



Developing a supervised machine learning model for predicting perioperative acute kidney injury in arthroplasty patients

Okke Nikkinen^{a,*,1}, Timo Kolehmainen^{b,1}, Toni Aaltonen^b, Elias Jämsä^b, Seppo Alahuhta^b, Merja Vakkala^{a,c}

^a Medical Research Center Oulu, Research Group of Surgery, Anesthesiology and Intensive Care Medicine, University of Oulu, Oulu University Hospital, Oulu, Finland

^b Faculty of Medicine, University of Oulu, Oulu, Finland

^c Department of Anaesthesiology, Oulu University Hospital, Oulu, Finland

ARTICLE INFO

Keywords:

Acute kidney injury
AKI
Supervised machine learning
RUSBoost
Arthroplasty

ABSTRACT

Background: Perioperative acute kidney injury (AKI) is challenging to predict and a common complication of lower limb arthroplasties. Our aim was to create a machine learning model to predict AKI defined by both serum creatinine (sCr) levels and urine output (UOP) and to investigate which features are important for building the model. The features were divided into preoperative, intraoperative, and postoperative feature sets.

Methods: This retrospective, register-based study assessed 648 patients who underwent primary knee or hip replacement at Oulu University Hospital, Finland, between January 2016 and February 2017. The RUSBoost algorithm was chosen to establish the models, and it was compared to Naïve/Kernel Bayes and support vector machine (SVM). Models of AKI classified by either sCr levels or UOP were established. All the models were trained and validated using a five-fold cross-validation approach. An external test set was not available at the time of this study.

Results: The performance of both the sCr level- and UOP-based AKI models improved when pre-, intra-, and postoperative features were used together. The best sCr level-based AKI model performed as follows: area under receiving operating characteristic (AUROC) of 0.91, (95% CI ± 0.02), area under precision-recall (AUPR) of 0.35 (95% CI ± 0.04) sensitivity of 0.88 (95% CI ± 0.03), specificity of 0.87 (95% CI ± 0.03), and precision of 0.35 (95% CI ± 0.03). This model correctly classified 22 out of 25 patients with AKI. The best UOP-based AKI model performed as follows: AUROC of 0.98 (95% CI ± 0.02), AUPR of 0.48 (95% CI ± 0.04), sensitivity of 0.88 (95% CI ± 0.02), specificity of 0.93 (95% CI ± 0.03), and precision of 0.34 (95% CI ± 0.04). This model correctly classified 23 out of 26 patients with AKI. In the sCr-AKI models, estimated glomerular filtration rate (eGFR)-related features were most important, and in the UOP-based AKI models, UOP-related features were most important. Other important and recurring features in the models were age, sex, body mass index, ASA status, operation type, preoperative eGFR, and preoperative sCr level. Naïve/Kernel Bayes performed similarly to RUSBoost. SVM performed poorly.

Conclusions: The performance of the models improved after the inclusion of intra- and postoperative features with preoperative features. The results of our study are not generalizable, and additional larger studies are needed. The optimal ML method for this kind of data is still an open research question.

1. Introduction

The onset of perioperative acute kidney injury (AKI) is generally a result of the accumulation of risk factors [1–5]. In patients who undergo hip and knee arthroplasties, the incidence of AKI varies between 0.5 and 24.0% [6–9]. Perioperative AKI is often determined by elevated serum

creatinine (sCr) levels alone, and the urine output (UOP) criterion is ignored. This situation is a common challenge for defining perioperative AKI, since UOP can physiologically diminish intraoperatively, and postoperatively, the measurement of UOP varies greatly [10]. Perioperative AKI is associated with both increased morbidity and mortality and lengthened hospital stay [9,11].

In AKI, elevated sCr levels appear relatively late after injury.

* Corresponding author. University of Oulu Graduate School/Department of Anaesthesiology Oulu University Hospital, Department of Anaesthesiology, Post Box 21, 90029, Oulu, Finland.

E-mail address: okke.nikkinen@student.oulu.fi (O. Nikkinen).

¹ Equal contributors.

Abbreviations

AI	artificial intelligence
AKI	acute kidney injury
ASA	American Society of Anesthesiologists
AUPR	area under precision-recall
AUROC	area under operating receiving characteristics
BMI	body mass index
CI	confidence interval
eGFR	estimated glomerular filtration rate
KDIGO	Kidney Disease Improving Global Outcomes
ML	machine learning
sCr	serum creatinine;
SVM	support vector machine;
UOP	urine output

Perioperative AKI is an example of a clinical problem that could benefit from machine learning (ML). Perioperative AKI is an adverse outcome, there are no certain, single risk factors or tests for predicting its onset, there is a vast amount of related data available, and the amount of data is impossible to interpret by a single clinician. It might be possible for an ML algorithm to allow more current risk evaluations or even predictions [12].

A common method for creating predictive algorithms is analyzing data retrospectively and trying to create as simple and accurate a model as possible by selecting artificial intelligence (AI)-algorithms or logistic regression [13]. Recent studies related to the generation of predictive algorithms have included patients who underwent cardiac and aortic surgery [14,15] or percutaneous coronary intervention [16] as well as major noncardiac patients [17,18], ICU patients, and hospitalized patients [19–21]. Most of these studies ignored the UOP criteria [14, 16–18,21]. These studies used different sets of features, and the timing of AKI diagnosis and prediction differed. Some studies focused on patient groups with a higher incidence of AKI, making the data less imbalanced [14,15,22]. A study of AKI and intra-abdominal surgeries showed that adding intraoperative data to preoperative data improved the performance of a risk prediction model [23].

The aim of this retrospective cohort study was to create a supervised ML model to predict the onset of perioperative AKI in lower limb arthroplasty patients. The aim was to create a model for AKI defined by both sCr levels and UOP. We aimed to create models using features from the pre-, intra-, and postoperative periods, assess the value of these different feature sets, and assess the most important features.

2. Materials and methods

2.1. Study design

The ethical committee of Northern Ostrobothnia approved this study in 2018 (EETTMK 11/2018). Approval from the Medical Director for the Northern Ostrobothnia Hospital District and the National Data Protection Ombudsman were obtained. Written informed consent was not required. We conducted a retrospective, hospital register-based study of patients who underwent either primary knee or hip arthroplasty between January 2016 and February 2017 at Oulu University Hospital, Oulu, Finland. The inclusion criteria were patients who underwent primary knee or hip arthroplasty, had available preoperative sCr level data, and had electronic anesthesia records. The operation codes used in the study were NFB10–90 and NGB10–90. This study was a continuation of a previous study in which we described the perioperative characteristics of the study cohort [24].

2.2. Data collection

The data were collected from the anesthesia and patient records. A list of the collected features is presented in Table 1.

2.3. Definition of acute kidney injury

AKI was evaluated based on the Kidney Disease Improving Global Outcomes (KDIGO) classification, except for the UOP criteria [25]. Due to the retrospective nature of this study and incomplete data for some patients, we had to modify the UOP-AKI criteria. We used the following criteria: an average UOP of <0.5 ml/kg/h over 48 h and the availability of at least one full day of UOP data afterward. Intraoperative UOP was not considered to be a diagnostic criterion. Instead, we used post-operative UOP. According to these criteria, a patient was classified as having either no AKI, AKI based on sCr level, or AKI based on UOP. A patient who met both criteria was classified into both the sCr and UOP groups.

2.4. Data preprocessing

Data were cleaned to address crucial missing values (Fig. 1). Patients who were transferred from the recovery room to a private hospital and, therefore, for whom all postoperative data were lacking were excluded. Patients who did not have UOP data from the operation and recovery room periods were excluded. Excluded patients were compared to included patients to assess the effect of exclusion. Patients who spent more than 40 h in the recovery room were excluded. After these exclusions, there were still some missing values in the register data, but no effort was made to impute them to develop the algorithms. Algorithms were developed using only the data that were realistically available. Body mass index (BMI) was categorized as <30, 30–34.99, 35–39.99 or >40 kg/m². The estimated glomerular filtration rate (eGFR) was calculated using the Chronic Kidney Disease Epidemiology Collaboration formula and categorized as ≥90, 60–89, and <60 ml/min/1.73 m² [26]. UOP and infused fluids were derived from raw data to ml/h and ml/kg/h features. A patient who met the sCr level-based AKI criteria was assigned to Target 1, and a patient who met the UOP-based AKI criteria was assigned to Target 2. The patients who met both criteria were assigned to both Targets 1 and 2.

An external test set was not available for testing the performance of the models at the time of the study. Due to the small sample size and high imbalance of the data, splitting the dataset into separate training and test sets was not feasible. Instead, a k-fold (k = 5) cross-validation method was used for the validation of all the ML models [27].

2.5. Statistical analyses

The data were analyzed using SPSS for Windows (IBM Corp. Released 2016. IBM SPSS Statistics for Windows, Version 24.0. Armonk, NY: IBM Corp.). Continuous features were analyzed using Mann–Whitney U tests when two groups were compared. Categorical features were analyzed using Fisher's exact tests or Pearson chi-square tests. Two-tailed p values are presented, and 95% confidence intervals (CIs) are presented for the ML algorithm performance characteristics.

2.6. Algorithm selection

The development environment was MATLAB (The MathWorks Inc, Natick, Massachusetts, United States) with the Statistics and Machine Learning Toolbox. MATLAB was chosen as the development environment because several algorithms can be easily compared. The MATLAB Statistics and Machine Learning Toolbox's Classification Learner application was used for the algorithm selection process. The algorithm selection was limited by what was available in MATLAB. The choices for the datasets included decision trees, logistic regression, Bayesian

Table 1

List of raw data features collected from the patient records and the percentage of missing data. Derivatives, such as urine output described in ml/kg/h and grouped features, such as preoperative eGFR ≥ 90 , 60–89, and < 60 ml/min/1.73 m² were derived from these listed features. eGFR = estimated glomerular filtration rate; BMI = body mass index; ASA-status = American society of anesthesiologist's status; NSAID = non-steroidal anti-inflammatory drug. * = Missing % is the percentage compared to all data. Tourniquet is only used in knee arthroplasties. ** = Patients discharged earlier have these values as missing. The median length of stay was 3 days in this material.

	Features	Type of feature input	Missing %		
Preoperative features	Sex	Nominal	0		
	Age (years)	Scale	0		
	Weight (kg)	Scale	0		
	BMI (kg/m ²)	Scale	0.9%		
	ASA-status	Nominal	0		
	Preoperative serum creatinine	Scale	0		
	Smoking	Nominal	0		
	Alcohol consumption	Nominal	0		
	Hypertension	Nominal	0		
	Cardiovascular diseases	Nominal	0		
	Kidney diseases	Nominal	0		
	NSAID medication	Nominal	0		
	Betablockers	Nominal	0		
	Diuretics	Nominal	0		
	ACE-inhibitors	Nominal	0		
	Sartans	Nominal	0		
	Prior major surgeries	Nominal	0		
	Prior radiological imaging with contrast medium	Nominal	0		
	Intraoperative features	Type of surgery, operation codes	Nominal	0	
Not used in models		Date of operation (date)	Scale	0.6%	
		Emergency/elective surgery	Nominal	0	
		Type of anesthesia	Nominal	0	
		Preoperative medication prior to surgery	Scale	0	
		Operation duration (hh:mm)	Scale	0	
		Operation and recovery room duration (hh:mm)	Scale	0	
		Tourniquet duration in knee operations (hh:mm)*	Scale	62.0%	
		Infused fluids (ml)	Scale	0	
		Urine output (ml)	Scale	0	
		Loss of blood (ml)	Scale	0	
		Infused blood products (ml)	Nominal/Scale	0	
		Vasoactive administration	Nominal	0	
		Mean arterial pressure trend (mmHg)	Nominal	0	
		Diuretics	Nominal	0	
		Antiemetics	Nominal	0	
		Pain medication	Nominal	0	
		Postoperative features	Serum creatinine, postop. day 1	Scale	15.6%
			Not used in models	Serum creatinine, postop. day 2–3	Scale
Serum creatinine, postop. day 4–7**	Scale			60.8%	
Serum creatinine, any value from postop. days 1–7	Scale			14.2%	
Urine output, postop. day 1 (ml)	Scale			4.8%	
Urine output, postop. day 2 (ml)	Scale			38.7%	
Urine output, postop. day 3 (ml)**	Scale			77.5%	
Fluid intake, postop. day 1	Scale			4.9%	
Fluid intake, postop. day 2	Scale			46.0%	
Fluid intake, postop. day 3 **	Scale			82.9%	

Table 1 (continued)

	Features	Type of feature input	Missing %
Not used in models	Intensive care	Nominal	0
	Length of stay (days)	Scale	0
	Location of follow-up treatment	Nominal	0.9%
	Time of death (date)	Scale	0

classifiers, support vector machines and ensemble tree classifiers. Because the minority class included only 4.0% of the patients in this study, only the RUS boosted trees algorithm yielded satisfactory results in preliminary testing. The imbalance of the data and the missing values limited the choices for the ML algorithm. RUSBoost was compared to Naïve/Kernel Bayes and support vector machine (SVM) [17,22]. RUSBoost has been used for ECG analysis, histopathology, and epileptic seizure detection [28–31]. The algorithm takes N, the number of members in the class with the fewest members in the training data, as the basic unit for sampling. Classes with more members are under-sampled by taking only N observations of every class. In other words, if there are K classes, then for each weak learner in the ensemble, RUSBoost takes a subset of the data with N observations from each of the K classes. The boosting procedure follows the procedure in Adaptive Boosting for Multiclass Classification for reweighting and constructing the ensemble [32,33].

2.7. Algorithm training and validation

All the algorithms were trained and validated using a five-fold cross-validation approach. The dataset was randomly divided into five folds, 80% for training and 20% for validation, and errors between the true classification and the predicted classification were averaged across all the test sets.

The algorithms were tuned through the selection of features for the model by testing performance with different feature sets. The optimization criterion was precision rather than accuracy. The former is considered more important than the latter in a clinical setting since the best model is a model with which the maximum number of patients who are likely to develop AKI can be identified and treated shortly after surgery. Automated hyperparameter optimization was attempted on several algorithms, but due to class imbalance, this did not improve precision, and hyperparameter optimization was not performed in further development. The performance metrics of the best models are reported.

Models for sCr level-based AKI (Target 1) and UOP-based AKI (Target 2) were generated separately using different sets of features, described as Trials. Trial A included only preoperative features, Trial B included pre- and intraoperative features, and Trial C included pre-, intra-, and selected Day 1 postoperative features, covering the time from the end of recovery room time to the end of postoperative Day 1. Intraoperative features comprised those available until the end of the time spent in the recovery room. From postoperative Day 1, we included only fluid balance data and the first sCr measurement if this value was measured.

The inclusion of the sCr level from the 1st postoperative day meant that in Target 1 Trial C, the diagnosis of AKI within the 1st postoperative day was self-explanatory, and these patients were automatically correctly classified. For sCr level-based AKI Trial C, a further model was generated, in which the patients who met the sCr-AKI criteria on postoperative Day 1 were excluded to examine the effect of the inclusion of these patients.

For all the models, the importance of the features was normalized to

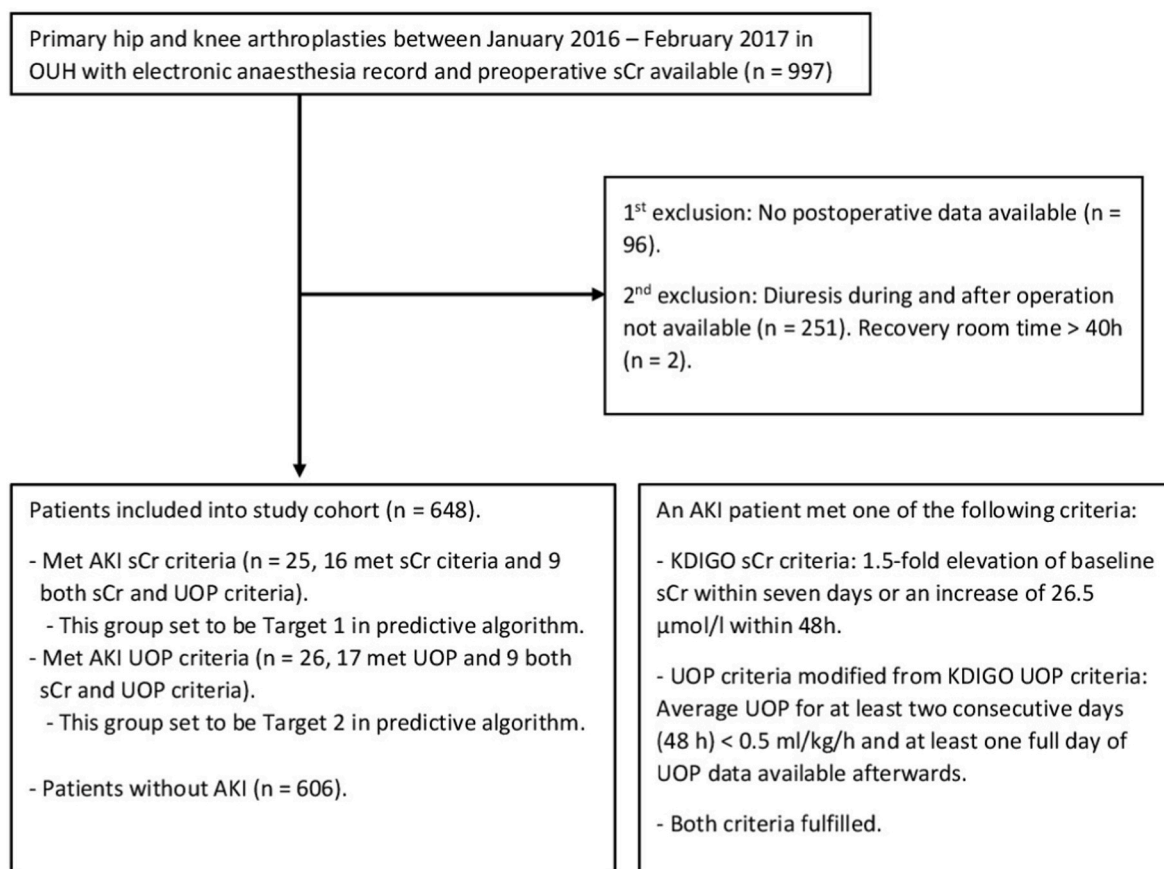


Fig. 1. Flow diagram. AKI = acute kidney injury; KDIGO = Kidney Disease Improving Global Outcomes; OUH = Oulu University Hospital; sCr = serum creatinine; UOP = urine output.

1. To assess the performance of fewer features, models using only the features with a relative importance of >0.1 were established for all TARGET 1 and 2 TRIALS A, B, C. The features were analyzed with RUSBoost models, and these same features were used in comparative algorithms. This resulted in a total of 14 different RUSBoost, Naïve/Kernel Bayes and SVM models.

Due to the large number of excluded patients, we studied the effect of increasing the sample size to its maximum. This was possible only for Target 1 Trial A by including the patients who were excluded due to missing UOP data but whose preoperative data and postoperative sCr were available. For other trials, it was not feasible to increase the sample size due to the lack of key intraoperative data.

3. Results

A total of 997 patients fulfilled the inclusion criteria. A total of 349 patients (35.0%) were excluded, mostly due to a lack of UOP data. The final sample included 648 patients of whom 16 met the criteria for sCr-AKI and 17 met the criteria for UOP-AKI. Nine patients met both the sCr and UOP criteria. The number of non-AKI patients was 606 (Fig. 1).

The 251 excluded patients with missing UOP data were analyzed and compared to the included patients (Supplemental Table 1). Most of the excluded patients underwent elective knee arthroplasty (87.9%), were younger (64 vs. 74%), had a lower ASA status, had a higher preoperative eGFR, received intraoperative nonsteroidal anti-inflammatory drugs more often, had a more stable intraoperative mean arterial pressure, experienced reduced blood loss, received fewer or no vasoactive drugs, had normal postoperative blood pressure more often, and were more often discharged home ($P = <0.001$ for all mentioned comparisons). The excluded patients were healthier than the patients with more accurate

UOP measurements.

Table 2 presents the main outcomes of patients with AKI. The patients with AKI according to either the sCr or UOP criteria had increased 1-year mortality rates. The mortality rate was 28% compared to 5.8% in the sCr group ($p = <0.001$) and 23.1% compared to 5.9% in the UOP group ($p = 0.005$). In both AKI groups, follow-up treatment usually did not occur at home ($p = <0.001$ for both groups). Most of the patients had stage 1 AKI, and there were only five stage 3 AKI patients.

3.1. Algorithm performance

The performance of the ML models is presented in Table 3. Nine sCr-based AKI cases occurred within the 1st postoperative day, while most of the sCr-AKI cases (64%) occurred between postoperative days two and seven. Most UOP-based AKI cases (92%) occurred on postoperative Day 2, and 2/26 occurred on Day 3.

The performance remained the same in RUSBoost TARGET 1 Trials A and B. In TARGET 2, the performance of the model was improved by including intraoperative features with preoperative features. Both TARGET performances were improved by including postoperative features. The performance for the RUSBoost Target 1 Trial C model was as follows: area under receiving operating characteristic (AUROC) of 0.91, (95% CI ± 0.02), area under precision-recall (AUPR) of 0.35 (95% CI ± 0.04) sensitivity of 0.88 (95% CI ± 0.03), specificity of 0.87 (95% CI ± 0.03), and precision of 0.22 (95% CI ± 0.03). The model correctly classified 22 out of 25 AKI patients and correctly classified 543 out of 623 patients who did not have sCr-based AKI. Excluding the nine sCr-based AKI patients who met the criteria on postoperative Day 1 slightly reduced the Trial C performance: AUROC of 0.92 (95% CI ± 0.02), AUPR of 0.20 (95% CI ± 0.03), sensitivity of 0.88 (95% CI \pm

Table 2

Outcomes of patients with and without acute kidney injury and the stages of acute kidney injury. AKI = acute kidney injury; sCr = serum creatinine; UOP = urine output; KDIGO-criteria for stage 1 AKI: An absolute increase of sCr by 26.5 µmol/l within 48h or a 1.5-fold increase of sCr within 7 days or UOP <0.5 ml/kg/h for >6 h. Stage 2: A 2-fold increase in sCr or UOP <0.5 ml/kg/h for >12h. Stage 3: A 3-fold increase of sCr or dialysis or UOP <0.3 ml/kg/h for >24h or anuria for >12h. The diagnosis criteria of AKI by UOP in this material was UOP <0.5 ml/kg/h for 48h at the postoperative ward for Stage 1 and < 0.3 ml/kg/h for 48h for Stage 3. P-values for comparisons for patients with either sCr-AKI or UOP-AKI and no-AKI patients are presented.

	No sCr-AKI, N = 623	sCr-AKI, N = 25	P-value	No UOP-AKI, N = 622	UOP-AKI, N = 26	P-value
Location of follow-up treatment						
Home N (%)	284 (46.0)	3 (12.0)	<0.001	280 (45.5)	7 (26.9)	<0.001
Health center ward N (%)	182 (28.3)	8 (32.0)		177 (28.7)	5 (19.2)	
Hospital N (%)	152 (24.6)	13 (52.0)		153 (24.8)	12 (46.2)	
Death during primary treatment period N (%)	1 (0.2)	1 (4.0)		0	2 (7.7)	
1-year mortality N (%)	36 (5.8)	7 (28.0)	0.001	37 (5.9)	6 (23.1)	0.005
Renal replacement therapy N (%)	0	0		0		
Intensive care N (%)	12 (1.9)	1 (4.0)	0.40	12 (1.9)	1 (3.8)	0.42
Length of stay, median (25th through 75th) (d)	3 (2-3)	3 (2-3.5)	0.65	3 (2-3)	3 (2-3)	0.54
AKI Stage						
Stage 1 N (%)		23 (92.0)		23 (88.5)		
Stage 2 N (%)		0 (0.0)		0 (0.0)		
Stage 3 N (%)		2 (8.0)		3 (11.5)		

0.03), specificity of 0.87 (95% CI ± 0.03), and precision of 0.15 (95% CI ± 0.03).

For the RUSBoost Target 2 Trial C model, the performance was as follows: AUROC of 0.95 (95% CI ± 0.02), AUPR of 0.31 (95% CI ± 0.04), sensitivity of 0.77 (95% CI ± 0.03), specificity of 0.91 (95% CI ± 0.03), and precision of 0.26 (95% CI ± 0.03). The model correctly classified 20 out of 26 AKI patients and correctly classified 566 out of 622 patients who did not have UOP-based AKI.

The overall performance of SVM was generally worse and inconsistent. However, Naïve Bayes had similar overall performance compared to RUSBoost, especially in TARGET 2 Trial C (Table 3). The AUROC and AUPR curves are presented in Supplemental Fig. 2.

The results of TARGET 1 TRIAL A, which included an increased sample size, are presented in Supplemental Table 2. No additional AKI cases occurred. The performance of both RUSBoost and Naïve Bayes slightly increased.

3.2. Importance of features

The relative importance of the features in the Target 1 Trial C and Target 2 Trial C models are shown in Figs. 2 and 3. The performance of the trained models using only features with a relative importance of >0.1 are presented in Table 4, and the features included in these models are listed in Table 5. Again, the performances of TARGET 1 TRIAL A and B were the same, and in TARGET 2, the performance was improved by including intraoperative features with preoperative features. The

inclusion of postoperative features improved both models.

Compared to the previous models with a larger number of features, the changes in the performance of both RUSBoost and Naïve Bayes models were small and inconsistent, with RUSBoost achieving better performance in TARGET 2 TRIAL C and Naïve Bayes achieving better performance in TARGET 1 TRIAL C. The RUSBoost TARGET 2 TRIAL C performed as follows: AUROC of 0.98 (95% CI ± 0.02), AUPR of 0.48 (95% CI ± 0.04), sensitivity of 0.88 (95% CI ± 0.02), specificity of 0.93 (95% CI ± 0.02), and precision of 0.34 (95% CI ± 0.04). This model correctly classified 23 out of 26 patients with AKI and 577 out of 622 patients who did not have UOP-based AKI.

In the sCr-AKI models, preoperative eGFR was the most important feature in Trials A and B. In trial C, four of the five most important features were directly related to eGFR. In trial B, UOP measurements were also among the most important features, but these UOP measurements had less importance than the preoperative eGFR values.

Similarly, in the UOP-AKI Trial C model, three out of seven features with a relative importance of >0.1 were UOP measurements. UOP on the operation day was also the most important feature in the UOP-AKI Trial B model. In Trial A, the eGFR was among the important features, but it was not as important as it was in sCr-AKI models.

Age and surgery type were among the most important features in all six models. Sex, BMI, ASA-status, preoperative eGFR, and preoperative sCr level were present in both Target 1 and 2 Trials A and B. Duration of operation and time spent in the recovery room were present in Trial B of both Targets and in Target 1 Trial C.

The model that predicted sCr-based AKI, but with the exclusion of nine patients who were shown to have AKI on postoperative Day 1, had 42 features with a relative importance of >0.1. The performance was as follows: AUROC of 0.92 (95% CI ± 0.02), AUPR of 0.20 (95% CI ± 0.03), sensitivity of 0.88 (95% CI ± 0.03), specificity of 0.87 (95% CI ± 0.03), and precision of 0.15 (95% CI ± 0.03). The model correctly classified 14 out of 16 AKI patients and 543 out of 623 patients who did not have sCr-based AKI.

The overall performance of SVM was worse. The performance of Naïve/Kernel Bayes was overall similar to RUSBoost in both TARGETs and Trials and was better in Target 1 Trials B and C. The AUROC and AUPR curves are presented in Supplemental Fig. 3.

4. Discussion

The main finding is that including both intraoperative and postoperative features with preoperative features improves the performance of predictive algorithms. In particular, the inclusion of postoperative features increases the performance. The results of using either a higher or lower number of features were inconsistent in this study, but since the performance was either almost the same or increased, it could be reasonable to focus on only the most important features.

Adding postoperative features clearly increased both the sensitivity and precision in both TARGETs and in both models using a large set of features or features with a relative importance of >0.1. The reason why the performance between TRIALS A and B in TARGET 1 did not improve is unclear, but it seems that the sCr and eGFR measurements on first postoperative day are very important for sCr-based AKI models. Although inclusion of the sCr level from the first postoperative day in TARGET 1 TRIAL C meant that the AKI cases that occurred on the first postoperative day were self-explanatory, the model could also identify cases of AKI that occurred days after the operation. This result remained after the exclusion of AKI patients diagnosed on postoperative Day 1.

Comparison between the models using many features and the models using features with a relative importance of >0.1 was inconsistent. Using only the features with a relative importance of >0.1 reduced the sensitivity of TARGET 1 Trial C to 0.84 from 0.88, while the precision of 0.22 remained the same. In TARGET 2, the sensitivity and precision both improved in Trial B and Trial C, with the improvement in the latter being particularly notable. In Trial C, the sensitivity increased from 0.77 to

Table 3
 Performance of the RUSBoost, Naïve/Kernel Bayes and Support Vector Machine models. Sensitivity/recall, specificity, accuracy, precision, area under receiving operating characteristic and area under precision-recall curve are presented for each model and Trial respectively. Results are presented in order Boost/Bayes/SVM. The #Predictors presents the number of features used in RUSBoost, Bayes and support vector machine models respectively. Trial A represents the model with only preoperative features, Trial B with pre-, and intraoperative features and Trial C with pre-, intra-, and postoperative features. Confidence intervals of 95% are presented. Pre-op = preoperative; Intra-op = intraoperative; Post.op. day 1 = postoperative features covering the day after the operation; AUPR = area under precision-recall curve; sCr = serum creatinine; Boost = RUSBoost; Bayes = Naïve/Kernel Bayes; SVM = support vector machine; AKI = acute kidney injury; Pred. = predicted; UOP = urine output.

Target 1 sCr	Trial A: Pre-Op						Trial B: Pre-Op, Intra-Op						Trial C: Pre-Op, Intra-Op, Post-Op-1day					
	Sens	Spec	Accur	Prec	AUROC	AUPR	Sens	Spec	Accur	Prec	AUROC	AUPR	Sens	Spec	Accur	Prec	AUROC	AUPR
Boost	0.68 ± 0.04	0.79 ± 0.03	0.79 ± 0.03	0.12 ± 0.02	0.81 ± 0.03	0.11 ± 0.02	0.68 ± 0.04	0.78 ± 0.03	0.78 ± 0.03	0.11 ± 0.02	0.79 ± 0.03	0.12 ± 0.03	0.88 ± 0.03	0.87 ± 0.03	0.87 ± 0.03	0.22 ± 0.03	0.91 ± 0.02	0.35 ± 0.04
Bayes	0.84 ± 0.03	0.63 ± 0.04	0.64 ± 0.04	0.08 ± 0.02	0.81 ± 0.03	0.17 ± 0.03	0.68 ± 0.04	0.77 ± 0.03	0.77 ± 0.03	0.11 ± 0.02	0.81 ± 0.03	0.11 ± 0.02	0.76 ± 0.03	0.87 ± 0.03	0.87 ± 0.03	0.20 ± 0.03	0.83 ± 0.03	0.15 ± 0.03
SVM	0.72 ± 0.03	0.74 ± 0.03	0.74 ± 0.03	0.10 ± 0.02	0.76 ± 0.03	NA	0.36 ± 0.04	0.76 ± 0.03	0.74 ± 0.03	0.06 ± 0.02	0.75 ± 0.03	NA	0.92 ± 0.02	0.09 ± 0.02	0.13 ± 0.03	0.04 ± 0.01	0.52 ± 0.04	NA
Confusion matrix N = 648	True "AKI" True "No AKI"		18/21/17 160/236/128		7/4/8 463/387/495		True "AKI" True "No AKI"		17/17/9 135/143/152		8/8/16 488/480/471		True "AKI" True "No AKI"		22/19/23 80/78/564		3/6/2 543/545/59	
	31/28/31		Pred. "AKI"		Pred. "No AKI"		#Predictors 45/45/45		Pred. "AKI"		Pred. "No AKI"		#Predictors 78/78/78		Pred. "AKI"		Pred. "No AKI"	
Target 2 UOP	Sens	Spec	Accur	Prec	AUROC	AUPR	Sens	Spec	Accur	Prec	AUROC	AUPR	Sens	Spec	Accur	Prec	AUROC	AUPR
Boost	0.62 ± 0.04	0.68 ± 0.04	0.68 ± 0.04	0.07 ± 0.02	0.68 ± 0.04	0.07 ± 0.02	0.65 ± 0.04	0.76 ± 0.03	0.76 ± 0.03	0.10 ± 0.02	0.77 ± 0.03	0.13 ± 0.03	0.77 ± 0.03	0.91 ± 0.02	0.90 ± 0.03	0.26 ± 0.03	0.95 ± 0.02	0.31 ± 0.04
Bayes	0.85 ± 0.03	0.41 ± 0.04	0.42 ± 0.04	0.06 ± 0.02	0.69 ± 0.04	0.06 ± 0.02	0.85 ± 0.03	0.56 ± 0.04	0.57 ± 0.04	0.07 ± 0.02	0.79 ± 0.03	0.14 ± 0.03	0.73 ± 0.03	0.91 ± 0.02	0.91 ± 0.02	0.26 ± 0.03	0.95 ± 0.02	0.33 ± 0.04
SVM	0.23 ± 0.03	0.87 ± 0.03	0.84 ± 0.03	0.07 ± 0.02	0.63 ± 0.04	NA	0.54 ± 0.04	0.85 ± 0.03	0.84 ± 0.03	0.13 ± 0.03	0.81 ± 0.03	NA	0.81 ± 0.03	0.31 ± 0.04	0.33 ± 0.04	0.05 ± 0.02	0.15 ± 0.03	NA
Confusion matrix N = 648	True "AKI" True "No AKI"		16/22/6 198/370/81		10/4/20 424/252/541		True "AKI" True "No AKI"		17/22/14 148/274/92		9/4/12 474/348/530		True "AKI" True "No AKI"		20/20/21 56/48/430		6/6/5 566/574/192	
	#Predictors 31/30/31		Pred."AKI"		Pred."No AKI"		#Predictors 33/27/28		Pred."AKI"		Pred."No AKI"		#Predictors 78/76/78		Pred."AKI"		Pred."No AKI"	

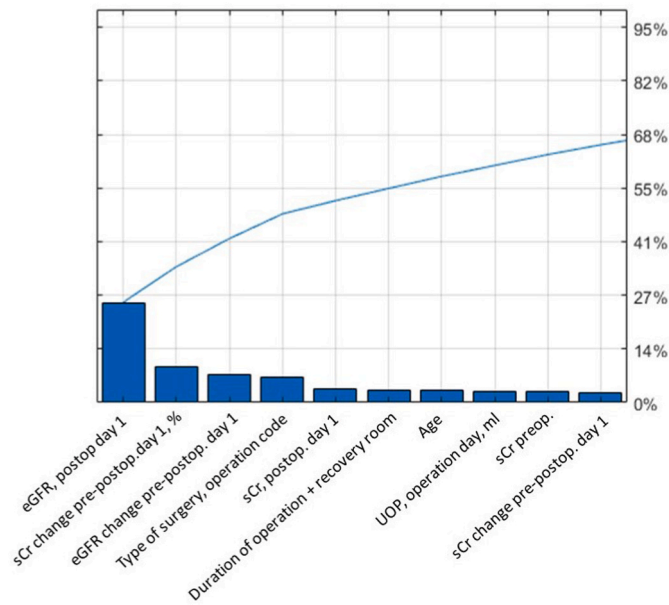


Fig. 2. The figure represents the relative importance of each feature in the RUSBoost model Target 1 Trial C predicting perioperative AKI, defined as serum creatinine criteria of the KDIGO classification. Preoperative, intraoperative and a few features from first postoperative day are included. eGFR = estimated glomerular filtration rate; sCr = serum creatinine; UOP = urine output.

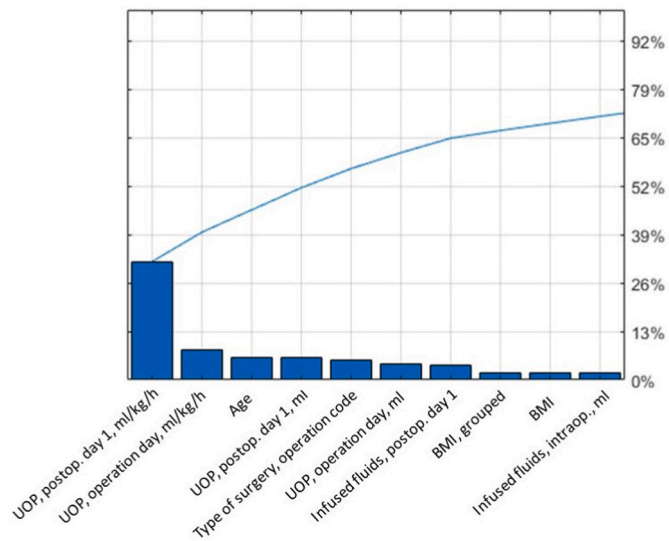


Fig. 3. The figure represents the relative importance of each feature in the RUSBoost model Target 2 Trial C predicting perioperative AKI, defined as postoperative UOP of <math><0.5\text{ ml/kg/h}</math> for at least 48 h. Preoperative, intraoperative and a few features from first postoperative day are included. eGFR = estimated glomerular filtration rate; sCr = serum creatinine; UOP = urine output.

0.88, and the precision increased from 0.26 to 0.34. The sensitivity increased in all the Naïve Bayes TARGET 1 Trials, and the precision increased in Trial C. Naïve Bayes TARGET 2 Trials showed much more inconsistency. The number of features was between 7 and 20 in the reduced feature models and 27 and 78 in the models that used more features. Some important features could be found in this larger set that could possibly explain these findings, but as it is often not feasible to use a vast set of features, we argue that using a reduced number of features could be sufficient.

In our imbalanced data, the AUROC and accuracy were less informative than precision-recall metrics since the accuracy of a model is increased by correctly classifying the majority class while neglecting the minority class [34]. In algorithm training, hyperparameter optimization did not improve precision. This result probably occurred due to the small minority class and a high class overlap [35,36].

Although RUSBoost was chosen as the main ML method in the preliminary testing, the Naïve/Kernel Bayes method performed similarly or, in some cases, better. Recent studies such as ours lack homogeneity in both defining AKI and in the methodology used in algorithms [13]. The optimal ML algorithms are still an open research question, and whether RUSBoost or Naïve/Kernel Bayes is one of these optimal ML algorithms is unclear. We were not able to rigorously test all the possible available options. Additionally, the predicted outcomes vary. Some studies have focused on more severe AKI (stage 2–3) or tried to predict the stage [20,21]. Our models only define whether the patient is likely to develop AKI regardless of the staging. This has advantages since mild AKI is also associated with increased mortality and morbidity.

Preoperatively decreased eGFR is a major risk factor for AKI [37]. Other important features recurring in the models (BMI, ASA status, operation type, preoperative sCr level and sex) are all risk factors for AKI [9,37]. The performance of both Target 1 and Target 2 Trial C probably occurred because both models are largely based on the same, or derivatives of the same, features that are used to diagnose either sCr-AKI or UOP-AKI. Our recent study with the same patient population showed that patients who met the UOP-AKI criteria had low UOP on both the operation day and postoperative Day 1 [24]. There could be similar small, but not diagnostic changes in the eGFR of patients who meet the sCr-AKI criteria on postoperative Day 1. The rate at which these key features were missing on postoperative Day 1 was 15.6% in patients who met the sCr criteria and 4.8% in patients who met the UOP criteria.

Trial B models included the operation UOP in values with units of ml/kg and ml/kg/h, whereas patient weight and operation duration were also included. The eGFR calculation included sCr level, age and sex, all of which recurred as impactful features. It has been shown that the diversity of the classifiers in an ensemble may increase the performance of an ensemble classifier [38]. In our study, we found that derived features promoted data diversity.

The sensitivity and precision of our Trial C models could be argued to be useful in clinical practice. However, there are several major concerns that contradict this argument. First, because testing the models on a new dataset was not possible, we were not able to assess the level of overfitting, which is likely, especially in Trial C models. Second, even if we were able to demonstrate that our modified UOP criteria for AKI identify oliguric patients who eventually might experience higher mortality rates, the whole concept of UOP-based AKI is under debate, and retrospectively, the diuresis measurements are largely not very reliable [39]. Finally, it is uncertain whether these kinds of AI models of AKI would work or be beneficial and offer decision-supporting information in the clinical use [13].

4.1. Limitations

This study was a retrospective study. We excluded many patients due to missing data. The UOP data were not complete, and thus training based on UOP data was not optimal. The definition of perioperative AKI by UOP is generally uncertain, and we modified the UOP criteria for this study. This study was a single-center study, which could reduce the usefulness of the results. External testing was not possible due to a lack of data, and our models are prone to overfitting. We were not able to use stratified cross-validation due to MATLAB limitations. Our algorithm selection process was not rigorous, and it is possible that neither RUSBoost nor Naïve/Kernel Bayes is the best method to use with this kind of data. To further validate and develop ML models, a larger study cohort is needed.

Table 4

Performance of the RUSBoost, Naïve/Kernel Bayes and support vector machine models with features of relative importance of >0.1. Sensitivity/recall, specificity, accuracy, precision, area under receiving operating characteristic and area under precision-recall curve are presented for each model and Trial respectively. Results are presented in order Boost/Bayes/SVM. The #Predictors presents the number of features used in RUSBoost, Bayes and support vector machine models respectively. Trial A represents the model with only preoperative features, Trial B with pre-, and intraoperative features and Trial C with pre-, intra-, and postoperative features. Confidence intervals of 95% are presented. Pre-op = preoperative; Intra-op = intraoperative; Post.op. day 1 = postoperative features covering the day after the operation; AUPR = area under precision-recall curve; sCr = serum creatinine; Boost = RUSBoost; Bayes = Naïve/Kernel Bayes; SVM = support vector machine; AKI = acute kidney injury; Pred. = predicted; UOP = urine output.

Target 1 sCr	Trial A: Pre-Op						Trial B: Pre-Op, Intra-Op						Trial C: Pre-Op, Intra-Op, Post-Op-1day					
	Sens	Spec	Accur	Prec	AUROC	AUPR	Sens	Spec	Accur	Prec	AUROC	AUPR	Sens	Spec	Accur	Prec	AUROC	AUPR
Boost	0.68 ± 0.04	0.80 ± 0.03	0.79 ± 0.03	0.12 ± 0.03	0.81 ± 0.03	0.15 ± 0.03	0.68 ± 0.04	0.80 ± 0.03	0.80 ± 0.03	0.12 ± 0.02	0.80 ± 0.03	0.12 ± 0.03	0.84 ± 0.03	0.88 ± 0.03	0.88 ± 0.03	0.22 ± 0.03	0.88 ± 0.03	0.27 ± 0.03
Bayes	0.92 ± 0.02	0.52 ± 0.04	0.54 ± 0.04	0.07 ± 0.02	0.85 ± 0.03	0.18 ± 0.03	0.92 ± 0.02	0.64 ± 0.04	0.65 ± 0.04	0.09 ± 0.02	0.82 ± 0.03	0.23 ± 0.03	0.88 ± 0.03	0.87 ± 0.03	0.87 ± 0.03	0.22 ± 0.03	0.94 ± 0.02	0.37 ± 0.03
SVM	0.48 ± 0.04	0.83 ± 0.03	0.81 ± 0.03	0.10 ± 0.02	0.67 ± 0.04	NA	0.60 ± 0.04	0.71 ± 0.04	0.71 ± 0.04	0.08 ± 0.02	0.73 ± 0.03	NA	0.80 ± 0.03	0.66 ± 0.04	0.67 ± 0.04	0.09 ± 0.02	0.90 ± 0.02	NA
Confusion matrix N = 648	True "AKI"		17/23/12		8/2/13		True "AKI"		17/23/15		8/2/10		True "AKI"		21/22/20		4/3/5	
	True "No AKI"		125/297/107		498/326/516		True "No AKI"		123/226/181		500/397/442		True "No AKI"		75/79/210		548/544/413	
	#Predictors 14/14/14		Pred. "AKI"		Pred. "No AKI"		#Predictors 20/20/20		Pred. "AKI"		Pred. "No AKI"		#Predictors 19/19/19		Pred. "AKI"		Pred. "No AKI"	
Target 2 UOP	Sens	Spec	Accur	Prec	AUROC	AUPR	Sens	Spec	Accur	Prec	AUROC	AUPR	Sens	Spec	Accur	Prec	AUROC	AUPR
Boost	0.62 ± 0.04	0.63 ± 0.04	0.63 ± 0.04	0.07 ± 0.02	0.65 ± 0.04	0.07 ± 0.02	0.69 ± 0.04	0.75 ± 0.03	0.75 ± 0.03	0.11 ± 0.02	0.80 ± 0.03	0.11 ± 0.02	0.88 ± 0.02	0.93 ± 0.02	0.93 ± 0.02	0.34 ± 0.04	0.98 ± 0.02	0.48 ± 0.04
Bayes	0.04 ± 0.01	0.97 ± 0.01	0.93 ± 0.02	0.05 ± 0.02	0.66 ± 0.04	0.06 ± 0.02	0.54 ± 0.04	0.82 ± 0.03	0.81 ± 0.03	0.11 ± 0.02	0.76 ± 0.03	0.12 ± 0.03	0.92 ± 0.02	0.84 ± 0.03	0.85 ± 0.03	0.20 ± 0.03	0.95 ± 0.02	0.54 ± 0.04
SVM	0.65 ± 0.04	0.61 ± 0.04	0.61 ± 0.04	0.07 ± 0.02	0.64 ± 0.04	NA	0.27 ± 0.03	0.93 ± 0.02	0.90 ± 0.02	0.14 ± 0.03	0.73 ± 0.03	NA	0.88 ± 0.02	0.82 ± 0.03	0.82 ± 0.03	0.17 ± 0.03	0.94 ± 0.02	NA
Confusion matrix N = 648	True "AKI"		16/1/17		10/25/9		True "AKI"		16/23/20		10/3/6		True "AKI"		23/24/23		3/2/3	
	True "No AKI"		228/20/244		394/602/378		True "No AKI"		269/170/141		481/353/452		True "No AKI"		45/98/113		577/524/509	
	#Predictors 11/11/11		Pred."AKI"		Pred."No AKI"		#Predictors 20/20/20		Pred."AKI"		Pred."No AKI"		#Predictors 7/7/7		Pred."AKI"		Pred."No AKI"	

Table 5

The features included in Naïve Bayes, Support Vector Machine and RUSBoost models using the features with relative importance of >0.1. The features with relative importance of >0.1 were ascertained from RUSBoost models. The first ten features are listed in order of relative importance, the first listed feature having a relative importance of 1 and if more than ten features, alphabetically. Colour annotations: white colour = preoperative feature, light grey = intraoperative feature, dark grey = postoperative feature. ASA = American Society of Anesthesiologists status; eGFR = estimated glomerular filtration rate; intraop. = intraoperative feature; preop. = preoperative feature; pre-postop. = a change of a feature value between preoperative and postoperative measurements; postop. = postoperative; sCr = serum creatinine; UOP = urine output.

Target 1 Trial A features	Target 1 Trial B features	Target 1 Trial C features
eGFR preop.	eGFR preop.	eGFR, postop. day 1
Age	Type of surgery, operation code	sCr change pre-postop.day 1, %
Type of surgery, operation code	Age	eGFR change pre-postop. day 1
Weight	Duration of operation	Type of surgery, operation code
sCr, preop.	Weight	sCr, postop. day 1
BMI	UOP, operation day, ml	Duration of operation + recovery room
Sex	sCr, preop.	Age
Prior major surgeries, N	BMI	UOP, operation day, ml
Contrast medium within 5 yr.	Duration of operation + recovery room	sCr preop.
Emergency hip arthroplasty	UOP, operation day, after surgery, ml	sCr change pre-postop.day 1
ASA	ASA	Infused fluids, operation day, after surgery, ml
BMI, grouped	BMI, grouped	Infused fluids, operation day, ml/kg/h
Contrast medium within 1 yr.	ACE inhibitor at home	Infused fluids, postop. day 1, ml
Type of prior major operations	Infused fluids, intraop., ml/h	Indused fluids, postop. day 1, ml/kg/h
	Infused fluids, intraop., ml/kg/h	Other fluid loss, postop. day 1, ml
	UOP, intraop., ml	UOP, operation day, after surgery, ml
	UOP, intraop., ml/h	UOP, operation day, ml/kg/h
	UOP, intraop., ml/kg/h	UOP, postop. day 1, ml
	UOP, operation day, ml/kg/h	UOP, postop. day 1, ml/kg/h
	Sex	
Target 2 Trial A features	Target 2 Trial B features	Target 2 Trial C features
Age	UOP, operation day, ml/kg/h	UOP, postop. day 1, ml/kg/h
Type of surgery, operation code	Age	UOP, operation day, ml/kg/h
Weight	UOP, operation day, ml	Age
BMI	Type of surgery, operation code	UOP, postop. day 1, ml
eGFR, preop.	BMI, grouped	Type of surgery, operation code
sCr, preop.	Duration of operation + recovery room	UOP, operation day, ml
BMI, grouped	UOP, intraop., ml/kg/h	Infused fluids, postop. day 1, ml
ASA	Infused fluids, intraop., ml/h	
Sex	ASA	
Sartan at home	UOP, intraop, ml	
Prior major operations within 5 yr, N.	ACE inhibitor at home	
	Contrast medium within 5 yr.	
	Diabetes	
	Emergency hip arthroplasty	
	Hypertension	

Table 5 (continued)

Target 1 Trial A features	Target 1 Trial B features	Target 1 Trial C features
		Prior major operations within 5 yr, N
		Sartan at home
		Sex
		sCr preop.
		UOP, operation day, ml/h

4.2. Conclusion

The inclusion of intra- and postoperative features with preoperative features improves the performance of predictive algorithms. Retrospective data could be used to create algorithms to predict perioperative AKI defined by both sCr level and UOP. The results of our study are not generalizable, and additional larger studies are needed. The optimal ML method for this kind of data is still an open research question.

Author contributions

ON designed the study, contributed to data collection, performed statistical analyses, contributed to ML development, and wrote and edited the manuscript. TK contributed to the study idea and designed the study, performed the ML analyses, and wrote and edited the manuscript. ON and TK contributed equally to the whole study. TA and EJ contributed to data collection. MV and SA contributed to the study protocol, study idea and design and editing of the manuscript. All the authors read and approved the final manuscript.

Funding

Data collection was supported by a small grant from the State Research Fund appointed to Department of Anaesthesiology, Oulu University Hospital. ML model development was supported by the European Regional Development Fund.

Declaration of interest

The authors declare no conflicts of interest.

Acknowledgments

We thank Pasi Ohtonen, biostatistician, Oulu University Hospital, for his comments concerning both performing and reporting of the statistical analyses.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.compbimed.2022.105351>.

References

- [1] A.S. Levey, J. Coresh, Chronic kidney disease, *Lancet* 379 (2012) 165–180, [https://doi.org/10.1016/S0140-6736\(11\)60178-5](https://doi.org/10.1016/S0140-6736(11)60178-5).
- [2] L.S. Chawla, P.W. Eggers, R.A. Star, P.L. Kimmel, Acute kidney injury and chronic kidney disease as interconnected syndromes, *N. Engl. J. Med.* 371 (2014) 58–66, <https://doi.org/10.1056/NEJMr1214243>.
- [3] J.T. Park, Postoperative acute kidney injury, *Korean J. Anesthesiol.* 70 (2017) 258–266, <https://doi.org/10.4097/kjae.2017.70.3.258>.
- [4] E. Borthwick, A. Ferguson, Perioperative acute kidney injury: risk factors, recognition, management, and outcomes, *Br. Med. J.* 341 (2010) 85–91, <https://doi.org/10.1136/bmj.c3365>.
- [5] Y.S. Louise, N.W. Duminda, A.T. Gordon, W. Sb, Association of intraoperative hypotension with acute kidney injury after elective noncardiac surgery, *Anesthesiology* 123 (2015), <https://doi.org/10.1097/ALN.0000000000000765>, 515–2.
- [6] C.J. Porter, I.K. Moppett, I. Juurlink, J. Nightingale, C.G. Moran, M.A.J. Devonald, Acute and chronic kidney disease in elderly patients with hip fracture: prevalence,

- risk factors and outcome with development and validation of a risk prediction model for acute kidney injury, *BMC Nephrol.* 18 (2017) 20, <https://doi.org/10.1186/s12882-017-0437-5>.
- [7] L.A. Kimmel, S. Wilson, J.D. Janardan, S.M. Liew, R.G. Walker, Incidence of acute kidney injury following total joint arthroplasty: a retrospective review by RIFLE criteria, *Clin. Kidney J.* 7 (2014) 546–551, <https://doi.org/10.1093/ckj/sfu108>.
- [8] A.B. Pedersen, C.F. Christiansen, H. Gammelager, J. Kahlert, H.T. Sørensen, Risk of acute renal failure and mortality after surgery for a fracture of the hip, *Bone Joint Lett. J* 98 (2016) 1112–1118, <https://doi.org/10.1302/0301-620X.98B8.37497>.
- [9] P. Jämsä, E. Jämsen, L. Lyytikäinen, J. Kalliovalkama, A. Eskelinen, N. Oksala, Risk factors associated with acute kidney injury in a cohort of 20,575 arthroplasty patients, *Acta Orthop.* 88 (2017) 370–376, <https://doi.org/10.1080/17453674.2017.1301743>.
- [10] S.D. Gumbert, F. Kork, M.L. Jackson, N. Vanga, S.J. Ghebremichael, C.Y. Wang, H. K. Eltzschig, Perioperative acute kidney injury, *Anesthesiology* 132 (2020) 180–204, <https://doi.org/10.1097/ALN.0000000000002968>.
- [11] M.E. O'Connor, R.W. Hewson, C.J. Kirwan, G.L. Ackland, R.M. Pearse, J.R. Prowle, Acute kidney injury and mortality 1 Year after major non-cardiac surgery, *Br. J. Surg.* 104 (2017) 868–876, <https://doi.org/10.1002/bjs.10498>.
- [12] C.W. Connor, Artificial intelligence and machine learning in anesthesiology, *Anesthesiology* 131 (2019) 1346–1359, <https://doi.org/10.1097/ALN.0000000000002694>.
- [13] J. Gameiro, T. Branco, J.A. Lopes, Artificial intelligence in acute kidney injury risk prediction, *J. Clin. Med.* 9 (3) (2020) 678, <https://doi.org/10.3390/jcm9030678>.
- [14] H.-C. Lee, H.-K. Yoon, K. Nam, Y.J. Cho, T.K. Kim, W.H. Kim, J.-H. Bahk, Derivation and validation of machine learning approaches to predict acute kidney injury after cardiac surgery, *J. Clin. Med.* 7 (10) (2018) 322, <https://doi.org/10.3390/jcm7100322>.
- [15] C. Zhou, R. Wang, W. Jiang, J. Zhu, Y. Liu, J. Zheng, X. Wang, W. Shang, L. Sun, Machine learning for the prediction of acute kidney injury and paraplegia after thoracoabdominal aortic aneurysm repair, *J. Card. Surg.* 35 (2019) 1, <https://doi.org/10.1111/jocs.14317>.
- [16] C. Huang, K. Murugiah, S. Mahajan, S.-X. Li, S.S. Dhruva, J.S. Haimovich, Y. Wang, W.L. Schulz, J.M. Testani, F.P. Wilson, C.I. Mena, F.A. Masoudi, J.S. Rumsfeld, J. A. Spertus, B.J. Mortazavi, H.M. Krumholz, Enhancing the prediction of acute kidney injury risk after percutaneous coronary intervention using machine learning techniques: a retrospective cohort study, *PLoS Med.* 15 (2018) 11, <https://doi.org/10.1371/journal.pmed.1002703>.
- [17] P. Thottakkara, T. Ozrazgat-Baslanti, B.B. Hupf, P. Rashidi, P. Pardalos, P. Momcilovic, A. Bihorac, Application of machine learning techniques to high-dimensional clinical data to forecast postoperative complications, *PLoS One* 11 (2016) 5, <https://doi.org/10.1371/journal.pone.0155705>.
- [18] V.J. Lei, T. Luong, E. Shan, X. Chen, M.D. Neuman, N.D. Eneanya, D.E. Polsky, K. G. Volpp, L.A. Fleisher, J.H. Holmes, A.S. Navathe, Risk stratification for postoperative acute kidney injury in major noncardiac surgery using preoperative and intraoperative data, *JAMA Netw. Open* 2 (2019) 12, <https://doi.org/10.1001/jamanetworkopen.2019.16921>.
- [19] Z. Zhang, K.M. Ho, Y. Hong, Machine learning for the prediction of volume responsiveness in patients with oliguric acute kidney injury in critical care, *Crit. Care* 12 (2019) 112, <https://doi.org/10.1186/s13054-019-2411-z>.
- [20] M. Flechet, S. Falini, C. Bonetti, F. Güiza, M. Schetz, G. Van den Berghe, G. Meyfroidt, Machine learning versus physicians' prediction of acute kidney injury in critically ill adults: a prospective evaluation of the AKIpredictor, *Crit. Care* 23 (2019) 282, <https://doi.org/10.1186/s13054-019-2563-x>.
- [21] J.L. Koyner, R. Adhikari, D.P. Edelson, M.M. Churpek, Development of a multicenter ward-based AKI prediction model, *Clin. J. Am. Soc. Nephrol.* 11 (2016) 11, <https://doi.org/10.2215/CJN.00280116>.
- [22] H.-C. Lee, S.B. Yoon, S.-M. Yang, W.H. Kim, H.-G. Ryu, C.-W. Jung, K.-S. Suh, K. H. Lee, Prediction of acute kidney injury after liver transplantation: machine learning approaches vs. Logistic regression model, *J. Clin. Med.* 7 (11) (2018) 428, <https://doi.org/10.3390/jcm7110428>.
- [23] M. Kim, G. Li, S. Mohan, Z.A. Turnbull, R.P. Kiran, G. Li, Intraoperative data enhance the detection of high-risk acute kidney injury patients when added to a baseline prediction model, *Anaesth Analg* 132 (2021) 430–441, <https://doi.org/10.1213/ane.0000000000005057>.
- [24] Jämsä Nikkinen, Alahuhta Aaltonen, Vakkala Ohtonen, Perioperative acute kidney injury and urine output in lower limb arthroplasties, *Acta Anaesthesiol. Scand.* 65 (2021) 1054–1064, <https://doi.org/10.1111/aas.13834>.
- [25] The Kidney Disease Improving Global Outcomes(KDIGO) Working Group, Definition and classification of acute kidney injury, *Kidney Int.* 2 (2012) 19–36, <https://doi.org/10.1038/kisup.2012.1>.
- [26] A.S. Levey, L.A. Stevens, C.H. Schmid, et al., A new equation to estimate glomerular filtration rate, *Ann. Intern. Med.* 150 (2009) 604–612.
- [27] H. Blockeel, J. Struyf, Efficient algorithms for decision tree cross-validation, *JMLR* (2002) 621–650.
- [28] C. Seiffert, T. Khoshgoftaar, J. Van Hulse, A. Napolitano, RUSBoost: Improving classification performance when training data is skewed, in: 2008. 19th International Conference on Pattern Recognition, 2008, pp. 1–4, <https://doi.org/10.1109/ICPR.2008.4761297>. Tampa, FL.
- [29] M. Hadjem, F. Naït-Abdesselam, A. Khokhar, ST-segment and T-wave anomalies prediction in an ECG data using RUSBoost, in: IEEE 18th International Conference on E-Health Networking, Applications and Services, Healthcom, Munich, 2016, pp. 1–6, <https://doi.org/10.1109/HealthCom.2016.7749493>.
- [30] J. Urdal, K. Engan, V. Kvikstad, E.A.M. Janssen, Prognostic Prediction of Histopathological Images by Local Binary Patterns and RUSBoost, 25th European Signal Processing Conference (EUSIPCO), Kos, 2017, pp. 2349–2353, 2017.
- [31] S. Amin, A.M. Kamboh, A robust approach towards epileptic seizure detection, in: IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP), Vietri sul Mare, 2016, pp. 1–6.
- [32] Matlab Documentation, Cited 3rd June, <https://se.mathworks.com/help/stats/ensemble-algorithms.html#bsw8avy>, 2020.
- [33] Y. Freund, A More Robust Boosting Algorithm, 2009. <https://arxiv.org/abs/0905.2138>. arXiv.org.
- [34] T. Saito, M. Rehmsmeier, The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets, *PLoS One* 10 (2015), <https://doi.org/10.1371/journal.pone.0118432>.
- [35] J. Kong, W. Kowalczyk, D.A. Nguyen, T. Bäck, S. Menzel, Hyperparameter optimisation for improving classification under class imbalance, *IEEE Symposium Series Comput. Intelligen.* (2019) 3072–3078. <https://doi.org/10.1109/SSCI44817.2019.9002679>.
- [36] P. Vuttipittayamongkol, E. Elyan, A. Petrovski, On the class problem in imbalanced data classification, *Knowl. Base Syst.* (2021) 212, <https://doi.org/10.1016/j.knsys.2020.106631>.
- [37] S. Khetarpal, K.K. Tremper, M. Heung, A.L. Rosenberg, M. Englesbe, A.M. Shanks, D.A. Campbell, Development and validation of an acute kidney injury risk index for patients undergoing general surgery: results from a national data set, *Anesthesiology* 110 (2009) 505–515, <https://doi.org/10.1097/ALN.0b013e3181979440>.
- [38] J.F. Díez-Pastor, J.J. Rodríguez, C.I. García-Osorio, L.I. Kuncheva, Diversity techniques improve the performance of the best imbalance learning ensembles, *Inf. Sci.* 325 (2015) 98–117, <https://doi.org/10.1016/j.ins.2015.07.025>.
- [39] A. Zarbock, J.L. Koyner, E.A.J. Hoste, J.A. Kellum, Update on perioperative acute kidney injury, *Anesth. Analg.* 127 (5) (2018) 1236–1245, <https://doi.org/10.1213/ANE.0000000000003741>.