Surgery xxx (2023) 1–6

Contents lists available at ScienceDirect

Surgery

journal homepage: www.elsevier.com/locate/surg

Machine learning versus logistic regression for the prediction of complications after pancreatoduodenectomy

Erik W. Ingwersen, MD[a,b,c], Wessel T. Stam, MD[a,b,c], Bono J.V. Meijs, MD[a,b,c], Joran Roor, MSc[d], Marc G. Besselink, MD, PhD[b,c,e], Bas Groot Koerkamp, MD, PhD[f], Ignace H.J.T. de Hingh, MD, PhD[g], Hjalmar C. van Santvoort, MD, PhD[h], Martijn W.J. Stommel, MD, PhD[i], Freek Daams, MD, PhD[a,b,*], for the Dutch Pancreatic Cancer Group

[a] Amsterdam UMC, location Vrije Universiteit Amsterdam, Department of Surgery, Amsterdam, the Netherlands
[b] Cancer Center Amsterdam, the Netherlands
[c] Amsterdam Gastroenterology Endocrinology and Metabolism, the Netherlands
[d] SAS institute B.V., Huizen, the Netherlands
[e] Department of Surgery, Amsterdam UMC, University of Amsterdam, the Netherlands
[f] Erasmus MC, University Medical Center Rotterdam, the Netherlands
[g] Catharina Cancer Institute, the Netherlands
[h] UMC Utrecht Cancer Center, the Netherlands
[i] Radboud University Medical Center, the Netherlands

A R T I C L E I N F O

Article history:
Accepted 20 March 2023
Available online xxx

A B S T R A C T

Background: Machine learning is increasingly advocated to develop prediction models for postoperative complications. It is, however, unclear if machine learning is superior to logistic regression when using structured clinical data. Postoperative pancreatic fistula and delayed gastric emptying are the two most common complications with the biggest impact on patient condition and length of hospital stay after pancreatoduodenectomy. This study aimed to compare the performance of machine learning and logistic regression in predicting pancreatic fistula and delayed gastric emptying after pancreatoduodenectomy.
Methods: This retrospective observational study used nationwide data from 16 centers in the Dutch Pancreatic Cancer Audit between January 2014 and January 2021. The area under the curve of a machine learning and logistic regression model for clinically relevant postoperative pancreatic fistula and delayed gastric emptying were compared.
Results: Overall, 799 (16.3%) patients developed a postoperative pancreatic fistula, and 943 developed (19.2%) delayed gastric emptying. For postoperative pancreatic fistula, the area under the curve of the machine learning model was 0.74, and the area under the curve of the logistic regression model was 0.73. For delayed gastric emptying, the area under the curve of the machine learning model and logistic regression was 0.59.
Conclusion: Machine learning did not outperform logistic regression modeling in predicting postoperative complications after pancreatoduodenectomy.

Introduction

Pancreatoduodenectomy is a complex surgical procedure with considerable morbidity and a negative influence on short-term quality of life.[1] It is a substantial burden for the health care resource use and health expenditure, especially in those with a complicated postoperative recovery.[2,3] Postoperative pancreatic fistula (POPF) and delayed gastric emptying (DGE) are the two most common, high-impact complications after pancreatoduodenectomy, with sizable effects on resource use and prolonged length of stay.[4,5]

Both patients and clinicians would benefit from an accurate prediction of POPF and DGE, and if available, a proper risk assessment would be adopted widely.[6] In the preoperative setting, individual risk

* Reprint requests: Freek Daams, MD, PhD, Amsterdam UMC, Department of Surgery, De Boelelaan 1117, 1081 HV Amsterdam, PO Box 7075, 1007 MB Amsterdam, The Netherlands.
E-mail address: f.daams@amsterdamUMC.nl (F. Daams).

https://doi.org/10.1016/j.surg.2023.03.012
0039-6060/

assessment could aid clinical decision-making. Perioperatively, this could lead to an early awareness of complications, allowing for selective preventive and timely therapeutic measures. Most of these models are based on traditional logistic regression models, in which the probability of an outcome is related to a certain number of predictors. These models are relatively easy to use and interpret but have some drawbacks.[7] First, the usual assumption for logistic regression is that there is a linear relation between the independent and dependent variables. Second, predictors are usually chosen using backward selection. This technique has some problems (eg, it cannot be re-entered once a variable has been eliminated).[8]

Machine learning algorithms are increasingly advocated as they are less prone to the above-mentioned problems.[7,9] Machine learning algorithms can detect non-linear relationships between independent and dependent variables and incorporate many variables.[10,11] These models are, however, more susceptible to overfitting and have the so-called black-box phenomenon, meaning that the models are not interpretable for humans. This can undermine the use of these models in daily clinical practice.[12]

Although both techniques have been used to develop risk models for postoperative complications, it is unclear if machine learning is superior to logistic regression when using structured data.[13–16] This study used structured clinical data from a nationwide audit to compare machine learning to logistic regression in predicting POPF and DGE after pancreatoduodenectomy.

Methods

Study design and population

This retrospective cohort study used data from the nationwide Dutch Pancreatic Cancer Audit (DPCA), including data from 16 centers. In the Netherlands, all patients undergoing a pancreatic resection are prospectively registered in the DPCA. This type of study does not require approval from an ethics committee. The scientific committee of the Dutch Pancreatic Cancer Group approved the study protocol.[17] The DPCA data has been verified regarding its completeness and accuracy of postoperative complication registration, showing that the data represents the Dutch pancreatic cancer population.[18]

All patients undergoing elective, open, and minimally invasive pancreatoduodenectomy were included. Patients with no information about the outcome and those below 18 years old were excluded. The STROBE guidelines were followed to ensure correct reporting of the study methods and results.[19]

Outcome and definition

The primary outcomes were POPF and DGE, defined by the International Study Group of Pancreatic Surgery.[20,21] Only grade B/C POPF and DGE were included. The area under the curve–receiver operating characteristic (AUROC) was calculated to evaluate the discriminative power of the logistic regression and machine learning model for the prediction of POPF and DGE. Four machine learning models were examined: a random forest, a neural network, a support vector machine, and gradient boosting. An AUROC between 0.50 and 0.60 was defined as bad, 0.60 and 0.70 as poor, 0.70 and 0.80 as moderate, and 0.80 and 0.90 as good.[22–24]

Variables

The following preoperative patient characteristics and preoperative data were collected: age, sex, body mass index (BMI), weight loss per week, Eastern Cooperative Oncology Group performance score, and the latest measurement of preoperative: serum bilirubin, hemoglobin, albumin, IgG4, carcinoembryonic antigen (CEA), and cancer-antigen 19.9 (CA19.9). American Society of Anesthesiologists classification, American Joint Committee of Cancer stage, age-adjusted Charlson Comorbidity Index,[25] tumor location (ie, pancreatic body, head, periampullary, duodenum), tumor size, vessel involvement, tumor involvement in other organs, lymph nodes >10 millimeter, remote metastases, the sum of worrisome features,[26] histological diagnoses, application of neoadjuvant therapy, biliary drainage, type of stent, complications and/or re-intervention after biliary drainage, and the Response Evaluation Criteria in Solid Tumors criteria.

The following intraoperative data were collected: type of surgical procedure (ie, pylorus-preserving pancreaticoduodenectomy or pylorus resecting pancreaticoduodenectomy), pancreatic duct diameter, perioperative evidence of residual macroscopic tumor, arterial vessel resection, venous vessel resection, additional resections (including spleen, transversal mesocolon, colon segment resection, hemicolectomy, gastric resection or other resections), the texture of the pancreas, single row anastomosis or non-single row anastomosis, other perioperative measures (ie, drains, stents, probes, somatostatins), length of surgical procedure, and blood loss.

Statistical analysis

The analysis was performed using SAS VIYA version 2021.2.3 (SAS Institute, Inc, Cary, NC). Baseline pre and intraoperative characteristics were described. Normally distributed continuous variables were presented as mean with SD and continuous variables with skew distribution as median with an IQR. Dichotomous variables were noted as numbers and percentages. Differences between patients with or without a POPF and DGE were compared using Pearson χ^2 statistic in dichotomous variables and a Student's *t* test (normally distributed) or Mann-Whitney *U* test (skew distribution) in continuous variables.

Logistic regression model

Variables were excluded if >80% of the data were missing. Data were imputed via predictive mean matching with 10 iterations.[27] The pooled outcome of the 10 iterations was used. The logistic regression model was developed using backward selection, including all variables with at least a *P* value < .1 in at least six of the ten imputed datasets. The associations between the predictors and outcome were displayed as odds ratios (ORs), 95% CIs, and *P* values. The models were validated using bootstrapping, with 250 bootstrapping samples. To evaluate and compare the discriminative power of the models, the median AUROC with an IQR and the Brier score of the 250 bootstrap samples was calculated. The Brier score is a method to verify the accuracy of a probability forecast. A Brier score of 0 means perfect accuracy, and a score of 1 means perfect inaccuracy.[28]

Machine learning model

Variables were excluded if >80% of the data were missing. Continuous variables were imputed after median imputation, and the categorical variables were imputed after count imputation. The hyperparameters of the best machine learning model were determined and chosen using the autotuning tool: this tool seeks the best hyperparameters while keeping the risk of overfitting low. This hyperparameter optimization tool is integrated into SAS VIYA (SAS Institute, Inc). The models were validated using bootstrapping, with 250 bootstrapping samples. Supplementary Appendices S1 and S2 show how many times variables were included in the 250 different bootstrapped of the best-performing machine learning

Table I
Pre and intraoperative characteristics

Preoperative characteristics	Total ($N = 4,912$)	POPF B/C ($n = 799$)	DGE B/C ($n = 943$)	Missing (%)
Age, y (mean)	66.7 ± 10.5	67.1 ± 10.3	67.6 ± 10.2	8 (0.2%)
Sex (male)	2,720 (55.4%)	486 (60.8%)	551 (58.4%)	2 (<0.1%)
BMI, kg/m^2 (mean)	25.3 ± 4.3	26.6 ± 4.3	25.3 ± 4.3	166 (3.4%)
ASA classification ≥3	1,335 (27.6%)	224 (28.6%)	268 (29.0%)	71 (1.4%)
ECOG score ≥2	370 (8.8%)	43 (6.3%)	72 (8.7%)	688 (14.0%)
Hemoglobin, mmol/l (mean)	7.8 ± 1.1	7.9 ± 1.1	7.8 ± 1.1	2,095 (42.7%)
Albumin, mmol/l (mean)	36.5 ± 10.2	36.8 ± 7.7	36.2 ± 7.8	2,812 (57.2%)
Weight loss per week, kg (median)	0.4 (0-1)	0.1 (0–0.88)	0.2 (0–1.0)	1,567 (31.9%)
Bilirubin, mmol/l (median)	17 (7–66)	11.5 (7.0–29.8)	10.0 (6.0–38.5)	2,972 (60.5%)
CA19.9, U/mL (median)	23 (0–160)	24.5 (9.0–99.4)	27.0 (16.5–327)	347 (7.1%)
CEA, μg/l (median)	3.1 (1.9–5.0)	2.7 (1.7–4.7)	3.0 (2.0–5.5)	2,383 (48.5%)
IgG4, g/l (median)	0 (0–0.29)	0 (0–0.4)	0.1 (0–0.42)	674 (13.7%)
Diameter of tumor on CT-scan, mm (mean)	27.7 ± 22.2	28.2 ± 27.0	28.5 ± 28.1	1,924 (39.2%)
Vessel involvement	1,213 (25.9%)	116 (15.3%)	200 (22.3%)	233 (4.7%)
Involvement of structures	457 (9.8%)	67 (8.8%)	107 (12.0%)	234 (4.8%)
Lymph node involvement	583 (12.4%)	101 (13.2%)	115 (12.8%)	229 (4.7%)
Biliary drainage	2,526 (53.7%)	362 (45.3%)	442 (48.4%)	207 (4.2%)
Application of neoadjuvant therapy	426 (14.8%)	20 (4.0%)	74 (13.5%)	2 028 (41.3%)
CCIa ≥4	2,940 (59.9%)	497 (62.2%)	589 (62.5%)	0
Surgical procedure				
PRPD	2,396 (48.8%)	435 (54.4%)	502 (53.2%)	
PPPD	2,516 (51.2%)	364 (45.6%)	441 (46.8%)	
Approach				87 (1.8%)
Open	4,001 (82.9%)	591 (74.0%)	741 (80.5%)	
Laparoscopic	298 (6.2%)	69 (8.8%)	60 (6.5%)	
Robotic	526 (10.9%)	126 (16.0%)	119 (12.9%)	
Location of tumor				1 671 (34.0%)
Duodenum	264 (8.1%)	76 (14.6%)	72 (11.8%)	
Head of pancreas	2,414 (74.5%)	287 (55.3%)	406 (66.8%)	
Periampullary	563 (17.4%)	156 (30.1%)	130 (21.4%)	
Arterial resection	74 (1.5%)	15 (1.9%)	21 (2.2%)	46 (0.9%)
Venous resection	710 (14.6%)	57 (7.2%)	116 (12.4%)	45 (0.9%)
Additional resection	464 (9.8%)	94 (12.2%)	135 (14.3%)	172 (3.5%)
Length of surgical procedure, min	340 ± 115	342 ± 113	341.6 ± 110.7	3,664 (74.6%)
Blood loss, ml (median)	450 (200–845)	450 (200–800)	500 (200–800)	3,545 (72.2%)
Soft pancreas	2 755 (61.7%)	606 (83.6%)	577 (67.2%)	450 (9.1%)
Diameter pancreatic duct	4.8 ± 4.6	3.5 ± 5.2	4.2 ± 4.5	1,442 (29.4%)
Single row pancreatic anastomosis	1,208 (24.6%)	145 (18.1%)	225 (23.9%)	494 (10.1%)
Intra-abdominal drains	4,725 (96.2%)	772 (97.1%)	901 (96.9%)	18 (0.4%)
Administration of somatostatin analog	2,990 (61.1%)	528 (66.4%)	589 (62.7%)	18 (0.4%)
Stent in pancreas anastomosis	1,619 (33.1%)	355 (44.7%)	349 (37.0%)	18 (0.4%)
Nasojejunal feeding tube	987 (20.2%)	106 (13.3%)	168 (17.9%)	18 (0.4%)

ASA, American Society of Anesthesiologists; *BMI;* body mass index; *CA19.9*, cancer-antigen 19.9; *CCIa*, adjusted comprehensive complication index*; CEA*, carcinoembryonic antigen; *CT*, computer tomography; *DGE*, delayed gastric emptying; *ECOG*, Eastern Cooperative Oncology Group; *IgG4*, immunoglobulin G4; *POPF*, postoperative pancreatic fistula; *PPPD*, pylorus preserving pancreaticoduodenectomy; PRPD, pylorus resecting pancreaticoduodenectomy.

model. The tables also show the mean importance of the variable and the total importance, which is the mean importance multiplied by the number included in the model. The mean importance of the variables was determined using the Gini Importance, which calculates each feature's importance as the sum over the number of splits that include the features proportionally to the number of samples it splits.[29] To evaluate and compare the discriminative power of the models, the median AUROC of the 250 bootstrap samples with an IQR was calculated.

Results

Between January 2014 and December 2020, 4,972 patients after pancreatoduodenectomy were registered in the DPCA. In total, 60 patients were excluded because of missing information on one of the outcomes. None of the included variables exceeded 80% of missing data. Among the 4,912 patients included, 799 (16.3%) patients developed a POPF, 943 (19.2%) patients DGE. Patients had a mean age of 66.7 (SD:10.5); the majority were male (55.4%) and underwent a procedure via laparotomy (82.9%). All pre- and intraoperative characteristics per postoperative outcomes are displayed in Table I.

Postoperative pancreatic fistula

The AUROC of the logistic regression for POPF after bootstrapping was .73 with a pooled Brier score of .12. Predictors in the logistic regression model were: soft pancreatic texture (OR: 2.70, $P < .001$), non-single row pancreatic anastomosis (OR: 1.41, $P < .002$), male sex (OR: 1.38, $P < .001$), a plastic stent for biliary drainage (OR: 1.12, $P = .043$), BMI (OR per kg/m^2 increase: 1.08, $P < .001$), bilirubin (per μmol/L increase: 0.99, $P = .053$), pancreatic duct size (OR per mm increase: 0.91, $P < .001$), weight loss (OR per kg weight loss: 0.89, $P = .012$), venous resection (OR: 0.74, $P = .003$), a nasojejunal feeding tube (OR: 0.61, $P < .001$), localization of the tumor, and surgical approach (Table II). The mean AUROC of the best-performing machine learning model (gradient boosting) for POPF after bootstrapping was .74 (IQR .73–.74), with a pooled Brier score of .13. The 10 variables with the highest sum of the relative importance were the following: BMI, the diameter of the pancreatic duct, pancreatic texture, CA19.9, hemoglobin, age, localization of the tumor, CEA, the diameter of tumor, and non-single row pancreatic anastomosis (Table III). Supplementary Appendix S1 shows the relative importance of all pre and intraoperative variables.

Table II
Predictors in the multivariate logistic regression model for POPF grade B/C

Variables	OR (95% CI)	*P* value
Soft pancreas	2.70 (2.19–3.33)	< .001
Non-single row anastomosis*	1.41 (1.40–1.76)	.002
Male sex	1.38 (1.17–1.62)	< .001
Plastic stent for biliary drainage	1.12 (1.00–1.26)	.043
BMI per kg/m^2 increase	1.08 (1.06–1.10)	< .001
Bilirubin per μmol/L increase	.99 (.99–1.00)	.053
Pancreatic duct size, per mm increase	.91 (.88–.95)	< .001
Weight loss per kg loss	.89 (.81–.97)	.012
Venous resection	.74 (.61–.91)	.003
Nasojejunal feeding tube	.61 (.48–.78)	< .001
Localization of tumor	–	–
Head of pancreas	Reference	–
Body of pancreas	1.15 (.65–2.02)	.640
Periampullary	1.74 (1.38–2.20)	< .001
Duodenum	1.61 (1.18–2.20)	.003
Approach	–	–
Open surgery	Reference	–
Laparoscopic	1.84 (1.37–2.49)	< .001
Robotic	1.28 (1.01–1.62)	.045

BMI, body mass index; *OR*, odds ratio; *POPF*, postoperative pancreatic fistula.
* This included the following techniques for PJ: Blumgart, duct-to-mucosa, and dunking/intussusception.[47]

Delayed gastric emptying

The AUROC of the logistic regression for DGE after bootstrapping was .59 with a pooled Brier score of .15. Predictors in the logistic regression model were: soft pancreatic texture (OR: 1.32, $P < .001$), a pylorus resecting pancreaticoduodenectomy (OR: 1.33, $P < .001$), the application of a somatostatin analog (OR: 1.18, $P = .002$), a plastic stent (OR: 1.15, $P = .010$), age (OR per year increase: 1.01, $P = .005$), bilirubin (OR per μmol/L increase: 0.99, $P = .007$), a nasojejunal feeding tube (OR 0.84, $P =$ 0..070), and localization of tumor (Table IV). The mean AUROC of the best-performing machine learning model (gradient boosting) for DGE after bootstrapping was 0.59 (IQR 0.58–0.60), with a pooled Brier score of 0.14. The 10 variables with the highest relative importance sum were BMI, age, CA19.9, pancreatic duct diameter, CEA, hemoglobin level, tumor diameter, albumin, weight loss, and non-single row anastomosis (Table V). Supplementary Appendix S2 shows the relative importance of the pre- and intraoperative variables.

Discussion

This first multicenter audit-based study to compare machine learning and logistic regression for prediction models found no superiority of the machine learning model for predicting POPF and DGE after pancreatoduodenectomy using structured pre- and intraoperative variables. For the machine learning and logistic regression prediction model, the predictive performance of the POPF model was moderate and was bad for DGE.

An accurate prediction of a POPF would be extremely helpful as it would support clinicians in identifying patients at risk, thereby improving perioperative decision-making.[30,31] Given the performance scores, the models developed in this study are not useful for daily clinical practice, especially considering several prediction models are available with higher AUROCs and fewer input variables. For example, the Fistula Risk Score, developed by Callery et al. This logistic regression model was based on four common POPF-related risk factors (small duct, soft pancreas, high-risk pathology, and excessive blood loss). After analysis and internal validation on 233 patients, the Fistula Risk Score performed moderately with an AUROC of 0.72.[32] Six years later, Mungroop et al developed an alternative Fistula Risk Score based on only 3 risk factors (small duct, soft pancreas, and high BMI). The group designed the model based on 1,924 patients, and the alternative score was externally validated and compared with the original fistula risk score based on 926 patients. The models achieved an AUROC of 0.78 for the alternative and 0.75 for the original fistula risk score in the external validation.[33] The models are easy to understand and convenient because they use only 3 or 4 predictors.

There are several reasons why machine learning did not outperform logistic regression, as was hypothesized in this study. As mentioned earlier, machine learning models can consider the nonlinear interaction between predictors and outcomes.[34] Most of the variables in this study were discrete values, although the potency of machine learning models lies in continuous data. These variables lie in a higher-dimensional space, and the predictors' relationship with the outcomes is more complex. The number of variables in this database may have been too small for machine learning to show its benefit. Moreover, only structured data was used. Machine learning showed its superiority over logistic regression when unstructured data was added.[35] Radiomic features derived from computed tomography or magnetic resonance images, for instance, might be able to provide a lot of unstructured data. The results from studies using radiomics to predict a POPF are promising, with an AUROC varying from 0.80 to 0.90 in several monocenter retrospective studies.[36–39] Future studies should further investigate its use and externally validate it, preferably using radiomics features and structured clinical data. It should also be acknowledged that not everything is predictable, and a perfect predictive performance is a utopia.[40] Although the etiology and the accompanying risk factors of POPF and DGE are partly understood,

Table III
The ten variables in the gradient boosting model with the highest sum of the mean relative importance in predicting POPF grade B/C

Variables	Included in prediction model	Mean relative importance	Sum
BMI	250	5.66	1,414.39
Diameter pancreatic duct	250	5.19	1,297.37
Soft pancreas	250	4.40	1,099.60
CA19.9	250	2.85	711.56
Hemoglobin	250	2.67	668.17
Age	250	2.44	609.17
Tumor location	250	2.27	566.33
CEA	250	1.97	491.36
Diameter of tumor	250	1.70	425.00
Non-single row anastomosis	250	1.62	404.51

Table IV shows the 10 most important variables in the gradient boosting model for predicting POPF. The entire model with all included variables and their importance is shown in Supplementary Appendix S1.
This included the following techniques for PJ: Blumgart, duct-to-mucosa, and dunking/intussusception.[47]
BMI; body mass index; *CA19.9*, cancer-antigen 19.9; *CEA*, carcinoembryonic antigen; *POPF*, postoperative pancreatic fistula.

Table IV
Predictors in the multivariate model for DGE grade B/C

Variables	OR (95% CI)	*P* value
Soft pancreas	1.32 (1.13–1.55)	< .001
Type surgical procedure–PRPD	1.33 (1.14–1.56)	< .001
Somatostatin analog	1.18 (1.07–1.31)	.002
Plastic stent	1.15 (1.03–1.27)	.010
Age per year increase	1.01 (1.00–1.02)	.005
Bilirubin per μmol/L increase	.99 (.99–1.00)	.007
Nasojejunal feeding tube	.84 (.70–1.01)	.070
Localization of tumor	–	–
Head of pancreas	Reference	–
Body of pancreas	.88 (.48–1.59)	.623
Periampullary	1.27 (1.03–1.57)	.029
Duodenum	1.34 (1.03–1.77)	.038

DGE, delayed gastric emptying; *OR*, odds ratio; *PRPD*; pylorus resecting pancreatoduodenectomy.

Table V
The ten variables in the gradient boosting model with the highest sum of the mean relative importance in predicting DGE

Variables	Included in prediction model	Mean relative importance	Sum
BMI	250	3.62	904.32
Age	250	3.02	755.54
CA19.9	250	2.69	672.80
Diameter pancreatic duct	250	2.34	584.68
CEA	250	2.12	530.58
Hemoglobin	250	2.11	527.92
Diameter of tumor	250	2.01	503.36
Albumin	250	1.69	422.43
Weight loss	250	1.62	404.62
Non-single row anastomosis	250	1.58	394.09

Table V shows the 10 most important variables in the gradient boosting model for predicting DGE. The entire model with all included variables and their importance is shown in Supplementary Appendix S2.
This included the following techniques for PJ: Blumgart, duct-to-mucosa, and dunking/intussusception.[47]
BMI; body mass index; *CA19.9*, cancer-antigen 19.9; *CEA*, carcinoembryonic antigen; *DGE*, delayed gastric emptying.

there might still be unknown and unrecognized factors (both in the pre and perioperative phase) playing roles of importance that are lacking in our current audit.

Despite the findings of this study, the use of machine learning models compared to logistic regression may have some advantages.[41,42] First, they are data-driven and can continuously improve because both algorithms can see patterns or features themselves without human intervention.[42] Second, each factor is individually reassessed where the complex interrelationships per patient between the different factors and outcomes are also considered. This is important in an individualized prediction because these interrelationships differ for each person, making some factors more relevant than others. Third, in the future, machine learning can easily be added to electronic health records, making it possible to easily and automatically predict complications based on the pre- and intraoperative information and unstructured information available.[42] All in all, accurate prediction of individual outcomes can reshape the future of postoperative management and offer the opportunity to develop a more individualized approach allowing for data-driven individualized medicine.[43–45]

Study Limitations

The results of this study should be interpreted considering several limitations. First, the proportion of missing data was large for some variables (≤75%), potentially causing bias and reduced efficiency. Because complete case analysis would drastically reduce the sample size, multiple imputations were performed. Madley-Dowd et al showed that imputation reduced bias even when the proportion of missingness was large (≤90%), and missing data was at random (as was the case in this study).[46] Variables were therefore only excluded if missingness exceeded 80%. Second, this study defined both POPF grades B and C as outcomes. Although grade B POPF is a complication that can significantly impact quality of life, it is not likely that the complication will determine whether a patient will undergo a pancreatoduodenectomy. A future prediction model should only predict a grade C POPF to accurately determine which patients are at high risk for this life-threatening complication. Third, pre- and intraoperative data were used to improve the outcomes' predictability. However, clinically it may be more helpful if the model uses only preoperative predictors. Finally, this study used multicenter data from a prospectively maintained audit. However, external and prospective validation is necessary to increase the generalizability of the findings. Strengths of this study include the large study population and its multicenter nationwide study design. Moreover, this study compared machine learning with logistic regression within the same database.

In conclusion, this study showed no difference in the predictive performance of machine learning compared with logistic regression in predicting POPF and DGE. Performances of both models are suboptimal and not useful for daily clinical practice. Future research should focus on adding radiomics to these models to objectively determine risks preoperatively.

Funding/Support

This research did not receive any specific funding from any agencies in the public, commercial, or not-for-profit areas.

Conflict of interest/Disclosure

The authors have no conflicts of interests or disclosures to report.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [https://doi.org/10.1016/j.surg.2023.03.012].

References

1. Nahm CB, Connor SJ, Samra JS, Mittal A. Postoperative pancreatic fistula: a review of traditional and emerging concepts. *Clin Exp Gastroenterol*. 2018;11:105–118.
2. Wang J, Ma R, Churilov L, et al. The cost of perioperative complications following pancreaticoduodenectomy: A systematic review. *Pancreatology*. 2018;18:208–220.
3. Wang J, Ma R, Eleftheriou P, et al. Health economic implications of complications associated with pancreaticoduodenectomy at a University Hospital: a retrospective cohort cost study. *HPB (Oxford)*. 2018;20:423–431.
4. Mirrielees JA, Weber SM, Abbott DE, Greenberg CC, Minter RM, Scarborough JE. Pancreatic fistula and delayed gastric emptying are the highest-impact complications after whipple. *J Surg Res*. 2020;250:80–87.
5. Smits FJ, Verweij ME, Daamen LA, et al. Impact of complications after pancreatoduodenectomy on mortality, organ failure, hospital stay, and readmission: analysis of a nationwide audit. *Ann Surg*. 2022;275:e222–e228.
6. Marchegiani G, Bassi C. Prevention, prediction, and mitigation of postoperative pancreatic fistula. *Br J Surg*. 2021;108:602–604.
7. Tu JV. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J Clin Epidemiol*. 1996;49:1225–1231.
8. Chowdhury MZI, Turin TC. Variable selection strategies and its importance in clinical prediction modelling. *Fam Med Community Health*. 2020;8:e000262.
9. Deo RC. Machine learning in medicine. *Circulation*. 2015;132:1920–1930.

10. Kuhn M, Johnson K. *Applied Predictive Modeling*. New York: Springer Science+Business Media LLC; 2013.
11. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436–444.
12. Guidi JL, Clark K, Upton MT, et al. Clinician perception of the effectiveness of an automated early warning and response system for sepsis in an academic medical center. *Ann Am Thorac Soc*. 2015;12:1514–1519.
13. Callery MP, Pratt WB, Kent TS, Chaikof EL, Vollmer Jr CM. A prospectively validated clinical risk score accurately predicts pancreatic fistula after pancreatoduodenectomy. *J Am Coll Surg*. 2013;216:1–14.
14. Han IW, Cho K, Ryu Y, et al. Risk prediction platform for pancreatic fistula after pancreatoduodenectomy using artificial intelligence. *World J Gastroenterol*. 2020;26:4453–4464.
15. 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*. 2020;24:1843–1851.
16. Stam WT, Goedknegt LK, Ingwersen EW, Schoonmade LJ, Bruns ERJ, Daams F. The prediction of surgical complications using artificial intelligence in patients undergoing major abdominal surgery: a systematic review. *Surgery*. 2022;171: 1014–1021.
17. Strijker M, Mackay TM, Bonsing BA, et al. Establishing and coordinating a nationwide multidisciplinary study group: lessons learned by the Dutch Pancreatic Cancer Group. *Ann Surg*. 2020;271:e102–e104.
18. van Rijssen LB, Koerkamp BG, Zwart MJ, et al. Nationwide prospective audit of pancreatic surgery: design, accuracy, and outcomes of the Dutch Pancreatic Cancer Audit. *HPB (Oxford)*. 2017;19:919–926.
19. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ*. 2007;335:806–808.
20. Bassi C, Marchegiani G, Dervenis C, et al. The 2016 update of the International Study Group (ISGPS) definition and grading of postoperative pancreatic fistula: 11 Years After. *Surgery*. 2017;161:584–591.
21. Welsch T, Borm M, Degrate L, Hinz U, Büchler MW, Wente MN. Evaluation of the International Study Group of Pancreatic Surgery definition of delayed gastric emptying after pancreatoduodenectomy in a high-volume centre. *Br J Surg*. 2010;97:1043–1050.
22. Tape TG. Interpretation of Diagnostic Tests. *Ann Intern Med*. 2001;135:72.
23. Trevethan R. Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice. *Front Public Health*. 2017;5:307.
24. Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol*. 2008;56:45–50.
25. de Groot V, Beckerman H, Lankhorst GJ, Bouter LM. How to measure comorbidity. a critical review of available methods. *J Clin Epidemiol*. 2003;56: 221–229.
26. Hipp J, Mohamed S, Pott J, et al. Management and outcomes of intraductal papillary mucinous neoplasms. *BJS Open*. 2019;3:490–499.
27. Morris TP, White IR, Royston P. Tuning multiple imputation by predictive mean matching and local residual draws. *BMC Med Res Methodol*. 2014;14:75.
28. Brier GW. Verification of forecasts expressed in terms of probability. *Mon Weather Rev*. 1950;78:1–3.
29. Nembrini S, König IR, Wright MN. The revival of the Gini importance? *Bioinformatics*. 2018;34:3711–3718.
30. Bootsma BT, Plat VD, van de Brug T, et al. Somatostatin analogues for the prevention of pancreatic fistula after open pancreatoduodenectomy: a nationwide analysis. *Pancreatology*. 2022;22:421–426.
31. Smits FJ, Henry AC, Besselink MG, et al. Algorithm-based care versus usual care for the early recognition and management of complications after pancreatic resection in the Netherlands: an open-label, nationwide, stepped-wedge cluster-randomised trial. *Lancet*. 2022;399:1867–1875.
32. Miller BC, Christein JD, Behrman SW, et al. A multi-institutional external validation of the fistula risk score for pancreatoduodenectomy. *J Gastrointest Surg*. 2014;18:172–179.
33. Mungroop TH, van Rijssen LB, van Klaveren D, et al. Alternative Fistula Risk Score for Pancreatoduodenectomy (a-FRS): design and international external validation. *Ann Surg*. 2019;269:937–943.
34. Gravesteijn BY, Nieboer D, Ercole A, et al. Machine learning algorithms performed no better than regression models for prognostication in traumatic brain injury. *J Clin Epidemiol*. 2020;122:95–107.
35. Zhang D, Yin C, Zeng J, Yuan X, Zhang P. Combining structured and unstructured data for predictive models: a deep learning approach. *BMC Med Inform Decis Mak*. 2020;20:280.
36. Capretti G, Bonifacio C, De Palma C, et al. A machine learning risk model based on preoperative computed tomography scan to predict postoperative outcomes after pancreatoduodenectomy. *Updates Surg*. 2022;74: 235–243.
37. Lin Z, Tang B, Cai J, et al. Preoperative prediction of clinically relevant postoperative pancreatic fistula after pancreaticoduodenectomy. *Eur J Radiol*. 2021;139:109693.
38. Skawran SM, Kambakamba P, Baessler B, et al. Can magnetic resonance imaging radiomics of the pancreas predict postoperative pancreatic fistula? *Eur J Radiol*. 2021;140:109733.
39. Zhang W, Cai W, He B, Xiang N, Fang C, Jia F. A radiomics-based formula for the preoperative prediction of postoperative pancreatic fistula in patients with pancreaticoduodenectomy. *Cancer Manag Res*. 2018;10:6469–6478.
40. Coggon DIW, Martyn CN. Time and chance: the stochastic nature of disease causation. *Lancet*. 2005;365:1434–1437.
41. Bertsimas D, Dunn J, Velmahos GC, Kaafarani HMA. Surgical risk is not linear: derivation and validation of a novel, user-friendly, and machine-learning-based Predictive OpTimal Trees in Emergency Surgery Risk (POTTER) calculator. *Ann Surg*. 2018;268:574–583.
42. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial intelligence in surgery: promises and perils. *Ann Surg*. 2018;268:70–76.
43. Beaulieu-Jones BK, Yuan W, Brat GA, et al. Machine learning for patient risk stratification: standing on, or looking over, the shoulders of clinicians? *NPJ Digit Med*. 2021;4:62.
44. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25:44–56.
45. Beam AL, Kohane IS. Translating artificial intelligence into clinical care. *JAMA*. 2016;316:2368–2369.
46. Madley-Dowd P, Hughes R, Tilling K, Heron J. The proportion of missing data should not be used to guide decisions on multiple imputation. *J Clin Epidemiol*. 2019;110:63–73.
47. Ratnayake CBB, Wells CI, Kamarajah SK, et al. Critical appraisal of the techniques of pancreatic anastomosis following pancreaticoduodenectomy: A network meta-analysis. *Int J Surg*. 2020;73:72–77.