Machine learning based preoperative analytics for the prediction of anastomotic insufficiency in colorectal surgery: a single-centre pilot study

Stephanie Taha-Mehlitz, MD; Larissa Wentzler, BMed; Fiorenzo Angehrn, MD; Ahmad Hendie; Vincent Ochs; Victor E. Staartjes, MD; Markus von Flüe, MD; Anas Taha*, MD; Daniel Steinemann*, MD

1. Clarunis, University Centre for Gastrointestinal and Liver Diseases, St. Clara Hospital and University Hospital, 4002 Basel, Switzerland
2. Medical Faculty, University Basel, 4056 Basel, Switzerland
3. Department of Computer Engineering, McGill University, Montreal H3A 0E9, Canada
4. Roche Innovation Centre Basel, Department of Pharma Research & Early Development, 4070 Basel, Switzerland
5. Machine Intelligence in Clinical Neuroscience (MICN) Laboratory, University Hospital Zurich, 8091 Zurich, Switzerland
6. Department of Biomedical Engineering, University of Basel, 4123 Allschwil, Switzerland

* A. Taha and D. Steinemann have contributed equally to this work and share senior authorship.

Corresponding author:
Stephanie Taha-Mehlitz, MD
Clarunis, University Centre for Gastrointestinal and Liver Diseases
St. Clara Hospital and University Hospital, Basel
Clarunis AG, Postfach, 4002 Basel, Switzerland
T +41 61 777 75 75; F +41 61 777 75 11
stephanie.taha@clarunis.ch

Sources of funding: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Previous communications: This manuscript has been submitted to medRxiv for preprint.

Ethical approval: The study was approved by the Northwestern and Central Ethics Committee Switzerland (BASEC-Nr 2020-02265).

Data Availability Statement: The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract

Machine learning based preoperative analytics for the prediction of anastomotic insufficiency in colorectal surgery: a single-centre pilot study

Introduction: Anastomotic insufficiency (AI) is a relatively common but grave complication after colorectal surgery. This study aims to determine whether AI can be predicted from simple preoperative data using machine learning (ML) algorithms.

Methods and analysis: In this retrospective analysis, patients undergoing colorectal surgery with creation of a bowel anastomosis from the University Hospital of Basel were included. Data was split into a training set (80%) and a test set (20%). The group of patients with AI was oversampled to a ratio of 50:50 in the training set and missing values were imputed. Known predictors of AI were included as inputs: age, BMI, smoking status, the Charlson Comorbidity Index, the American Society of Anesthesiologists score, type of operation, indication, haemoglobin and albumin levels, and renal function.

Results: Of the 593 included patients, 88 experienced AI. At internal validation on unseen patients from the test set, area under the curve (AUC) was 0.61 (95% confidence interval [CI]: 0.44-0.79), calibration slope was 0.16 (95% CI: -0.06-0.39) and calibration intercept was 0.06 (95% CI: 0.02-0.11). We observed a specificity of 0.67 (95% CI: 0.58-0.76), sensitivity of 0.36 (95% CI: 0.08-0.67), and accuracy of 0.64 (95% CI: 0.55-0.72).

Conclusion: By using 10 patient-related risk factors associated with AI, we demonstrate the feasibility of ML-based prediction of AI after colorectal surgery. Nevertheless, it is crucial to include multicenter data and higher sample sizes to develop a robust and generalisable model, which will subsequently allow for deployment of the algorithm in a web-based application.
Strengths and limitations of this study

• To the best of our knowledge, this is the first study to establish a risk prediction model for anastomotic insufficiency in a perioperative setting in colon surgery.

• Data from all patients that underwent colon surgery within 8 years at University Hospital Basel were included.

• We evaluated the feasibility of developing a machine learning model that predicts the outcome by using well-known risk factors for anastomotic insufficiency.

• Although our model showed promising results, it is crucial to validate our findings externally before clinical practice implications are possible.

Keywords: anastomotic insufficiency, anastomotic leakage, machine learning, colorectal surgery, prediction tool, prediction of anastomotic insufficiency
Introduction

Anastomotic insufficiency (AI) is a severe complication following gastrointestinal surgery. According to Rahbari et al., AI is defined as a defect at the anastomotic site which leads to a connection between intra- and extraluminal compartments. AI is itself considered to be an independent risk factor for adverse clinical and oncological outcomes like decreased survival of cancer patients and increased readmission rates after surgery. The approximated incidence of AI is 3.3% after colon anastomosis and 8.6% after colorectal anastomosis in specialised centres. However, these rates are likely considerably higher in centres lacking dedicated colorectal surgery teams and after emergency surgery. In fact, depending on diagnostic criteria, AI rates of over 10% have been reported in the literature. Hospital stay is extended by twelve days on average, and healthcare-related expenses are increased by up to 30,000 USD in patients who experience AI.

In previous publications, a multitude of risk factors for AI has been identified. Among these risk factors, age, gender, body mass index (BMI), Charlson Comorbidity Index (CCI), American Society of Anesthesiologists (ASA) score, leukocytosis, anaemia, hypoalbuminemia, steroid use, renal function, previous abdominal surgery, active smoking, and alcohol abuse, liver metastasis, indication, type of surgery, emergency surgery, surgical approach, anastomotic technique, and defunctioning ileostomy were identified more or less consistently. Integrating all of these risk factors into one holistic clinical prediction of AI is a very challenging task, even for experienced physicians. Indeed, even experienced surgeons were reported to systematically underestimate the risk of AI by clinical assessment. Undoubtedly, the ability to preoperatively predict AI precisely would allow for better resource allocation and enhanced patient preparation, and an improved patient-physician relationship due to the improved quality of informed consent.
Machine learning (ML) algorithms can be exceptionally competent at integrating various patient variables into a unified risk model, which is able to generate predictions specifically for each patient. Development and rigorous validation of clinical prediction models require large amounts of multicentre data as well as external validation. Before embarking on said multicentre data collection, piloting a modelling strategy to assess feasibility and to identify the most valuable inputs is crucial. Consequently, this pilot study with data from a single centre aims to assess whether AI can be predicted from simple patient-related risk factors.

**Methods**

**Overview and Data collection**

Data was extracted retrospectively from the patient registry of the University Hospital of Basel. Patients who underwent colon anastomosis for various reasons including neoplasia, diverticulitis, ischaemia, iatrogenic or traumatic perforation, or inflammatory bowel disease between 1st of January 2012 and 31st of December 2019 were eligible. This study was completed based on the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) statement checklist for the development of clinical prediction models. Utilizing the aforementioned data, we developed ML models with the aim of predicting AI, and internally validated the models on held-out test data.

**Ethical Considerations**

Ethical approval was provided for this study by the Northwestern and Central Ethics Committee Switzerland (BASEC-Nr 2020-02265).

**Predictors and Outcome Measures**

AI was defined according to Gessler et al. and Rahbari et al. as any clinical sign of leakage, confirmed by radiological examination, endoscopy, clinical examination of the anastomosis, or upon reoperation.
Recorded variables included risk factors that already have been reported in literature such as age, gender, BMI, active smoking, alcohol abuse (>2 alcoholic beverages per day), prior abdominal surgery, preoperative leucocytosis (≥10,000 per mm$^3$), preoperative steroid use, CCI, ASA Score, renal function (CKD Stages G1 to G5), albumin (g/dl), and haemoglobin level (g/dl), liver metastasis, indication, type of surgery, emergency surgery, surgical approach (laparoscopic or open), anastomotic technique (hand-sewn or stapler) and defunctioning ileostomy.

**Model Development**

Continuous variables and categorical variables were recorded as mean ± standard deviation (SD). Moreover, categorical variables as decimal numbers instead of percentages. Python 3.6.9 was used to perform all analyses. The Keras library was used for the artificial neural networks, the XGBoost library was used for XGBs, and the scikit-learn library was used for all other architectures.

Data was randomly split into two sets: 80% of the data was used as training data, while the remaining 20% was utilized as a test set to validate the models internally. Initially, the two sets had similar class distribution. However, the training set’s minority class (patients with AI) was oversampled using random oversampling until a 50:50 class distribution was achieved to prevent the models from overpredicting the majority class due to class imbalance. Recursive feature elimination was applied to determine the optimum set of relevant features predicting the outcome in the training data set.

A k-nearest neighbor (KNN) imputer was co-trained on the pre-oversampling training data to impute any missing data. We trialled multiple model architectures, including generalised linear models (GLMs), support vector machines (SVMs), naïve Bayes classifiers, random forests, extreme gradient boosting machines (XGB), and artificial neural networks. Training occurred in repeated 5-fold cross validation with 10 repetitions to select the optimal model architecture and hyperparameters based on the area under the
receiver operating characteristics curve (AUC). Recalibration was applied using Platt’s method based on the pre-oversampling training set, and the cut-off for binary classification was selected based on AUC on the training set using the closest-to-(0,1)-criterion.

The models were subsequently internally validated using the held-out test set, with ninety-five percent confidence intervals (CIs) being generated using 1000 bootstrap repetitions.

**Results**

**Cohort**

In the training set, a total of 474 patients were included in the training set, of which 77 (16.2%) suffered from AI. The mean age was 64.6 ± 15.4. Furthermore, 225 (47.5%) patients were male and 249 (52.5%) were female. The mean BMI was 25.1 ± 5.3 and 152 (32.1%) patients were active smokers. **Table 1** provides an overview of the cohorts.

As for the test set, a total of 119 patients were included, of which 11 (9.2%) suffered from AI. Furthermore, 58 (48.8%) patients were male, and 61 (51.2%) were female. The mean BMI in the test set was 25.0 ± 5.1 and 30 (25.2%) patients were active smokers.

**Model Performance**

The performance of all six models was assessed by internal validation using the test data set. This performance is reported in detail along with resampled training performance in **Table 2**. The random forest provided the highest AUC with a value of 0.61 (95% CI: 0.44 – 0.79) while also producing acceptable probabilities based on the calibration slope [0.16 (95% CI: -0.06 – 0.39)], and intercept [0.06 (95% CI: 0.02 - 0.11)]. Moreover, the random forest exhibited a sensitivity of 0.36 (95% CI: 0.08 – 0.67), specificity of 0.67 (95% CI: 0.58 – 0.76) and accuracy of 0.64 (95% CI: 0.55 – 0.72). Specific feature importance within the
random forest is displayed in Table 3. The calibration and AUC curves for the random forest classifier are displayed in Figures 1 and 2.

**Discussion**

In a pilot study using data from a single reference centre for colorectal surgery, we assess the feasibility of predicting AI accurately from simple data using ML techniques. Our findings demonstrate that modelling AI to a certain extent is feasible and identifies the most important input variables, laying the groundwork for more extensive multicentre modelling and external validation.

Even though a plethora of studies have analysed risk factors for AI, up to this day, no reliable clinical prediction model for AI has been established. To the best of the authors’ knowledge, this study is the first one attempting to solve the classification problem of AI (‘will my patient suffer from AI after colorectal surgery?’) by application of ML algorithms.

There is a widespread misunderstanding that variable importance measures gleaned from clinical prediction models can discover correlations and causalities in the same way that explanatory modelling does (prediction versus explanation). Indeed, this common misconceptualisation exists because predictive and explanatory modelling are often not as explicitly distinguished as attempted here in this study. Indeed, the interchangeable use of the concepts of in-sample correlation and out-of-sample generalisation can lead to false clinical decision making. While those variables identified as having high feature importance in this study may indeed be the most crucial ones for precise and generalisable prediction of AI, it cannot safely be concluded that these variables are necessarily also important independent risk factors for AI in their own right.
Another separate question is the initial choice of input variables for clinical prediction modelling, which can be achieved in various ways: In any case, a balance between performance through the inclusion of many variables and between the goal of arriving at parsimonious models that truly generalise needs to be struck. The choice of variables for this study was focused on patient-related risk factors to minimize the statistical noise from differing standard procedures in distinct clinical centres. The following section aims to clarify why the risk factors chosen have not only been correlated to AI previously but are also part of plausible pathomechanisms.

The CCI and the ASA classification system are two commonly used scoring systems assessing the patients’ pre-anaesthesia medical comorbidities and expected mortality risk, respectively. Even though the ability to predict peri- and postoperative complications is claimed for both scores, the ASA score and the CCI give valuable insight into the patients’ general medical conditions and thus can be used to generate more specific scoring systems. Indeed, the CCI and ASA score represent the most important features in our final model.

Serum levels of albumin and other hepatic proteins have been established as easily obtainable laboratory parameters for malnourishment. Since malnourishment itself is a risk factor for impaired wound healing in general, it is conceivable that hypoalbuminemia could be factor predisposing patients for AI. In fact, hypoalbuminemia has already been linked to increased postoperative morbidity and mortality in patients with colon cancer.

The most important pathophysiological basis for AI is believed to be gastrointestinal ischaemia, decreasing the amount of oxygen available to the recovering intestinal wall. Consequently, the intestinal tissue becomes more dysfunctional, which may lead to delayed wound healing. These suspected pathophysiological processes align well with the factors determined relevant for the prediction of AI risk.
in our model. In a similar way, haemoglobin levels determine the blood’s capacity to supply peripheral tissues with oxygen\textsuperscript{57,58}.

Smoking has been consistently evaluated as a risk factor for AI\textsuperscript{15,31}. Furthermore, it is well established that tobacco smoking is among the top risk factors for vascular disease\textsuperscript{59}, making it a logical contributor to vascular insufficiency leading to subsequently impaired wound healing and postoperative AI risk.

Our model seems to be more specific than sensitive. This indicates that our model is proficient at ruling in high-risk AI with a positive prediction. Conversely, our model is less precise in ruling out low-risk AI. However, a rule-in model could prove to be of great value for clinicians by simply identifying the very high-risk group. Furthermore, the random forest model seems fairly well-calibrated. Well-calibrated predicted probabilities are arguably more important in clinical practice (‘How likely is it that I am going to experience AI?’ – ‘Your probability is 17%.’) instead of binary predictions (‘Am I going to suffer from AI?’ – ‘The model predicts yes/no’). Physicians are experts at dealing with uncertainty and risks, and probabilities are thus more appreciated by patients and physicians than a mere yes or no answer – apart from the fact that patients are never binary but instead represent a spectrum of risk\textsuperscript{60}.

Another difficulty in clinical prediction modelling is choosing the appropriate sample size. According to a common rule of thumb, there should be at least 10 minority class observations in a dataset per feature\textsuperscript{61}. This study relies on 10 patient-related risk factors, thus making a total number of 1000 patients with AI who would be necessary for training, while we finally were able to include 88 patients with AI. Other architectures such as random forests, artificial neural networks, and SVMs seem to require much more data per feature\textsuperscript{62}. Therefore, it is conceivable that including more patients might further refine the model. Consequently, it is crucial to recruit more patient data from other hospitals in a multicentre study.
Clinical prediction models can facilitate assessing individual risks and making more informed decisions based on predictive analytics that are tailored to each patient. However, especially in colorectal surgery, the indication for surgery is only rarely truly elective. Therefore, a prediction model can only help decide whether an intervention should be postponed to improve the risk profile or, especially for emergency interventions, whether a patient would benefit from a diverting stoma to minimize and modify risk factors before re-joining the colon. On the other hand, a comprehensive predictive model may also increase a patient’s acceptance of the primary placement of a protective stoma. Thus, such a model could potentially also help to improve the physician-patient relationship through enhanced patient education.

The results of a predictive model cannot be seen as a clear recommendation pro or contra an intervention as the risk profile it mirrors is tailored only to a specific endpoint and thus does not entirely reflect the patient’s global situation. Indeed, components of decision-making such as the psychological distress of a patient with chronic diverticulitis are not included in the model and have a decisive influence on the indication. Consequently, prediction models should be seen only as adjunctive information to be used in a complementary way for informed shared decision-making. Nevertheless, the necessity for evidence-based clinical prediction models becomes clear when considering the relative inability of even experienced clinicians in predicting clinical outcomes\(^\text{38}\), while the ethical implications of an ‘AI doctor’ technology independent from human control have to be taken into account, too\(^\text{63}\). Consequently, ML-based clinical prediction models could be deemed a contemporary optimal trade-off between the clinical experience of human experts and the exploitation of big data by learning algorithms.

**Limitations**

The relatively high AI rate of 14.8% in our cohort can be seen as a limitation. Similarly, the difference in AI incidence among training versus test set represents an additional hurdle that is realistic, as AI rate is described inconsistently in the literature\(^\text{4,6-10}\). The patient population at the University Hospital Basel with
32.9% emergency cases and a cohort that includes transplanted and immunosuppressed patients is expected to have higher complication rates. Nevertheless, such a difference to other hospitals should be reflected in the ASA score, the CCI, and blood values and thus also in our results. By including patient data from other institutions in future analyses, this number will be balanced out, and a differentiated breakdown according to emergency interventions, immunosuppressant use, previous radio-/chemotherapy, and cancer diagnosis, which additionally reflect a patient’s health status, is conceivable and could be implemented in our ML-algorithm.

One further caveat of any model is the danger of overfitting. In the context of clinical prediction modelling, overfitting means that an algorithm adheres too strictly to the training data, especially its inherent variance and possible noise factors (e.g. noise generated by a hospital’s standardized procedures). With enough training, the algorithm will perform extremely well on the training data while losing its generalisation capability towards new data from other centres. Indeed, it is not unlikely that this study might suffer from slight overfitting due to standardized hospital procedures because the data was exclusively collected in one clinical centre. However, this weakness could be addressed by recruiting more patient data from other hospitals. Lastly, we have purposefully not yet deployed our model for clinical application in, e.g. a web-app, as any single-centre and not yet externally validated clinical prediction model is not yet recommended for clinical use.

Conclusion

In this pilot study, by using 10 patient-related risk factors associated with AI, we demonstrate the feasibility of ML-based prediction of AI after colorectal surgery. Nevertheless, it is crucial to include multicentre data and higher sample sizes to develop a robust and generalisable model, which will subsequently allow for deployment of the algorithm in a web-based application.
References


### TABLE 1. Overview of patient characteristics including missingness.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Missing</th>
<th>Total n = 593</th>
<th>Training set n = 474</th>
<th>Test set n = 119</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>0%</td>
<td>64.6 ± 15.4</td>
<td>64.6 ± 15.5</td>
<td>64 ± 15.3</td>
</tr>
<tr>
<td>Gender</td>
<td>0%</td>
<td>Female - 310 (52.3%)</td>
<td>Female - 249 (52.5%)</td>
<td>Female - 61 (51.2%)</td>
</tr>
<tr>
<td>BMI (kg/m^2)</td>
<td>1.69%</td>
<td>25.1 ± 5.3</td>
<td>25.1 ± 5.3</td>
<td>25.0 ± 5.1</td>
</tr>
<tr>
<td>Active smoking</td>
<td>0%</td>
<td>Yes – 182 (30.7%)</td>
<td>Yes – 152 (32.1%)</td>
<td>Yes – 30 (25.2%)</td>
</tr>
<tr>
<td>Alcohol abuse</td>
<td>0%</td>
<td>Yes – 83 (14.0%)</td>
<td>Yes – 64 (13.5%)</td>
<td>Yes – 19 (16%)</td>
</tr>
<tr>
<td>CCI</td>
<td>0%</td>
<td>4.9 ± 3.7</td>
<td>4.9 ± 3.8</td>
<td>4.9 ± 3.5</td>
</tr>
<tr>
<td>ASA</td>
<td>0.51%</td>
<td>1 – 10 (1.6%)</td>
<td>1 – 7 (1.6%)</td>
<td>1 – 3 (2.6%)</td>
</tr>
<tr>
<td>Renal function</td>
<td>0.17%</td>
<td>G1 – 225 (38.0%)</td>
<td>G1 – 183 (38.2%)</td>
<td>G1 – 42 (36.5%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G2 – 201 (34.0%)</td>
<td>G2 – 163 (34.4%)</td>
<td>G2 – 38 (33.0%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G3 – 145 (24.5%)</td>
<td>G3 – 110 (23.2%)</td>
<td>G3 – 35 (30.4%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G4 – 20 (3.4%)</td>
<td>G4 – 17 (3.5%)</td>
<td>G4 – 3 (2.6%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G5 – 1 (0.2%)</td>
<td>G5 – 1 (0.2%)</td>
<td>G5 – 0 (0%)</td>
</tr>
<tr>
<td>Albumin (g/dl)</td>
<td>1.35%</td>
<td>3.2 ± 0.7</td>
<td>3.2 ± 0.7</td>
<td>3.2 ± 0.6</td>
</tr>
<tr>
<td>Haemoglobin (g/dl)</td>
<td>0.34%</td>
<td>12.9 ± 8.7</td>
<td>12.9 ± 8.3</td>
<td>13.2 ± 10.0</td>
</tr>
<tr>
<td>Leukocytosis (≥10,000 per mm^3)</td>
<td>0.34%</td>
<td>Yes – 171 (28.9%)</td>
<td>Yes – 135 (29.5%)</td>
<td>Yes – 36 (30.5%)</td>
</tr>
<tr>
<td>Preoperative steroid use</td>
<td>0%</td>
<td>Yes – 41 (6.9%)</td>
<td>Yes – 34 (7.2%)</td>
<td>Yes – 7 (5.9%)</td>
</tr>
<tr>
<td>Prior abdominal surgery</td>
<td>0%</td>
<td>Yes – 322 (54.7%)</td>
<td>Yes – 253 (53.4%)</td>
<td>Yes – 69 (58.0%)</td>
</tr>
<tr>
<td>Liver metastasis</td>
<td>0.34%</td>
<td>Yes – 37 (6.3%)</td>
<td>Yes – 31 (6.6%)</td>
<td>Yes – 6 (0.5%)</td>
</tr>
<tr>
<td>Indication</td>
<td>0%</td>
<td>tumour – 270 (45.5%)</td>
<td>tumour – 216 (45.6%)</td>
<td>tumour – 54 (45.4%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IBD – 25 (4.2%)</td>
<td>IBD – 20 (4.2%)</td>
<td>IBD – 5 (4.2%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>diverticular disease – 151 (25.5%)</td>
<td>diverticular disease – 119 (25.1%)</td>
<td>diverticular disease – 32 (26.9%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ischemia/ileus – 49 (8.3%)</td>
<td>ischemia/ileus – 43 (9.1%)</td>
<td>ischemia/ileus – 6 (5.0%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>other – 98 (16.5%)</td>
<td>other – 76 (16.0%)</td>
<td>other – 22 (18.5%)</td>
</tr>
<tr>
<td>Perforation</td>
<td>0.34%</td>
<td>Yes – 71 (12.0%)</td>
<td>Yes – 68 (14.3%)</td>
<td>Yes – 13 (1.3%)</td>
</tr>
<tr>
<td>Emergency surgery</td>
<td>0%</td>
<td>Yes – 195 (32.9%)</td>
<td>Yes – 160 (33.8%)</td>
<td>Yes – 35 (29.4%)</td>
</tr>
<tr>
<td>Surgical procedure</td>
<td>0%</td>
<td>(extended) left hemicolectomy – 57 (9.6%)</td>
<td>(extended) left hemicolectomy – 47 (9.9%)</td>
<td>(extended) left hemicolectomy – 10 (8.4%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(extended) right hemicolectomy – 178 (30.0%)</td>
<td>(extended) right hemicolectomy – 149 (31.4%)</td>
<td>(extended) right hemicolectomy – 29 (24.4%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ileocecal resection – 84 (14.2%)</td>
<td>ileocecal resection – 67 (14.1%)</td>
<td>ileocecal resection – 17 (14.3%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>transverse colectomy – 19 (3.2%)</td>
<td>transverse colectomy – 16 (3.4%)</td>
<td>transverse colectomy – 3 (2.5%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>rectosigmoid resection/sigmoidectomy – 211 (35.6%)</td>
<td>rectosigmoid resection/sigmoidectomy – 160 (33.8%)</td>
<td>rectosigmoid resection/sigmoidectomy – 51 (42.9%)</td>
</tr>
<tr>
<td>Procedure</td>
<td>Group 1</td>
<td>Group 2</td>
<td>Group 3</td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>Hartmann’s reversal or reversal of colostomy</td>
<td>44 (7.4%)</td>
<td>35 (7.4%)</td>
<td>9 (7.6%)</td>
<td></td>
</tr>
<tr>
<td>Laparoscopy</td>
<td>0.17%</td>
<td>Yes – 227 (38.3%)</td>
<td>Yes – 179 (37.8%)</td>
<td>Yes – 48 (40.7%)</td>
</tr>
<tr>
<td>Technique (hand-sewn)</td>
<td>0.34%</td>
<td>Yes – 247 (41.8%)</td>
<td>Yes – 209 (44.3%)</td>
<td>Yes – 38 (31.9%)</td>
</tr>
<tr>
<td>Defunctioning ileostomy</td>
<td>0%</td>
<td>Yes – 64 (10.8%)</td>
<td>Yes – 52 (11.0%)</td>
<td>Yes – 12 (10.1%)</td>
</tr>
<tr>
<td>Anastomotic insufficiency</td>
<td>0%</td>
<td>Yes – 88 (14.8%)</td>
<td>Yes – 77 (16.2%)</td>
<td>Yes – 11 (9.2%)</td>
</tr>
</tbody>
</table>

Absolute numbers and percentages for categorical or mean ± SD for continuous variables are presented. BMI body mass index, CCI Charlson Comorbidity Index, ASA American Society of Anesthesiologists score, IBD inflammatory bowel disease.
**TABLE 2.** Performance evaluation of the final random forest classifier. Resampled training performance and test set performance are reported.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Training Set</th>
<th>Test Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUC</td>
<td>0.87 (0.77-0.95)</td>
<td>0.61 (0.44-0.79)</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.83 (0.76-0.9)</td>
<td>0.64 (0.55-0.72)</td>
</tr>
<tr>
<td>Recall</td>
<td>0.74 (0.53-0.92)</td>
<td>0.36 (0.08-0.67)</td>
</tr>
<tr>
<td>Precision</td>
<td>0.48 (0.3-0.67)</td>
<td>0.1 (0.02-0.2)</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.85 (0.78-0.91)</td>
<td>0.67 (0.58-0.76)</td>
</tr>
<tr>
<td>Brier</td>
<td>0.09 (0.05-0.12)</td>
<td>0.11 (0.08-0.15)</td>
</tr>
<tr>
<td>Calibration slope</td>
<td>1.01 (0.74-1.24)</td>
<td>0.16 (-0.06-0.39)</td>
</tr>
<tr>
<td>Calibration intercept</td>
<td>-0.01 (-0.04-0.03)</td>
<td>0.06 (0.02-0.11)</td>
</tr>
</tbody>
</table>

*AUC* area under recall-precision curve
TABLE 3. Predictor importance according to random forest classifier.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMI</td>
<td>0.17</td>
</tr>
<tr>
<td>CCI</td>
<td>0.15</td>
</tr>
<tr>
<td>Albumin</td>
<td>0.14</td>
</tr>
<tr>
<td>Active smoking</td>
<td>0.12</td>
</tr>
<tr>
<td>Haemoglobin</td>
<td>0.11</td>
</tr>
<tr>
<td>Age</td>
<td>0.11</td>
</tr>
<tr>
<td>ASA</td>
<td>0.1</td>
</tr>
<tr>
<td>Renal function</td>
<td>0.05</td>
</tr>
<tr>
<td>Type of surgery</td>
<td>0.04</td>
</tr>
<tr>
<td>Indication</td>
<td>0.02</td>
</tr>
</tbody>
</table>

*BMI* body mass index, *CCI* Charlson Comorbidity Index, *ASA* American Society of Anesthesiologists score
FIGURE 1. Resampled training (a) and test (b) set calibration curves of the random forest classifier.

Figure 1

a

b
FIGURE 2. Resampled training (a) and test (b) set area under the receiver operating characteristics curves (AUC) of the random forest classifier.

Figure 2