



Machine Learning Algorithms for Predicting Surgical Outcomes after Colorectal Surgery: A Systematic Review

Mustafa Bektaş¹ · Jurriaan B. Tuynman¹ · Jaime Costa Pereira² · George L. Burchell³ · Donald L. van der Peet¹

Accepted: 22 August 2022 / Published online: 15 September 2022
© The Author(s) 2022

Abstract

Background Machine learning (ML) has been introduced in various fields of healthcare. In colorectal surgery, the role of ML has yet to be reported. In this systematic review, an overview of machine learning models predicting surgical outcomes after colorectal surgery is provided.

Methods Databases PubMed, EMBASE, Cochrane, and Web of Science were searched for studies using machine learning models for patients undergoing colorectal surgery. To be eligible for inclusion, studies needed to apply machine learning models for patients undergoing colorectal surgery. Absence of machine learning or colorectal surgery or studies reporting on reviews, children, study abstracts were excluded. The Probst risk of bias tool was used to evaluate the methodological quality of machine learning models.

Results A total of 1821 studies were analysed, resulting in the inclusion of 31 articles. A vast proportion of ML algorithms have been used to predict the course of disease and response to neoadjuvant chemoradiotherapy. Radiomics have been applied most frequently, along with predictive accuracies up to 91%. However, most studies included a retrospective study design without external validation or calibration.

Conclusions Machine learning models have shown promising potential in predicting surgical outcomes after colorectal surgery. However, large-scale data is warranted to bridge the gap between calibration and external validation. Clinical implementation is needed to demonstrate the contribution of ML within daily practice.

Introduction

Colorectal cancer is estimated to have approximately 2 million new cases and 1 million deaths per year [1]. Appendicitis cases appeared to be approximately 18 million in the last few years [2]. Performing colorectal surgical procedures come with several risks, such as postoperative bleeding, anastomotic leakage, or fistulas [3]. These complications could become a burden for surgeons because they lead to readmissions of patients and require revision surgery. Additionally, in patients with colorectal cancer, tumor recurrence or metastasis are commonly discovered, causing a decrease in survival for these patients [4]. Although chemotherapy has already demonstrated improvements in survivability for colorectal cancer

✉ Mustafa Bektaş
m.bektas@amsterdamumc.nl

¹ Department of Surgery, Amsterdam UMC Location Vrije Universiteit Amsterdam, De Boelelaan 1117, 1081 HV Amsterdam, The Netherlands

² Department of Computer Science, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

³ Medical Library, Amsterdam UMC Location Vrije Universiteit Amsterdam, De Boelelaan 1117, 1081 HV Amsterdam, The Netherlands

patients, it is still difficult to predict which patients will completely respond to chemotherapy [5]. Therefore, risk stratification of patients with colorectal cancer remains challenging. Artificial Intelligence (AI) could support surgeons with this risk stratification by predicting postoperative complications, response to chemotherapy, and overall survival of colorectal cancer patients.

Recently, machine learning (ML), an essential branch of AI, has already been used for several complex tasks within healthcare. Examples of these tasks are the detection of tumors on radiologic images and prediction of biomarkers [6]. Due to its ability to train on large datasets and recognize patterns within data, machine learning algorithms are able to improve the accuracy of their prediction model [7]. Based on this capacity, machine learning models could be used to predict surgical outcomes prior to colorectal surgery [8]. By assessing several surgical outcomes with AI, surgeons could preoperatively decide the most efficient clinical pathway for patients undergoing colorectal surgery [9]. Currently, there are several machine learning algorithms available to make these predictions, an overview of algorithms is presented in Table 1.

Although machine learning algorithms have shown major potential to improve surgical outcomes, the current status and quality of machine learning models within colorectal surgery have not been evaluated in recent literature. However, it is essential to bridge this gap in order to understand the extent of predicted surgical outcomes, generalizability, and validity of current machine learning algorithms applied in colorectal surgery. Therefore, this systematic review aims to provide a comprehensive

overview of machine learning algorithms that have been used to predict any surgical outcome after general colorectal surgery. This review also evaluates the area under the curve and/or accuracy of included machine learning models.

Materials and methods

Literature was retrieved and systematically reviewed in accordance with the Cochrane Handbook for Systematic Reviews of Interventions version 6.0 and PRISMA guidelines.

Literature search strategy

A systematic search was performed in the databases: PubMed, Embase.com, Clarivate Analytics/Web of Science Core Collection and the Wiley/Cochrane Library. The timeframe within the databases was from inception to the 7th of July 2021 and conducted by G.L.B. and M.B. The search included keywords and free text terms for (synonyms of) 'machine learning' combined with (synonyms of) 'digestive system surgical procedures'. This search strategy was peer-reviewed by an information specialist (G.L.B.), using the PRESS checklist. A full overview of the search terms per database can be found in the supplementary information (see Appendix 1 as ESM). No limitations on date were applied in the search. Studies reporting on conference proceedings, book chapters, editorials,

Table 1 Terminology of AI subsets

Machine learning (ML)	Algorithms that are able to improve the prediction accuracies by training on large data [10]
Decision tree	A model that consists of nodes and branches, representing variables and related outcomes. Various combinations of outcomes give several predictions. The end model will be the smallest tree that fits the data best [11]
Gradient boosting (GBM)	Builds models that focus on inaccuracies of preceding models and improves these parts until the most accurate model is formed [12]
Random forest	Combines multiple decision trees to build the final accurate prediction model [13]
Support vector machine (SVM)	Finds the optimal border in the dataset to classify outcomes in two groups [14]
Artificial neural networks (ANNs)	Trains by using various processing layers to automatically find relevant features for the prediction. Additionally, weights of the extracted features are adjusted to form the most accurate model [15]
Convolutional neural networks (CNNs)	Similar to ANNs, except these models use filters instead of weight for extracted features [16]
Deep learning	Deep learning algorithms function similarly to neural networks, however, deep learning models have more layers or depth than neural networks [17]
Radiomics	Extracts quantitative features of clinical images to construct predictive or prognostic associations with the predicted medical outcomes [18]

ML machine learning, *SVM* support vector machine, *GBM* gradient boosting machine, *RF* random forest, *ANN* artificial neural networks, *CNN* convolutional neural networks

errata, letters, notes, surveys, or tombstones were excluded from the search.

Eligibility criteria

Studies were only eligible if they specifically met the following criteria: (i) described machine learning methods, (ii) involved patients undergoing any type of colorectal surgery, (iii) reported predictive performance of the machine learning model, (iv) clinical study. Regression models could be seen as machine learning. Nonetheless, regression models have existed in healthcare for many years. As this review is addressing new machine learning models only, regression models are therefore excluded from this review. In addition, appendectomy procedures were considered as colorectal surgery. Studies were excluded if they (i) were not written in English, (ii) reported on reviews, editorials, letters, or study abstracts. No specific study design or setting was preferred in the inclusion criteria.

Study selection

Two reviewers (M.B. & J.C.P.) independently performed the title and abstract screening in conformity with the inclusion and exclusion criteria. Eligible articles were read in full text, and duplicate studies were eliminated. The full-text screening of the retrieved articles was performed by the same two reviewers (M.B. & J.C.P.) to secure they comply with the inclusion criteria. Disagreements were resolved by discussions between two reviewers, resulting in consensus.

Risk of bias assessment

The Probst risk of bias tool was independently applied to each study by two reviewers (M.B. & J.C.P.) to assess the methodological quality of included machine learning models [19]. This tool is able to evaluate the overall risk of bias based on four bias domains: participant selection, predictors, outcomes, and analysis.

Data collection process

A table was formed for the extraction of all data. All data aspects were independently extracted and double-checked by two of the authors (M.B. & J.C.P.). Conflicts were resolved by consensus between the two authors. No additional processes were required for this data.

Data items

An inventory of data items was formed according to the Cochrane guidance for data collection, and the CHARMS checklist [20]. The following information was extracted from each study: first author, publication year, country of research, number of patients, mean age, study design, surgical procedure, intervention, surgical outcome, internal validation method, external validation, predictive performance (discrimination, and calibration). For studies involving multiple machine learning models, predictive performance of each model was described separately.

Data synthesis

A descriptive summary was used to represent the type of machine learning models, predicted surgical outcomes, risk of bias assessment, and model validation. To illustrate the predictive performance of machine learning models, results of machine learning studies were reported for each predicted outcome. To represent the discriminative ability, the range of mean accuracy (ACC) and area under the curve (AUC) was described for machine learning models of each predicted outcome. Additionally, the proportion of machine learning models that have applied calibration was described, along with the calibration method. A comparative meta-analysis of machine learning models was not possible, due to heterogeneity in study methodology, and the report on outcomes.

Results

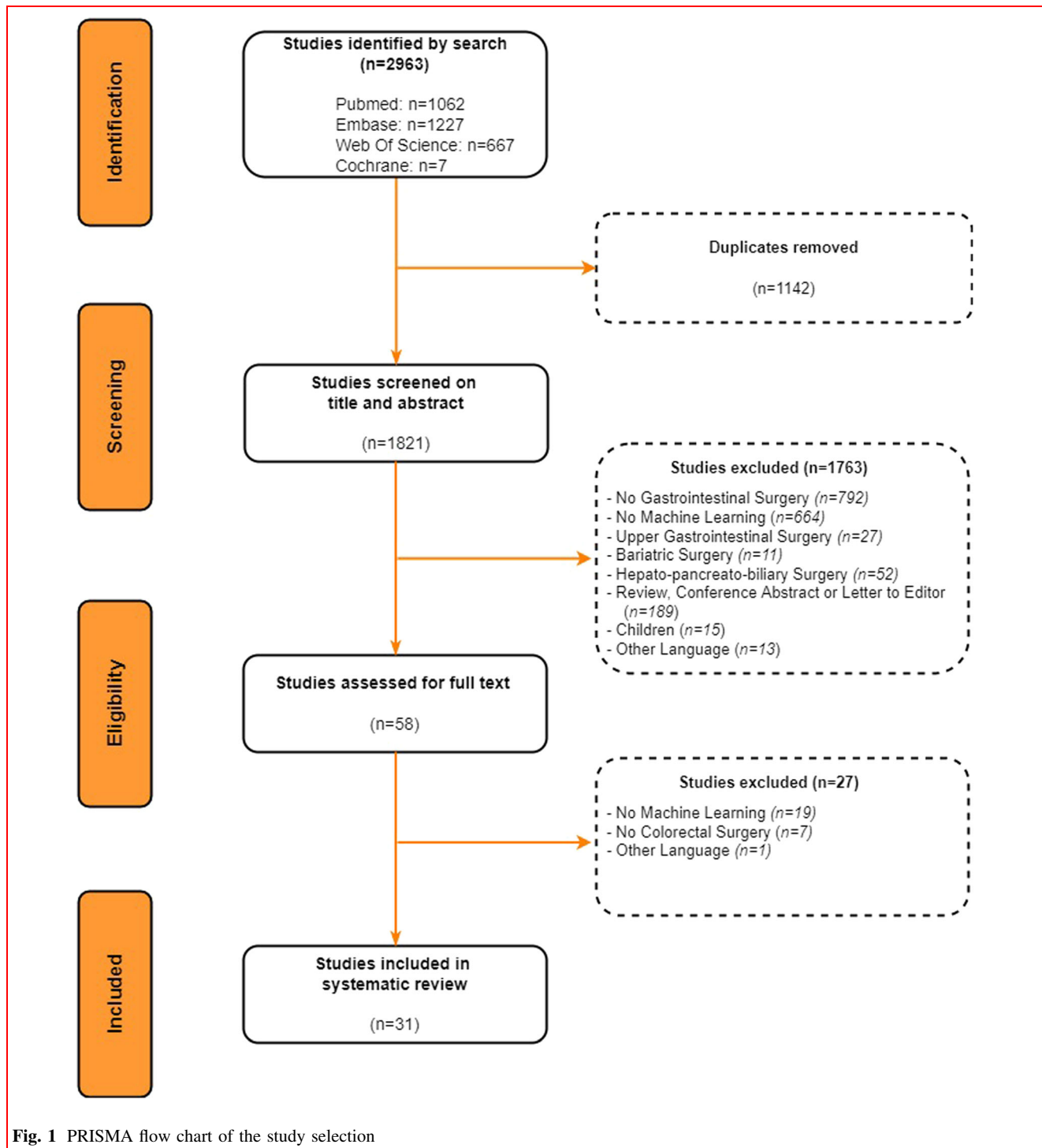
The search strategy provided a total of 1821 studies after removal of duplicates (Fig. 1). Therefore, 1821 studies were screened for eligibility based on the title and abstract. After excluding 1763 studies, 58 studies remained for a full-text assessment. In the end, 31 studies were included in this systematic review.

Machine learning models

Various machine learning algorithms have been applied to patients undergoing colorectal surgery. The frequencies of applied machine learning models were as follows: radiomics ($n = 13$), neural networks ($n = 7$), multiple machine learning ($n = 6$), random forest ($n = 4$), gradient boosting ($n = 1$).

Surgical outcomes

Surgical outcomes of these machine learning models predominantly included prediction of the clinical staging and



prognosis ($n = 9$), chemoradiotherapy response ($n = 7$), and postoperative complications ($n = 7$). Remaining studies involved prediction of diagnosis ($n = 4$), success of intervention ($n = 2$), and pre-and postoperative management ($n = 2$). An overview of key study characteristics is presented in Table 2.

Methodological quality assessment

Based on the Probst tool, the majority of studies received a low risk of bias score for the predictors and outcome domains. For most studies, the participants and analysis domains have received unclear or high risk of bias scores due to inappropriate inclusion criteria or measures to

Table 2 Key characteristics of included studies

Author	Year	Country	Patients	Age (mean)	Study design	Surgical procedures	Intervention (machine learning model)	Surgical outcomes	Internal validation method	External validation (yes/no)	Discrimination (ACC/AUC)	Calibration (yes/no)
Alvarez-Jimenez et al.	2020	USA	94	62	Retrospective Cohort	Total mesorectal excision	Radiomics	Prediction of tumour differentiation	Cross-validation	Yes	-0.73	No
Lee et al.	2020	Korea	2019	62	Retrospective Cohort	Colectomy	CNN; RF	Prediction of metastasis	Cross-validation	No	-0.70	No
Zhang et al.	2020	China	94	58	Retrospective Cohort	NS	Radiomics	Prediction of metastasis	Cross-validation	No	-0.88	Yes
Eresen et al.	2020	USA	390	63	Retrospective Cohort	Colectomy	Radiomics; SVM	Prediction of metastasis	Cross-validation	No	-0.83	No
Nakanishi et al.	2020	Japan	247	60	Retrospective Cohort	Total mesorectal excision	Radiomics	Prediction of metastasis	Bootstrapping	Yes	/0.91	Yes
Li et al.	2019	China	48	61	Retrospective Cohort	Colectomy	Radiomics; SVM	Prediction of metastasis	Cross-validation	No	0.86/0.87	No
Liang et al.	2019	China	108	55	Retrospective Cohort	Total mesorectal excision	Radiomics; SVM	Prediction of metastasis	Cross-validation	No	0.72/0.83	No
Li et al.	2020	China	207	62	Retrospective Cohort	Colectomy	Radiomics; SVM	Prediction of metastasis	Cross-validation	No	0.73/0.83	No
Dimitriou et al.	2018	UK	173	67	Retrospective Cohort	Colectomy	Multiple machine learning methods	Prediction of survival	Cross-validation	No	SVM-0.95 RF-0.92	No
Antunes et al.	2020	USA	104	63	Retrospective Cohort	Proctectomy	Radiomics; RF	Prediction of chemotherapy response	Cross-validation	No	0.71/0.71	No
Ferrari et al.	2019	Italy	55	65	Retrospective Cohort	Total mesorectal excision	Radiomics; RF	Prediction of chemotherapy response	Random split of dataset	No	-0.86	No
Yuan et al.	2020	USA	91	56	Retrospective Cohort	Total mesorectal excision	Radiomics	Prediction of chemotherapy response	Random split of dataset	No	0.84/-	No
Boyne et al.	2020	Canada	1378	64	Retrospective Cohort	Colectomy	RF	Prediction of chemotherapy response	Bootstrapping	No	-0.8	Yes
Fu et al.	2020	USA	43	54	Retrospective Cohort	Total mesorectal excision	Radiomics; CNN	Prediction of chemotherapy response	Cross-validation	No	-0.73	No

Table 2 continued

Author	Year	Country	Patients	Age (mean)	Study design	Surgical procedures	Intervention (machine learning model)	Surgical outcomes	Internal validation method	External validation (yes/no)	Discrimination (ACC/AUC)	Calibration (yes/no)
Shaish et al.	2020	USA	132	63	Retrospective Cohort	Total mesorectal excision	Radiomics	Prediction of chemotherapy response	Cross-validation	No	-0.80	No
Yi et al.	2019	China	134	52	Retrospective Cohort	Total mesorectal excision	Radiomics	Prediction of chemotherapy response	Random split of dataset	No	-0.91	No
Weller et al.	2018	USA	4773	NS	Retrospective Cohort	NS	Multiple machine learning methods	Prediction of postoperative complications	Cross-validation	No	GBM-0.64 RF-0.6 SVM-0.65	No
Chen et al.	2019	USA	13,399	58	Retrospective Cohort	NS	GBM	Prediction of postoperative complications	Cross-validation	No	-0.82	No
Azimi et al.	2020	USA	208	58	Retrospective Cohort	Colectomy	Multiple machine learning methods	Prediction of postoperative complications	Bootstrapping	No	DT0.86/- RF0.93/- SVM0.81/- ANN 0.88/-	No
Bunn et al.	2020	USA	223,214	40	Retrospective Cohort	Appendectomy	Multiple machine learning methods	Prediction of postoperative complications	Random split of dataset	No	GBM-0.93 RF-0.96 SVM-0.5	No
Adams et al.	2013	UK	76	61	Retrospective Cohort	NS	ANN	Prediction of postoperative complications	Random split of dataset	Yes	-0.89	No
Wen et al.	2021	USA	5220	59	Retrospective Cohort	Anterior resection + Total mesorectal excision	RF	Prediction of postoperative complications	Cross-validation	Yes	-0.85	No
Cao et al.	2020	Sweden	157	76	Retrospective Cohort	Emergency colectomy	RF	Prediction of mortality	Cross-validation	No	0.81/0.93	No
Akmese et al.	2020	Turkey	128	NS	Retrospective Cohort	Appendectomy	Multiple machine learning methods	Prediction of diagnosis	Random split of dataset	No	DT0.81/- GBM0.95/- RF 0.93/- SVM 0.80/- ANN 0.65/-	No

Table 2 continued

Author	Year	Country	Patients	Age (mean)	Study design	Surgical procedures	Intervention (machine learning model)	Surgical outcomes	Internal validation method	External validation (yes/no)	Discrimination (ACC/AUC)	Calibration (yes/no)
Hsieh et al.	2010	Taiwan	180	40	Retrospective Cohort	NS	Multiple machine learning methods	Prediction of diagnosis	Cross-validation	No	RF 0.96/0.98 SVM 0.93/0.96 ANN 0.91/0.91	No
Yoldas et al.	2011	Turkey	156	30	Retrospective Cohort	NS	ANN	Prediction of diagnosis	Random split of dataset	No	-0.95	No
Prabhudesai et al.	2007	UK	60	25	Retrospective Cohort	NS	ANN	Prediction of diagnosis	Random split of dataset	No	0.98/-	No
Gardiner et al.	2004	UK	72	NS	Retrospective Cohort	Anterior sphincter repair	ANN	Prediction of intervention success	Random split of dataset	No	0.93/-	No
Mamilich et al.	2012	USA	3754	38	Retrospective Cohort	proctocolectomy	RF	Prediction of intervention success	Random split of dataset	No	NS	No
Curtis et al.	2019	UK	668	70	Retrospective Cohort	NS	ANN	Prediction of pre- and postoperative management	Cross-validation	No	-0.86	No
Francis et al.	2015	UK	275	NS	Retrospective Cohort	NS	ANN	Prediction of pre- and postoperative management	Random split of dataset	No	-0.75	No

ACC accuracy, AUC area under the curve, NA not applicable, NS not specified

account for overfitting and missing data. Therefore, a low overall bias was given for 29% of the studies, whereas 48% of the studies received an unclear overall bias. Additionally, a high overall bias was decided for 23% of the studies (Fig. 2).

Model validation

For internal validation of machine learning models, most studies used cross-validation ($n = 17$), a random split of the dataset ($n = 11$), or bootstrapping ($n = 3$). External validation was performed in four studies (13%), including two radiomics, one ANN, and one random forest model. The discriminative ability (AUCs) of these models ranged between 0.64 and 0.9, and the calibration was reported for one machine learning model.

Predictive performance

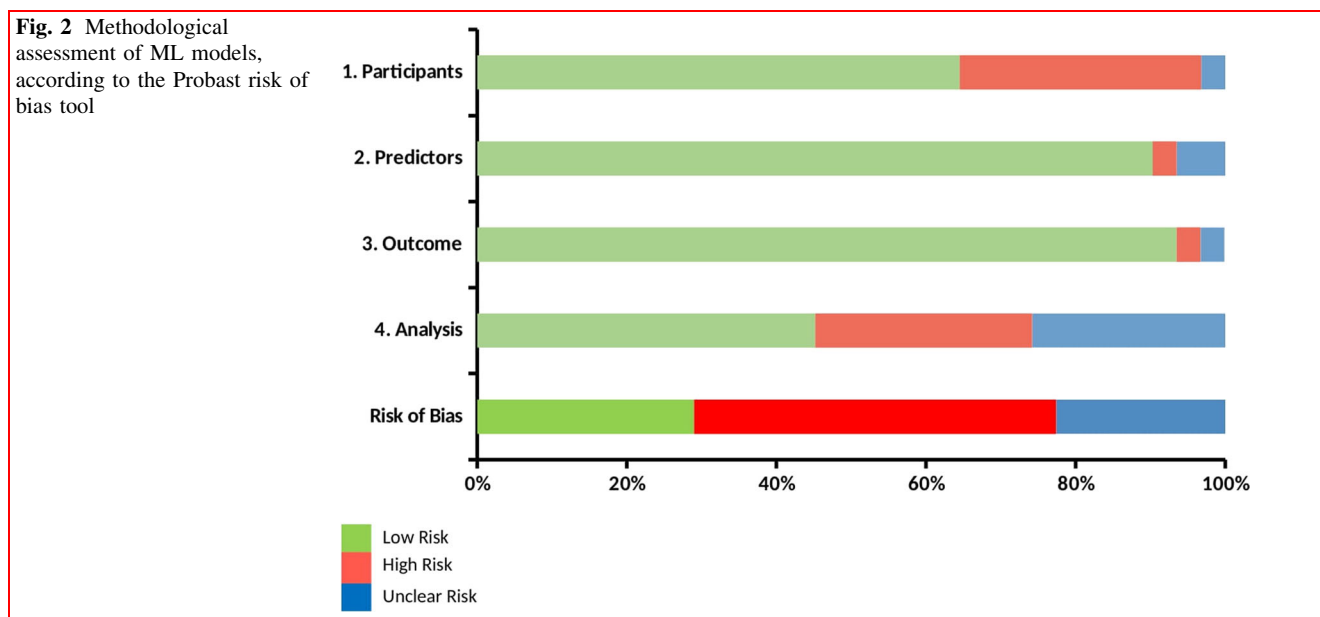
The AUCs of machine learning models for predicting clinical staging and prognosis have ranged between 0.7 and 0.95, whereas the accuracies were discovered to be between 72 and 86% [21–29]. For the prediction of chemotherapy response, AUCs between 0.71 and 0.91 were discovered, accuracies varied between 71 and 84% [30–36]. In the machine learning group for prediction of postoperative complications, AUCs have ranged from 0.6 and 0.96, additionally, accuracies were found to be between 81 and 93% [37–43]. For machine learning models predicting appendicitis cases, AUCs varied from 0.91 to 0.95, and accuracies ranged between 65 and 98% [44–47]. In the group for prediction of intervention success, AUCs were up to 0.93 [48, 49]. For prediction of pre-and

postoperative management, machine learning models showed AUCs up to 0.86 [50, 51]. Calibration was described for three models (10%), in which two studies used a Hosmer–Lemeshow test, and one study used a calibration plot only. Additionally, one study did not use AUC or accuracies to describe predictive performance of the machine learning model.

Discussion

This review illustrates the capabilities of machine learning in predicting several surgical outcomes for patients undergoing colorectal surgery. In this study, promising discriminative abilities of applied ML models have been discovered, especially for radiomic models.

Nine studies have used machine learning algorithms to predict the course of disease with accuracies ranging between 70 and 90%. Radiomics models have shown highest accuracies in these predictions. Theoretically, the use of ML could improve pre-operative decision-making for patients undergoing colorectal surgery, eventually enabling individualized surveillance for patients. For patients with high risks of metastasis, treatment decision such as minimal or aggressive surgery could be reconsidered for optimal surgical outcomes. However, most studies included small cohorts, this might give rise to the problem of overfitting, in which the ML model is overly adjusted to the training dataset and is unable to perform well on the test set [52, 53]. Although measures such as cross-validation and feature selection might help, this problem could be solved by including an external validation cohort [54].



Seven studies have applied machine learning to predict response to neoadjuvant chemoradiation therapy (nCRT) with accuracies between 71 and 91%. Radiomics appeared to perform this prediction with the highest accuracies. Although chemoradiotherapy has already shown improved outcomes for patients with advanced rectal cancer, incomplete therapy response and overtreatment of nCRT could occur [55]. Surgeons experience difficulties in determining patients who would completely respond to nCRT [56]. By using machine learning, surgeons could improve risk stratification, and decide to tailor therapy to patients with predicted nCRT response. This might eventually enable personalized decision-making for every patient, preventing unnecessary hospital stays and costs.

Seven studies have attempted to predict postoperative complications. Accuracies of ML models have ranged from 47 and 96%, in which random forests had the best predictive performance. Ideally, colorectal surgeons could use machine learning models to accurately predict postoperative complications for every patient. Subsequently, early discharge, enhanced monitoring or prophylactic steps could be implemented based on the predicted risk of complications. In addition, one study developed a predictive model for mortality in patients undergoing acute abdominal surgery [43]. This could potentially be helpful for clinical decision-making in acute surgery. Nonetheless, these ML studies have primarily included preoperative risk factors for postoperative complications. Previous studies have already indicated that postoperative complications are dependent upon several preoperative, intraoperative, and postoperative risk factors [57]. Therefore, more datasets are required to reveal essential intraoperative and postoperative factors for the prediction of postoperative complications.

For predicting patients with acute appendicitis, ML models have performed with accuracies up to 98%. Akmes et al. have demonstrated that ML could be applied with web-based interfaces, with internet as the only necessary criteria. Prabhudesai et al. have discovered that neural networks are able to predict appendicitis cases better than clinicians. These two findings may suggest that ML models could be practical and accurate tools for improving surgical decision-making. With proper use, surgeons could diagnose faster and prevent unnecessary appendectomies.

Although high accuracies have been found for machine learning models within this review, it seems that some uncertainties are still present. External validation was missing in most of the studies (87%), indicating that most machine learning models have not been applied to data from external hospital settings. However, external validation is crucial to demonstrate the generalizability of machine learning models [58]. Additionally, calibration was not reported in most studies (90%), while calibration

reflects the similarity between predicted risks and the true observed risks [59]. Poor calibration indicates that the machine learning model is under- or overestimating the desired outcome.

This review has some limitations. Due to the heterogeneity in methodologies of studies, a comparative meta-analysis of ML models was not possible. Additionally, a number of studies have not described predictive performances of ML models in ACC or AUCs, possibly leading to an over- or underrepresentation of actual discriminative abilities.

Future studies should focus on the external validation of ML models. Since external validation is important for the generalizability of machine learning algorithms, gaining this validation could facilitate the introduction of machine learning in daily clinical practice. However, large-scale datasets are required for this external validation, existing patient databases could be used to fulfill this need [60]. With proper use of these data, surgeons may achieve personalized decision-making for patients undergoing colorectal surgery. In addition, the calibration of machine learning models should be demonstrated in future studies to represent the extent of consensus between predicted outcomes and outcomes in the clinics.

In conclusion, this review shows the promising potential of ML in predicting various surgical outcomes for patients undergoing colorectal surgery. However, clinical implementation is required to demonstrate the contribution of ML within daily practice. The use of large patient databases may be required to fulfill the need for calibration and external validation.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00268-022-06728-1>.

Acknowledgements M Bektaş: participated in the design of the study, data collection and interpretation, wrote and submitted the manuscript. JB Tuynman: participated in the design of the study, and revised the manuscript critically. J Costa Pereira: participated in the design of the study, and interpretation of data. GL Burchell: performed the literature search and wrote parts of the manuscript. DL van der Peet: participated in the design of the study, revised the manuscript critically, and submitted the manuscript. All authors approved the final version of the manuscript.

Funding No grants or funding were received for this research.

Declarations

Conflict of interest The authors declare that they have no conflicts of interest.

Human or animal rights This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent No informed consent was required for this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Sung H, Ferlay J, Siegel RL et al (2021) Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 71(3):209–249
- Wickramasinghe DP, Xavier C, Samarasekera DN (2021) The worldwide epidemiology of acute appendicitis: an analysis of the global health data exchange dataset. *World J Surg* 45(7):1999–2008
- Pallan A, Dedelaite M, Mirajkar N et al (2021) Postoperative complications of colorectal cancer. *Clin Radiol* 76(12):896–907
- Xi Y, Xu P (2021) Global colorectal cancer burden in 2020 and projections to 2040. *Trans Oncol* 14(10):101174
- Jin M, Frankel WL (2018) Lymph node metastasis in colorectal cancer. *Surg Oncol Clin N Am* 27(2):401–412
- Yu KH, Beam AL, Kohane IS (2018) Artificial intelligence in healthcare. *Nat Biomed Eng* 2(10):719–731
- Visvikis D, Cheze Le Rest C, Jaouen V et al (2019) Artificial intelligence, machine (deep) learning and radio(geno)mics: definitions and nuclear medicine imaging applications. *Eur J Nucl Med Mol Imaging* 46(13):2630–2637
- Noorbakhsh-Sabet N, Zand R, Zhang Y et al (2019) Artificial intelligence transforms the future of health care. *Am J Med* 132(7):795–801
- Hashimoto DA, Rosman G, Rus D et al (2018) Artificial intelligence in surgery: promises and perils. *Ann Surg* 268(1):70–76
- El-Naqa I, Murphy MJ (2015) What is machine learning? Machine learning in radiation oncology. Springer, Cham, pp 3–11
- Song YY, Lu Y (2015) Decision tree methods: applications for classification and prediction. *Shanghai Arch Psychiatr* 27(2):130–135
- Friedman JH (2002) Stochastic gradient boosting. *Comput Stat Data Anal* 38(4):367–378
- Breiman L, Forests R (2001) In: *Machine learning*. Springer, New York, pp 5–32
- Zhang L, Zhou W, Jiao L (2004) Wavelet support vector machine. *IEEE Trans Syst Man Cybern B (Cybernetics)* 34(1):34–39
- Abraham A, Networks AN (2005) In: *Handbook of measuring system design*. Wiley, New Jersey, pp 901–908
- Albawi S, Mohammed TA, Al-Zawi S (2017) Understanding of a convolutional neural network. In: *International conference on engineering and technology (ICET)*, pp 1–6
- Rusk N (2016) *Deep learning*. In: *Nat methods*, New York
- Kumar V, Gu Y, Basu S et al (2012) Radiomics: the process and the challenges. *Magn Reson Imaging* 30(9):1234–1248
- Moons K, Wolff RF, Riley RD et al (2019) PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Ann Int Med* 170(1):W1–W33
- Moons KG, de Groot JA, Bouwmeester W et al (2014) Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med* 11(10):e1001744
- Alvarez-Jimenez C, Antunes JT, Talasila N et al (2020) Radiomic texture and shape descriptors of the rectal environment on post-chemoradiation T2-weighted MRI are associated with pathologic tumor stage regression in rectal cancers: a retrospective, multi-institution study. *Cancers* 12(8):2027
- Lee S, Choe EK, Kim SY et al (2020) Liver imaging features by convolutional neural network to predict the metachronous liver metastasis in stage I-III colorectal cancer patients based on pre-operative abdominal CT scan. *BMC Bioinform* 21(Suppl 13):382
- Zhang Y, He K, Guo Y et al (2020) A novel multimodal radiomics model for preoperative prediction of lymphovascular invasion in rectal cancer. *Front Oncol* 10:457
- Eresen A, Li Y, Yang J et al (2020) Preoperative assessment of lymph node metastasis in Colon Cancer patients using machine learning: a pilot study. *Cancer Imaging* 20(1):30
- Nakanishi R, Akiyoshi T, Toda S et al (2020) Radiomics approach outperforms diameter criteria for predicting pathological lateral lymph node metastasis after neoadjuvant (Chemo)radiotherapy in advanced low rectal cancer. *Ann Surg Oncol* 27(11):4273–4283
- Li Y, Eresen A, Shangguan J et al (2019) Establishment of a new non-invasive imaging prediction model for liver metastasis in colon cancer. *Am J Cancer Res* 9(11):2482–2492
- Liang M, Cai Z, Zhang H et al (2019) Machine learning-based analysis of rectal cancer MRI radiomics for prediction of metachronous liver metastasis. *Acad Radiol* 26(11):1495–1504
- Li Y, Eresen A, Shangguan J et al (2020) Preoperative prediction of perineural invasion and KRAS mutation in colon cancer using machine learning. *J Cancer Res Clin Oncol* 146(12):3165–3174
- Dimitriou N, Arandjelović O, Harrison DJ et al (2018) A principled machine learning framework improves accuracy of stage II colorectal cancer prognosis. *NPJ Dig Med* 1:52
- Antunes JT, Ofshteyn A, Bera K et al (2020) Radiomic features of primary rectal cancers on baseline T₂-weighted MRI are associated with pathologic complete response to neoadjuvant chemoradiation: a multisite study. *J Magnet Reson Imaging JMIR* 52(5):1531–1541
- Ferrari R, Mancini-Terracciano C, Voena C et al (2019) MR-based artificial intelligence model to assess response to therapy in locally advanced rectal cancer. *Eur J Radiol* 118:1–9
- Yuan Z, Frazer M, Zhang GG et al (2020) CT-based radiomic features to predict pathological response in rectal cancer: a retrospective cohort study. *J Med Imaging Radiat Oncol* 64(3):444–449
- Boyne DJ, Brenner DR, Sajobi TT et al (2020) Development of a model for predicting early discontinuation of adjuvant chemotherapy in stage III Colon Cancer. *JCO Clin Cancer Inform* 4:972–984
- Fu J, Zhong X, Li N et al (2020) Deep learning-based radiomic features for improving neoadjuvant chemoradiation response prediction in locally advanced rectal cancer. *Phys Med Biol* 65(7):075001
- Shaish H, Aukerman A, Vanguri R et al (2020) Radiomics of MRI for pretreatment prediction of pathologic complete response, tumor regression grade, and neoadjuvant rectal score in patients with locally advanced rectal cancer undergoing neoadjuvant chemoradiation: an international multicenter study. *Eur Radiol* 30(11):6263–6273

36. Yi X, Pei Q, Zhang Y et al (2019) MRI-based radiomics predicts tumor response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Front Oncol* 9:552
37. Weller GB, Lovely J, Larson DW et al (2018) Leveraging electronic health records for predictive modeling of post-surgical complications. *Stat Methods Med Res* 27(11):3271–3285
38. Chen D, Afzal N, Sohn S et al (2018) Postoperative bleeding risk prediction for patients undergoing colorectal surgery. *Surgery* 164(6):1209–1216
39. Azimi K, Honaker MD, Chalil Madathil S et al (2020) Postoperative infection prediction and risk factor analysis in colorectal surgery using data mining techniques: a pilot study. *Surg Infect* 21(9):784–792
40. Bunn C, Kulshrestha S, Boyda J et al (2021) Application of machine learning to the prediction of postoperative sepsis after appendectomy. *Surgery* 169(3):671–677
41. Adams K, Papagrigroriadis S (2014) Creation of an effective colorectal anastomotic leak early detection tool using an artificial neural network. *Int J Colorectal Dis* 29(4):437–443
42. Wen R, Zheng K, Zhang Q et al (2021) Machine learning-based random forest predicts anastomotic leakage after anterior resection for rectal cancer. *J Gastrointest Oncol* 12(3):921–932
43. Cao Y, Bass GA, Ahl R et al (2020) The statistical importance of P-POSSUM scores for predicting mortality after emergency laparotomy in geriatric patients. *BMC Med Inform Decis Mak* 20(1):86
44. Akmese OF, Dogan G, Kor H et al (2020) The use of machine learning approaches for the diagnosis of acute appendicitis. *Emerg Med Int* 2020:7306435. <https://doi.org/10.1155/2020/7306435>
45. Hsieh CH, Lu RH, Lee NH et al (2011) Novel solutions for an old disease: diagnosis of acute appendicitis with random forest, support vector machines, and artificial neural networks. *Surgery* 149(1):87–93
46. Yoldaş Ö, Tez M, Karaca T (2012) Artificial neural networks in the diagnosis of acute appendicitis. *Am J Emerg Med* 30(7):1245–1247
47. Prabhudesai SG, Gould S, Rekhraj S et al (2008) Artificial neural networks: useful aid in diagnosing acute appendicitis. *World J Surg* 32(2):305–311
48. Gardiner A, Kaur G, Cundall J et al (2004) Neural network analysis of anal sphincter repair. *Dis Colon Rectum* 47(2):192–197
49. Manilich E, Remzi FH, Fazio VW et al (2012) Prognostic modeling of preoperative risk factors of pouch failure. *Dis Colon Rectum* 55(4):393–399
50. Curtis NJ, Dennison G, Salib E et al (2019) Artificial neural network individualised prediction of time to colorectal cancer surgery. *Gastroenterol Res Pract* 2019:1285931. <https://doi.org/10.1155/2019/1285931>
51. Francis NK, Luther A, Salib E et al (2015) The use of artificial neural networks to predict delayed discharge and readmission in enhanced recovery following laparoscopic colorectal cancer surgery. *Tech Coloproctol* 19(7):419–428
52. Lee JG, Jun S, Cho YW et al (2017) Deep learning in medical imaging: general overview. *Korean J Radiol* 18(4):570–584
53. Wagner MW, Namdar K, Biswas A et al (2021) Radiomics, machine learning, and artificial intelligence-what the neuroradiologist needs to know. *Neuroradiology* 63(12):1957–1967
54. Ho SY, Phua K, Wong L et al (2020) Extensions of the external validation for checking learned model interpretability and generalizability. *Patterns* 1(8):100129
55. Valadão M, Dias JA, Araújo R et al (2020) Do we have to treat all T3 rectal cancer the same way?. *Clin Colorectal Cancer* 19(4):231–235
56. Body A, Prenen H, Lam M et al (2021) Neoadjuvant therapy for locally advanced rectal cancer: recent advances and ongoing challenges. *Clin Colorectal Cancer* 20(1):29–41
57. McDermott FD, Heeney A, Kelly ME et al (2015) Systematic review of preoperative, intraoperative and postoperative risk factors for colorectal anastomotic leaks. *Br J Surg* 102(5):462–479
58. Steyerberg EW, Harrell FE (2016) Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol* 69:245–247
59. Alba AC, Agoritsas T, Walsh M et al (2017) Discrimination and calibration of clinical prediction models: users' guides to the medical literature. *JAMA* 318(14):1377–1384
60. Chand M, Ramachandran N, Stoyanov D et al (2018) Robotics, artificial intelligence and distributed ledgers in surgery: data is key! *Tech Coloproctol* 22(9):645–648

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.