BJU Int 2020; 126: 350–358 doi:10.1111/bju.15087

BJUI
BJU International

Original Article

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

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[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] and Craig Rogers[1]

[1]Vattikuti Urology Institute, Henry Ford Hospital, Detroit, MI, USA, [2]RediMinds Inc., Southfield, MI, USA, [3]Swedish Medical Centre, Seattle, WA, USA, [4]OLV Vattikuti Institute, Aalst, Belgium, [5]MRC Centre of Transplantation, King's College London, [6]Guy's and St Thomas' Hospitals, London, UK, [7]Ohio Health Dublin Methodist Hospital, Dublin, OH, USA, [8]Yonsei University, Seoul, Korea, [9]Sylvester Comprehensive Cancer Centre, University of Miami Miller School of Medicine, 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, [13]Rajiv Gandhi Cancer Institute and Research Centre, New Delhi, India, [14]Peter MacCallum Cancer Centre, Melbourne, Vic., Australia, [15]Humanitas Research Hospital, Milan, Italy, 16Apollo Hospitals, Chennai, India, [17]Centre 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, and [21]Vattikuti Foundation, Southfield, MI, USA

Abstract

Objective

To predict intra-operative (IOEs) and postoperative events (POEs) consequential to the derailment of the ideal clinical course of patient recovery.

Materials and Methods

The Vattikuti Collective Quality Initiative is a multi-institutional dataset of patients who underwent robot-assisted partial nephectomy for kidney tumours. Machine-learning (ML) models were constructed to predict IOEs and POEs using logistic regression, random forest and neural networks. The models to predict IOEs used patient demographics and preoperative data. In addition to these, intra-operative data were used to predict POEs. Performance on the test dataset was assessed using area under the receiver-operating characteristic curve (AUC-ROC) and area under the precision-recall curve (PR-AUC).

Results

The rates of IOEs and POEs were 5.62% and 20.98%, respectively. Models for predicting IOEs were constructed using data from 1690 patients and 38 variables; the best model had an AUC-ROC of 0.858 (95% confidence interval [CI] 0.762, 0.936) and a PR-AUC of 0.590 (95% CI 0.400, 0.759). Models for predicting POEs were trained using data from 1406 patients and 59 variables; the best model had an AUC-ROC of 0.875 (95% CI 0.834, 0.913) and a PR-AUC 0.706 (95% CI, 0.610, 0.790).

Conclusions

The performance of the ML models in the present study was 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

Published by John Wiley & Sons Ltd. www.bjui.org

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.

Keywords

deep learning, intra-operative complications, machine learning, postoperative complications, postoperative morbidity, robot-assisted partial nephrectomy

Introduction

Current guidelines recommend partial nephrectomy for T1a masses, as and when indicated. Robot-assisted surgery has gained popularity among surgeons [1], but is associated with an overall complication rate of up to 30% and a major complications (Clavien–Dindo grade ≥3) rate of 3–6% [2]. The morbidity of partial nephrectomy is attributed to the combination of tumour complexity, patient-related comorbidities [3], tumour 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 the 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 utilised to improve clinical diagnosis and decision-making in 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 ML, data mining and data engineering techniques, offers a promising opportunity to build models that could translate to valuable clinical practice tools for the personalized care of a patient.

The objective of the present retrospective study was to apply ML/AI models to predict the probability of a patient experiencing intra-operative (IOEs) and postoperative events (POEs) that are likely to adversely impact the clinical course for the patient undergoing robot-assisted partial nephrectomy (RAPN). We further propose a design for the potential deployment of these models for clinical validation in a prospective clinical setting.

Materials and Methods

Dataset

The Vattikuti Collective Quality Initiative (VCQI), a multi-institutional dataset of patients who underwent RAPN for kidney tumours, was used for the present study. Eighteen centres from around the world contributed to the dataset. Ethics committee/institutional review board clearance for data collection was obtained by each centre contributing to the VCQI database.

Intra-operative Events

Gross violation of tumour bed, major bleeding from the tumour bed, injury to major vessels, injury to abdominal organs, conversion to open surgery, and intra-operative blood transfusion ≥1 unit were defined as significant IOEs that are likely to delay the recovery of a patient (Table 1). After accounting for missing values and outliers, the resultant dataset for IOEs comprised 1690 patients and 38 predictors. The data processing steps are presented below. The median (Q1–Q3) age of patients was 58.0 (48.0–66.0) years, and the median (Q1–Q3) body mass index (BMI) was 27.3 (24.3–30.1). The POE rate was 5.62%. More than 90% of the patients had clinical staging of T1 tumours. Tables S1 and S2 show the summary statistics for demographic variables and preoperative variables (Table 2).

Postoperative Events

Clavien–Dindo grade ≥3, and length of hospital stay greater than the 75th percentile (>4 days in our database) within 30 days of surgery were considered as POEs (Table 1). After accounting for missing values and outliers, the resultant dataset comprised 1406 patients and 59 predictors. The data processing steps are presented below. The median (Q1–Q3) age of patients was 57.0 (48.0–66.0) years, and the median

Table 1 Composition of intra-operative and postoperative events.

	Patients, *n* (%)
Intra-operative events	
Gross violation of the tumour bed	70 (4.14)
Major bleeding from the tumour bed	14 (0.83)
Injury to major vessels	5 (0.30)
Injury to abdominal organs	3 (0.18)
Intra-operative blood transfusion >1 unit	3 (0.18)
Conversion to open	2 (0.12)
Postoperative events	
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 intra-operative and postoperative events.

Feature group	Features
Demographics	Age at surgery, BMI, gender, marital status, race, education
Preoperative	Clinical size (mm), Charlson score, haemoglobin (preoperative), haematocrit (preoperative), leukocyte count (preoperative), creatinine (preoperative), number of lesions, centre codes, centre volume, symptoms, solitary kidney, bilaterality of tumour, side of tumour, side of surgery, face, tumour location, Padua risk score, polar location, rim location, renal sinus, exophytic rate, clinical size group, clinical staging – tumour, clinical staging – regional lymph nodes, RENAL nephrometry risk stratification, radius (cm), nearness of tumour (mm), anterior/posterior, location to polar line, ASA score, partial nephrectomy indication, multifocality
Intra-operative	Operating time (min), ischaemia time (min), blood loss (mL), access, da Vinci model, robotic arms, assistant trocars, dual console, ischaemia, 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-operative), intra-operative events

BMI, body mass index. Definitions: demographics, patient demographic data; preoperative, data collected prior to surgery; intra-operative, data collected during surgery. Demographics and preoperative data were used for predicting intra-operative events. In addition to above, intra-operative data were used to predict postoperative events.

(Q1–Q3) BMI was 27.2 (24.3–30.0) kg/m^2. The male to female ratio was 66:34. The POE rate was 20.98%. More than 90% of the patients had a clinical staging of T1 tumours. Tables S3–S5 illustrate summary statistics for demographic, preoperative and intra-operative variables (Table 2).

Model Development

We trained classification models using logistic regression (LR), random forests classifier (RFC), and neural networks (NN). LR has been traditionally used because of its ease of result interpretation and analysis. RFC and NN are non-parametric models, capable of modelling complex non-linear relationships.

Data Processing

Missing values for numeric variables were imputed using 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 was 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.

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.

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 trees makes RFC 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.

Neural Network Architecture

Neural networks are composed of units of calculation called neurons and are capable of modelling complex patterns present in the data. NN can contain many types of layers that perform calculations on input received from the previous layer. Our NN models (Fig. S1) comprised 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-day array. After the input layer, a two-layered dense network with a dropout layer after each dense layer was implemented. Each dense layer comprised neurons and used a rectified linear unit for non-linear activation. The output layer comprised 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.

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 nine sets and evaluated on the remaining one set. The model with the best average performance over 10 folds was selected as the final model. We used Python (Python Software Foundation, Wilmington, DE, USA) with Scikit-learn [18] package and TensorFlow [19] for analysis.

Model Evaluation and Statistical Analysis

We assessed the comparative performance of our models using area under the receiver-operating characteristic curve (AUC-ROC) and area under the 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 CIs 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% 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% CI was constructed.

Results

Three models each were trained for predicting IOEs and POEs, resulting in a total of six models. The models were trained using LR, RF classifier (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.

Intra-operative Events

The models developed for predicting IOEs were denoted as LR_IOE, RFC_IOE and NN_IOE for LR, RFC and NN, respectively. The performance of the models on the test dataset is shown in Fig. 1. The sensitivity, specificity, positive predictive value, negative predictive value and F-1 score are reported in Table 3.

The AUC-ROC and PR-AUC for model RFC_IOE were 0.858 (95% CI 0.762, 0.936) and 0.590 (95% CI 0.400, 0.759), respectively. The AUC-ROCs for model LR_IOE and NN_IOE were 0.826 (95% CI 0.731, 0.905) and 0.856 (95% CI 0.779, 0.923), respectively. The PR-AUCs for model LR_IOE and NN_IOE were 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 ($P = 0.035$), and between RFC_IOE and NN_IOE it was 0.208 ($P = 0.067$). RFC_IOE outperformed LR_IOE and NN_IOE.

Postoperative Events

The models constructed for predicting POE were denoted as LR_POE, RFC_POE and NN_POE for LR, RFC and NN, respectively. The performance of the models on the test dataset is shown in Fig. 1. The sensitivity, specificity, positive predictive value, negative predictive value 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 ($P = 0.068$), and between RFC_IOE and NN_IOE was 0.057 ($P = 0.21$). RFC_POE outperformed LR_POE and NN_POE.

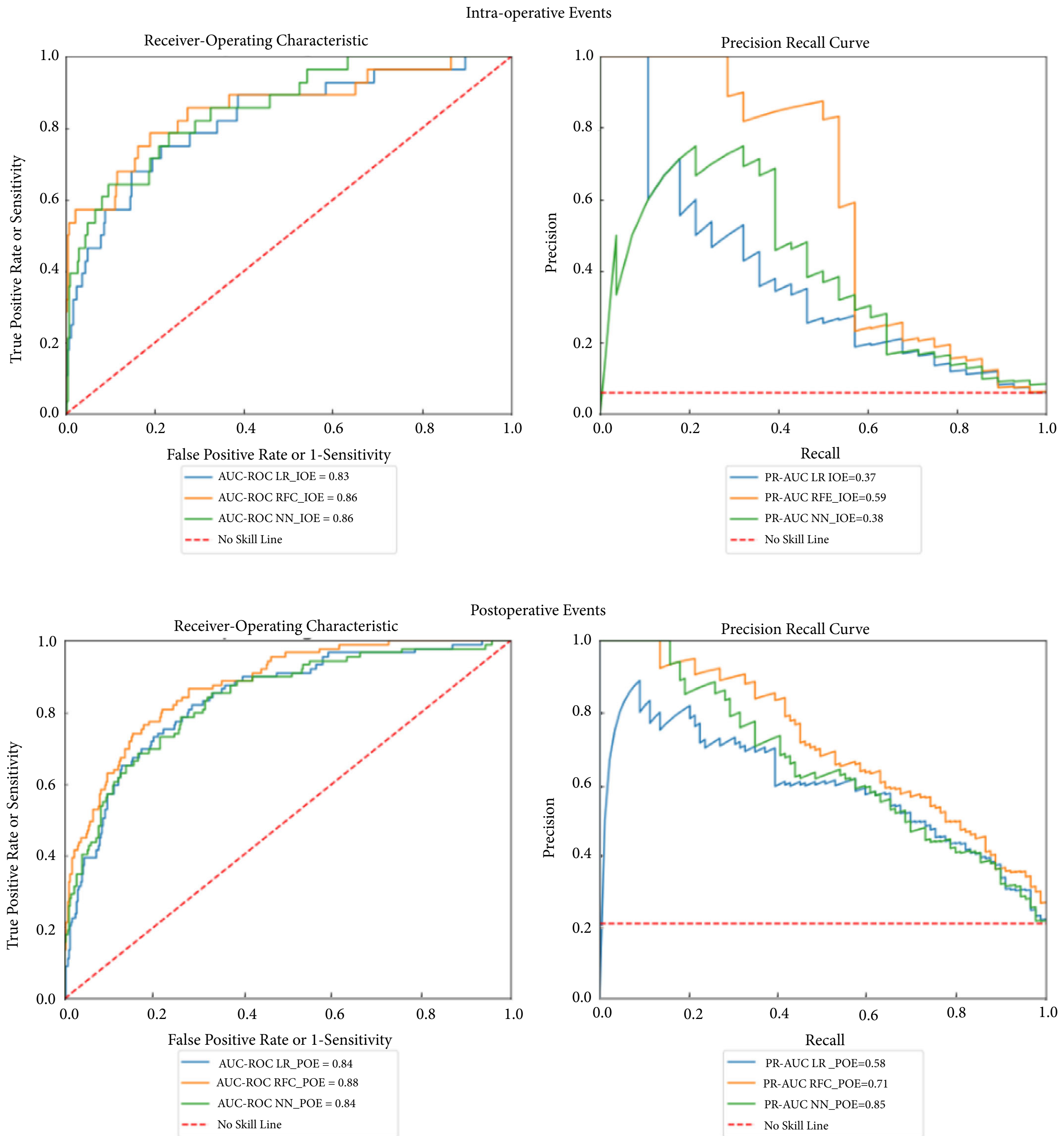
Feature Importance

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 is reported in Tables S6 and S7, respectively. Multifocality, clinical staging – regional lymph nodes, centre code and solitary kidney – were observed to be the most important features for RF_IOE (Table S6). Centre code, haemostatic agents, centre volume, ischaemia time (min) and race were observed to be top five important features for RF_POE (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 it was centre 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.

Discussion

Surgeons always endeavour to bring predictability to their actions and decisions taken to manage patients under their care. With the successful application of ML, it is possible to

Fig. 1 Comparative performance on test set for models trained for predicting intra-operative events and postoperative events. AUC-ROC, area under the receiver- operating characteristic curve; IOE, intra-operative event; LR, logistic regression; NN, neural network; POE, postoperative event; PR-AUC, area under precision-recall curve; RFC, random forest classifier.



predict many IOEs and POEs, 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. RAPN for small T1 renal tumours is establishing itself as a safe, day-

Table 3 Model performance for intra-operative and postoperative events.

Outcomes	Model	Sensitivity	Specificity	PPV	NPV	F-1 score
IOEs	LR_IOE	0.679	0.827	0.186	0.978	0.292
	NN_IOE	0.643	0.831	0.182	0.975	0.283
	RFC_IOE	0.357	0.996	0.833	0.964	0.500
POEs	LR_POE	0.730	0.793	0.485	0.917	0.583
	NN_POE	0.281	0.985	0.833	0.837	0.420
	RFC_POE	0.427	0.967	0.776	0.863	0.551

IOE, intra-operative event; LR, logistic regression; NN, neural network; NPV, negative predictive value; POE, postoperative event; PPV, positive predictive value (precision); RFC, random forest classifier; Sensitivity, recall or true positive rate; Specificity, selectivity or true negative rate.

Table 4 Model fit summary for intra-operative and postoperative events.

Outcomes	Model	AUC-ROC (95% CI)	PR-AUC (95% CI)
IOEs	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)
	RFC_IOE	0.858 (0.762, 0.936)	0.590 (0.400, 0.759)
POEs	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)
	RFC_POE	0.875 (0.834, 0.913)	0.706 (0.610, 0.790)

AUC-ROC, area under the receiver-operating characteristic curve; IOE, intra-operative event; LR, logistic regression; NN, neural network; POE, postoperative event; PR-AUC, area under the precision-recall curve; RFC, random forest classifier.

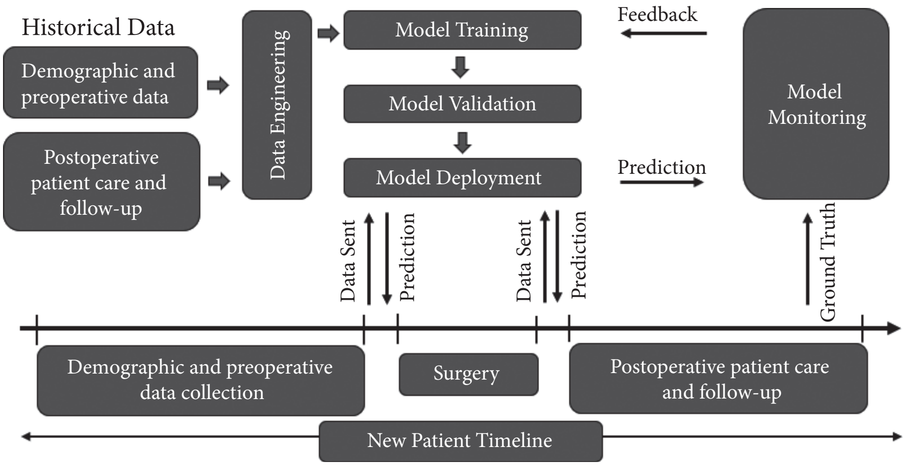
case surgery procedure. For a strict case selection, the surgeon needs accurate predictive tools to preempt consequential events as a precursor to emergency readmission of the patient. Currently, different risk profile tools are in practice to identify high-risk cohorts.

The existing surgical risk predictors such as PADUA, MAP and RENAL scores played a significant role in assisting clinicians in the risk assessment of a patient, but their application needs to be enhanced with the development of tools prompting an intervention to prevent complications and improve overall patient outcomes. The 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 RAPN [12]. AI-driven predictive models hold promise to fill in the gap in the field, and open an opportunity to assess the potential utility of ML 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. [26] 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, sepsis mortality decreased by 53%, and the 30-day readmission rate dropped from 19.08% to 13.21%. Recently, in urology, ML 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 RCC [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. [30] report that conversion to open surgery was a significant reason for a prolonged hospital stay, which increases 30-day morbidity. Accurate prediction of prolonged length of stay could improve case selection for RAPN as an outpatient procedure and to counsel the patient and the family with certainty. While deciding on the cut-off limit of 4 days (>75th percentile) as a POE, we agree with Sperling et al. [31], who included hospital stay greater than 75th percentile in their dataset as a secondary outcome in their study. Shumate et al. [30] reported that 80% of patients were discharged within 3 days of the surgery. Centre volume was calculated as the number of surgeries performed by each centre before operating on a given patient, and was based on the information available in our database.

In the present study, AI/ML algorithms were used to construct models to predict consequential IOEs and POEs. 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 the VCQI dataset, 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 has been shown to be more informative in the evaluation of model performance [20,21]. AUC-ROC takes into account the true positive rate and false positive rate. 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 false positive rate. 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. true positive rate and precision (ratio of number true positives divided by the sum of true positives and false negatives). Precision indicates how many of the positive

Fig. 2 Model deployment proposal overview. Historical data: comprise demographic, preoperative, intra-operative and follow-up data from patients who underwent robot-assisted partial nephrectomy. Data engineering: historical data are pre-processed, and features are engineered to enable training of the data with machine-learning algorithms. Model training and model validation models are trained, and the model with the best performance on unseen test data is selected. Model deployment: the best performing model is deployed in a clinical setting for further evaluation. Model monitoring: model prediction is compared against ground truth to determine if the model is performing as expected.



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, centre experience and solitary kidney were observed to be the most important features for a model constructed for predicting IOEs using RF, i.e. RF_IOE. Centre experience, haemostatic agents, centre volume, ischaemia time (min) and race were observed to be the most important features for a model constructed to predict POEs using RF, i.e. RF_POE. In recent studies, surgeons with high volumes were reported to have fewer adverse peri-operative outcomes [5] and significant variability was observed between surgeon for outcomes of partial nephrectomy [6]. Centre code was used as a surrogate for centre experience. In another study, ischaemia 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 in the use of partial nephrectomy across hospitals and postoperative complications after partial nephrectomy [40,41].

The present study has some limitations. First, the variables in the dataset do not account for the temporal shift in patient characteristics during the admission. Second, intra-operative outcomes are known to vary significantly among surgeons. Unfortunately, surgeon information was not available for many patients; therefore, it could not be used for constructing the models in this study. Third, patient data, such as imaging data and caregiver notes, were not available. Fourth, predicting individual IOEs or POEs (Clavien–Dindo grade ≥3) was not possible because there were very few occurrences of these individual events. Fifth, the cohort of patients was small, with 1690 patients included for IOEs and 1406 patients for POEs. Further studies will be required with larger cohorts to validate the findings of the present study.

We propose deploying the models at participating centres contributing to the database. The deployment process will include the following steps: (i) re-validating models with new data in the VCQI database; (ii) deploying models in coordination with respective centres; and (iii) monitoring performance and retraining models when performance falls below a predetermined threshold (Fig. 2). This would enable the centres to use the models to identify patients likely to have a complication during or after surgery and to 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.

In 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 IOEs and 30-day POEs. 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.

Acknowledgements

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

Conflicts of Interest

None declared.

References

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

2 **Tanagho YS, Kaouk JH, Allaf ME et al.** Perioperative complications of robot-assisted partial nephrectomy: analysis of 886 patients at 5 United States centers. *Urology* 2013; 81: 573–80

3 **Ficarra V, Bhayani S, Porter J et al.** Predictors of warm ischemia time and perioperative complications in a multicenter, international series of robot-assisted partial nephrectomy. *Eur Urol* 2012; 61: 395–402

4 **Khene Z-E, Peyronnet B, Kocher NJ et al.** Predicting morbidity after robotic partial nephrectomy: the effect of tumor, environment, and patient-related factors. *Urol Oncol* 2018; 338.e19–26

5 **Khandwala YS, Jeong IG, Kim JH et al.** The impact of surgeon volume on perioperative outcomes and cost for patients receiving robotic partial nephrectomy. *J Endourol* 2017; 31: 851–7

6 **Dagenais J, Bertolo R, Garisto J et al.** Variability in partial nephrectomy outcomes: does your surgeon matter? *Eur Urol* 2019; 75: 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: 844–53

8 **Ficarra V, Novara G, Secco S 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: 786–93

9 **Davidiuk AJ, Parker AS, Thomas CS et al.** Mayo adhesive probability score: an accurate image-based scoring system to predict adherent perinephric fat in partial nephrectomy. *Eur Urol* 2014; 66: 1165–71

10 **Kapoor A**. Renal tumour scoring systems: how useful are they? *Can Urol Assoc J* 2015; 9: 46–7

11 **Kumar RM, Lavallée LT, Desantis D et al.** Are renal tumour scoring systems better than clinical judgement at predicting partial nephrectomy complexity? *Can Urol Assoc J* 2017; 11: 199

12 **Winoker JS, Paulucci DJ, Anastos H et al.** Predicting complications following robot-assisted partial nephrectomy with the ACS NSQIP® universal surgical risk calculator. *J Urol* 2017; 198: 803–9

13 **Yu K-H, Beam AL, Kohane IS.** Artificial intelligence in healthcare. *Nat Biomed Eng* 2018; 2: 719–31

14 **Abbod MF, Catto JWF, Linkens DA, Hamdy FC**. Application of artificial intelligence to the management of urological cancer. *J Urol* 2007; 178: 1150–6

15 **Kotu V, Deshpande B**. *Deep Learning. Data Science*. Amsterdam, The Netherlands: Elsevier, 2019: 307–42

16 **Géron A.** Hands-on machine learning with scikit-learn, keras, and tensorflow: concepts, tools, and techniques to build intelligent systems. *J Mach Learn Res*, 2019; 10: 279–325.

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: 1929–58

18 **Pedregosa F, Varoquaux G, Gramfort A et al.** Scikit-learn: machine learning in python. *J Mach Learn Res* 2011; 12: 2825–30

19 **Abadi M, Barham P, Chen 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 at: https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf. Accessed October 2019

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: ACM, 2006. p. 233–40. (ICML '06). Available at: http://portal.acm.org/citation.cfm?doid=1143844.1143874. Accessed July 2019

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* 2015; 10: e0118432

22 **Harutyunyan H, Khachatrian H, Kale DC, Ver Steeg G, Galstyan A**. Multitask learning and benchmarking with clinical time series data. *Sci Data* 2019; 6: 1–18.

23 **Rajpurkar P, Irvin J, Zhu Ket al.**CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning [Internet]. Eprint arXiv:1711.05225. 2017. p. arXiv:1711.05225. Available at: http://arxiv.org/abs/1711.05225. Accessed July 2019

24 **Breiman L**. Random forests. *Mach Learn* 2001; 45: 5–32

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: 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: 88–95

27 **Hung AJ, Chen J, Ghodoussipour S 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* 2019; 124: 487–95

28 **Wong NC, Lam C, Patterson L, Shayegan B**. Use of machine learning to predict early biochemical recurrence after robot-assisted prostatectomy. *BJU Int* 2019; 123: 51–7

29 **Guo Y, Braga L, Kapoor A**. PD07-08 machine learning to predict recurrence of localized renal cell carcinoma. *J Urol* 2019; 201 (Suppl. 4): 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* 2019; 26: 9726–32

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: 1321–5

33 **Fritz BA, Chen Y, Murray-Torres TM et al.** Using machine learning techniques to develop forecasting algorithms for postoperative complications: protocol for a retrospective study. *BMJ Open* 2018; 8: 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; 33: 549–56

35 **Thompson RH, Blute ML**. At what point does warm ischemia cause permanent renal damage during partial nephrectomy? *Eur Urol* 2007; 52: 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 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: 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: 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* 2018; 36: 339.e17–23
40 **Kiechle JE**, **Abouassaly R**, **Gross CP et al.** Racial disparities in partial nephrectomy persist across hospital types: results from a population-based cohort. *Urology* 2016; 90: 69–75
41 **Chung BI**, **Leow JJ**, **Gelpi-Hammerschmidt F et al.** Racial disparities in postoperative complications after radical nephrectomy: a population-based analysis. *Urology* 2015; 85: 1411–6

Correspondence: Mahendra Bhandari, Suit 700, 1000 Town Center, Southfield, MI 48075, USA.

e-mail: mahendra@vattikutifoundation.com

Abbreviations: AI, artificial intelligence; AUC-ROC, area under the receiver operating charecterictic curve; IOE, intra-operative event; LR, logistic regression; ML, machine learning; NN, neural networks; POE, postoperative event; PR-AUC, area under the precision-recall curve; RAPN, robot-assisted partial nephrectomy; RF, random forests; RFC, random forest classifier; VCQI, Vattikuti Collective Quality Initiative.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Illustration of neural network.
Table S1. Demographic characteristics.
Table S2. Preoperative characteristics.
Table S3. Demographic characteristics.
Table S4. Preoperative variables.
Table S5. Intra-operative variables.
Table S6. Summary of feature importance for random forest classifier model for predicting intra-operative events (RF_IOE).
Table S7. Summary of feature importance for random forest classifier model constructed for predicting postoperative events (RF_POE).

