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

Stephanie Taha-Mehlitz[1], MD; Larissa Wentzler[2], BMed; Fiorenzo Angehrn[1], MD; Ahmad Hendie[3]; Vincent Ochs[4]; Victor E. Staartjes[5], MD; Markus von Flüe[1], MD; Anas Taha[6]*, MD; Daniel Steinemann[1]*, MD

1. Clarunis, University Centre for Gastrointestinal and Liver Diseases, St. Clara Hospital and University Hospital, Basel, Switzerland
2. Medical Faculty, University Basel, Basel, Switzerland
3. Department of Computer Engineering, McGill University, Montreal, 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, Allschwil, Switzerland

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

Abstract

Background: 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: 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, gender, BMI, smoking status, alcohol abuse, prior abdominal surgery, leukocytosis, haemoglobin and albumin levels, steroid use, the Charlson Comorbidity Index, the American Society of Anesthesiologists score, 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 0.64 (95% confidence interval [CI]: 0.44-0.82), calibration slope was 0.21 (95% CI: -0.02-0.46) and calibration intercept was 0.06 (95% CI: 0.01-0.1). We observed a specificity of 0.76 (95% CI: 0.68-0.84), sensitivity of 0.36 (95% CI: 0.08-0.7), and accuracy of 0.72 (95% CI: 0.65-0.8).

Conclusion: By using 13 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.

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 external funding.

Author Approval: All authors have seen and approved the manuscript.

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.

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 [1]. According to Rahbari et al. [2], 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 [3–5]. The approximated incidence of AI is 3.3% after colon anastomosis and 8.6% after colorectal anastomosis in specialised centres [4]. 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 [6–10]. 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 [1,2].

In previous publications, a multitude of risk factors for AI has been identified [11,12]. Among these risk factors, age [13], gender [7,14–19], body mass index (BMI) [20,21], Charlson Comorbidity Index (CCI) [14,22], American Society of Anesthesiologists (ASA) score [13,21–24], leukocytosis [23], anaemia [25], hypoalbuminemia [15,26,27], steroid use [28,29], renal function [23], previous abdominal surgery [16], active smoking [6,30,31], and alcohol abuse [6,32] 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 [33]. 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 2013 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 [34]. 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. [35] and Rahbari et al. [2] 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).

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 [36] was used for the artificial neural networks, the XGBoost library [37] was used for XGBs, and the scikit-learn library [38] 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 [39,40].

A *k*-nearest neighbor (KNN) imputer was co-trained on the pre-oversampling training data to impute any missing data [41]. 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 [42] 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 [43].

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. 152 (32.1%) patients were active smokers, and 64 (13.5.0%) suffered from alcohol abuse. **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. 30 (25.2%) patients were active smokers, and 19 (16.0%) suffered from alcohol abuse.

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 0.64 (95% CI: 0.44 – 0.82) while also producing acceptable probabilities based on the calibration slope [0.21 (95% CI: -0.02 – 0.46)], and intercept [0.06 (95% CI: 0.01 - 0.1)]. Moreover, the random forest exhibited a sensitivity of 0.36 (95% CI: 0.08 – 0.7), specificity of 0.76 (95% CI: 0.68 – 0.84) and accuracy of 0.72 (95% CI: 0.65 – 0.8). 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 [11–32]. 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) [44]. 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 [45]. 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 [46]: 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 [47,48]. 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 [49]. Since malnourishment itself is a risk factor for impaired wound healing in general [50], 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 [27,51].

The most important pathophysiological basis for AI is believed to be gastrointestinal ischemia, decreasing the amount of oxygen available to the recovering intestinal wall [52]. 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 [53,54].

Smoking has been consistently evaluated as a risk factor for AI [15,31]. Furthermore, it is well established that tobacco smoking is among the top risk factors for vascular disease [55], making it a logical contributor to vascular insufficiency leading to subsequently impaired wound healing and postoperative AI risk.

Similarly, the feature importance for alcohol consumption corresponds to data from previous studies identifying daily alcohol consumption as a risk factor for AI [6,32]. Moreover, alcohol is a well-established negative modulator of wound healing processes that inhibits re-epithelialisation, angiogenesis, collagen production, and wound closure [50]. In addition, steroids have been established to impede wound healing by inhibiting the formation of adequate granulation tissue [56] and the production of hypoxia-inducible factor-1 [57].

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 arguable 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 [58].

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 [59]. This study relies on 13 patient-related risk factors, thus making a total number of 130 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 [60]. 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 make 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 [33], while the ethical implications of an 'AI doctor' technology independent from human control have to be taken into account, too [61]. 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 [4,6–10]. The patient population at the University Hospital Basel with 32.8% emergency cases and a cohort that includes transplanted and immunosuppressed patients is expected to have higher complication rates [62]. 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 [63].

Conclusion

In this pilot study, by using 13 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

1. Lee SW, Gregory D, Cool CL. Clinical and economic burden of colorectal and bariatric anastomotic leaks. Surg Endosc. 2020 Oct;34(10):4374–81.

2. Rahbari NN, Weitz J, Hohenberger W, Heald RJ, Moran B, Ulrich A, et al. Definition and grading of anastomotic leakage following anterior resection of the rectum: a proposal by the International Study Group of Rectal Cancer. Surgery. 2010 Mar;147(3):339–51.

3. Mirnezami A, Mirnezami R, Chandrakumaran K, Sasapu K, Sagar P, Finan P. Increased local recurrence and reduced survival from colorectal cancer following anastomotic leak: systematic review and meta-analysis. Ann Surg. 2011 May;253(5):890–9.

4. Krell RW, Girotti ME, Fritze D, Campbell DA, Hendren S. Hospital readmissions after colectomy: a population-based study. J Am Coll Surg. 2013 Dec;217(6):1070–9.

5. Hammond J, Lim S, Wan Y, Gao X, Patkar A. The burden of gastrointestinal anastomotic leaks: an evaluation of clinical and economic outcomes. J Gastrointest Surg. 2014 Jun;18(6):1176–85.

6. Sørensen LT, Jørgensen T, Kirkeby LT, Skovdal J, Vennits B, Wille-Jørgensen P. Smoking and alcohol abuse are major risk factors for anastomotic leakage in colorectal surgery. Br J Surg. 1999 Jul;86(7):927–31.

7. Law WI, Chu KW, Ho JW, Chan CW. Risk factors for anastomotic leakage after low anterior resection with total mesorectal excision. Am J Surg. 2000 Feb;179(2):92–6.

8. Kang CY, Halabi WJ, Chaudhry OO, Nguyen V, Pigazzi A, Carmichael JC, et al. Risk factors for anastomotic leakage after anterior resection for rectal cancer. JAMA Surg. 2013 Jan;148(1):65–71.

9. Eberl T, Jagoditsch M, Klingler A, Tschmelitsch J. Risk factors for anastomotic leakage after resection for rectal cancer. Am J Surg. 2008 Oct;196(4):592–8.

10. Kruschewski M, Rieger H, Pohlen U, Hotz HG, Buhr HJ. Risk factors for clinical anastomotic leakage and postoperative mortality in elective surgery for rectal cancer. Int J Colorectal Dis. 2007 Aug;22(8):919–27.

11. Komen N, Dijk J-W, Lalmahomed Z, Klop K, Hop W, Kleinrensink G-J, et al. After-hours colorectal surgery: a risk factor for anastomotic leakage. Int J Colorectal Dis. 2009 Jul;24(7):789–95.

12. Kryzauskas M, Bausys A, Degutyte AE, Abeciunas V, Poskus E, Bausys R, et al. Risk factors for anastomotic leakage and its impact on long-term survival in left-sided colorectal cancer surgery. World J Surg Oncol. 2020 Aug 14;18(1):205.

13. Bakker IS, Grossmann I, Henneman D, Havenga K, Wiggers T. Risk factors for anastomotic leakage and leak-related mortality after colonic cancer surgery in a nationwide audit. Br J Surg. 2014 Mar;101(4):424–32; discussion 432.

14. Trencheva K, Morrissey KP, Wells M, Mancuso CA, Lee SW, Sonoda T, et al. Identifying important predictors for anastomotic leak after colon and rectal resection: prospective study on 616 patients. Ann Surg. 2013 Jan;257(1):108–13.

15. Suding P, Jensen E, Abramson MA, Itani K, Wilson SE. Definitive risk factors for anastomotic leaks in elective open colorectal resection. Arch Surg. 2008 Sep;143(9):907–11; discussion 911–2.

16. Lipska MA, Bissett IP, Parry BR, Merrie AEH. Anastomotic leakage after lower gastrointestinal anastomosis: men are at a higher risk. ANZ J Surg. 2006 Jul;76(7):579–85.

17. Park JS, Choi G-S, Kim SH, Kim HR, Kim NK, Lee KY, et al. Multicenter analysis of risk factors for anastomotic leakage after laparoscopic rectal cancer excision: the Korean laparoscopic colorectal surgery study group. Ann Surg. 2013 Apr;257(4):665–71.

18. Hamabe A, Ito M, Nishigori H, Nishizawa Y, Sasaki T. Preventive effect of diverting stoma on anastomotic leakage after laparoscopic low anterior resection with double stapling technique reconstruction applied based on risk stratification. Asian J Endosc Surg. 2018 Aug;11(3):220–6.

19. Tanaka K, Okuda J, Yamamoto S, Ito M, Sakamoto K, Kokuba Y, et al. Risk factors for anastomotic leakage after laparoscopic surgery with the double stapling technique for stage 0/I rectal carcinoma: a subgroup analysis of a multicenter, single-arm phase II trial. Surg Today. 2017 Oct;47(10):1215–22.

20. Silva-Velazco J, Stocchi L, Costedio M, Gorgun E, Kessler H, Remzi FH. Is there anything we can modify among factors associated with morbidity following elective laparoscopic sigmoidectomy for diverticulitis? Surg Endosc. 2016 Aug;30(8):3541–51.

21. Buchs NC, Gervaz P, Secic M, Bucher P, Mugnier-Konrad B, Morel P. Incidence, consequences, and risk factors for anastomotic dehiscence after colorectal surgery: a prospective monocentric study. Int J Colorectal Dis. 2008 Mar;23(3):265–70.

22. Tan WP, Talbott VA, Leong QQ, Isenberg GA, Goldstein SD. American Society of Anesthesiologists class and Charlson's Comorbidity index as predictors of postoperative colorectal anastomotic leak: a single-institution experience. J Surg Res. 2013 Sep 1;184(1):115–9.

23. Alves A, Panis Y, Trancart D, Regimbeau J-M, Pocard M, Valleur P. Factors associated with clinically significant anastomotic leakage after large bowel resection: multivariate analysis of 707 patients. World J Surg. 2002 Apr;26(4):499–502.

24. Choi H-K, Law W-L, Ho JWC. Leakage after resection and intraperitoneal anastomosis for colorectal malignancy: analysis of risk factors. Dis Colon Rectum. 2006 Nov;49(11):1719–25.

25. Hayden DM, Mora Pinzon MC, Francescatti AB, Saclarides TJ. Patient factors may predict anastomotic complications after rectal cancer surgery: Anastomotic complications in rectal cancer. Ann Med Surg (Lond). 2015 Mar;4(1):11–6.

26. Ionescu D, Tibrea C, Puia C. Pre-operative hypoalbuminemia in colorectal cancer patients undergoing elective surgery—a major risk factor for postoperative outcome. Chirurgia . 2013;108(6):822–8.

27. Lai C-C, You J-F, Yeh C-Y, Chen J-S, Tang R, Wang J-Y, et al. Low preoperative serum albumin in colon cancer: a risk factor for poor outcome. Int J Colorectal Dis. 2011 Apr;26(4):473–81.

28. Slieker JC, Komen N, Mannaerts GH, Karsten TM, Willemsen P, Murawska M, et al. Long-term and perioperative corticosteroids in anastomotic leakage: a prospective study of 259 left-sided colorectal anastomoses. Arch Surg. 2012 May;147(5):447–52.

29. Konishi T, Watanabe T, Kishimoto J, Nagawa H. Risk factors for anastomotic leakage after surgery for colorectal cancer: results of prospective surveillance. J Am Coll Surg. 2006 Mar;202(3):439–44.

30. Kim MJ, Shin R, Oh H-K, Park JW, Jeong S-Y, Park J-G. The impact of heavy smoking on anastomotic leakage and stricture after low anterior resection in rectal cancer patients. World J Surg. 2011 Dec;35(12):2806–10.

31. Kwak HD, Kim S-H, Kang DW, Baek S-J, Kwak JM, Kim J. Risk Factors and Oncologic Outcomes of Anastomosis Leakage After Laparoscopic Right Colectomy. Surg Laparosc Endosc Percutan Tech. 2017 Dec;27(6):440–4.

32. Mäkelä JT, Kiviniemi H, Laitinen S. Risk factors for anastomotic leakage after left-sided colorectal resection with rectal anastomosis. Dis Colon Rectum. 2003 May;46(5):653–60.

33. Karliczek A, Harlaar NJ, Zeebregts CJ, Wiggers T, Baas PC, van Dam GM. Surgeons lack predictive accuracy for anastomotic leakage in gastrointestinal surgery. Int J Colorectal Dis. 2009 May;24(5):569–76.

34. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD Statement. Br J Surg. 2015 Feb;102(3):148–58.

35. Gessler B, Eriksson O, Angenete E. Diagnosis, treatment, and consequences of anastomotic leakage in colorectal surgery. Int J Colorectal Dis. 2017 Apr;32(4):549–56.

36. Chollet F. others,"Keras", GitHub. 2015.

37. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: Association for Computing Machinery; 2016. p. 785–94. (KDD '16).

38. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. the Journal of machine Learning research. 2011;12:2825–30.

39. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. J Artif Intell Res. 2002 Jun 1;16:321–57.

40. Staartjes VE, Schröder ML. Letter to the Editor. Class imbalance in machine learning for neurosurgical outcome prediction: are our models valid? J Neurosurg Spine. 2018 Nov 1;29(5):611–2.

41. Templ M, Alfons A, Kowarik A, Prantner B. VIM: Visualization and Imputation of Missing Values, 2012. R package version. 2019;1(0).

42. Staartjes VE, Regli L, Serra C, editors. Machine learning in clinical neuroscience: Foundations and applications. 1st ed. Cham, Switzerland: Springer Nature; 2021. 355 p. (Acta Neurochirurgica Supplement).

43. Perkins NJ, Schisterman EF. The Inconsistency of "Optimal" Cutpoints Obtained using Two Criteria based on the Receiver Operating Characteristic Curve [Internet]. Vol. 163, American Journal of Epidemiology. 2006. p. 670–5. Available from: http://dx.doi.org/10.1093/aje/kwj063

44. Shmueli G. To explain or to predict? Stat Sci. 2010 Aug 1;25(3):289–310.

45. Yarkoni T, Westfall J. Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. Perspect Psychol Sci. 2017 Nov;12(6):1100–22.

46. Chowdhury MZI, Turin TC. Variable selection strategies and its importance in clinical prediction modelling. Family Medicine and Community Health. 2020 Feb 1;8(1):e000262.

47. Sparreboom CL, van Groningen JT, Lingsma HF, Wouters MWJM, Menon AG, Kleinrensink G-J, et al. Different Risk Factors for Early and Late Colorectal Anastomotic Leakage in a Nationwide Audit. Dis Colon Rectum. 2018 Nov;61(11):1258–66.

48. Froehner M, Koch R, Litz R, Heller A, Oehlschlaeger S, Wirth MP. Comparison of the American Society of Anesthesiologists Physical Status classification with the Charlson score as predictors of survival after radical prostatectomy [Internet]. Vol. 62, Urology. 2003. p. 698–701. Available from: http://dx.doi.org/10.1016/s0090-4295(03)00570-3

49. Fuhrman MP, Charney P, Mueller CM. Hepatic proteins and nutrition assessment. J Am Diet Assoc. 2004 Aug;104(8):1258–64.

50. Guo S, Dipietro LA. Factors affecting wound healing. J Dent Res. 2010 Mar;89(3):219–29.

51. Kwag S-J, Kim J-G, Kang W-K, Lee J-K, Oh S-T. The nutritional risk is a independent factor for postoperative morbidity in surgery for colorectal cancer. Ann Surg Treat Res. 2014 Apr;86(4):206–11.

52. Rutegård M, Rutegård J. Anastomotic leakage in rectal cancer surgery: The role of blood perfusion. World J Gastrointest Surg. 2015 Nov 27;7(11):289–92.

53. Yudin J, Verhovsek M. How we diagnose and manage altered oxygen affinity hemoglobin variants. Am J Hematol. 2019 May;94(5):597–603.

54. Shepherd JRA, Dominelli PB, Roy TK, Secomb TW, Hoyer JD, Oliveira JL, et al. Modelling the relationships between haemoglobin oxygen affinity and the oxygen cascade in humans. J Physiol. 2019 Aug;597(16):4193–202.

55. the Surgeon General O of, Others. The Health Consequences of Smoking-50 Years of Progress: A Report of the Surgeon General. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention (US). National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health https://www surgeongeneral gov/library/reports/50-years-of-progress/full-report pdf. 2014;

56. Franz MG, Steed DL, Robson MC. Optimizing Healing of the Acute Wound by Minimizing Complications [Internet]. Vol. 44, Current Problems in Surgery. 2007. p. 691–763. Available from: http://dx.doi.org/10.1067/j.cpsurg.2007.07.001

57. Wagner AE, Huck G, Stiehl DP, Jelkmann W, Hellwig-Bürgel T. Dexamethasone impairs hypoxia-inducible factor-1 function. Biochem Biophys Res Commun. 2008 Jul 25;372(2):336–40.

58. Staartjes VE, Kernbach JM. Letter to the Editor. Importance of calibration assessment in machine learning–based predictive analytics. J Neurosurg Spine. 2020 Feb 21;32(6):985–7.

59. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996 Dec 1;49(12):1373–9.

60. van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. BMC Med Res Methodol. 2014 Dec 22;14:137.

61. Keskinbora KH. Medical ethics considerations on artificial intelligence. J Clin Neurosci. 2019 Jun;64:277–82.

62. McDermott FD, Heeney A, Kelly ME, Steele RJ, Carlson GL, Winter DC. Systematic review of preoperative, intraoperative and postoperative risk factors for colorectal anastomotic leaks. Br J Surg. 2015 Apr;102(5):462–79.

63. Staartjes VE, Kernbach JM. Significance of external validation in clinical machine learning: let loose too early? Spine J. 2020 Jul;20(7):1159–60.

Tables and Figures

TABLE 1. Overview of patient characteristics including missingness.

Variable	Missing	Total n = 593	Training set n = 474	Test set n = 119
Age (years)	0%	64.6 ± 15.4	64.6 ± 15.5	64.4 ± 15.3
Gender	0%	Female - 310 (52.3%)	Female – 249 (52.5%)	Female – 61 (51.2%)
BMI (kg/m^2)	1.69%	25.1 ± 5.3	25.1 ± 5.3	25.0 ± 5.1
Active smoking	0%	Yes – 182 (30.7%)	Yes – 152 (32.1%)	Yes – 30 (25.2%)
Alcohol abuse	0%	Yes – 83 (14.0%)	Yes – 64 (13.5%)	Yes – 19 (16%)
Prior abdominal surgery	0%	Yes – 322 (54.7%)	Yes – 253 (53.4%)	Yes – 69 (58.0%)
Leukocytosis (≥10.000 per mm^3)	0.34%	Yes – 171 (28.9%)	Yes – 135 (29.5%)	Yes – 36 (30.5%)
Preoperative steroid use	0%	Yes – 41 (6.9%)	Yes – 34 (7.2%)	Yes – 7 (5.9%)
CCI	0%	4.9 ± 3.7	4.9 ± 3.8	4.9 ± 3.5
ASA	0.51%	1 – 10 (1.6%)	1 – 7 (1.6%)	1 – 3 (2.6%)
		2 – 173 (29.3%)	2 – 144 (29.3%)	2 – 29 (24.8%)
		3 – 343 (58.1%)	3 – 273 (58.1%)	3 – 70 (59.8%)
		4 – 61 (10.3%)	4 – 46 (10.3%)	4 – 15 (12.8%)
		5 – 3 (0.5%)	5 – 3 (0.5%)	5 – 0 (0%)
Renal function	0.17%	G1 – 225 (38.0%)	G1 – 183 (38.2%)	G1 – 42 (36.5%)
		G2 – 201 (34.0%)	G2 – 163 (34.4%)	G2 – 38 (33.0%)
		G3 – 145 (24.5%)	G3 – 110 (23.2%)	G3 – 35 (30.4%)
		G4 – 20 (3.4%)	G4 – 17 (3.5%)	G4 – 3 (2.6%)
		G5 – 1 (0.2%)	G5 – 1 (0.2%)	G5 – 0 (0%)
Albumin (g/dl)	1.35%	3.2 ± 0.7	3.2 ± 0.7	3.2 ± 0.6
Haemoglobin (g/dl)	0.34%	12.9 ± 8.7	12.9 ± 8.3	13.2 ± 10.0
Anastomotic insufficiency	0%	Yes – 88 (14.8%)	Yes – 77 (16.2%)	Yes – 11 (9.2%)

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.

TABLE 2. Performance evaluation of the final random forest classifier. Resampled training performance and test set performance are reported.

Metric	Training Set	Test Set
AUC	0.85 (0.75-0.93)	0.64 (0.44-0.82)
Accuracy	0.82 (0.74-0.88)	0.72 (0.65-0.8)
Recall	0.68 (0.47-0.88)	0.36 (0.08-0.7)
Precision	0.45 (0.27-0.64)	0.13 (0.03-0.26)
Specificity	0.77 (0.77-0.91)	0.76 (0.68-0.84)
Brier	0.09 (0.06-0.13)	0.1 (0.07-0.14)
Calibration Slope	1.0 (0.72-1.29)	0.21 (-0.02-0.46)
Calibration intercept	-0.01 (-0.02-0.07)	0.06 (0.01-0.1)

AUC area under recall-precision curve

TABLE 3. Predictor importance according to random forest classifier.

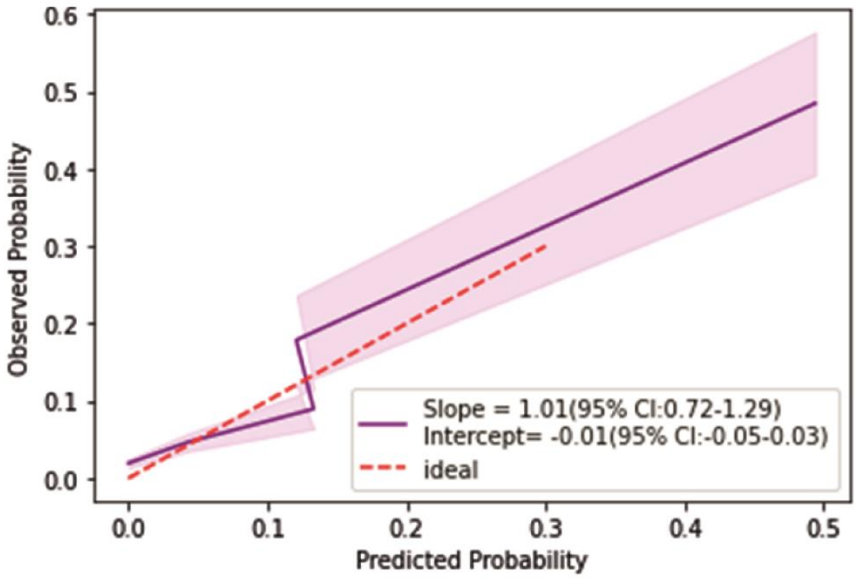
Feature	Importance (95% CI)
CCI	0.17 (0.09 - 0.25)
ASA	0.14 (0.05 - 0.24)
Haemoglobin	0.13 (0.06 - 0.2)
BMI	0.13 (0.06 - 0.23)
Albumin	0.11 (0.04 - 0.17)
Active Smoking	0.1 (0.03 - 0.2)
Age	0.07 (0.04 - 0.11)
Renal function	0.05 (0.01 - 0.1)
Alcohol Abuse	0.02 (0.0 - 0.06)
Leukocytosis	0.02 (0.0 - 0.06)
Gender	0.02 (0.0 - 0.07)
Steroids	0.02 (0.0 - 0.07)
Prior abdominal surgery	0.01 (0.0 - 0.02)

CI Confidence interval, *CCI* Charlson Comorbidity Index, *ASA* American Society of Anesthesiologists score, *BMI* body mass index

FIGURE 1. Resampled training (**a**) and test (**b**) set calibration curves of the random forest classifier.

Figure 1

a



b

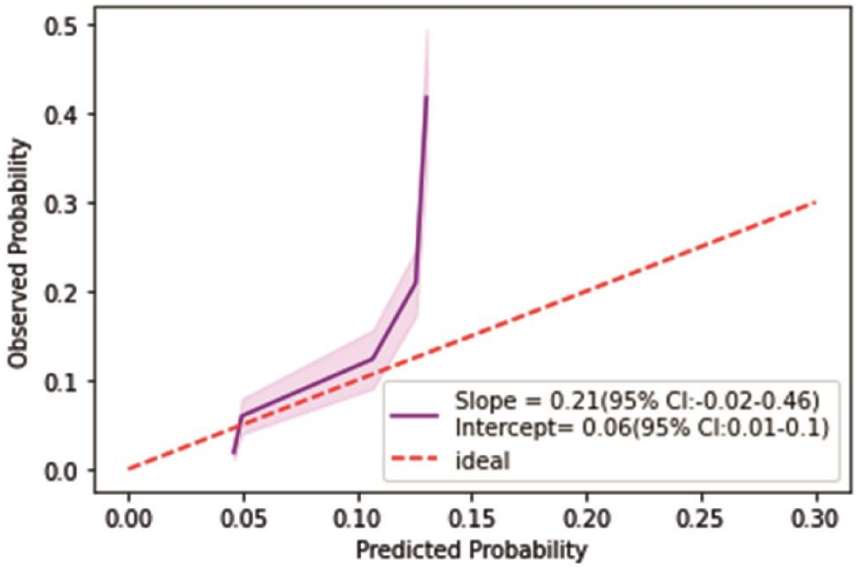
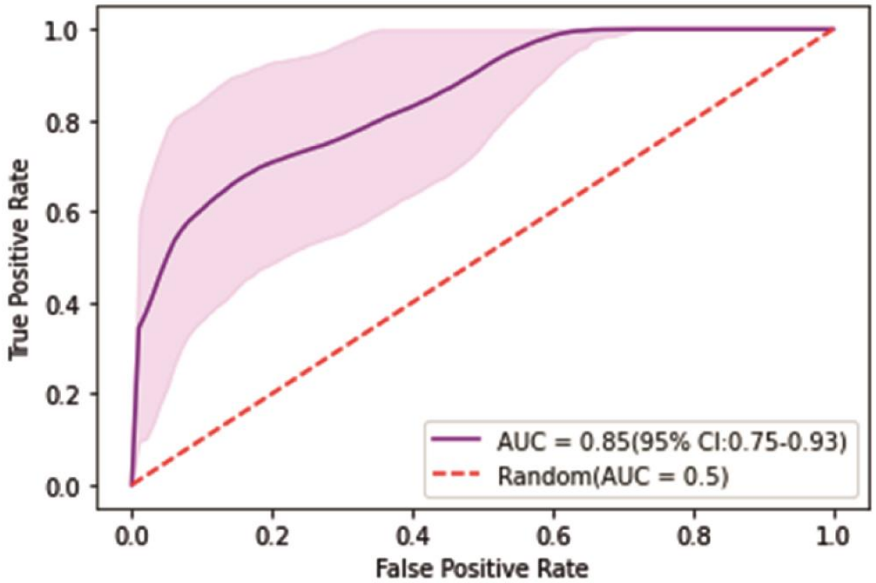
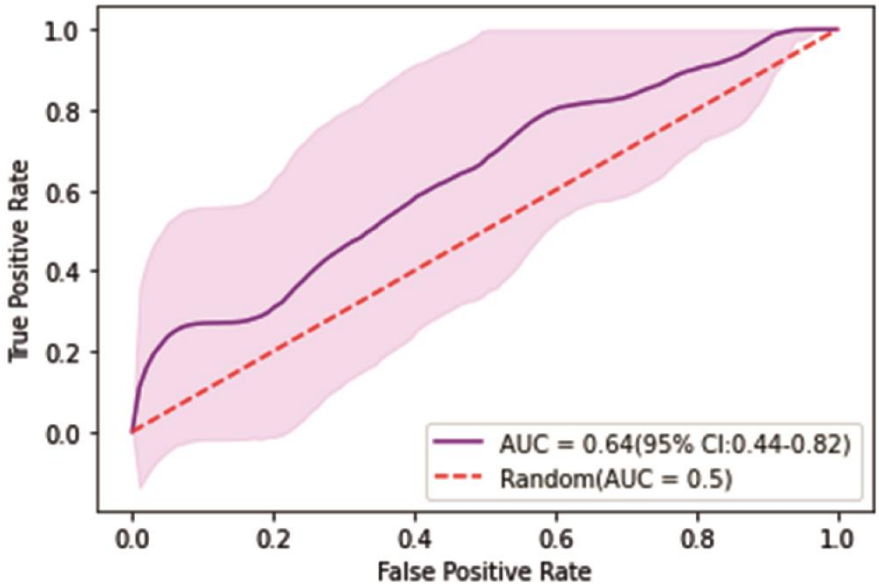


FIGURE 2. Resampled training (**a**) and test (**b**) set area under the receiver operating characteristics curves (AUC) of the random forest classifier.

Figure 2

a



b