



Original Investigation | Surgery

Development of a Clinical Prediction Model for Anastomotic Leakage in Colorectal Surgery

Vincent Ochs; Stephanie Taha-Mehlitz, MD; Joël L. Lavanchy, MD; Daniel Steinemann, MD; Beat P. Müller, MD; Robert Rosenberg, MD; Mariana Flaifel; Bassey Enodien, MD; Daniel M. Frey, MD; Martin Baltuonis; Stephan Vorburger, MD; Sven Flemming, MD; Lars Kollmann, MD; Johan Friso Lock, MD; Victor E. Staartjes, MD, PhD; Julia Kristin Grass, MD; Nathaniel Melling, MD; Michael Rousek, MD; Austėja Elzbieta Degutyte, MD; Vilius Abeciūnas, MD; Tomas Poskus, MD; Florian Ponholzer, MD; Dietmar Öfner, MD; Alaa El-Hussuna, MD; Ovunc Bardakcioglu, MD; Katerina Neumann, MD, PhD; Richard T. Spence, MD; Tia Sutton, MD; Michael Drew Honaker, MD; Philippe C. Cattin, PhD; Anas Taha, MD; for the 2015 European Society of Coloproctology Collaborating Group

Abstract

IMPORTANCE Anastomotic leakage is a severe complication after colorectal surgery that can significantly affect clinical outcomes and patient prognosis. Although multiple risk factors for anastomotic leakage have been identified, predicting the incidence of anastomotic leakage remains difficult.

OBJECTIVE To develop a machine learning predictive model that combines logits from different base models to predict the incidence of anastomotic leakage more accurately.

DESIGN, SETTING, AND PARTICIPANTS This retrospective prognostic study was conducted across 13 centers in Europe and North America between January 1, 2012, and December 31, 2020, and with data from the European Society of Coloproctology (ESCP). The study population included patients 18 years or older who underwent colon resection with anastomosis and had a follow-up of at least 6 months. Patients with metastatic disease and insufficient follow-up were excluded. A total of 6079 patients from 13 centers across different countries were included in the analysis with an additional 3041 patients from the ESCP. Data analysis was conducted from October 2024 to January 2025.

EXPOSURE Colon anastomosis for various reasons, including neoplasia, diverticulitis, ischemia, iatrogenic or traumatic perforation, or inflammatory bowel disease.

MAIN OUTCOMES AND MEASURES The primary outcome of this study was to predict anastomotic leakage using a cross-attention-based meta-model, which integrated predictions from base models and patient-level clinical features. The F1 score was used to assess the prediction performance of the models.

RESULTS A total of 9120 patients (mean [SD] age, 61.26 [15.71] years; 4636 [50.8%] male) from the 513 centers and the ESCP were included in the analysis. There was no difference in distribution between the 13 centers and the ESCP dataset. The model's predictive performance achieved overall F1 scores of 87% (95% CI, 78%-95%) and 70% across tests using a cross-validation process and external validation test set, respectively. The meta-model's performance was better than that of its component parts (eg, CatBoost: cross-validation F1 score, 87% [95% CI, 78%-95%]; external validation F1 score, 70%).

CONCLUSIONS AND RELEVANCE In this prognostic study of patients who underwent colon resection with anastomosis, a meta-model was found to improve predictive accuracy. Additional

(continued)

Key Points

Question Can a predictive model that integrates multiple base models' predictions and patient-specific features accurately predict the incidence of anastomotic leakage?

Findings In this retrospective prognostic study involving 9120 patients, the predictive performance of a meta-model achieved an overall F1 score of 87% and 70% using the cross-validation and external validation test sets, respectively.

Meaning These results suggest that a predictive model integrating multisource information can predict anastomotic leakage more accurately than individual base models, demonstrating the potential of advanced machine learning techniques to enhance clinical decision-making.

+ [Invited Commentary](#)

+ [Supplemental content](#)

Author affiliations and article information are listed at the end of this article.

Open Access. This is an open access article distributed under the terms of the CC-BY-NC-ND License, which does not permit alteration or commercial use, including those for text and data mining, AI training, and similar technologies.

Abstract (continued)

prospective studies are needed to assess the clinical utility of the model in real-time settings and its integration into surgical workflows.

JAMA Network Open. 2025;8(10):e2538267. doi:10.1001/jamanetworkopen.2025.38267

Introduction

Anastomotic leakage (AL) is among the most severe complications after colorectal surgery, with varying incidence rates of 2.8% to 30%.¹⁻⁴ It is defined as a defect in the intestinal wall at the anastomotic site, resulting in communication between the intraluminal and extraluminal space.⁵ AL is associated with higher reoperative rates, longer hospital stays, and increased morbidity and mortality, with rates equivalent to 20.0% to 35.0% and 2.0% to 16.4% respectively.^{1,6} Several risk factors are implicated in the development of AL, including disease characteristics, surgical indication, and preoperative and intraoperative surgical techniques.⁷ A deeper understanding of the complex relationship and correlation among these factors can assist in predicting the risk of AL more accurately. Identifying modifiable risk factors can potentially optimize resource allocation, enhance patient preparation, and guide surgeons toward the optimal surgical technique at the patient level.⁸ Despite these efforts, objectively and accurately predicting the incidence of AL for patients remains difficult even for experienced surgeons.⁹

Machine learning (ML) is a rapidly evolving tool with the potential to develop more accurate predictive models, which in turn could help decrease the rates of AL. ML offers a framework that can analyze a large multicenter dataset and integrate diverse patient variables into unified risk models for patient-specific predictions.¹⁰ Unlike traditional methods, ML can analyze complex, nonlinear relationships among variables, enabling more precise predictions.¹¹ Such tools can successfully support preoperative planning, allowing surgeons to identify high-risk patients and tailor their surgical techniques accordingly.¹² Identifying the optimal model is difficult because several predictive models with varying predictive accuracy have been developed.^{9,12} An overarching model that integrates and analyzes the predictions of the separate models can potentially create a more accurate and generalizable prediction of AL. This study aims to develop and evaluate a meta-model that integrates logits from different existing predictive models combined with patient-specific features to predict AL risk after colorectal surgery.

Methods

Study Design and Participants

This retrospective prognostic study was conducted across centers in Europe and North America and with data from the European Society of Coloproctology (ESCP). The ESCP cohort is a prospective dataset that has been collected through international collaboration led by the ESCP Cohort Studies Working Group.¹³ Data were collected from 13 cohorts of adult patients undergoing colonic resection and anastomosis for various reasons, including neoplasia, diverticulitis, ischemia, iatrogenic or traumatic perforation, or inflammatory bowel disease, between January 1, 2012, and December 31, 2020, with a follow-up of at least 6 months. Participating centers are listed in the eAppendix 1 in Supplement 1. Data collection was conducted by consultants, surgical residents, or medical students under supervision (eTable 5 in Supplement 1). Patients younger than 18 years, those with insufficient follow-up (<3 months), and those with metastatic, recurrent, or unresectable disease were excluded. The patient flowchart is available in eFigure 1 in Supplement 1. Additional training and validation were conducted using data from participants in a registered, previously published, prospective audit, multicenter trial conducted by the 2015 European Society of Coloproctology Collaborating Group.¹³ Details about cohort data collection can be found in the Descriptive Statistics section of eTables 1 and

2 in Supplement 1). All centers obtained institutional review board and ethics approval for their respective institutions. This study was conducted in accordance with the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting guideline for developing clinical prediction models¹⁴ as well as the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline.¹⁵

Outcomes

The primary outcome of this study was to predict AL using a cross-attention-based meta-model that integrated predictions from base models and patient-specific clinical features. By incorporating both model-derived logits and original feature representations, the approach aimed to provide robust predictions across diverse clinical settings. Furthermore, the generalizability of the model was assessed by validating its performance on an external dataset, ensuring its applicability across various health care environments.

Model Development

We developed and tested the predictive model using a specific cross-clinic model validation in addition to a 5-fold cross-validation and external evaluation to ensure consistency of the model performance. AL was diagnosed based on clinical sign of leakage, confirmed by radiologic imaging or endoscopy or identified at reoperation. Recorded variables included 34 preoperative risk factors that have been reported in the literature, such as age, sex, body mass index (BMI), active smoking, pack-years, alcohol abuse (no, 1-2 alcoholic beverages per day, or >2 alcoholic beverages per day), prior abdominal surgery, preoperative leukocytosis (white blood cell count >10 000/ μ L [to convert to $\times 10^9$ /L, multiply by 0.001]), preoperative steroid use, American Society of Anesthesiologists score (I-IV), Charlson Comorbidity Index, kidney function (chronic kidney disease stages G1-G5), albumin level, hemoglobin level, C-reactive protein level, neoadjuvant therapy, intake of immunosuppressive drugs, dosage of steroids, use of nonsteroidal anti-inflammatory drugs, preoperative blood transfusion, tumor necrosis factor α inhibitors, liver metastasis (at the time of surgery proven by radiologic imaging or biopsy preoperatively), indication (eg, tumor, diverticular disease, ileus, ischemia, volvulus, Crohn disease, and others), type of surgery (eg, sigmoid resection, left or extended left hemicolectomy, right or extended right hemicolectomy, transverse colectomy, Hartman reversal, colon segment resection, and ileocecal resection), emergency surgery, bowel perforation, surgical approach (laparoscopic, robotic, or open), anastomotic technique (hand-sewn or stapler), conversion to open, surgeon's experience (consultant or teaching operation), nutrition status, diverting ostomy, type of anastomosis (colon-colon anastomosis, colorectal anastomosis, or ileocolonic anastomosis), and anastomotic configuration (end to end, side to end, side to side, or end to side).

We started the preprocessing phase by thoroughly analyzing the baseline features of all patient data. We excluded any variable with 30% or more missing values from the dataset to minimize potential bias from missing data. For the remaining data, k -nearest neighbors imputation was used, ensuring the preservation of the distribution of the overall dataset.

Categorical features were converted to numerical formats using ordinal encoding. Minimum to maximum scaling normalized all numerical features to a range of (0,1), preventing any single feature from dominating due to scale differences. After that, we applied feature engineering by generating polynomial features with a degree of 2 for capturing higher-order associations between variables.

To address class imbalance within the dataset, we implemented the synthetic minority over-sampling technique (SMOTE), which generates synthetic instances for the minority class based on the k -nearest neighbors of existing samples. This method improved the model's ability to learn from minority class examples. Additionally, we applied Tomek links to eliminate noisy or borderline synthetic instances, refining the dataset and enhancing the quality of the training data.

The training process included a section in which 4 models (CatBoost, LightGBM, random forest, and bagging classifier) were trained to extract logits. This information was subsequently fed into a

cross-attention-based neural network serving as a meta-model, which takes those logits and processed features as inputs for final predictions. A summary of the workflow is shown in eFigure 2 in [Supplement 1](#). The multihead self-attention layer plays a crucial role in analyzing embeddings by dividing them into multiple attention heads, with each head independently computing attention. This layer uses a series of essential operations, including query, key, and value-dense layers, a scaled dot-product attention mechanism, multihead processing with concatenation, and a final dense layer to merge the outputs of all attention heads into a cohesive representation.

Building on this foundation, the cross-attention layer leverages the principles of the multihead self-attention mechanism to compute associations between the base model predictions and feature representations. By aligning and combining these inputs, the model effectively captures interactions and dependencies that are critical for accurate prediction.

The meta-model processing workflow integrates these components into a seamless pipeline. First, the base model predictions are embedded into a higher-dimensional space. Self-attention is then applied to these embedded predictions. Simultaneously, cross-attention is computed between the base predictions and the original features, enabling the model to dynamically align and merge information from both sources. Following this, the combined representations are processed through global average pooling to reduce dimensionality and enhance generalization. Finally, dense layers with regularization and dropout are applied, ensuring robust and reliable predictions.

A schematic diagram illustrating the architecture of the meta-model is provided in eFigure 3 in [Supplement 1](#), offering a visual representation of the interplay among its various components. The model parameter of the final model can be found in eTable 3 in [Supplement 1](#). The choice of base learners, preprocessing steps, and feature engineering was based on an extensive ablation study evaluating 144 model combinations (eAppendix 2 and eFigures 4 and 5 in [Supplement 1](#)), identifying the optimal configuration that yielded the highest F1 score.

Model Validation

The validation process was carefully designed to mitigate bias and ensure consistent performance across various data subsets. This approach involved several essential steps, beginning with 5-fold stratified cross-validation, which divided the dataset into 5-folds while maintaining class proportions. This method preserved the representation of both majority and minority classes in training and validation splits, promoting model reliability.

In addition to cross-validation, a per-center evaluation was conducted to test the robustness of the model in different clinical settings. For each center, the data from that specific center were designated as the test set, whereas the base models were trained on data from the other centers. Predictions generated by the base models for the test center were combined with the original features and used to train the meta-model. This center-specific approach provided insights into the model's adaptability and generalizability.

To enhance predictive accuracy, threshold optimization was performed for each model on the validation set. This step identified the optimal decision threshold, ensuring a balanced trade-off between sensitivity and specificity. After threshold optimization, a center validation was performed using an independent dataset from the ESCP, demonstrating the model's capability to generalize to novel datasets. The feature importance has also been assessed.

Statistical Analysis

Patient characteristics were summarized using mean (SD) for continuous variables and number (percentage) for categorical variables. To ensure robust hypothesis testing, statistical analyses were conducted using Shapiro-Wilk tests for normality, Levene tests for homogeneity of variance, 2-tailed, unpaired *t* tests for comparisons of means, and Mann-Whitney *U* tests for nonparametric comparisons. These analyses were performed from October 2024 to January 2025 with a 95% confidence level to validate the hypothesis that the meta-model's performance significantly exceeded that of the individual base models.

The computational workflow for the analysis was implemented using Python, version 3.10 (Python Software Foundation). The TensorFlow library was used for constructing and training the neural network-based meta-model, whereas Scikit-Learn (Python Software Foundation) supported preprocessing, feature engineering, and model evaluation tasks. These tools ensured a reliable and efficient analysis pipeline. Data analysis was conducted from January 2015 to March 2015.

Results

A total of 9120 patients (mean [SD] age, 61.26 [15.71] years; 4636 [50.8%] male and 4484 [49.2%]) were studied, including 6079 patients from 13 centers across different countries and an additional 3041 patients from the ESCP holdout dataset. An exploratory data analysis was conducted to profile patients, focusing on demographic features and the distribution of comorbidities. The mean (SD) age distribution for the 13 clinics and the holdout set from the ESCP 5 were slightly different (58.7 years [14.9] vs 66.3 [15.9] years, respectively). The mean BMI (calculated as weight in kilograms divided by height in meters squared) of the 13 participating clinics was slightly higher than for the ESCP dataset (27.4 vs 26.3). In the main dataset, 1021 patients (16.8%) were smokers vs 942 (31%) for the ESCP dataset.

Of the 34 remaining variables, the features contributing most to the attention layer were the logits of the 4 models, C-reactive protein levels, Charlson Comorbidity Index, prior abdominal surgery, albumin level, BMI, indication, surgical approach, surgeon's experience, emergency surgery, and some interaction terms between the features. The importance of each feature is visualized in **Figure 1** as the Shapley Additive Explanation values.

The model's predictive performance was assessed using a 10-fold cross-validation process, achieving an overall F1 macro score of 87% (95% CI, 78%-95%) across test centers (with the final model chosen based on the F1 score performance on external validation). The detailed results are presented in **Table 1**. Additionally, the model was evaluated on an external validation test set, achieving an F1 score of 70% (Table 1). The F1 scores of the external validation can be found in Table 1. The different thresholds of the models were also evaluated (**Figure 2**; eFigure 6 in [Supplement 1](#)). The calibration plots can be found in eFigure 7 in [Supplement 1](#).

The performance of the meta-model was compared against alternative approaches, including its individual components (CatBoost, LightGBM, random forest, and bagging classifier). These comparisons were conducted using cross-validation and external validation. Benchmark results are summarized in Table 1.

The final model was selected based on its superior F1 score performance during external validation. The performance of the meta-model was also evaluated and compared against the individual models on all the clinics separately, always leaving one clinic out for testing (**Table 2**). To ensure that the meta-model performed significantly better than the individual base models, statistical significance was assessed using McNemar tests for paired comparisons of predicted outcomes. These tests were repeated across all centers using cross-validation and the external validation set to confirm the robustness of the results. The statistical analysis demonstrated a significant improvement in performance for the meta-model compared with individual models in both validation scenarios (details provided in eAppendix 3 in [Supplement 1](#)). The class distributions of the 2 datasets (from the participating clinics and from the holdout set), combined with statistical tests to show the domain shift between the 2 datasets, are shown in eFigure 8 in [Supplement 1](#).

Finally, the ML meta-model was integrated into a user-friendly web application prototype designed to predict the probability of AL based on preoperative clinical data. The prototype application allows clinicians to simulate various scenarios by modifying patient information, surgical approach, and surgeon experience, providing tailored recommendations with associated probabilities of AL. This tool has the potential to support clinical decision-making by optimizing patient-specific surgical strategies.¹⁶

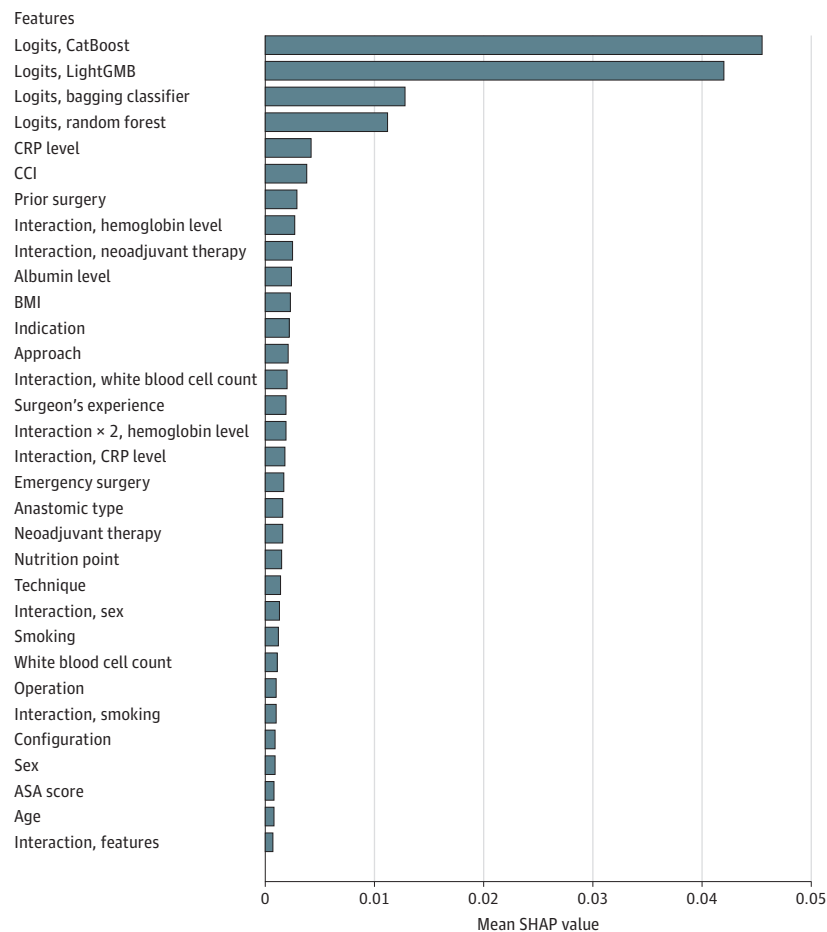
Discussion

AL remains one of the most devastating complications of colorectal surgery, associated with increased mortality, morbidity, reduced survival, and increased cancer recurrence rates.¹⁷ Additionally, AL significantly increases the financial burden on health care systems due to intensive care unit admissions, longer hospitalization, and resource use.¹⁷ To our knowledge, this is the first study to develop a meta-model that combines logits from different existing predictive models with a cross-attention mechanism to predict AL after colorectal surgery. By incorporating baseline characteristics, risk factors, and intraoperative factors, the model guides toward the optimal surgery to minimize AL risk for each patient. Thus, our tool might help reduce morbidity, mortality, and the future financial burden associated with AL.

Although surgical decisions are often straightforward in patients with extreme risk profiles, it becomes more complex in intermediate-risk patients, such as younger patients with few comorbidities or nonemergency indications. As detailed in eTable 4 in Supplement 1, we constructed 5 fictional cases to illustrate the model's predictive capacity and ability to capture nuanced clinical variation. It enables differentiation between cases that might appear similar but differ in subtle risk profiles, highlighting its potential to support actual decision-making. The "black-box" phenomenon of a ML algorithm applies here, making it challenging to understand how slight changes of variables affect the model's output.

Although clinicians are experts in dealing with risks and uncertain situations, surgeons are poor in predicting AL risk.¹⁸⁻²⁰ A cutoff value to define intermediate-risk patients for AL would be of

Figure 1. Shapley Additive Explanation (SHAP) Value Graphic for the Meta-Model



This graphic displays the importance of each feature in predicting anastomotic leakage risk. Features with longer bars indicate higher SHAP values, signifying a greater impact on the model's predictions. Features are ranked by their importance, with the most influential features listed at the top. ASA indicates American Society of Anesthesiologists; BMI, body mass index; CCI, Charlson Comorbidity Index; CRP, C-reactive protein.

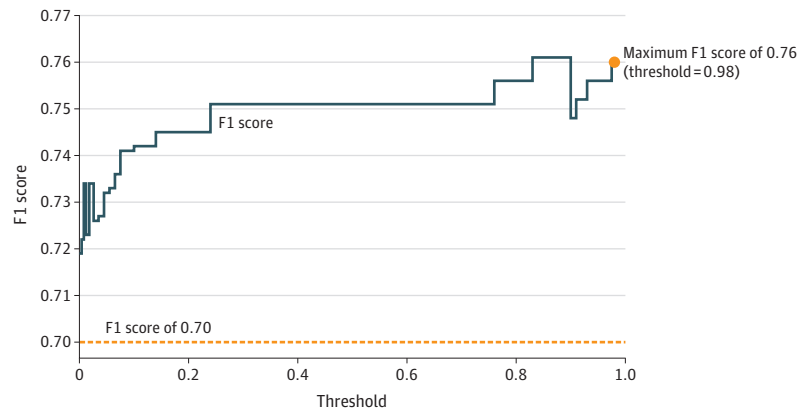
utmost importance to clinical decision-making. There is no universally accepted threshold or definition of what constitutes a patient as intermediate risk for AL. However, this is the cohort of interest who clinicians aim to identify before surgery to adapt the patient’s preoperative condition or adjust the intraoperative surgical strategy. Currently, the decision of whether a patient receives an

Table 1. Performance Evaluation of All Models Used

Model and metric	Cross-validation	External validation
Meta-model		
Accuracy	0.97 (0.80-0.98)	0.88
AUC	0.89 (0.83-0.95)	0.78
Precision	0.91 (0.85-0.95)	0.73
Recall	0.82 (0.81-0.93)	0.66
F1 score	0.87 (0.78-0.95)	0.70
CatBoost		
Accuracy	0.96 (0.90-0.97)	0.87
AUC	0.89 (0.87-0.91)	0.67
Precision	0.87 (0.80-0.88)	0.69
Recall	0.81 (0.79-0.88)	0.61
F1 score	0.85 (0.80-0.87)	0.65
LightGBM		
Accuracy	0.96 (0.82-0.97)	0.87
AUC	0.88 (0.86-0.89)	0.71
Precision	0.87 (0.70-0.90)	0.71
Recall	0.80 (0.70-0.83)	0.67
F1 score	0.85 (0.79-0.89)	0.69
Random forest		
Accuracy	0.94 (0.79-0.97)	0.82
AUC	0.88 (0.89-0.90)	0.75
Precision	0.85 (0.81-0.88)	0.65
Recall	0.81 (0.78-0.83)	0.60
F1 score	0.85 (0.84-0.86)	0.63
Bagging classifier		
Accuracy	0.95 (0.89-0.97)	0.86
AUC	0.81 (0.80-0.84)	0.66
Precision	0.84 (0.78-0.91)	0.68
Recall	0.79 (0.72-0.84)	0.64
F1 score	0.82 (0.77-0.90)	0.65

Abbreviation: AUC, area under the receiver operating characteristic curve.

Figure 2. Example Plot for the Threshold Optimization for 1 Participating Clinic



anastomosis or a diverting stoma is up to the surgeon's estimation of AL in the particular patient. This decision is driven by known risk factors, personal experience, and the intraoperative situs. A surgeon's intuition should not be underestimated because it often influences decision-making. However, this gut feeling is difficult to measure and objectify and therefore challenging to incorporate into objective tools. At the same time, surgeons are prone to bias; for example, being criticized for an AL in a recent case or loss of a patient due to AL has an influence on objective decision-making on colorectal surgeons.²¹ Thus, scores and widely accepted tools within the colorectal surgical community provide a foundation upon which surgeons can base and, if necessary, justify their decisions legally. This underlines the importance of establishing clear cutoffs and expert opinions to stratify patients at intermediate risk for AL, reflecting the at-risk subgroup.

Risk perception is also subjective; a 5% risk may seem low to a surgeon but high to a patient. In some cultures, complications after surgery are seen as unfortunate but acceptable, whereas in other regions blame and legal action can be associated. This can also affect a surgeon's point of view; practicing in a highly litigious environment may cause a surgeon to view lower risks as more significant because legal and reputational consequences are feared.²²

In addition, outcome severity clearly skews perception. A 5% risk of a mild complication resulting, for example, in antibiotic application feels different than a 5% risk of a reoperation and

Table 2. Performance Evaluation of the Meta-Model for the Participating Clinics^a

Test clinic	F1 scores, %				
	Meta-model	CatBoost	LightGBM	Random forest	Bagging classified
European and North American centers					
University Hospital of Würzburg	91	90	91	88	85
University Medical Center Hamburg-Eppendorf	84	82	85	82	84
Kirk Kerkorian School of Medicine at UNLV	94	91	91	94	97
University Hospital of Basel	95	95	95	96	91
Cantonal Hospital Basel-Landschaft	53	49	49	68	49
GZO Spital Wetzikon	83	79	83	72	72
Brody School of Medicine, East Carolina University	75	76	67	71	65
Military University Hospital Prague	91	91	91	85	82
Teaching Hospital Emmental	83	82	81	80	74
Vilnius University	95	90	92	89	87
Claraspital	100	100	100	100	100
Dalhousie University	95	93	93	92	95
Mean (SD)	87 (0.13)	85 (0.13)	85 (0.14)	82 (0.10)	70 (0.15)
European Society of Coloproctology					
University Hospital of Würzburg	71	65	72	60	57
University Medical Center Hamburg-Eppendorf	69	63	70	61	66
Kirk Kerkorian School of Medicine at UNLV	68	65	67	65	70
University Hospital of Basel	70	62	70	63	67
Cantonal Hospital Basel-Landschaft	69	67	70	60	70
GZO Spital Wetzikon	70	63	67	68	62
Brody School of Medicine, East Carolina University	73	68	70	64	59
Military University Hospital Prague	67	64	67	65	61
Teaching Hospital Emmental	69	66	71	62	66
Vilnius University	70	64	70	61	66
Claraspital	70	66	68	65	66
Dalhousie University	70	64	70	66	66
Mean (SD)	65 (0.01)	69 (0.02)	63 (0.02)	63 (0.03)	65 (0.04)

^a Data were analyzed using a leave-one-clinic-out cross-validation approach. Each row corresponds to a test clinic, where the models were trained on all clinics except for the one in that row and then tested on that excluded clinic. For example, in the University Hospital of Würzburg row, all models were trained on data from all clinics except the University Hospital of Würzburg and then evaluated on the University Hospital of Würzburg dataset. In the second row, the same process was applied but with University Medical Center Hamburg-Eppendorf as the excluded test clinic and so on for each clinic. The same leave-one-clinic-out training procedure was used for the European Society of Coloproctology rows, but instead of testing only on the left-out clinic, the trained models were additionally evaluated on the European Society of Coloproctology external validation dataset. This dual evaluation setup allows for assessing how well the models generalize to unseen clinical environments. The variation in performance across clinics and the generally lower scores on the European Society of Coloproctology dataset indicate potential domain shifts and highlight the importance of external validation.

stoma placement after elective surgery. For lifesaving surgery a higher risk may be acceptable but not in an elective setting with alternative treatment.^{23,24} Emotions such as fear can affect how a risk is perceived. Although patients with anxiety are naturally more risk averse, others may be more optimistic and risk tolerant.²³ How a patient perceives a risk also depends on trust in the health care professional and communication style. Understanding that 1 of 20 patients is affected may sound worse than a 95% chance of not experiencing an AL.²⁵

Risk perception and trust in decision-making based on an artificial intelligence tool seems to be well received by patients who believe that the surgeon remains the primary decision maker.²⁶ Our model on AL prediction offers personalized, numeric risk estimates to guide individualized surgical planning, informed patient discussions, and expectation management for patients.

Unlike prior studies²⁷⁻³⁰ that relied predominantly on linear regression or single ML algorithms, our approach leverages ML to account for nonlinear associations and dependencies among features.²⁷ In terms of performance, our model can achieve an F1 score of 87% in cross-validation and 70% in the external test set, respectively, which is higher than the values of traditional statistical models.^{14,27,31} The cross-attention mechanism enables better feature extraction and decision-making compared with conventional ML approaches, which are limited by the inability to generalize to the clinical setting.¹² This implies that the meta-model demonstrates high performance in predicting AL risk. Furthermore, our methods ensure that the model's performance metrics are not inflated by overfitting or localized biases.

Limitations

Our study has limitations. Despite the success of our meta-model in predicting AL and integrating multisource information, the model focused only on patient-specific features and preoperative risk factors, excluding postoperative factors. Thus, future studies should consider additional factors that may influence AL risk, which could potentially refine the predictive accuracy, particularly intraoperative and real-time data, such as operative time, blood loss, hypothermia, vasopressor application, or fluid administration.³²⁻³⁴ Other variables, which are less convenient to measure or quantify, such as tension on the anastomosis, could further enhance intraoperative AL risk calculation.³⁵ Although these variables were outside the scope of this study, advancements in real-time monitoring technologies could enable their integration into predictive models. Such enhancements would provide a more comprehensive approach to AL prediction. Furthermore, an additional predictive tool for identifying patients at risk for AL in the postoperative setting could aid in surveillance and resource allocation. Patients discovered at an early stage of leakage may be spared reoperation or takedown of the anastomosis, and early-stage sepsis can be targeted to reduce additional morbidity.

Further prospective studies are needed to refine our meta-model and evaluate its application in varied clinical environments. Additionally, the usability of this tool in clinical settings depends on the practicality of data input. Although the model uses 34 input features, we recognize that this may be impractical in many settings. Reducing the number of variables through feature selection led to a decrease in predictive performance. However, the tool allows partial input, and clinicians can prioritize entering all available data. A trade-off between completeness and usability represents a key area for future improvement.

Conclusions

In this prognostic study, we highlighted the potential of advanced ML methods to enhance surgical outcome prediction. By addressing key challenges such as domain shifts and leveraging diverse clinical datasets, our approach provided a scalable and robust framework for AL prediction. Future work should focus on prospective validation, assessment of the clinical utility of the model in real-time settings, and integration into surgical workflows. This combination of data-driven insights and clinical expertise could substantially improve predictive accuracy.

ARTICLE INFORMATION

Accepted for Publication: July 17, 2025.

Published: October 20, 2025. doi:10.1001/jamanetworkopen.2025.38267

Open Access: This is an open access article distributed under the terms of the [CC-BY-NC-ND License](#), which does not permit alteration or commercial use, including those for text and data mining, AI training, and similar technologies. © 2025 Ochs V et al. *JAMA Network Open*.

Corresponding Author: Anas Taha, Department of Biomedical Engineering, Faculty of Medicine, University of Basel, Hegenheimerweg 167C, 4123 Allschwil, Switzerland (anas.taha@unibas.ch).

Author Affiliations: Department of Biomedical Engineering, Faculty of Medicine, University of Basel, Allschwil, Basel, Switzerland (Ochs, Lavanchy, Cattin, Taha); Clarunis, University Center for Gastrointestinal and Liver Diseases, St Clara Hospital and University Hospital Basel, Basel, Switzerland (Taha-Mehlitz, Lavanchy, Steinemann, Müller); Department of Visceral Surgery, Cantonal Hospital Basel-Landschaft, Liestal, Switzerland (Rosenberg, Taha); School of Medicine, St George's University of London, London, United Kingdom (Flaifel); Department of Surgery, Cantonal Hospital Glarus, Glarus, Switzerland (Enodien); Department of Surgery, GZO Spital Wetzikon, Wetzikon, Switzerland (Frey); Department of Surgery, Teaching Hospital Emmental, Berne, Switzerland (Baltuonis, Vorburger); Department of General, Visceral, Transplant, Vascular, and Pediatric Surgery, University Hospital of Würzburg, Würzburg, Germany (Flemming, Kollmann, Lock); Machine Intelligence in Clinical Neuroscience Laboratory, Department of Neurosurgery, University Hospital Zurich, Zurich, Switzerland (Staartjes); Department of General, Visceral and Thoracic Surgery, University Medical Center Hamburg-Eppendorf, Hamburg, Germany (Grass, Melling); Military University Hospital Prague, Prague, Czech Republic (Rousek); Faculty of Medicine, Department of Surgery, Vilnius University, Vilnius, Lithuania (Degutyte, Abeciūnas, Poskus); Department of Visceral, Transplant, and Thoracic Surgery, Center of Operative Medicine, Medical University of Innsbruck, Innsbruck, Austria (Ponholzer, Öfner); Medical Faculty, Copenhagen University, Copenhagen, Denmark (El-Hussuna); Open Source Research Collaboration, Aalborg, Denmark (El-Hussuna); Division of Colon and Rectal Surgery, Kirk Kerkorian School of Medicine at UNLV, Las Vegas, Nevada (Bardakcioglu); Department of Surgery, Dalhousie University, Halifax, Nova Scotia, Canada (Neumann, Spence); Department of Surgery, Brody School of Medicine, East Carolina University, Greenville, North Carolina (Sutton, Honaker, Taha).

Author Contributions: Dr Taha had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Mr Ochs and Dr Taha-Mehlitz share the first authorship. Dr Taha and Prof Cattin share the last authorship.

Concept and design: Ochs, Taha-Mehlitz, Steinemann, Müller, Rosenberg, Enodien, Lock, Staartjes, El-Hussuna, Bardakcioglu, Taha, Cattin.

Acquisition, analysis, or interpretation of data: Ochs, Lavanchy, Rosenberg, Flaifel, Enodien, Frey, Baltuonis, Vorburger, Flemming, Kollmann, Lock, Staartjes, Grass, Melling, Rousek, Degutyte, Abeciūnas, Poskus, Ponholzer, Öfner, El-Hussuna, Bardakcioglu, Spence, Neumann, Sutton, Honaker, Cattin.

Drafting of the manuscript: Ochs, Taha-Mehlitz, Rosenberg, Enodien, Kollmann, El-Hussuna, Bardakcioglu, Taha.

Critical review of the manuscript for important intellectual content: Ochs, Lavanchy, Steinemann, Müller, Rosenberg, Flaifel, Enodien, Frey, Baltuonis, Vorburger, Flemming, Kollmann, Lock, Staartjes, Grass, Melling, Rousek, Degutyte, Abeciūnas, Poskus, Ponholzer, Öfner, El-Hussuna, Bardakcioglu, Spence, Neumann, Sutton, Honaker, Taha, Cattin.

Statistical analysis: Ochs, Poskus.

Obtained funding: Enodien, Vorburger, Poskus, Cattin.

Administrative, technical, or material support: Taha-Mehlitz, Steinemann, Rosenberg, Enodien, Vorburger, Kollmann, Lock, Staartjes, Grass, Degutyte, Abeciūnas, Poskus, Ponholzer, Spence, Sutton, Honaker, Taha, Cattin.

Supervision: Lavanchy, Steinemann, Müller, Rosenberg, Frey, Vorburger, Kollmann, Lock, Staartjes, Poskus, El-Hussuna, Bardakcioglu, Neumann, Honaker, Taha, Cattin.

Conflict of Interest Disclosures: Dr Lavanchy reported receiving funding from the Swiss National Science Foundation, Novartis Foundation, and Vontobel Foundation outside the submitted work. Dr Neumann reported receiving grants from GlaxoSmithKline outside the submitted work. No other disclosures were reported.

Funding/Support: This study was funded by grant A1813455- SY from Medtronic and by Fluid AI.

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Group Information: The 2015 European Society of Coloproctology Collaborating Group members are listed in [Supplement 2](#).

Disclaimer: This article reflects the authors' view, and no associated partners are responsible for any use that may be made of the information contained herein.

Data Sharing Statement: See Supplement 3.

REFERENCES

- Chiarello MM, Fransvea P, Cariati M, Adams NJ, Bianchi V, Brisinda G. Anastomotic leakage in colorectal cancer surgery. *Surg Oncol*. 2022;40:101708. doi:10.1016/j.suronc.2022.101708
- Ellis CT, Maykel JA. Defining anastomotic leak and the clinical relevance of leaks. *Clin Colon Rectal Surg*. 2021;34(6):359-365. doi:10.1055/s-0041-1735265
- Sciuto A, Merola G, De Palma GD, et al. Predictive factors for anastomotic leakage after laparoscopic colorectal surgery. *World J Gastroenterol*. 2018;24(21):2247-2260. doi:10.3748/wjg.v24.i21.2247
- Tamini N, Cassini D, Giani A, et al. Computed tomography in suspected anastomotic leakage after colorectal surgery: evaluating mortality rates after false-negative imaging. *Eur J Trauma Emerg Surg*. 2020;46(5):1049-1053. doi:10.1007/s00068-019-01083-8
- Rahbari NN, Weitz J, Hohenberger W, 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;147(3):339-351. doi:10.1016/j.surg.2009.10.012
- Hyman N, Manchester TL, Osler T, Burns B, Cataldo PA. Anastomotic leaks after intestinal anastomosis: it's later than you think. *Ann Surg*. 2007;245(2):254-258. doi:10.1097/01.sla.0000225083.27182.85
- Zarnescu EC, Zarnescu NO, Costea R. Updates of risk factors for anastomotic leakage after colorectal surgery. *Diagnostics (Basel)*. 2021;11(12):2382. doi:10.3390/diagnostics11122382
- Tsalikidis C, Mitsala A, Mentonis VI, et al. Predictive factors for anastomotic leakage following colorectal cancer surgery: where are we and where are we going? *Curr Oncol*. 2023;30(3):3111-3137. doi:10.3390/curroncol30030236
- Mazaki J, Katsumata K, Ohno Y, et al. A novel predictive model for anastomotic leakage in colorectal cancer using auto-artificial intelligence. *Anticancer Res*. 2021;41(11):5821-5825. doi:10.21873/anticancer.15400
- Taha-Mehlitz S, Wentzler L, Angehrn F, et al. Machine learning-based preoperative analytics for the prediction of anastomotic leakage in colorectal surgery: a Swiss pilot study. *Surg Endosc*. 2024;38(7):3672-3683. doi:10.1007/s00464-024-10926-4
- Ryo M, Rillig MC. Statistically reinforced ML for nonlinear patterns and variable interactions. *Ecosphere*. 2017;8(11):e01976. doi:10.1002/ecs2.1976
- Bektaş M, Tuynman JB, Costa Pereira J, Burchell GL, van der Peet DL. Machine learning algorithms for predicting surgical outcomes after colorectal surgery: a systematic review. *World J Surg*. 2022;46(12):3100-3110. doi:10.1007/s00268-022-06728-1
- 2015 European Society of Coloproctology Collaborating Group. Predictors for anastomotic leak, postoperative complications, and mortality after right colectomy for cancer: results from an international snapshot audit. *Dis Colon Rectum*. 2020;63(5):606-618. doi:10.1097/DCR.0000000000001590
- 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;102(3):148-158. doi:10.1002/bjs.9736
- von Elm E, Altman DG, Egger M, Pocock SJ, Gotsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *J Clin Epidemiol*. 2008;61(4):344-349. doi:10.1016/j.jclinepi.2007.11.008
- AL prediction app. Accessed September 3, 2025. <https://meta-model-for-anastomotic-leakage-prediction.streamlit.app/>
- Tonini V, Zanni M. Impact of anastomotic leakage on long-term prognosis after colorectal cancer surgery. *World J Gastrointest Surg*. 2023;15(5):745-756. doi:10.4240/wjgs.v15.i5.745
- 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;24(5):569-576. doi:10.1007/s00384-009-0658-6
- Sammour T, Lewis M, Thomas ML, Lawrence MJ, Hunter A, Moore JW. A simple web-based risk calculator (www.anastomoticleak.com) is superior to the surgeon's estimate of anastomotic leak after colon cancer resection. *Tech Coloproctol*. 2017;21(1):35-41. doi:10.1007/s10151-016-1567-7
- Moskowitz AJ, Kuipers BJ, Kassirer JP. Dealing with uncertainty, risks, and tradeoffs in clinical decisions: a cognitive science approach. *Ann Intern Med*. 1988;108(3):435-449. doi:10.7326/0003-4819-108-3-435

21. Moug SJ, Henderson N, Tiernan J, et al; Edinburgh Delphi Collaborative Group. The colorectal surgeon's personality may influence the rectal anastomotic decision. *Colorectal Dis*. 2018;20(11):970-980. doi:10.1111/codi.14293
22. Smith TR, Hulou MM, Yan SC, et al. Defensive medicine in neurosurgery: the Canadian experience. *J Neurosurg*. 2016;124(5):1524-1530. doi:10.3171/2015.6.JNS15764
23. El-Toukhy S. Parsing susceptibility and severity dimensions of health risk perceptions. *J Health Commun*. 2015;20(5):499-511. doi:10.1080/10810730.2014.989342
24. Wilson RS, Zwickle A, Walpole H. Developing a broadly applicable measure of risk perception. *Risk Anal*. 2019;39(4):777-791. doi:10.1111/risa.13207
25. Fagerlin A, Zikmund-Fisher BJ, Ubel PA. Helping patients decide: ten steps to better risk communication. *J Natl Cancer Inst*. 2011;103(19):1436-1443. doi:10.1093/jnci/djr318
26. Ben Hmido S, Abder Rahim H, Ploem C, et al. Patient perspectives on AI-based decision support in surgery. *BMJ Surg Interv Health Technol*. 2025;7(1):e000365. doi:10.1136/bmjsit-2024-000365
27. Rajula HSR, Verlato G, Manchia M, Antonucci N, Fanos V. Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. *Medicina (Kaunas)*. 2020;56(9):455. doi:10.3390/medicina56090455
28. Litchinko A, Buchs N, Balaphas A, et al. Score prediction of anastomotic leak in colorectal surgery: a systematic review. *Surg Endosc*. 2024;38(4):1723-1730. doi:10.1007/s00464-024-10705-12
29. Lee JK, Mishra N. Predicting anastomotic leak: can we? *Semin Colon Rectal Surg*. 2014;25(2):74-78. doi:10.1053/j.scrs.2014.04.003
30. Venn ML, Hooper RL, Pampiglione T, Morton DG, Nepogodiev D, Knowles CH. Systematic review of preoperative and intraoperative colorectal Anastomotic Leak Prediction Scores (ALPS). *BMJ Open*. 2023;13(7):e073085-e073085. doi:10.1136/bmjopen-2023-073085
31. Rullier E, Laurent C, Garrelon JL, Michel P, Saric J, Parneix M. Risk factors for anastomotic leakage after resection of rectal cancer. *Br J Surg*. 1998;85(3):355-358. doi:10.1046/j.1365-2168.1998.00615.x
32. Reudink M, Huisman DE, van Rooijen SJ, et al; LekCheck Study Group. Association between intraoperative blood glucose and anastomotic leakage in colorectal surgery. *J Gastrointest Surg*. 2021;25(10):2619-2627. doi:10.1007/s11605-021-04933-2
33. Ozmen I, Grupa VEM, Bedrikovetski S, et al; LekCheck Study Group. Risk nomogram does not predict anastomotic leakage after colon surgery accurately: results of the multi-center LekCheck study. *J Gastrointest Surg*. 2022;26(4):900-910. doi:10.1007/s11605-021-05119-6
34. Huisman DE, Bootsma BT, Ingwersen EW, et al; LekCheck Study group. Fluid management and vasopressor use during colorectal surgery: the search for the optimal balance. *Surg Endosc*. 2023;37(8):6062-6070. doi:10.1007/s00464-023-09980-1
35. Rutegård M, Svensson J, Segelman J, Matthiessen P, Lydrup ML, Park J; RectoLeak study group. Splenic flexure mobilization and anastomotic leakage in anterior resection for rectal cancer: a multicentre cohort study. *Scand J Surg*. 2023;112(4):246-255. doi:10.1177/14574969231181222

SUPPLEMENT 1.

eAppendix 1. Participating Centers

eAppendix 2. Ablation Study

eAppendix 3. Statistical Analysis Using McNemar Test

eFigure 1. Flowchart of Included Patients From All Hospitals

eFigure 2. Workflow Diagram

eFigure 3. Overview of Architecture of the Meta-Model

eFigure 4. Ablation Study: F1 Score by Single Model

eFigure 5. Ablation Study: Meta-Model by Oversampling and Scaler

eFigure 6. Threshold Optimization Plots

eFigure 7. Calibration Plots

eFigure 8. Distributions and Domain-Shift Analysis

eTable 1. Baseline Characteristics of the Training Data

eTable 2. Baseline Characteristics of the Test Data (ESCP)

eTable 3. Model Parameter of the Final Model

eTable 4. Model-Predicted Anastomotic Leakage Risk on Hypothetical Patients

eTable 5. List of Authors in Charge of the Access and Verification of Each Cohort

SUPPLEMENT 2.

Nonauthor Collaborators

SUPPLEMENT 3.

Data Sharing Statement