A, De Caro R, et al. Preoperative aspects and dimensions used for an anatomical (PADUA) classification of renal tumours in patients who are candidates for nephron-sparing surgery. Eur Urol. 2009;56(5):786–93.

9. Davidiuk AJ, Parker AS, Thomas CS, Leibovich BC, Castle EP, Heckman MG, et al. Mayo adhesive probability score: an accurate image-based scoring system to predict adherent perinephric fat in partial nephrectomy. Eur Urol. 2014;66(6):1165–71.

10. Kapoor A. Renal tumour scoring systems: How useful are they? Can Urol Assoc journal= J l'Association des Urol du Canada. 2015;9(1–2):46–7.

11. Kumar RM, Lavallée LT, Desantis D, Cnossen S, Mallick R, Cagiannos I, et al. Are renal tumour scoring systems better than clinical judgement at predicting partial nephrectomy complexity? Can Urol Assoc J. 2017;11(6):199.

12. Winoker JS, Paulucci DJ, Anastos H, Waingankar N, Abaza R, Eun DD, et al. Predicting complications following robot-assisted partial nephrectomy with the ACS NSQIP®universal surgical risk calculator. J Urol. 2017;198(4):803–9.

13. Yu K-H, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng [Internet]. 2018;2(10):719–31. Available from: https://doi.org/10.1038/s41551-018-0305-z

14. Abbod MF, Catto JWF, Linkens DA, Hamdy FC. Application of Artificial Intelligence to the Management of Urological Cancer. J Urol [Internet]. 2007;178(4):1150–6. Available from:

http://www.sciencedirect.com/science/article/pii/S0022534707013936

15. Kotu V, Deshpande B. Deep Learning. Data Science. MIT press; 2019. 307–342 p.

16. Géron A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. " O'Reilly Media, Inc."; 2019.

17. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. J Mach Learn Res. 2014;15(1):1929–58.

18. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in {P}ython. J Mach Learn Res. 2011;12:2825–30.

19. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. TensorFlow: A system for large-scale machine learning. In: 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16) [Internet]. 2016. p. 265–83. Available from: https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf

20. Davis J, Goadrich M. The relationship between Precision-Recall and ROC curves. In: Proceedings of the 23rd international conference on Machine learning - ICML '06 [Internet]. New York, NY, USA: ACM; 2006. p. 233–40. (ICML '06). Available from: http://portal.acm.org/citation.cfm?doid=1143844.1143874

21. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One [Internet]. 2015 Mar 4;10(3):e0118432–e0118432. Available from: https://www.ncbi.nlm.nih.gov/pubmed/25738806

22. Harutyunyan H, Khachatrian H, Kale DC, Ver Steeg G, Galstyan A. Multitask learning and benchmarking with clinical time series data. Sci Data. 2019 Mar;6(1):arXiv:1703.07771.

23. Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, et al. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning [Internet]. eprint arXiv:1711.05225. 2017. p. arXiv:1711.05225. Available from:

http://arxiv.org/abs/1711.05225

24. Breiman L. Random Forests. Mach Learn [Internet]. 2001;45(1):5–32. Available from: https://doi.org/10.1023/A:1010933404324

25. Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. Nat Med. 2018;24(11):1716.

26. Manaktala S, Claypool SR. Evaluating the impact of a computerized surveillance algorithm and decision support system on sepsis mortality. J Am Med Informatics Assoc. 2016;24(1):88–95.

27. Hung AJ, Chen J, Ghodoussipour S, Oh PJ, Liu Z, Nguyen J, et al. A deep-learning model using automated performance metrics and clinical features to predict urinary continence recovery after robot-assisted radical prostatectomy. BJU Int [Internet]. 2019;124(3):487–95. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/bju.14735

28. Wong NC, Lam C, Patterson L, Shayegan B. Use of machine learning to predict early biochemical recurrence after robot-assisted prostatectomy. BJU Int [Internet]. 2019;123(1):51–7. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/bju.14477

29. Guo* Y, Braga L, Kapoor A. PD07-08 MACHINE LEARNING TO PREDICT RECURRENCE OF LOCALIZED RENAL CELL CARCINOMA. J Urol. 2019;201(Supplement 4):e145–e145.

30. Shumate AM, Roth G, Ball CT, Custer K, Thiel DD. Factors associated with prolonged length of stay following robotic-assisted partial nephrectomy. Can J Urol [Internet]. 2019 Apr;26(2):9726—9732. Available from: http://europepmc.org/abstract/MED/31012837

31. Sperling CD, Xia L, Berger IB, Shin MH, Strother MC, Guzzo TJ. Obesity and 30-Day Outcomes Following Minimally Invasive Nephrectomy. Urology. 2018;121:104–11.

32. Bazzi WM, Sjoberg DD, Grasso AAC, Bernstein M, Parra R, Coleman JA. Predicting length

of stay after robotic partial nephrectomy. Int Urol Nephrol. 2015;47(8):1321–5.

33. Fritz BA, Chen Y, Murray-Torres TM, Gregory S, Abdallah A Ben, Kronzer A, et al. Using machine learning techniques to develop forecasting algorithms for postoperative complications: protocol for a retrospective study. BMJ Open. 2018;8(4):e020124.

34. Harris AM, James A, Dugan A, Bylund J. Increased Operative Duration in Minimally Invasive Partial Nephrectomy Is Associated with Significantly Increased Risk of 30-Day Morbidity. J Endourol. 2019;

35. Thompson RH, Blute ML. At what point does warm ischemia cause permanent renal damage during partial nephrectomy? Eur Urol. 2007;52(4):961.

36. Secin FP. Importance and limits of ischemia in renal partial surgery: experimental and clinical research. Adv Urol. 2008;2008:102461.

37. Sun M, Bianchi M, Trinh Q-D, Abdollah F, Schmitges J, Jeldres C, et al. Hospital volume is a determinant of postoperative complications, blood transfusion and length of stay after radical or partial nephrectomy. J Urol. 2012;187(2):405–10.

38. Monn MF, Bahler CD, Flack CK, Dube HT, Sundaram CP. The impact of hospital volume on postoperative complications following robot-assisted partial nephrectomy. J Endourol. 2014;28(10):1231–6.

39. Arora S, Keeley J, Pucheril D, Menon M, Rogers CG. What is the hospital volume threshold to optimize inpatient complication rate after partial nephrectomy? Urol Oncol Semin Orig Investig [Internet]. 2018;36(7):339.e17-339.e23. Available from: http://www.sciencedirect.com/science/article/pii/S1078143918301315

40. Kiechle JE, Abouassaly R, Gross CP, Dong S, Cherullo EE, Zhu H, et al. Racial Disparities in Partial Nephrectomy Persist Across Hospital Types: Results From a Population-based Cohort. Urology [Internet]. 2016;90:69–75. Available from: http://www.sciencedirect.com/science/article/pii/S0090429515011371

41. Chung BI, Leow JJ, Gelpi-Hammerschmidt F, Wang Y, Giudice F Del, De S, et al. Racial

Disparities in Postoperative Complications After Radical Nephrectomy: A Population-based Analysis. Urology [Internet]. 2015;85(6):1411–6. Available from: http://www.sciencedirect.com/science/article/pii/S0090429515002411

Table 1. Composition of Intraoperative Events and Postoperative Events

Intraoperative Events	**Patients no. (%)**
Gross Violation of the Tumor Bed	70 (4.14%)
Major Bleeding from the Tumor Bed	14 (0.83%)
Injury to Major Vessels	5 (0.30%)
Injury to Abdominal Organs	3 (0.18%)
Intraoperative Blood Transfusion > 1unit	3 (0.18%)
Conversion to Open	2 (0.12%)
Postoperative Events	**Patients no. (%)**
Length of Stay >4 days	290 (20.63%)
Clavien-Dindo Grade ≥3	14 (1.00%)

*Patients may have more than one complication

Table 2. List of features included in the model for predicting Intraoperative events and Postoperative Events. Demographics and Preoperative data were used for predicting intraoperative events. In addition to above, intraoperative data was used to predict Postoperative Events.

Feature Group	**Features**
Demographics	Age at surgery, BMI, Gender, Marital status, Race, Education
Preoperative	Clinical size (mm), Charlson score, Hemoglobin (Pre-op), Hematocrit (Pre-op), Leukocytes count (Pre-op), Creatinine (Pre-op), No of lesions, Center codes, Center volume, Symptoms, Solitary Kidney, Bilaterality of tumor, Side of tumor, Side of surgery, Face, Tumor location, Padua Risk score, Polar location, Rim location, Renal sinus, Exophytic rate, Clinical size group, Clinical Staging - Tumor, Clinical Staging – Regional Lymph Nodes, RENAL Nephro risk stratification, Radius (cm), Nearness of tumor (mm), Anterior/Posterior, Location to polar line, ASA score, Partial nephro indication, Multifocality
Intraoperative	Operative time (min), Ischemia time (min), Blood loss (ml), Access, Davinci model, Robotic arms, Assistant trocars, Dual console, Ischemia, Arterial clamping, Selective arterial clamping, Vein clamping, Early unclamping, Fluorescence, Inner renorrhaphy, Outer renorrhaphy, Urinary calyceal system repair, Haemostatic agents, Lymph node dissection, Blood Transfusion (Intra-op), Intraoperative events

Demographics – Patient demographic data.

Preoperative – Data collected prior to surgery.

Intraoperative – Data collected during surgery

Table 3. Model performance for Intraoperative Events and Postoperative Events.

Outcomes	Model	Sensitivity	Specificity	PPV	NPV	F-1 Score
Intraoperative Events (IOE)	LR_IOE	0.679	0.827	0.186	0.978	0.292
	NN_IOE	0.643	0.831	0.182	0.975	0.283
	RFR_IOE	0.357	0.996	0.833	0.964	0.500
Postoperative Events (POE)	LR_POE	0.730	0.793	0.485	0.917	0.583
	NN_POE	0.281	0.985	0.833	0.837	0.420
	RFR_POE	0.427	0.967	0.776	0.863	0.551

LR – Logistic Regression

RFC – Random Forest Classifier

NN – Neural Network

IOE – Intraoperative Events

POE – Postoperative Events

Sensitivity - Recall or True Positive Rate

Specificity – Selectivity or True Negative Rate

PPV – Positive Predictive Value (Precision)

NPV – Negative Predictive Value

Table 4. Model fit summary for Intraoperative Events and Postoperative Events.

Outcomes	Model	AUC-ROC (CLI-95%)	PR-AUC (CLI-95%)
Intraoperative Events (IOE)	LR_IOE	0.826 (0.731, 0.905)	0.372 (0.189, 0.552)
	NN_IOE	0.856 (0.779, 0.923)	0.398 (0.212, 0.605)
	RFR_IOE	0.858 (0.762, 0.936)	0.590 (0.400, 0.759)
Postoperative Events (POE)	LR_POE	0.837 (0.786, 0.882)	0.591 (0.477, 0.701)
	NN_POE	0.837 (0.786, 0.883)	0.649 (0.549, 0.739)
	RFR_POE	0.875 (0.834, 0.913)	0.706 (0.610, 0.790)

LR – Logistic Regression

RFC – Random Forest Classifier

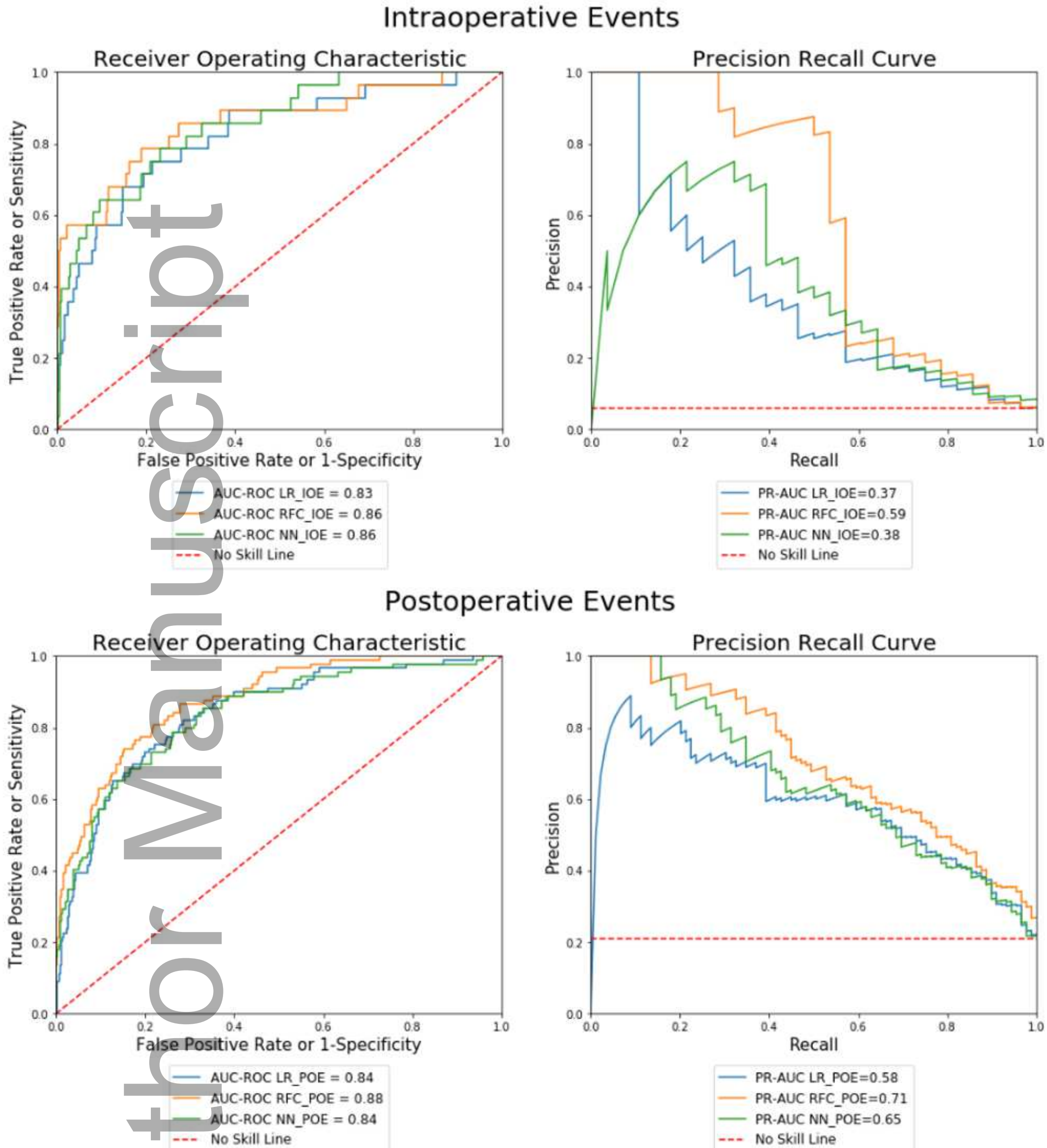
NN – Neural Network

IOE – Intraoperative Events

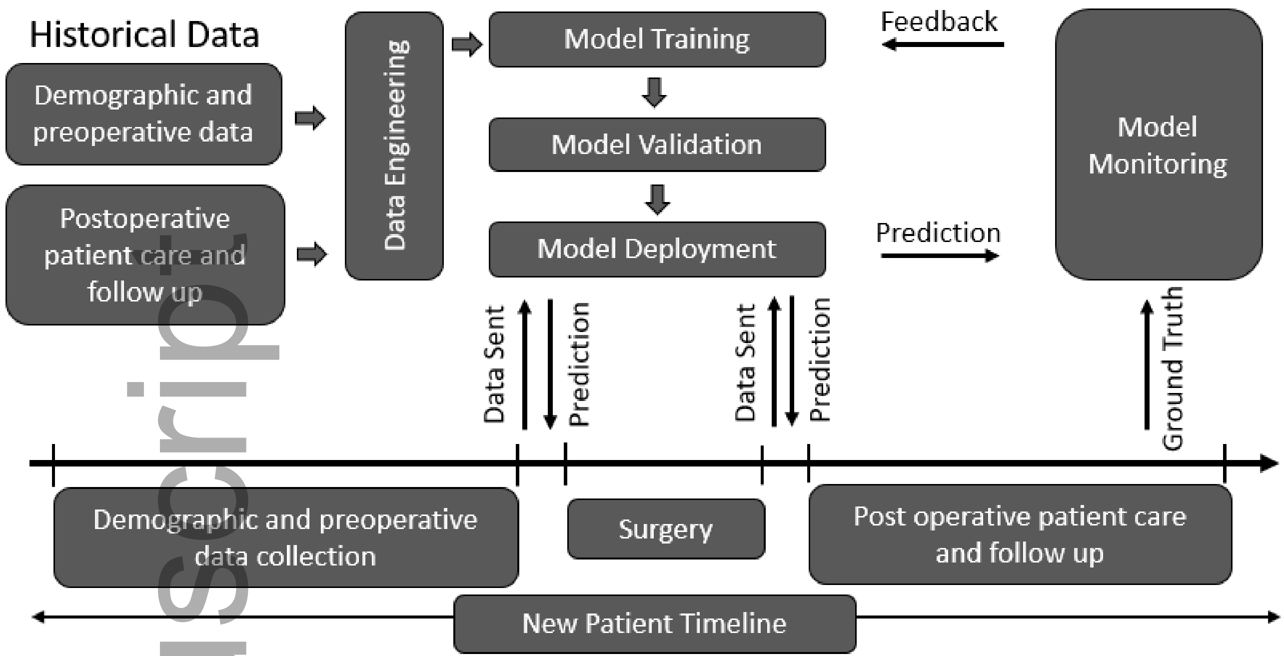
POE – Postoperative Events

AUC-ROC – Area Under Receiver Operating Curve

PR-AUC – Area Under Precision-Recall Curve



bju_15087_f1.png



bju_15087_f2.png

Author/s:
Bhandari, M; Nallabasannagari, AR; Reddiboina, M; Porter, JR; Jeong, W; Mottrie, A; Dasgupta, P; Challacombe, B; Abaza, R; Rha, KH; Parekh, DJ; Ahlawat, R; Capitanio, U; Yuvaraja, TB; Rawal, S; Moon, DA; Buffi, NM; Sivaraman, A; Maes, KK; Porpiglia, F; Gautam, G; Turkeri, L; Meyyazhgan, KR; Patil, P; Menon, M; Rogers, C

Title:
Predicting intra-operative and postoperative consequential events using machine-learning techniques in patients undergoing robot-assisted partial nephrectomy: a Vattikuti Collective Quality Initiative database study

Date:
2020-05-18

Citation:
Bhandari, M., Nallabasannagari, A. R., Reddiboina, M., Porter, J. R., Jeong, W., Mottrie, A., Dasgupta, P., Challacombe, B., Abaza, R., Rha, K. H., Parekh, D. J., Ahlawat, R., Capitanio, U., Yuvaraja, T. B., Rawal, S., Moon, D. A., Buffi, N. M., Sivaraman, A., Maes, K. K. ,... Rogers, C. (2020). Predicting intra-operative and postoperative consequential events using machine-learning techniques in patients undergoing robot-assisted partial nephrectomy: a Vattikuti Collective Quality Initiative database study. BJU INTERNATIONAL, 126 (3), pp.350-358. https://doi.org/10.1111/bju.15087.

Persistent Link:
http://hdl.handle.net/11343/275763

File Description:
Accepted version