

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

Super Learner Enhances Postoperative Complication Prediction in Colorectal Surgery

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:

Violante, T., Ferrari, D., Novelli, M., Perry, W.R., Mathis, K.L., Dozois, E.J., et al. (2025). Super Learner Enhances Postoperative Complication Prediction in Colorectal Surgery. ANNALS OF SURGERY, NA (Online first), 1-25 [10.1097/sla.0000000000006847].

Availability:

This version is available at: <https://hdl.handle.net/11585/1022206> since: 2025-09-01

Published:

DOI: <http://doi.org/10.1097/sla.0000000000006847>

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).
When citing, please refer to the published version.

(Article begins on next page)

Super Learner Enhances Postoperative Complication Prediction in Colorectal Surgery

Tommaso Violante M.D.*^{1,2}, Davide Ferrari M.D.*^{1,3}, Marco Novelli M.S.⁴, William R. Perry M.B., Ch.B.¹, Kellie L. Mathis M.D.¹, Eric J. Dozois M.D.¹, David W. Larson M.D., M.B.A.¹

*Contributed equally

¹ Department of Colon and Rectal Surgery, Mayo Clinic, Rochester, MN, USA

² Alma Mater Studiorum Università di Bologna, School of General Surgery, Bologna, Italy

³ General Surgery Residency Program, University of Milan, Milan, Italy

⁴ Department of Statistics, University of Bologna, Bologna, Italy

Corresponding author

David W. Larson, M. D., M. B. A.

Division of Colon and Rectal Surgery
Mayo Clinic
200 First St. Southwest
Rochester, Minnesota, USA 55905
Phone: (507) 774- 1503,
Fax: (507) 284-1794,

E-mail: Larson.David2@mayo.edu

DECLARATION:

Funding: This research did not receive external funding.

Conflicts of interest: The authors declare no conflict of interest.

Acknowledgements: Data supporting the research are available upon request to protect the privacy of study participants

ABSTRACT

Objective: To determine if a Super Learner (SL) machine learning approach could improve the predictive accuracy of the American College of Surgeons Risk Calculator (ACS-RC) for postoperative complications in patients undergoing colorectal surgery.

Summary of Background Data: Machine learning (ML) has shown significant potential to advance medical fields, including surgical risk prediction. Current tools, like the ACS-RC which uses logistic regression and extreme gradient boosting, are standard but may be enhanced by more advanced ML ensembles.

Methods: This retrospective study analyzed colorectal surgery cases from the 2018-2022 ACS National Surgical Quality Improvement Program (NSQIP) database. An SL model, which combines multiple ML algorithms, was developed to predict fourteen postoperative outcomes. Its performance was compared against traditional logistic regression (LOG) and extreme gradient boosting (XGB) models. Key performance metrics included discrimination (AUROC, AUPRC) and calibration (Brier score, Hosmer-Lemeshow test).

Results: The SL model demonstrated superior performance across all predicted complications when compared to both LOG and XGB. It showed superior discrimination for severe outcomes,

achieving an AUROC greater than 0.94 for predicting mortality. The SL model was also more accurate in predicting infectious complications and length of stay, and its calibration metrics indicated a better overall fit and accuracy.

Conclusions: The Super Learner model enhances the accuracy of postoperative risk prediction in colorectal surgery. Its superior performance suggests it is a promising tool for improving personalized patient counseling, aiding clinical decision-making, and optimizing resource allocation.

INTRODUCTION

Machine learning (ML), subset of artificial intelligence (AI), empowers computers to learn from data and make accurate predictions or decisions without explicit programming or human input. By analyzing vast datasets, ML algorithms identify patterns and relationships, enabling them to adapt and improve their performance over time. This adaptability has shown promise in transforming various fields, particularly medicine¹.

ML algorithms identify subtle patterns and anomalies in medical imaging, aiding in diagnostics and prognostics, especially in oncology²⁻⁵. They contribute to risk assessment, disease staging, and treatment prediction, ultimately enhancing patient care^{6, 7}. ML also plays a pivotal role in clinical decision support by integrating diverse patient data to improve risk stratification, diagnosis, and treatment recommendations^{8, 9}. Additionally, ML-powered digital health interventions are transforming the management of chronic conditions like depression and stress^{10, 11}. Furthermore, wearable sensors equipped with ML algorithms enable continuous health monitoring, providing real-time feedback and early warnings for various health conditions¹².

Machine learning is advancing rapidly in surgery, with a notable application being the ability to predict postoperative complications in patients before their procedures¹³⁻¹⁶. This has proven particularly valuable in colorectal surgery: ML algorithms have demonstrated superior precision compared to traditional methods in predicting risks of complications such as ileus, anastomotic leak, surgical site infections, and readmission rates¹⁷⁻²².

Given the rapid evolution of ML, with new models emerging daily, this article explores the potential of a new approach called Super Learner²³. This study investigates its ability to enhance

the American College of Surgeons Risk Calculator (ACS-RC), based on an extreme gradient boosting (XGB)-ML algorithm, using a vast colorectal patient cohort from the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) databases.

METHODS

Patients 'selection

This retrospective analysis followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement guidelines for reporting observational studies²⁴.

ACS-NSQIP Participant Use Data Files (PUFs) were compiled from 2018 to 2022. Colorectal cases were identified using the Current Procedural Terminology (CPT) codes ranging from 44005 to 46999: this encompassed a wide array of procedures performed on the intestine (44005-44799), mesentery and Meckel's diverticulum (44800-44899), appendix (44900-44979), colon (45000-45999), and anus (46020-46999). All other codes were excluded. This refined dataset formed the basis for subsequent analysis and modeling efforts.

Variables used to model the Super Learner

The following patient characteristics, as reported according to the ACS-RC methodology and adhering to the ACS-NSQIP terminology, were used as risk predictors: age group, gender, functional status, emergent status, American Society of Anesthesiologists (ASA) class, steroid use, ascites, sepsis category, ventilator dependent, disseminated cancer, diabetes, hypertension, congestive heart failure, dyspnea, smoker, COPD, dialysis, acute renal failure, body mass index (BMI) group, and Planned Principal Operation (CPT Code).

Outcomes predicted

Fourteen outcomes were identified to model: thirteen binary 30-day outcomes (death, morbidity, serious morbidity, cardiac complication, pneumonia, venous thromboembolism, renal failure, urinary tract infection, surgical site infection, sepsis, return to the operating room, unplanned readmission, discharge to nursing or rehab facility) and one continuous outcome, length of stay (LOS). The SL model generates individual probability estimates (e.g., a value between 0 and 100) for each predicted outcome, providing patient-specific risk scores directly applicable to preoperative counseling. Like the ACS-RC, each of these 14 outcomes in this SL model was implemented using its predictive model, built upon the most recent 5 years of historical data, as detailed in the patient selection process^{25, 26}.

The Super Learner

Super Learner is an advanced ensemble learning method designed to optimize predictive accuracy by combining multiple machine learning algorithms. In recent years, there has been a growing stream of articles proposing SL applications for various prediction tasks, from mortality prediction²⁷ and waste generation²⁸ to predicting prison violence²⁹. By leveraging the strengths of multiple models, called base learners, the SL creates a highly accurate predictive algorithm, often outperforming individual techniques^{30, 31}.

The base learners can be chosen from any machine learning algorithms, such as generalized additive models, random forests, or neural networks, and they can be selected based on their individual strengths and/or the specific characteristics of the dataset. This confers excellent flexibility to the procedures by allowing the use of several diverse ML techniques. Each algorithm is trained on the dataset, and their predictions are combined using a meta-learner, typically a non-negative least squared-based method, or another machine learning algorithm targeted to minimize some pre-specified performance metric³².

The meta-learner assigns weights to the predictions of the base learners based on their performance, thereby creating a more robust and accurate predictive model. This strategy allows the SL model to adapt to different data types and complex relationships within the dataset, providing superior predictive accuracy compared to single algorithm models. The flexibility and robustness of the SL model have made it particularly valuable in clinical settings, where it can be used to predict patient outcomes and complications and personalize treatment plans with greater precision, ultimately enhancing patient care and risk management³³⁻³⁵.

Statistical analysis

The predictive ability of the SL was assessed by comparing it to traditional logistic regression (LOG) and the extreme gradient boosting (XGB) used in the ACS-R. For the length of stay outcome, instead of the logistic regression the negative binomial regression (NBR) approach is used.

The dataset was randomly split into 80% of the records for the train data and 20% for the test data. The performance evaluation metrics for the three approaches are based on results from the test set.

Discrimination, namely the ability to reliably classify patients into those that will or will not experience the event, was evaluated using the area under the receiver operating characteristic curve (AUROC) and the area under the precision-recall curve (AUPRC) for the 13 binary outcomes.

For both AUROC and AUPRC, higher values indicate better model performance. The mean squared error (MSE) was instead used to evaluate the accuracy of length of stay predictions; for this metric lower values reflects more accurate predictions.

For AUROC, AUPRC, and the MSE bootstrap, confidence intervals based on 5000 bootstrap samples are also reported.

Calibration, which measures the accuracy of predicted risk across different predicted risk scores, was evaluated using the Hosmer-Lemeshow (HL) test with 10 groups, while general accuracy using the Brier score³⁶ and the Log-Loss metric, also known as Cross-Entropy Loss.³⁷

HL is a general and well-known test calibration of predicted probabilities, higher p-values (conventionally greater than 0.05) indicates a better calibration. The Brier score and the Log-Loss instead aim to assess the overall accuracy of the predictions. The Brier score is a calibration-sensitive metric which measures mean squared difference between predicted probabilities and actual outcomes. The Log-Loss is a likelihood-based metric that strongly penalizes incorrect predictions made with high confidence and can be used for ranking models on probabilistic performance. For both Brier score and Log-Loss lower values are to be preferred.

Differences in performance between SL and XGB, as well as between SL and LOG, are assessed using the DeLong test³⁸ for AUROC and non-parametric bootstrap procedures for the remaining metrics. Specifically, 95% confidence intervals for the differences in performance metrics (SL vs XGB and SL vs LOG) are reported, along with one-tailed p-values testing the superiority of the SL approach over the other two methods. To control the false discovery rate, the Benjamini–Yekutieli correction has been applied to the p-values.³⁹

Regression modeling for the LOG and XGB competitors closely resembles that used in the original RC calculator and that proposed by Liu et al.²⁶.

For the base learners of the SL, the following algorithms were selected: Bayesian Additive Regression Trees⁴⁰, Gradient Tree Boosting⁴¹, Generalized Additive Models⁴², Multivariate Adaptive Regression Splines⁴³, Faster Random Forest⁴⁴, Feed-Forward Neural Networks⁴⁵ and Multivariate Adaptive Polynomial Spline Regression⁴⁶. This set of learners has been carefully chosen to guarantee maximum accuracy and flexibility through parametric and non-parametric techniques. As for the meta learner algorithm, the Nelder-Mead method⁴⁷ targeted at maximizing the AUROC has been adopted for dichotomous outcomes, while for the length of stay variables, the non-negative least squares method⁴⁸ has been used. The SL parameter specification and XGB hyperparameters have been tuned with 10-fold cross-validation. All the analyses were performed using R version 4.3.0⁴⁹, while the SL has been implemented using the *SuperLearner* and *sl3* packages^{50, 51}.

RESULTS

The baseline characteristics, used as predictors of the 667,444 patients included in the study are detailed in **Supplementary Table 1**, Supplemental Digital Content 1, <http://links.lww.com/SLA/F544>. These individual colorectal patient datasets were sourced from the PUF files from 2018-2022, further categorized by procedure location: anus (16,279), appendix (213,856), intestine (406,078), colon (30,784), mesentery and Meckel's diverticulum (447). **Supplementary Figure 1**, Supplemental Digital Content 1, <http://links.lww.com/SLA/F544> presents a diagram outlining the patient selection process. The evaluation of the three approaches is based on their performance on the test set.

Postoperative complication prediction analysis is presented in **Tables 1** and **2**. The SL model consistently outperformed both XGB and LOG across all complications. This superior performance was evident in both discriminative abilities, as measured by the Area Under the

Receiver Operating Characteristic curve (AUROC), and predictive power, as measured by the Area Under the Precision-Recall curve (AUPRC).

Formal statistical testing confirmed that the SL model's improvements in AUROC and AUPRC were statistically significant for the majority of outcomes when compared to both XGB and LOG (**Supplementary Table 2** and **Supplementary Table 3**, Supplemental Digital Content 1, <http://links.lww.com/SLA/F544>).

The SL model was superior in predicting severe postoperative complications. For the outcome of death, the SL model achieved an Area Under the Receiver Operating Characteristic curve (AUROC) exceeding 0.94 and an Area Under the Precision-Recall curve (AUPRC) greater than 0.38. The model also exhibited high AUROC values (0.88-0.91) for other severe complications, such as postoperative renal failure, cardiac complication, and unplanned return to the operating room. Furthermore, it performed well in predicting infectious complications like surgical site infection and sepsis, with AUROC values ranging from 0.78 to 0.83.

While the SL model, along with XGB and LOG, faced challenges in predicting less frequent complications like urinary tract infection (AUROC 0.69) and deep vein thrombosis (AUROC 0.76), it still maintained a slight advantage over the other models in terms of AUROC and AUPRC. This suggests the SL model's ability to integrate multiple learning algorithms may provide a more comprehensive and accurate patient risk assessment, even for less common postoperative events. These results underscore the potential of the SL approach as a valuable tool for predicting a wide range of postoperative complications, particularly for severe events where early identification and intervention are crucial. While SL's performance varied across different complications, its consistent superiority over other established machine learning approaches

highlights its potential for enhancing patient care and risk stratification in the postoperative setting.

The Hosmer-Lemeshow test results, shown in **Supplementary Table 4**, Supplemental Digital Content 1, <http://links.lww.com/SLA/F544>, assess the goodness of fit of the models. The SL model fits well with most binary outcomes, with high p-values (above 0.05) in many cases. However, some outcomes, such as Unplanned readmission, Surgical site infection, Urinary tract infection, and Death, showed poor fit ($p < 0.05$).

These results are also corroborated by the calibration plots presented in **Supplementary Figure 2**, Supplemental Digital Content 1, <http://links.lww.com/SLA/F544>, which consistently demonstrates that the SL data points lie closer to the diagonal compared to those from the XGB and the regression methods, indicating that SL achieves superior calibration performance.

Supplementary Table 5, Supplemental Digital Content 1, <http://links.lww.com/SLA/F544> depicts the Brier score, which measures the accuracy of probabilistic predictions; the values are multiplied by 1000 to increase readability. Lower values are preferred since they point to better calibration and discrimination, namely a smaller distance between observed and predicted probabilities of the event. The results indicated that the SL model outperformed XGB and LOG, suggesting better predictive performance for all binary outcomes.

To verify this, statistical comparisons were performed. These tests confirmed that the SL model's improvement in the Brier score was significant across all thirteen outcomes (**Supplementary Table 6**, Supplemental Digital Content 1, <http://links.lww.com/SLA/F544>), while the improvement in Log-Loss was significant in most, but not all, comparisons (**Supplementary Table 7**, Supplemental Digital Content 1, <http://links.lww.com/SLA/F544>).

The performance metric for predicting length of stay is the mean squared error (MSE); the SL model showed the lowest value, indicating the highest accuracy among the three models (**Supplementary Table 8**, Supplemental Digital Content 1, <http://links.lww.com/SLA/F544>).

Graphical representations of the ROC curves for the binary postoperative outcomes can be found in **Figures 1 to 3**, and **Supplementary Figures 3 to 12**, Supplemental Digital Content 1, <http://links.lww.com/SLA/F544>.

DISCUSSION

The SL model predicted the postoperative length of stay and a range of short-term postoperative outcomes with a high degree of accuracy. The SL model showed statistically significant improvements in predictive performance over the XGB and logistic regression models for all evaluated postoperative outcomes.

Machine learning techniques like Super Learner offer a valuable tool for surgeons to enhance patient care by providing personalized risk assessments and predictions for surgical outcomes¹⁴. Such models can identify a spectrum of patient-specific factors that contribute to postoperative risk.

The Super Learner model's ability to identify these modifiable risk factors empowers surgeons to implement targeted, evidence-based interventions that can meaningfully impact surgical outcomes. Unlike static risk factors, these modifiable elements allow for proactive risk reduction strategies, fostering patient engagement in their surgical preparation and potentially improving both short-term and long-term outcomes.

Several key modifiable risk factors can be addressed through this personalized approach.

Smoking cessation, for instance, represents one of the most impactful modifiable risk factors identified by predictive models. Evidence shows that even short-term smoking cessation (as brief as 4-8 weeks preoperatively) can significantly reduce postoperative complications, including wound healing problems, respiratory complications, and cardiovascular events. The Super Learner model's capacity to identify high-risk smokers enables targeted counseling and pharmacological interventions to support cessation efforts⁵².

Similarly, glycemic control in diabetic patients presents another critical opportunity for preoperative optimization. Poor glycemic control is clearly associated with increased rates of surgical site infections, delayed wound healing, and longer hospital stays. Patients identified as high-risk through predictive modeling can therefore benefit from intensified diabetes management, including endocrinology consultation and adherence to perioperative glucose protocols⁵³. Furthermore, nutritional status and prehabilitation have emerged as increasingly important modifiable factors. Malnutrition, often identified through albumin levels and BMI, significantly impacts surgical outcomes. The Super Learner model can effectively identify patients who would benefit most from preoperative nutritional optimization, protein supplementation, or formal prehabilitation programs that combine exercise, nutrition, and psychological preparation⁵⁴.

Lastly, cardiac risk stratification and optimization represent a well-established area for preoperative intervention. Patients identified as high-risk for cardiac complications can benefit from cardiology consultation, medication optimization, and in some cases, cardiac intervention before surgery⁵⁵.

This personalized approach to risk assessment and modification represents a paradigm shift from traditional "one-size-fits-all" preoperative preparation to individualized optimization strategies. By identifying which patients are most likely to benefit from specific interventions, the Super Learner model enables more efficient allocation of resources and more targeted patient counseling, ultimately enhancing the quality of perioperative care.

The potential economic impact of implementing the SL model in colorectal surgery is a crucial aspect to consider. The SL model could significantly enhance resource allocation by providing accurate preoperative risk assessments. For instance, patients identified as high-risk could be prioritized for intensive care unit beds or closer postoperative monitoring. In contrast, lower-risk patients could be allocated to step-down units, earlier discharge, or same-day discharge⁵⁶⁻⁵⁸, optimizing bed utilization and potentially reducing overall healthcare costs⁵⁹.

Furthermore, the SL model's ability to predict specific complications like surgical site infections, sepsis, unplanned readmissions, and reoperations could enable targeted interventions⁶⁰. By proactively identifying high-risk patients, the model may facilitate clinicians' early recognition and intervention of these complications, potentially leading to shorter hospital stays, fewer readmissions, and decreased healthcare utilization⁶¹. These reductions could translate into substantial cost savings for healthcare providers and payers.

This study presents several limitations, primarily from its retrospective design and reliance on administrative data. While the dataset is extensive, it may not fully capture the nuances of individual patient risk factors or surgical details that could influence outcomes. Additionally, the models were developed and validated using data from a specific healthcare system, potentially limiting the generalizability of our findings to other settings. The timeframe of the dataset (2018-

2022) may not fully reflect the most recent advancements in surgical techniques or patient care, potentially impacting the long-term predictive accuracy of the model.

Notably, while the SL model demonstrated superior performance to other models, it's crucial to acknowledge that even the most advanced models are not infallible. Therefore, the predictions generated by this model should be viewed as a valuable tool to augment clinical judgment rather than a definitive replacement for it.

CONCLUSION

The SL model shows potential for enhancing the accuracy of postoperative risk prediction in colorectal surgery. Given its superior performance compared to existing models, we will integrate it into our clinical practice during routine preoperative visits. This integration will enable us to offer tailored strategies for minimizing postoperative complications, thereby enhancing personalized patient care.

Looking ahead, we plan to automate this process by incorporating the algorithm into our electronic chart review program. This automation will allow for the automatic assessment of colorectal patient data and the generation of personalized risk profiles, further improving our preoperative risk assessment capabilities and optimizing resource allocation.

Subsequently, the next step will be to analyze surgeons' strategies based on these personalized risk profiles. This analysis will inform the development of an additional machine learning module that can automatically suggest optimal strategies for each patient, supporting surgeons in their decision-making process.

We remain committed to monitoring the real-world impact of the SL model in our clinical setting. Findings from this ongoing evaluation will be published, contributing to the broader knowledge base and aiming to improve preoperative risk assessment in colorectal surgery.

Looking further ahead, once we have thoroughly assessed and refined its application in colorectal surgery, we plan to extend SL model implementation to all surgical specialties at an institutional level. This expansion represents our commitment to leveraging advanced predictive modeling across the full spectrum of surgical care.

REFERENCES

1. Sarker IH. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*. 2021/03/22 2021;2(3):160. doi:10.1007/s42979-021-00592-x
2. Cuocolo R, Caruso M, Perillo T, Ugga L, Petretta M. Machine Learning in oncology: A clinical appraisal. *Cancer Lett*. Jul 1 2020;481:55-62. doi:10.1016/j.canlet.2020.03.032
3. Varlamova EV, Butakova MA, Semyonova VV, et al. Machine Learning Meets Cancer. *Cancers (Basel)*. Mar 8 2024;16(6)doi:10.3390/cancers16061100
4. Swanson K, Wu E, Zhang A, Alizadeh AA, Zou J. From patterns to patients: Advances in clinical machine learning for cancer diagnosis, prognosis, and treatment. *Cell*. Apr 13 2023;186(8):1772-1791. doi:10.1016/j.cell.2023.01.035
5. Kayikcioglu E, Onder AH, Bacak B, Serel TA. Machine learning for predicting colon cancer recurrence. *Surg Oncol*. Jun 2024;54:102079. doi:10.1016/j.suronc.2024.102079
6. Sammut SJ, Crispin-Ortuzar M, Chin SF, et al. Multi-omic machine learning predictor of breast cancer therapy response. *Nature*. Jan 2022;601(7894):623-629. doi:10.1038/s41586-021-04278-5
7. Huang C, Clayton EA, Matyunina LV, et al. Machine learning predicts individual cancer patient responses to therapeutic drugs with high accuracy. *Sci Rep*. Nov 6 2018;8(1):16444. doi:10.1038/s41598-018-34753-5
8. Peiffer-Smadja N, Rawson TM, Ahmad R, et al. Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. *Clin Microbiol Infect*. May 2020;26(5):584-595. doi:10.1016/j.cmi.2019.09.009
9. Susanto AP, Lyell D, Widyantoro B, Berkovsky S, Magrabi F. Effects of machine learning-based clinical decision support systems on decision-making, care delivery, and patient outcomes: a scoping review. *J Am Med Inform Assoc*. Nov 17 2023;30(12):2050-2063. doi:10.1093/jamia/ocad180
10. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. *Lancet Oncol*. May 2019;20(5):e262-e273. doi:10.1016/s1470-2045(19)30149-4
11. Triantafyllidis AK, Tsanas A. Applications of Machine Learning in Real-Life Digital Health Interventions: Review of the Literature. *J Med Internet Res*. Apr 5 2019;21(4):e12286. doi:10.2196/12286
12. Shehab M, Abualigah L, Shambour Q, et al. Machine learning in medical applications: A review of state-of-the-art methods. *Comput Biol Med*. Jun 2022;145:105458. doi:10.1016/j.combiomed.2022.105458
13. Elfanagely O, Toyoda Y, Othman S, et al. Machine Learning and Surgical Outcomes Prediction: A Systematic Review. *J Surg Res*. Aug 2021;264:346-361. doi:10.1016/j.jss.2021.02.045
14. Mahajan A, Esper S, Oo TH, et al. Development and Validation of a Machine Learning Model to Identify Patients Before Surgery at High Risk for Postoperative Adverse Events. *JAMA Netw Open*. Jul 3 2023;6(7):e2322285. doi:10.1001/jamanetworkopen.2023.22285
15. Xue B, Li D, Lu C, et al. Use of Machine Learning to Develop and Evaluate Models Using Preoperative and Intraoperative Data to Identify Risks of Postoperative Complications. *JAMA Netw Open*. Mar 1 2021;4(3):e212240. doi:10.1001/jamanetworkopen.2021.2240
16. Merath K, Hyer JM, Mehta R, et al. Use of Machine Learning for Prediction of Patient Risk of Postoperative Complications After Liver, Pancreatic, and Colorectal Surgery. *J Gastrointest Surg*. Aug 2020;24(8):1843-1851. doi:10.1007/s11605-019-04338-2

17. Nwaiwu CA, Rivera Perla KM, Abel LB, et al. Predicting Colonic Neoplasia Surgical Complications: A Machine Learning Approach. *Dis Colon Rectum*. May 1 2024;67(5):700-713. doi:10.1097/dcr.0000000000003166
18. Moaven O, Tavolara TE, Valenzuela CD, et al. Machine Learning Models for Predicting the Outcomes of Surgical Treatment of Colorectal Liver Metastases. *J Am Coll Surg*. Apr 1 2023;236(4):884-893. doi:10.1097/xcs.0000000000000573
19. Traeger L, Bedrikovetski S, Hanna JE, Moore JW, Sammour T. Machine learning prediction model for postoperative ileus following colorectal surgery. *ANZ J Surg*. May 2 2024;doi:10.1111/ans.19020
20. Chen T, Shang C, Su P, et al. A Decision Tree-Initialised Neuro-fuzzy Approach for Clinical Decision Support. *Artif Intell Med*. Jan 2021;111:101986. doi:10.1016/j.artmed.2020.101986
21. 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*. Dec 2022;46(12):3100-3110. doi:10.1007/s00268-022-06728-1
22. Taha-Mehlitz S, Däster S, Bach L, et al. Modern Machine Learning Practices in Colorectal Surgery: A Scoping Review. *J Clin Med*. Apr 26 2022;11(9)doi:10.3390/jcm11092431
23. Laan MJvd, Polley EC, Hubbard AE. Super Learner. *Statistical Applications in Genetics and Molecular Biology*. 2007;6(1)doi:doi:10.2202/1544-6115.1309
24. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med*. Oct 16 2007;147(8):573-7. doi:10.7326/0003-4819-147-8-200710160-00010
25. Cohen ME, Liu Y, Hall BL, Ko CY. ACS NSQIP Risk Calculator Performance Across Multiple Domains of Operative Risk and Risk-associated Features. *Ann Surg*. May 13 2025;doi:10.1097/sla.0000000000006753
26. Liu Y, Ko CY, Hall BL, Cohen ME. American College of Surgeons NSQIP Risk Calculator Accuracy Using a Machine Learning Algorithm Compared with Regression. *J Am Coll Surg*. May 1 2023;236(5):1024-1030. doi:10.1097/xcs.0000000000000556
27. Pirracchio R, Petersen ML, Carone M, Rigon MR, Chevret S, van der Laan MJ. Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study. *Lancet Respir Med*. Jan 2015;3(1):42-52. doi:10.1016/s2213-2600(14)70239-5
28. Jassim MS, Coskuner G, Sultana N, Hossain SMZ. Forecasting domestic waste generation during successive COVID-19 lockdowns by Bidirectional LSTM super learner neural network. *Applied Soft Computing*. 2023/01/01/ 2023;133:109908. doi:https://doi.org/10.1016/j.asoc.2022.109908
29. Baćak V, Kennedy EH. Principled Machine Learning Using the Super Learner: An Application to Predicting Prison Violence. *Sociological Methods & Research*. 2019;48(3):698-721. doi:10.1177/0049124117747301
30. Dudoit S, van der Laan MJ. Asymptotics of cross-validated risk estimation in estimator selection and performance assessment. *Statistical Methodology*. 2005/07/01/ 2005;2(2):131-154. doi:https://doi.org/10.1016/j.stamet.2005.02.003
31. Polley ECavdL, Mark J. Super Learner In Prediction. *UC Berkeley Division of Biostatistics Working Paper Series*. 2010;Working Paper 266

32. Phillips RV, van der Laan MJ, Lee H, Gruber S. Practical considerations for specifying a super learner. *International Journal of Epidemiology*. 2023;52(4):1276-1285. doi:10.1093/ije/dyad023
33. Rose S. Mortality risk score prediction in an elderly population using machine learning. *Am J Epidemiol*. Mar 1 2013;177(5):443-52. doi:10.1093/aje/kws241
34. Pirracchio R, Petersen ML, van der Laan M. Improving propensity score estimators' robustness to model misspecification using super learner. *Am J Epidemiol*. Jan 15 2015;181(2):108-19. doi:10.1093/aje/kwu253
35. Pirracchio R, Carone M. The Balance Super Learner: A robust adaptation of the Super Learner to improve estimation of the average treatment effect in the treated based on propensity score matching. *Stat Methods Med Res*. Aug 2018;27(8):2504-2518. doi:10.1177/0962280216682055
36. BRIER GW. VERIFICATION OF FORECASTS EXPRESSED IN TERMS OF PROBABILITY. *Monthly Weather Review*. 01 Jan. 1950 1950;78(1):1-3. doi:https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2
37. Bishop CM, Nasrabadi NM. *Pattern recognition and machine learning*. vol 4. Springer; 2006.
38. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;837-845.
39. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*. 2001:1165-1188.
40. Chipman HA, George EI, McCulloch RE. BART: Bayesian additive regression trees. *The Annals of Applied Statistics*. 2010;4(1):266-298, 33.
41. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016; San Francisco, California, USA. https://doi.org/10.1145/2939672.2939785
42. Hastie T, Tibshirani R. Generalized Additive Models. *Statistical Science*. 1986;1(3):297-310, 14.
43. Friedman JH. Multivariate Adaptive Regression Splines. *The Annals of Statistics*. 1991;19(1):1-67, 67.
44. Wright MN, Ziegler A. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software*. 03/31 2017;77(1):1 - 17. doi:10.18637/jss.v077.i01
45. Ripley BD. *Pattern Recognition and Neural Networks*. Cambridge University Press; 1996.
46. Kooperberg C, Bose S, Stone CJ. Polychotomous Regression. *Journal of the American Statistical Association*. 1997/03/01 1997;92(437):117-127. doi:10.1080/01621459.1997.10473608
47. Nelder JA, Mead R. A Simplex Method for Function Minimization. *Comput J*. 1965;7:308-313.
48. Goldfarb D, Idnani A. A numerically stable dual method for solving strictly convex quadratic programs. *Mathematical Programming*. 1983/09/01 1983;27(1):1-33. doi:10.1007/BF02591962
49. Team RC. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. 2014.

50. Polley E LE, Kennedy C, van der Laan M. SuperLearner: Super Learner Prediction. R package version 2.0-29, <<https://CRAN.R-project.org/package=SuperLearner>>. 2024;
51. Coyle J HN, Malenica I, Phillips R, Sofrygin O. sl3: Pipelines for Machine Learning and Super Learning. 2024;doi:10.5281/zenodo.1342293
52. Thomsen T, Villebro N, Møller AM. Interventions for preoperative smoking cessation. *Cochrane Database Syst Rev*. Mar 27 2014;2014(3):Cd002294. doi:10.1002/14651858.CD002294.pub4
53. 16. Diabetes Care in the Hospital: Standards of Care in Diabetes-2025. *Diabetes Care*. Jan 1 2025;48(Supplement_1):S321-s334. doi:10.2337/dc25-S016
54. Larson DW, Abd El Aziz MA, Perry W, et al. Additional Value of Preoperative Albumin for Surgical Risk Stratification among Colorectal Cancer Patients. *Ann Nutr Metab*. 2020;76(6):422-430. doi:10.1159/000514058
55. Thompson A, Fleischmann KE, Smilowitz NR, et al. 2024 AHA/ACC/ACS/ASNC/HRS/SCA/SCCT/SCMR/SVM Guideline for Perioperative Cardiovascular Management for Noncardiac Surgery: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *J Am Coll Cardiol*. Nov 5 2024;84(19):1869-1969. doi:10.1016/j.jacc.2024.06.013
56. Ferrari D, Violante T, Moriarty JP, et al. Same-Day Ileostomy Closure Discharge Reduces Costs without Compromising Outcomes: An Economic Analysis. *Ann Surg*. Mar 28 2024;doi:10.1097/sla.0000000000006285
57. Ferrari D, Violante T, Bin Zubair A, et al. Rethinking postoperative care: same-day ileostomy closure discharge improves patient outcomes. *Journal of Gastrointestinal Surgery*. 2024/05/01/ 2024;28(5):667-671. doi:<https://doi.org/10.1016/j.gassur.2024.02.017>
58. Ferrari D, Violante T, Sassun R, et al. The safety and trends of same-day and short-stay stoma closure: are we really on the verge of a paradigm shift? *Journal of Gastrointestinal Surgery*. 2025/05/28/ 2025:102096. doi:<https://doi.org/10.1016/j.gassur.2025.102096>
59. Bihorac A, Ozrazgat-Baslanti T, Ebadi A, et al. MySurgeryRisk: Development and Validation of a Machine-learning Risk Algorithm for Major Complications and Death After Surgery. *Ann Surg*. Apr 2019;269(4):652-662. doi:10.1097/sla.0000000000002706
60. Mamlook REA, Wells LJ, Sawyer R. Machine-learning models for predicting surgical site infections using patient pre-operative risk and surgical procedure factors. *Am J Infect Control*. May 2023;51(5):544-550. doi:10.1016/j.ajic.2022.08.013
61. Chen KA, Joisa CU, Stem JM, Guillem JG, Gomez SM, Kapadia MR. Improved Prediction of Surgical-Site Infection After Colorectal Surgery Using Machine Learning. *Dis Colon Rectum*. Mar 1 2023;66(3):458-466. doi:10.1097/dcr.0000000000002559

Figure 1: SL: Super Learner, LOG: Logistic Regression, XGB: XGBoost

Figure 1: Receiver Operating Characteristic curve for Return to the Operating Room



Figure 2: SL: Super Learner, LOG: Logistic Regression, XGB: XGBoost

Figure 2: Receiver Operating Characteristic curve for Morbidity

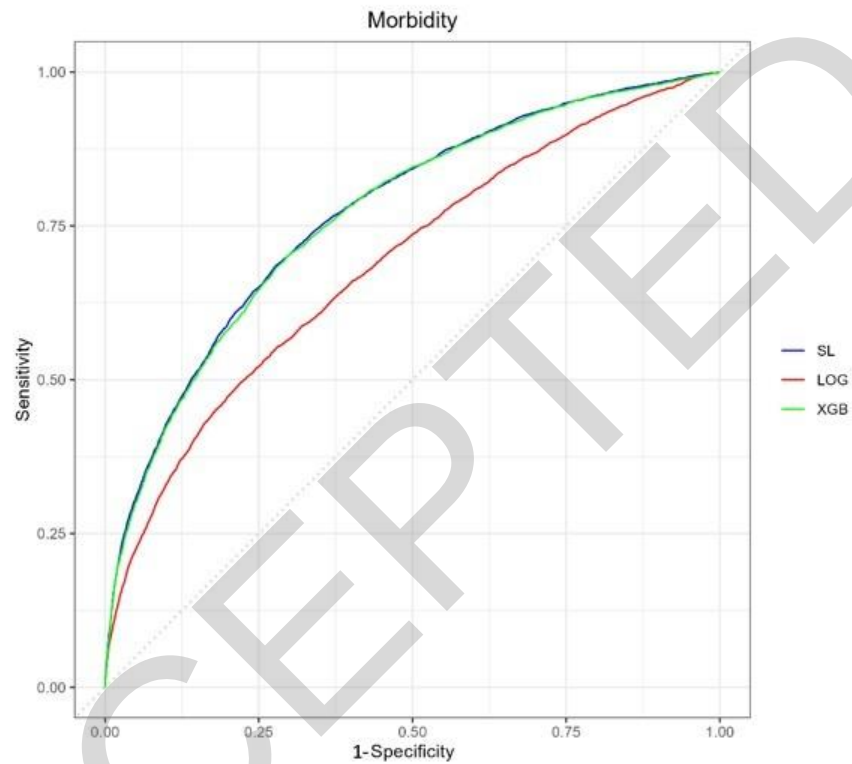


Figure 3: SL: Super Learner, LOG: Logistic Regression, XGB: XGBoost

Figure 3: Receiver Operating Characteristic curve for Serious Morbidity



Table 1: Area Under the Receiver Operating Characteristic Curve (AUROC) values for postoperative complications

	SL Value	CI	XGB Value	CI	LOG Value	CI
Pneumonia	0.928	(0.899, 0.951)	0.913	(0.882, 0.938)	0.927	(0.902, 0.949)
Unplanned readmission	0.68	(0.669, 0.691)	0.668	(0.657, 0.68)	0.644	(0.633, 0.656)
Venous thromboembolism	0.759	(0.732, 0.786)	0.721	(0.692, 0.749)	0.743	(0.711, 0.774)
Surgical site infection	0.712	(0.694, 0.731)	0.702	(0.683, 0.72)	0.645	(0.625, 0.666)
Cardiac complication	0.88	(0.853, 0.906)	0.86	(0.834, 0.887)	0.867	(0.838, 0.895)
Urinary tract infection	0.693	(0.669, 0.716)	0.681	(0.657, 0.706)	0.661	(0.636, 0.686)
Return to operating room	0.714	(0.699, 0.729)	0.695	(0.68, 0.71)	0.675	(0.659, 0.692)
Death	0.947	(0.939, 0.954)	0.942	(0.933, 0.95)	0.934	(0.925, 0.943)
Renal failure	0.884	(0.85, 0.913)	0.859	(0.823, 0.89)	0.869	(0.832, 0.902)
Sepsis	0.832	(0.819, 0.844)	0.826	(0.814, 0.839)	0.681	(0.667, 0.695)
Serious morbidity	0.781	(0.772, 0.788)	0.777	(0.769, 0.785)	0.697	(0.688, 0.707)
Morbidity	0.771	(0.763, 0.779)	0.768	(0.76, 0.776)	0.689	(0.68, 0.698)
Discharge to nursing or rehab facility	0.901	(0.894, 0.908)	0.9	(0.892, 0.907)	0.876	(0.868, 0.884)

SL: Super Learner, XGB: XGBoost, LOG: logistic regression, Value: AUROC value, CI: lower and upper bounds it of 95% Confidence Interval for AUROC. For AUROC higher values indicate better model performance.

Table 2: Area Under the Precision-Recall Curve (AUPRC) for postoperative complications

	SL		XGB		LOG	
	Value	CI	Value	CI	Value	CI
Pneumonia	0.095	(0.063, 0.14)	0.067	(0.046, 0.099)	0.083	(0.056, 0.129)
Unplanned readmission	0.145	(0.136, 0.155)	0.143	(0.133, 0.153)	0.126	(0.118, 0.135)
Venous thromboembolism	0.024	(0.019, 0.033)	0.021	(0.016, 0.03)	0.022	(0.018, 0.028)
Surgical site infection	0.059	(0.052, 0.069)	0.056	(0.049, 0.065)	0.044	(0.039, 0.049)
Cardiac complication	0.056	(0.037, 0.094)	0.043	(0.028, 0.073)	0.049	(0.032, 0.08)
Urinary tract infection	0.027	(0.023, 0.032)	0.028	(0.023, 0.033)	0.024	(0.021, 0.029)
Return to operating room	0.079	(0.07, 0.091)	0.072	(0.064, 0.082)	0.068	(0.061, 0.077)
Death	0.383	(0.341, 0.426)	0.355	(0.315, 0.399)	0.323	(0.284, 0.365)
Renal failure	0.07	(0.046, 0.113)	0.054	(0.034, 0.086)	0.06	(0.039, 0.096)
Sepsis	0.262	(0.24, 0.286)	0.252	(0.231, 0.276)	0.09	(0.083, 0.098)
Serious morbidity	0.412	(0.396, 0.43)	0.402	(0.386, 0.419)	0.308	(0.293, 0.324)
Morbidity	0.449	(0.433, 0.465)	0.442	(0.426, 0.458)	0.356	(0.341, 0.371)
Discharge to nursing or rehab facility	0.443	(0.417, 0.469)	0.428	(0.403, 0.455)	0.381	(0.356, 0.407)

SL: Super Learner, XGB: XGBoost, LOG: logistic regression, Value: AUPRC value, CI: lower and upper bounds of 95% Confidence Interval for AUPRC. For AUPRC, higher values indicate better model performance.