BMC Medicine

RESEARCH

Open Access

Development and external validation of a machine learning-based model to predict postoperative recurrence in patients with duodenal adenocarcinoma: a multicenter, retrospective cohort study

Xu Liu[2†], Qifeng Xiao[1†], Zongting Gu[3], Xin Wu[4], Chunhui Yuan[5], Xiaolong Tang[6], Fanbin Meng[7], Dong Wang[8], Ren Lang[9], Kaiqing Guo[10], Xiaodong Tian[11], Yu Zhang[12], Enhong Zhao[13], Zhenzhou Wu[14], Jingyong Xu[15], Ying Xing[16], Feng Cao[17], Chengfeng Wang[1*] and Jianwei Zhang[1*]

Abstract

Background Duodenal adenocarcinoma (DA) has a high recurrence rate, making the prediction of recurrence after surgery critically important.

Methods Our objective is to develop a machine learning-based model to predict the postoperative recurrence of DA. We conducted a multicenter, retrospective cohort study in China. 1830 patients with DA who underwent radical surgery between 2012 and 2023 were included. Wrapper methods were used to select optimal predictors by ten machine learning learners. Subsequently, these ten learners were utilized for model development. The model's performance was validated using three separate cohorts, and assessed by the concordance index (C-index), time-dependent calibration curve, time-dependent receiver operating characteristic curves, and decision curve analysis.

Results After selecting predictors, ten feature subsets were identified. And ten feature subsets were combined with the ten machine learning learners in a permutation, resulting in the development of 100 predictive models, and the Penalized Regression + Accelerated Oblique Random Survival Forest model (PAM) exhibited the best predictive performance. The C-index for PAM was 0.882 (95% CI 0.860–0.886) in the training cohort, 0.747 (95% CI 0.683–0.798) in the validation cohort 1, 0.736 (95% CI 0.649–0.792) in the validation cohort 2, and 0.734 (95% CI 0.674–0.791) in the validation cohort 3. A publicly accessible web tool was developed for the PAM.

Conclusions The PAM has the potential to identify postoperative recurrence in DA patients. This can assist clinicians in assessing the severity of the disease, facilitating patient follow-up, and aiding in the formulation of adjuvant treatment strategies.

Keywords Duodenal adenocarcinoma, Cancer recurrence, Surgery, Machine learning-based model

†Xu Liu and Qifeng Xiao contributed equally to this work.

*Correspondence:
Chengfeng Wang
wangchengfeng62@163.com
Jianwei Zhang
panchutong@163.com
Full list of author information is available at the end of the article

Background

Small intestinal adenocarcinomas are rare, with a lower prevalence than adenocarcinomas in other parts of the gastrointestinal tract. Nevertheless, there has been a reported increase in the incidence of duodenal adenocarcinoma (DA) [1–3]. Pathologically, DA originates from the intestinal mucosal epithelium, most commonly affecting the descending duodenum [4]. Although the prognosis of DA is reportedly better than that of other periampullary malignancies, the 5-year survival rate is still low, at about 50% [5–7]. Additionally, the recurrence rate of DA reaches 30–60% [8]. Therefore, the treatment and prognosis of DA should receive more attention.

Due to the rarity of DA, the prognostic and recurrence risk factors for this condition have not yet been determined. Only limited studies indicate that non-R0 resection, vascular invasion, postoperative complications, lymph node metastasis, and peritoneal invasion are adverse prognostic factors for DA [7, 9–13]. Furthermore, the predictive capability of the AJCC 8th edition staging system for prognosis is poor [14]. Therefore, there is an urgent need to further explore the prognostic factors of DA and to develop models that can accurately predict patient outcomes.

The progress in big data and high-performance computing technologies has made machine learning an essential component in medical research and clinical practice [15]. Machine learning technology can sift through vast amounts of data to uncover subtle correlations that may not be immediately apparent to human analysts, thereby offering new insights into disease mechanisms, patient outcomes, and the effectiveness of various treatments [16–18].

Therefore, our research aims to develop and validate a machine learning-based model to predict recurrence in DA patients. By employing a variety of sophisticated algorithms, we strive to offer a more precise and individualized risk assessment tool, thereby enabling early interventions and improving treatment outcomes. In addition, we intend to create a web-based version of the risk calculator to broaden the reach of our model.

Methods

This study strictly adhered to the Prediction model Risk Of Bias Assessment Tool (PROBAST) standards and a checklist for useful clinical prediction tools reported by Florian Markowetz, and followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) Checklist for reporting (Additional file 1) [19–21]. The complete research process of this study is shown in Fig. 1.

Study population

This multicenter, retrospective cohort study involved patients with DA who underwent radical surgery between 2012 and 2023 at 16 Chinese hospitals which are all tertiary grade A hospitals in China. Prior to surgery and the subsequent follow-up survey, each patient provided informed consent. The surgical procedures were expertly conducted by seasoned senior surgeons. Adherence to ethical standards was ensured by conducting the study in compliance with the Declaration of Helsinki (revised in 2013), and ethical approval was secured from the Hospital Ethics Committee of the National Cancer Center (No. 22/213–3415). Patients were divided into a training cohort (National Cancer Center; Beijing Tiantan Hospital; Peking University First Hospital; Shanxi Cancer Hospital; Affiliated Hospital of Chengde Medical University; Shaanxi Provincial People's Hospital; Beijing Friendship Hospital; Chinese Pla General Hospital; Qilu Hospital of Shandong University; Xuanwu Hospital Capital Medical University; The Second Hospital of Hebei Medical University; Beijing Hospital; And The First Hospital of China Medical University) and three independent external validation cohorts (Peking University Third Hospital (validation cohort 1), Beijing Chao-Yang Hospital (validation cohort 2)), and Zhejiang Provincial People's Hospital (validation cohort 3)).

Inclusion and exclusion criteria

Adult participants with DA who underwent Pancreaticoduodenectomy or Pylorus-preserving pancreaticoduodenectomy were included. The exclusion criteria were as follows: (1) died during the perioperative period; (3) lost to follow-up; (3) insufficient clinical data.

Data collection and blind assessment

Before starting the data collection process, training sessions on the data extraction form were held for all staff involved at the participating centers. The medical records of eligible patients were reviewed. Given the variations in units across participating centers, the units for all laboratory measurements were standardized. All preoperative laboratory measurements should be the most recent measurements conducted prior to the surgery, and all postoperative laboratory measurements should be conducted on the first day after the surgery.

For assessing predictive outcomes for recurrence, and to preserve the integrity of predictor assessment, all evaluations were performed by researchers blinded to the participants' predicted outcomes and the values of other predictors. This method aimed to minimize any potential biases or preconceived notions that might influence the evaluation results. An independent assessment team was established specifically for this purpose. The team

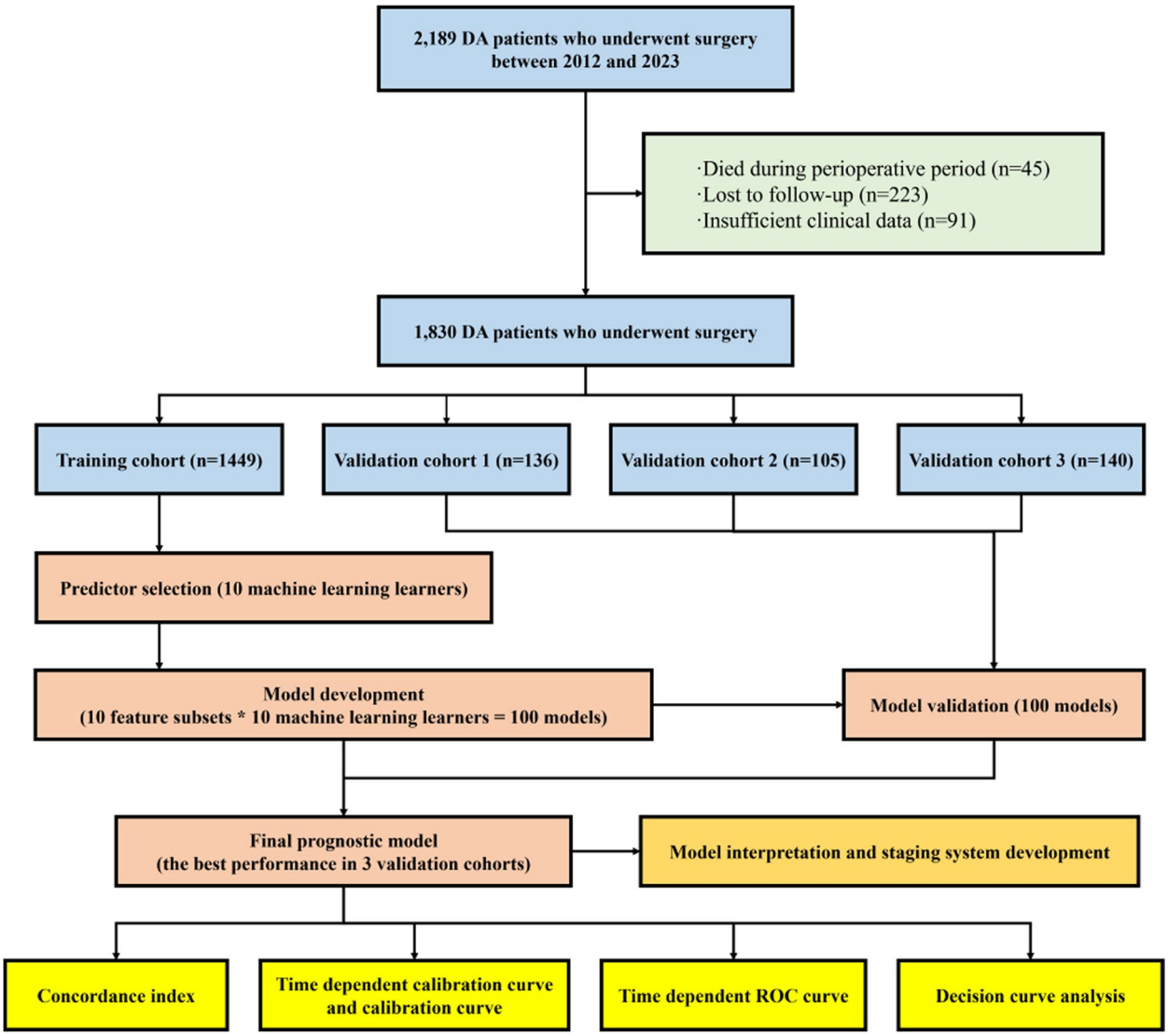

Fig. 1 Flow diagram of the study

members were deliberately kept uninformed about the machine learning model's recurrence predictions and the input predictors. This was to ensure that the verification of recurrence was based purely on the clinical data of the participants, free from any influence of the model's predictive outcomes.

Predictor selection

In the training cohort, over 35% of missing parameters were excluded from the analysis. 53 variables of the training cohort, including sex, age, body mass index (BMI), stomach ache, abdominal bloating, jaundice, fever, poor appetite, emaciation, gastrointestinal obstruction, biliary obstruction, coronary heart disease, stroke, hypertension, diabetes, personal cancer history, liver dysfunction, smoking history, drinking history, family cancer history, preoperative white blood cell (WBC) count, preoperative hemoglobin level, preoperative platelet (PLT) count, preoperative CA19.9 level, preoperative CEA level, preoperative total bilirubin (TBIL) level, preoperative direct bilirubin (DBIL) level, operation time, surgical approach, surgical procedure, intraoperative blood loss, volume of intraoperative blood transfusion, maximum diameter of the tumor, nerve invasion, vascular tumor thrombus, tumor site, invasion of the muscularis propria, invasion of adjacent organs or structures, number of positive lymph nodes, postoperative alanine aminotransferase (ALT) level, postoperative aspartate aminotransferase (AST) level, postoperative alkaline phosphatase (ALP) level, postoperative TBIL level, postoperative DBIL level, postoperative pancreatic fistula,

postoperative biliary fistula, postoperative gastroparesis, abdominal hemorrhage, gastrointestinal bleeding, abdominal infection, postoperative ileus, incisional infection, and secondary surgery, were included in ten machine learning learners to select predictors.

We chose widely recognized machine learning learners capable of handling both continuous and categorical variables. These included: Akritas estimator (AKE), Gradient Boosting (GB), Boosted Generalized Additive Model (GAMB), Boosted Generalized Linear Model (GLMB), Survival Tree (ST), Conditional Inference Tree (CIT), Random Survival Forest (RSF), Conditional Random Forest (CRF), Accelerated Oblique Random Survival Forest (AORSF), Penalized Regression (PR). These learners were all sourced from the "mlr3proba" R package [22].

The AKE is a non-parametric estimation method used in regression analysis to handle censored or truncated dependent variables. It estimates the relationship between the dependent and independent variables through ranking and regression techniques [23]. GB is an ensemble learning algorithm that iteratively adds weak learners (typically decision trees) to fit the residuals of the previous step, progressively optimizing the loss function to enhance the overall predictive performance of the model [24]. The GAMB and GLMB are ensemble learning methods that enhance the predictive performance and adaptability of a model by iteratively adding and adjusting a series of base learners from Generalized Additive Models or Generalized Linear Models [25]. ST is a decision tree model used in survival analysis that recursively splits data to group individuals based on their survival times and the risks of event occurrence, thereby providing insights into survival probabilities [26]. Besides, ST can also be used to identify cut-offs. The CIT is a method for constructing decision trees that use statistical tests to select splitting variables and points, thereby reducing bias and overfitting in the model during the splitting process in a non-parametric and conditional inference-based manner [27]. RSF is an ensemble learning method used in survival analysis that constructs multiple survival trees and uses the results from these trees to estimate the survival functions and hazard ratios for individuals, thereby offering a powerful analysis of survival time data [28]. The CRF is an ensemble learning algorithm that builds multiple decision trees and uses conditional inference tests to select variables and split points. This approach enhances the robustness and accuracy of the model while reducing bias in variable selection [29]. The AORSF is a survival analysis model that combines the ensemble learning techniques of random forests with the approach of oblique decision trees. This integration enhances the accuracy and efficiency in handling complex survival data [30]. PR is a regression analysis technique that controls model complexity by adding penalty terms (such as L1 or L2 regularization) to the loss function. This approach helps to prevent overfitting and enhances the model's generalization ability [31].

We use Wrapper methods (WM) for the predictor selection. WM work by fitting models on selected feature subsets, evaluating their performance, and ultimately selecting the feature subset that performs best for that learner. The entire predictor selection process using WM is as follows: (1) the learner selects a feature subset (iteratively adding features to the model in sequential forward selection); (2) a ten-fold cross-validation resampling strategy is used to develop a pre-model and calculate the concordance index (C-index) of the pre-model for that feature subset; (3) repeat the above process until the C-index for all feature subsets has been calculated; (4) select and output the feature subset with the highest C-index as the result of the WM for that learner. The predictor selection result for each learner is a feature subset that includes several clinical features.

After separately calculating the above ten machine learning learners, we obtained a total of ten feature subsets for subsequent model development.

Development and validation of machine learning model

The ten machine learning learners, combined with ten feature subsets, were used to develop models. These models were trained using the training cohort, culminating in 100 prediction models. These models were subsequently validated within three validation cohorts.

The C-index was utilized to assess model performance. As a statistical measure for evaluating the predictive capability of survival analysis models, the C-index is widely used in medical research. It gauges the congruence between model predictions and actual outcomes, with its value ranging from 0 to 1. A higher C-index indicates superior predictive accuracy of the model. The model with the highest average C-index of the three validation cohorts was chosen for further investigation. Calibration curves were generated to evaluate the correspondence between predicted and actual non-incidence rates of recurrence at 1, 3, and 5 years. Time-dependent calibration curves were used to reflect the degree of calibration over an entire time range. The area under the time-dependent receiver operating characteristic (ROC) curves (AUC) served to compare the predictive accuracy and discriminative power of the model and its components. Decision curve analysis (DCA) was conducted to determine the clinical utility of the model, assessing the clinical benefits for patients at 1, 3, and 5 years. Risk scores were calculated in the training and validation cohorts using the machine learning model.

To examine how different features influence model performance over time, a time-dependent feature importance analysis method was employed. The significance of each predictor was evaluated by computing the model's Brier score loss after permuting feature values, with this process repeated through a tenfold cross-validation resampling strategy for statistical reliability. This approach enabled the identification of which features' importance for model predictions varies over time, providing insights crucial for time-sensitive clinical decision-making.

Development of a web risk calculator and a staging system

The ST was employed to determine the cut-off value for the risk score, thereby classifying patients into high-risk, medium-risk, and low-risk groups. Furthermore, a web-based application (https://drlx0721.shinyapps.io/PAMforDA/) was developed to make these predictive models accessible online, utilizing the R package "shiny" for its development [32].

Statistical analysis

Kolmogorov–Smirnov test was used to assess whether the data followed a normal distribution. For normally-distributed continuous variables, the data were described as mean ± standard deviation and compared using the *t*-test. If continuous variables did not conform to a normal distribution, the Mann–Whitney *U* test was used, and results were presented as median (interquartile range). Categorical data were presented as numbers and frequencies, and either the chi-square test or Fisher's exact test was used for comparisons. The cumulative risk curves were utilized to assess differences in recurrence rate across the high-risk, intermediate-risk, and low-risk groups. All statistical tests were two-sided, with *P*-values < 0.05 indicating statistical significance. All figure illustrations and statistical analyses were conducted using R version 4.2.3.

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

Patient characteristics

A total of 2189 patients with DA who underwent radical surgery were identified. Among them, 359 patients who did not meet the inclusion criteria were excluded. Consequently, 1830 patients were included in the analysis.

The baseline characteristics of the training cohort ($n = 1449$), validation cohort 1 ($n = 136$), validation cohort 2 ($n = 105$), and validation cohort 3 ($n = 140$) are presented in Table 1. In the training cohort, the median age is 61 (IQR 54.0; 67.0), the median follow-up time is 24 (IQR 12.0; 43.0) months, and the recurrence rate is 30.8%. Among them, 58.2% were male. In the validation cohort 1, the median age is 64 (IQR 56.0; 69.0), the median follow-up time is 19 (IQR 6.00; 37.0) months, and the recurrence rate is 49.3%. Among them, 64.7% were male. In the validation cohort 2, the median age is 66 (IQR 59.0; 73.0), the median follow-up time is 20 (IQR 10.0; 61.0) months, and the recurrence rate is 56.2%. Among them, 53.3% were male. In the validation cohort 3, the median age is 66 (IQR 58.0; 71.0), the median follow-up time is 21 (IQR 12.0; 39.2) months, and the recurrence rate is 25.0%. Among them, 59.3% were male.

Predictor selection

Fifty-three variables from the training cohort were included in the selection of predictors. Using ten machine learning learners and applying the WM, we obtained ten feature subsets, the details of which are displayed in Table S1 (Additional file 2: Table S1).

Model development, validation, and evaluation

For the development of the model using the training cohort, ten feature subsets were processed through ten machine learning learners and subsequently validated across three validation cohorts. This process resulted in a total of 100 machine learning models specifically designed for predicting recurrence.

The initial evaluation focused on the C-index, with rankings among the top 50 average C-indexes out of all 100 prediction models displayed in Fig. 2. The PR + AORSF model (PAM) showcased the highest average C-index of three validation cohorts at 0.739, making it the most effective model among all. The C-index for PAM was 0.882 (95% CI 0.860–0.886) in the training cohort, 0.747 (95% CI 0.683–0.798) in the validation cohort 1, 0.736 (95% CI 0.649–0.792) in the validation cohort 2, and 0.734 (95% CI 0.674–0.791) in the validation cohort 3, marking the highest values compared to other models. PAM incorporates 17 predictors: age, jaundice, preoperative WBC count, preoperative hemoglobin level, preoperative CA19.9 level, preoperative CEA level, operation time, volume of intraoperative blood transfusion, maximum diameter of the tumor, nerve invasion, vascular tumor thrombus, invasion of the muscularis propria, invasion of adjacent organs or structures, number of positive lymph nodes, postoperative ALP level, gastrointestinal bleeding, and abdominal hemorrhage.

The time-dependent calibration curves, along with the 1, 3, and 5-year calibration curves for PAM, demonstrate that PAM achieved moderate calibration in the training cohort

Table 1 Baseline characteristics of the cohort

	Training cohort, $n=1449$	Validation cohort 1, $n=136$	Validation cohort 2, $n=105$	Validation cohort 3, $n=140$	*P* value
Sex					0.336
Female	605 (41.8%)	48 (35.3%)	49 (46.7%)	57 (40.7%)	
Male	844 (58.2%)	88 (64.7%)	56 (53.3%)	83 (59.3%)	
Age (years)	61.0 (54.0;67.0)	64.0 (56.0;69.0)	66.0 (59.0;73.0)	66.0 (58.0;71.0)	< 0.001
Stomach ache					< 0.001
No	1096 (75.6%)	76 (55.9%)	90 (85.7%)	90 (64.3%)	
Yes	353 (24.4%)	60 (44.1%)	15 (14.3%)	50 (35.7%)	
Abdominal bloating					0.045
No	1230 (84.9%)	115 (84.6%)	99 (94.3%)	115 (82.1%)	
Yes	219 (15.1%)	21 (15.4%)	6 (5.71%)	25 (17.9%)	
Jaundice					0.024
No	655 (45.2%)	53 (39.0%)	45 (42.9%)	79 (56.4%)	
Yes	794 (54.8%)	83 (61.0%)	60 (57.1%)	61 (43.6%)	
Fever					0.276
No	1309 (90.3%)	117 (86.0%)	98 (93.3%)	126 (90.0%)	
Yes	140 (9.66%)	19 (14.0%)	7 (6.67%)	14 (10.0%)	
Poor appetite					0.452
No	1399 (96.5%)	130 (95.6%)	103 (98.1%)	138 (98.6%)	
Yes	50 (3.45%)	6 (4.41%)	2 (1.90%)	2 (1.43%)	
Emaciation					< 0.001
No	1433 (98.9%)	95 (69.9%)	105 (100%)	138 (98.6%)	
Yes	16 (1.10%)	41 (30.1%)	0 (0.00%)	2 (1.43%)	
Coronary heart disease					0.054
No	1381 (95.3%)	124 (91.2%)	96 (91.4%)	135 (96.4%)	
Yes	68 (4.69%)	12 (8.82%)	9 (8.57%)	5 (3.57%)	
Stroke					0.306
No	1418 (97.9%)	130 (95.6%)	104 (99.0%)	138 (98.6%)	
Yes	31 (2.14%)	6 (4.41%)	1 (0.95%)	2 (1.43%)	
Hypertension					< 0.001
No	1106 (76.3%)	95 (69.9%)	60 (57.1%)	88 (62.9%)	
Yes	343 (23.7%)	41 (30.1%)	45 (42.9%)	52 (37.1%)	
Diabetes					0.070
No	1263 (87.2%)	110 (80.9%)	90 (85.7%)	114 (81.4%)	
Yes	186 (12.8%)	26 (19.1%)	15 (14.3%)	26 (18.6%)	
Personal cancer history					< 0.001
No	1396 (96.3%)	123 (90.4%)	105 (100%)	139 (99.3%)	
Yes	53 (3.66%)	13 (9.56%)	0 (0.00%)	1 (0.71%)	
Liver dysfunction					0.001
No	1403 (96.8%)	126 (92.6%)	105 (100%)	140 (100%)	
Yes	46 (3.17%)	10 (7.35%)	0 (0.00%)	0 (0.00%)	
Smoking history					0.010
No	1009 (69.6%)	99 (72.8%)	77 (73.3%)	116 (82.9%)	
Yes	440 (30.4%)	37 (27.2%)	28 (26.7%)	24 (17.1%)	
Drinking history					0.262
No	1111 (76.7%)	111 (81.6%)	84 (80.0%)	115 (82.1%)	
Yes	338 (23.3%)	25 (18.4%)	21 (20.0%)	25 (17.9%)	
Family cancer history					0.304
No	1348 (93.0%)	128 (94.1%)	98 (93.3%)	136 (97.1%)	
Yes	101 (6.97%)	8 (5.88%)	7 (6.67%)	4 (2.86%)	
Body mass index (kg/m^2)					0.004

Table 1 (continued)

	Training cohort, $n=1449$	Validation cohort 1, $n=136$	Validation cohort 2, $n=105$	Validation cohort 3, $n=140$	*P* value
< 18.5	77 (5.31%)	12 (8.82%)	2 (1.90%)	12 (8.57%)	
18.5–23.9	815 (56.2%)	70 (51.5%)	50 (47.6%)	89 (63.6%)	
> 23.9	557 (38.4%)	54 (39.7%)	53 (50.5%)	39 (27.9%)	
Gastrointestinal obstruction					< 0.001
No	1398 (96.5%)	127 (93.4%)	81 (77.1%)	130 (92.9%)	
Yes	51 (3.52%)	9 (6.62%)	24 (22.9%)	10 (7.14%)	
Biliary obstruction					< 0.001
No	477 (32.9%)	44 (32.4%)	19 (18.1%)	77 (55.0%)	
Yes	972 (67.1%)	92 (67.6%)	86 (81.9%)	63 (45.0%)	
Preoperative WBC count (*10^9/L)	6.00 (4.93;7.32)	6.30 (5.03;7.97)	6.15 (5.05;7.36)	5.85 (4.76;7.30)	0.278
Preoperative hemoglobin level (g/L)	122 (108;134)	122 (105;134)	115 (103;126)	118 (104;129)	0.001
Preoperative PLT count (*10^9/L)	249 (207;306)	259 (212;335)	256 (201;298)	229 (184;304)	0.038
Preoperative CA199 level (U/ml)	49.6 (18.2;145)	49.6 (21.9;154)	37.8 (14.2;211)	32.3 (12.6;99.3)	0.024
Preoperative CEA level (ng/ml)	2.30 (1.53;3.54)	2.30 (1.67;3.95)	1.85 (1.11;3.09)	2.40 (1.70;4.03)	0.002
Preoperative TBIL level (μmol/L)	45.2 (15.1;114)	53.4 (14.3;163)	47.1 (14.9;165)	26.4 (12.9;88.3)	0.020
Preoperative DBIL level (μmol/L)	30.6 (6.70;88.3)	28.9 (3.68;112)	30.9 (8.20;129)	11.4 (2.88;48.7)	< 0.001
Operation time (min)	300 (240;360)	372 (315;420)	540 (480;645)	360 (295;431)	< 0.001
Surgical approach					< 0.001
Open	991 (68.4%)	85 (62.5%)	105 (100%)	11 (7.86%)	
Laparoscope	202 (13.9%)	34 (25.0%)	0 (0.00%)	107 (76.4%)	
Robotic	256 (17.7%)	17 (12.5%)	0 (0.00%)	22 (15.7%)	
Surgical procedure					< 0.001
Local resection	18 (1.24%)	2 (1.47%)	0 (0.00%)	3 (2.14%)	
Pancreaticoduodenectomy	1316 (90.8%)	134 (98.5%)	105 (100%)	122 (87.1%)	
Pyloric-preserving pancreaticoduodenectomy	115 (7.94%)	0 (0.00%)	0 (0.00%)	15 (10.7%)	
Intraoperative blood loss (ml)	200 (100;300)	200 (100;300)	500 (300;600)	200 (100;400)	< 0.001
Volume of intraoperative blood transfusion (ml)	0.00 (0.00;0.00)	0.00 (0.00;0.00)	0.00 (0.00;800)	0.00 (0.00;412)	< 0.001
Maximum diameter of the tumor (cm)	2.20 (1.50;3.00)	2.45 (1.50;3.50)	2.30 (1.80;3.00)	2.50 (1.50;3.23)	0.304
Nerve invasion					< 0.001
No	1148 (79.2%)	73 (53.7%)	57 (54.3%)	90 (64.3%)	
Yes	301 (20.8%)	63 (46.3%)	48 (45.7%)	50 (35.7%)	
Vascular tumor thrombus					< 0.001
No	1179 (81.4%)	84 (61.8%)	60 (57.1%)	100 (71.4%)	
Yes	270 (18.6%)	52 (38.2%)	45 (42.9%)	40 (28.6%)	
Tumor site					< 0.001
Ampullary	696 (48.0%)	112 (82.4%)	61 (58.1%)	29 (20.7%)	
Non-ampullary	753 (52.0%)	24 (17.6%)	44 (41.9%)	111 (79.3%)	
Invasion of the muscularis propria					0.024
No	205 (14.1%)	7 (5.15%)	17 (16.2%)	21 (15.0%)	
Yes	1244 (85.9%)	129 (94.9%)	88 (83.8%)	119 (85.0%)	
T stage					< 0.001
T1	205 (14.1%)	7 (5.15%)	17 (16.2%)	21 (15.0%)	
T2	439 (30.3%)	41 (30.1%)	39 (37.1%)	64 (45.7%)	
T3	669 (46.2%)	61 (44.9%)	49 (46.7%)	47 (33.6%)	
T4	136 (9.39%)	27 (19.9%)	0 (0.00%)	8 (5.71%)	
Lymph node metastasis					< 0.001
No	1048 (72.3%)	80 (58.8%)	60 (57.1%)	99 (70.7%)	
Yes	401 (27.7%)	56 (41.2%)	45 (42.9%)	41 (29.3%)	
AJCC stage					< 0.001
I	536 (37.0%)	37 (27.2%)	44 (41.9%)	69 (49.3%)	

Table 1 (continued)

	Training cohort, $n=1449$	Validation cohort 1, $n=136$	Validation cohort 2, $n=105$	Validation cohort 3, $n=140$	*P* value
II	481 (33.2%)	37 (27.2%)	16 (15.2%)	29 (20.7%)	
III	432 (29.8%)	62 (45.6%)	45 (42.9%)	42 (30.0%)	
Invasion of adjacent organs or structures					< 0.001
No	625 (43.1%)	55 (40.4%)	56 (53.3%)	125 (89.3%)	
Yes	824 (56.9%)	81 (59.6%)	49 (46.7%)	15 (10.7%)	
Number of positive lymph nodes	0.00 (0.00;1.00)	0.00 (0.00;2.00)	0.00 (0.00;3.00)	0.00 (0.00;1.00)	< 0.001
Postoperative ALT level (U/L)	74.0 (43.0;127)	71.0 (71.0;71.0)	39.0 (29.0;58.0)	70.0 (41.0;127)	< 0.001
Postoperative AST level (U/L)	63.0 (41.0;97.1)	63.0 (57.8;96.5)	39.0 (28.0;69.0)	63.0 (56.8;99.2)	< 0.001
Postoperative ALP level (U/L)	149 (96.0;266)	182 (85.2;330)	144 (97.0;176)	32.9 (28.8;43.8)	< 0.001
Postoperative TBIL level (μmol/L)	33.6 (17.9;79.6)	52.8 (21.8;132)	32.4 (18.7;98.1)	31.9 (20.3;62.3)	0.007
Postoperative DBIL level (μmol/L)	20.8 (8.80;59.3)	19.9 (19.9;19.9)	21.6 (8.99;80.9)	13.1 (6.40;33.5)	0.002
Postoperative pancreatic fistula					< 0.001
No	798 (55.1%)	60 (44.1%)	70 (66.7%)	0 (0.00%)	
A	530 (36.6%)	55 (40.4%)	29 (27.6%)	118 (84.3%)	
B	121 (8.35%)	21 (15.4%)	6 (5.71%)	22 (15.7%)	
Postoperative biliary fistula					< 0.001
No	1360 (93.9%)	127 (93.4%)	103 (98.1%)	120 (85.7%)	
Yes	89 (6.14%)	9 (6.62%)	2 (1.90%)	20 (14.3%)	
Postoperative gastroparesis					0.005
No	1350 (93.2%)	131 (96.3%)	94 (89.5%)	121 (86.4%)	
Yes	99 (6.83%)	5 (3.68%)	11 (10.5%)	19 (13.6%)	
Abdominal hemorrhage					< 0.001
No	1377 (95.0%)	132 (97.1%)	100 (95.2%)	115 (82.1%)	
Yes	72 (4.97%)	4 (2.94%)	5 (4.76%)	25 (17.9%)	
Gastrointestinal bleeding					< 0.001
No	1419 (97.9%)	128 (94.1%)	95 (90.5%)	137 (97.9%)	
Yes	30 (2.07%)	8 (5.88%)	10 (9.52%)	3 (2.14%)	
Abdominal infection					< 0.001
No	1255 (86.6%)	100 (73.5%)	98 (93.3%)	107 (76.4%)	
Yes	194 (13.4%)	36 (26.5%)	7 (6.67%)	33 (23.6%)	
Postoperative ileus					0.517
No	1421 (98.1%)	131 (96.3%)	103 (98.1%)	137 (97.9%)	
Yes	28 (1.93%)	5 (3.68%)	2 (1.90%)	3 (2.14%)	
Incisional infection					0.088
No	1380 (95.2%)	127 (93.4%)	103 (98.1%)	138 (98.6%)	
Yes	69 (4.76%)	9 (6.62%)	2 (1.90%)	2 (1.43%)	
Secondary surgery					< 0.001
No	1391 (96.0%)	132 (97.1%)	97 (92.4%)	122 (87.1%)	
Yes	58 (4.00%)	4 (2.94%)	8 (7.62%)	18 (12.9%)	
Recurrence					< 0.001
No	1002 (69.2%)	69 (50.7%)	46 (43.8%)	105 (75.0%)	
Yes	447 (30.8%)	67 (49.3%)	59 (56.2%)	35 (25.0%)	
Follow-up time (months)	24.0 (12.0;43.0)	19.0 (6.00;37.0)	20.0 (10.0;61.0)	21.0 (12.0;39.2)	0.029

WBC White blood cell, *PLT* Platelet, *TBIL* Total bilirubin, *DBIL* Direct bilirubin, *ALT* Alanine aminotransferase, *AST* Aspartate aminotransferase, *ALP* Alkaline phosphatase

and the three validation cohorts (Fig. 3 and Additional file 3: Fig. S1).

The time-dependent ROC curves demonstrate the strong predictive accuracy of PAM over 1, 3, and 5 years in the training cohort and the three validation cohorts. In the training cohort, PAM achieved an AUC of 0.91 (95% CI 0.89–0.93) at 1 year, 0.95 (95% CI 0.94–0.97) at 3 years, and 0.93 (95% CI 0.91–0.95) at 5 years (Fig. 4A). In the validation cohort 1, PAM achieved an AUC of 0.83 (95% CI 0.76–0.90) at 1 year, 0.83 (95% CI 0.75–0.90) at

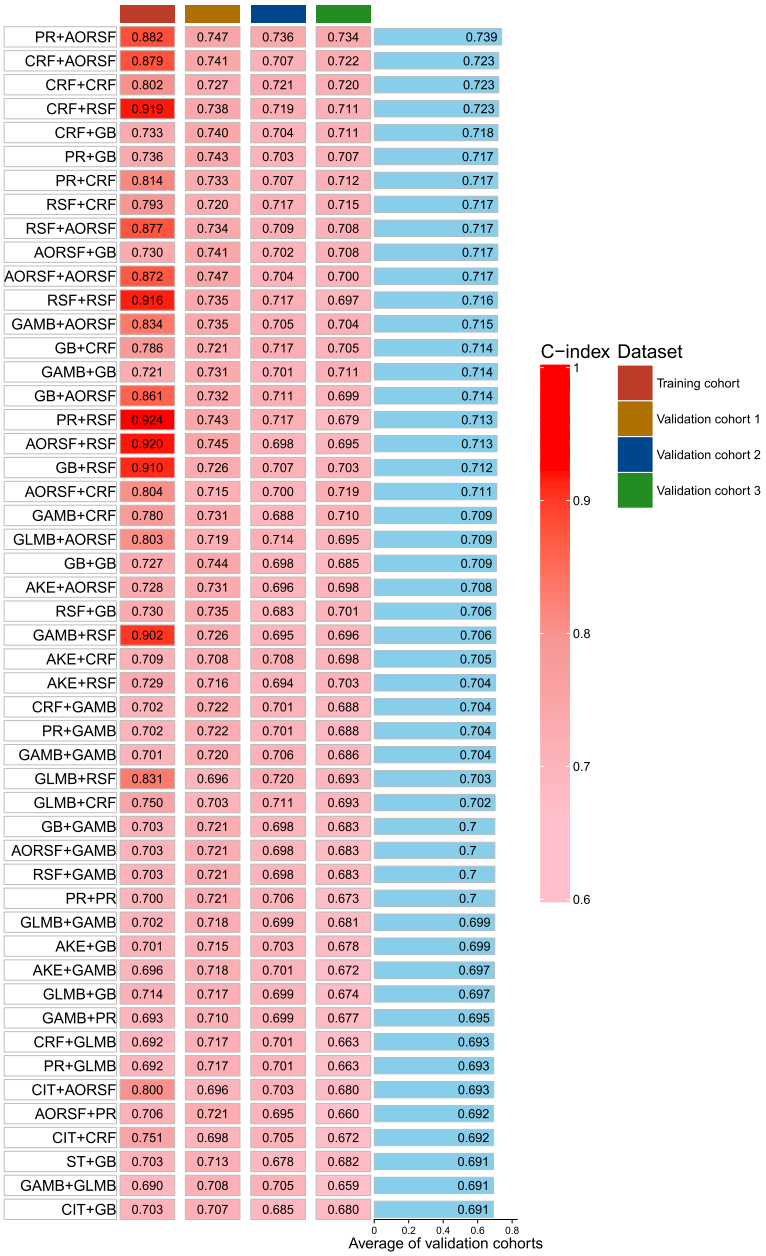

Fig. 2 Concordance index of top 50 machine learning models. The C-index for the top 50 out of 100 machine learning models was calculated for the training cohort and three validation cohorts. Ranking of the models was based on the average C-index of three validation cohorts. AKE, Akritas estimator; GB, Gradient Boosting, GAMB, Boosted Generalized Additive Model; GLMB, Boosted Generalized Linear Model; ST, Survival Tree; CIT, Conditional Inference Tree; RSF, Random Survival Forest; CRF, Conditional Random Forest; AORSF, Accelerated Oblique Random Survival Forest; PR, Penalized Regression; C-index, concordance index

3 years, and 0.82 (95% CI 0.71–0.91) at 5 years (Fig. 4B). In the validation cohort 2, PAM achieved an AUC of 0.74 (95% CI 0.62–0.84) at 1 year, 0.76 (95% CI 0.65–0.85) at 3 years, and 0.77 (95% CI 0.66–0.86) at 5 years (Fig. 4C). In the validation cohort 3, PAM achieved an AUC of 0.77 (95% CI 0.66–0.86) at 1 year, 0.81 (95% CI 0.71–0.89) at 3 years, and 0.77 (95% CI 0.62–0.89) at 5 years (Fig. 4D).

The DCA for the PAM demonstrated a consistent net benefit in the training cohort and three validation cohorts over a range of threshold probabilities (Additional file 4: Fig. S2). In all four cohorts, the PAM outperformed the 'treat none' and 'treat all' strategies, indicating that it had practical utility in decision-making.

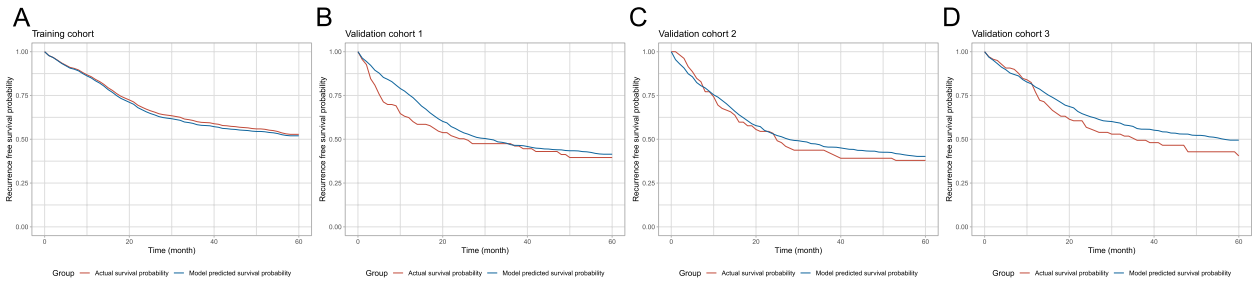

Fig. 3 Evaluating the calibration of PAM by time-dependent calibration curves. **A** Training cohort, **B** validation cohort 1, **C** validation cohort 2, **D** Validation cohort 3

Model interpretation and development of stage system

The time-dependent feature importance curves show the varying importance of each predictor in PAM over time (Fig. 5). The results indicate that the number of positive lymph nodes, invasion of adjacent organs or structures, and the maximum diameter of the tumor are the most important factors over the follow-up period. To further enhance the usability of the model, we developed a web-based risk calculator (Methods). Additionally, we developed a staging system based on PAM, utilizing the ST learner to categorize DA patients into three risk groups according to PAM's risk scores. Cumulative risk curves demonstrate statistically significant differences in recurrence rate among the high-risk, medium-risk, and low-risk groups across the four cohorts (Additional file 5: Fig. S3).

Discussion

In this study, we developed and validated a machine learning-based model to predict recurrence in postoperative DA patients. The prognostic model exhibited accuracy in the training and three validation cohorts. In terms of predictive values, PAM generally exhibits a moderate level C-index and AUC, indicating the model's accuracy and stability in predicting cancer recurrence. Additionally, the time-dependent calibration curves and DCA demonstrate the moderate calibration and clinical net benefit of PAM. Our study indicates that PAM has the potential to identify postoperative recurrence in DA patients. This can assist clinicians in assessing the severity of the disease, facilitating patient follow-up, and aiding in the formulation of adjuvant treatment strategies. To our knowledge, this study is the first to develop a prognostic model for DA, and the predictive performance of the model is exceptionally good.

Recently, the development of predictive models has gained significant attention among clinical scientists. Therefore, the standardized development and validation of predictive models are crucial, and our study adhered rigorously to these standards. Finhn et al. noted that the proliferation of predictive models has been accompanied by an increasing awareness of the need for standards to ensure their accuracy. A significant milestone was the publication of the TRIPOD guidelines nearly a decade ago [21, 33]. Wolff et al. developed a tool to assess the risk of bias and the applicability of prediction model studies [19]. This tool includes 20 signal questions designed to enable researchers to self-assess their studies. Florian Markowetz proposed a checklist for useful clinical prediction tools aimed at making clinical prediction models impactful for patients [20]. The aforementioned checklist and tools were used to standardize our research.

PAM has a clear clinical decision point: the end of the perioperative period, where the patient transitions to the

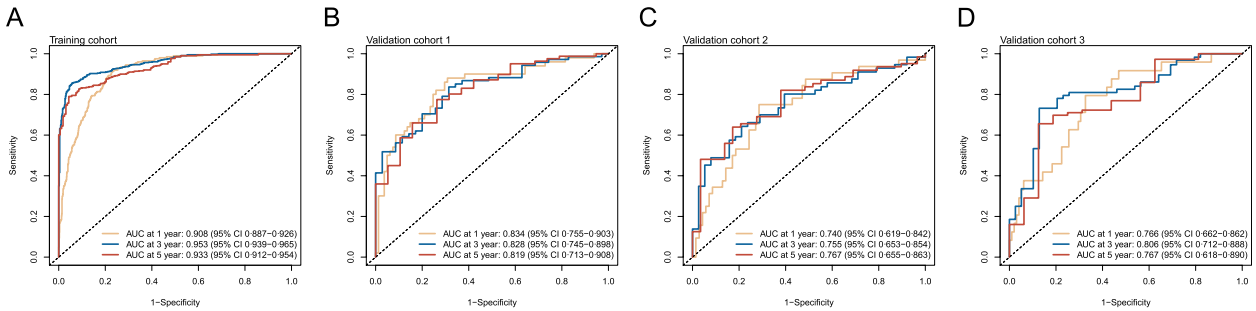

Fig. 4 Evaluating the predictive accuracy of PAM by time-dependent ROC curves. **A** Training cohort, **B** validation cohort 1, **C** validation cohort 2, **D** validation cohort 3

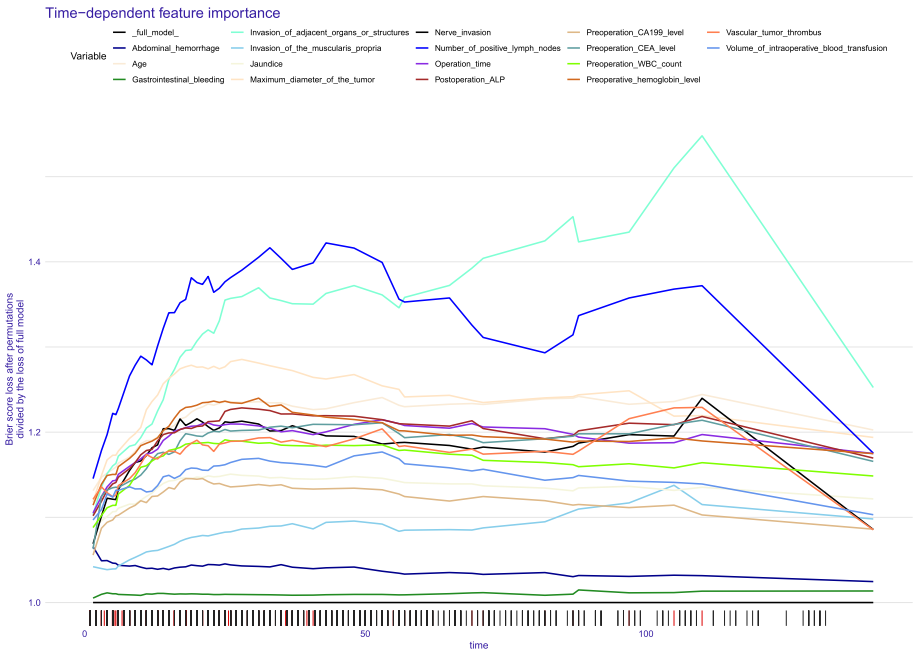

Fig. 5 Interpretation the PAM by time-dependent feature importance curves

next phase of treatment or enters long-term follow-up. PAM outputs the patient's risk score, risk group, and the probability of recurrence at a selected time, which aids doctors in making decisions about further treatment plans. Due to the lack of standardized adjuvant treatment methods for DA after surgery, clinical trials are the preferred option for all patients considering adjuvant therapy [4]. Therefore, providing guidance for postoperative adjuvant therapy for DA patients is crucial, which precisely constitutes the greatest clinical value of PAM.

In the DCA, we identified threshold probability ranges for recurrence of approximately 10–60% at 1 year, 12–90% at 3 years, and 15–90% at 5 years. By using these thresholds, PAM enables clinicians to make more informed treatment decisions [34]. For patients whose predicted recurrence risk exceeds the lower limits of these thresholds, more aggressive treatment or closer monitoring may be recommended, such as chemotherapy or encouraging participation in clinical trials. Conversely, for those whose recurrence risk falls below the lower limits of these thresholds, a conservative approach may be preferred to avoid unnecessary interventions and improve quality of life. This strategy can also assist in resource allocation, ensuring that high-risk patients receive prioritized care and intensive follow-up when needed, while low-risk patients can be monitored with standard care protocols. This not only tailors follow-up schedules and treatments to individual needs but also provides patients and their families with clearer expectations for future health management.

Feature selection was employed to enhance the model's interpretability and clinical utility. By reducing the number of predictors, the model becomes easier to understand and apply, streamlining the data collection process for clinical practice. Additionally, we used a Wrapper Method to ensure that only predictors contributing to model performance were retained, improving the model's efficiency and accuracy.

All predictors used in PAM are commonly employed in clinical practice. The patient's age can be determined from their identification documents, and the presence of jaundice can be ascertained from the physical examination upon admission. Preoperative levels of WBC, hemoglobin, CEA, and CA19.9 can be obtained from preoperative laboratory records. Operation time and intraoperative blood transfusion volumes are available in surgical records. The maximum diameter of the tumor, nerve invasion, vascular tumor thrombus, invasion of the muscularis propria, invasion of adjacent organs or structures, and the number of positive lymph nodes can be determined from postoperative pathology reports. Postoperative ALP levels can be obtained from postoperative laboratory records. Postoperative complications such as gastrointestinal bleeding and abdominal hemorrhage can be identified from perioperative medical records.

It is important to note that due to differences in T staging between DA of the ampulla and non-ampulla, the T

stage was not included in the model development process [14]. Instead, the T stage was divided into two variables: whether there was an invasion into the muscularis propria and whether there was an invasion into adjacent organs or structures, which helped reduce information entropy. Additionally, the number of positive lymph nodes was used instead of the N stage because it provides more detailed information that is more beneficial for model fitting.

As patients age, a decline in immune function and cellular repair mechanisms may increase the risk of cancer recurrence [35–37]. Abnormal preoperative levels of WBC, hemoglobin, CEA, and CA19.9 may indicate systemic inflammation, anemia, or higher tumor burden, all of which are associated with a higher likelihood of recurrence [38–42]. Similarly, prolonged surgery and significant blood transfusions may lead to postoperative immune suppression, potentially facilitating residual cancer cell survival [43–48]. Tumor-related factors, such as larger tumor size, nerve invasion, vascular tumor thrombi, and muscularis propria invasion, reflect more aggressive biological behavior and are linked to poorer prognoses [49–55]. Postoperative indicators, including elevated ALP levels and gastrointestinal or abdominal bleeding, can signify compromised health status, which indirectly impacts tumor management and recovery outcomes [56–59].

The model's performance in the validation cohorts, with an average C-index of approximately 0.750, indicates moderate predictive accuracy. This level of performance, while not exceptional, still offers clinical value in guiding treatment planning for postoperative DA patients. However, several factors likely contribute to this moderate performance in the validation cohorts. In our multicenter study, we included four cohorts with varying degrees of clinical baseline differences. These variations improve the model's generalizability by testing it across diverse patient populations and settings, but they may also account for the moderate C-index observed in the validation sets. It is noteworthy, however, that achieving a C-index of approximately 0.750 in cohorts with significant differences in baseline characteristics and clinical settings demonstrates the robustness of our model. Despite only moderate predictive accuracy, the model's reliability is reinforced by its performance across large sample sizes in external multicenter validation. This certainly provides DA, which lacks effective predictive tools, with a valuable new resource.

While our study yielded promising results, it is crucial to recognize its limitations. Firstly, the PAM was developed primarily using data from Chinese patients, requiring further validation across different racial groups to confirm its generalizability. Secondly, the absence of standardized treatment guidelines meant that data on potentially prognostic variables like adjuvant chemotherapy and radiotherapy were missing, thereby limiting their inclusion in this study and potentially restricting the model's predictive capacity. Thirdly, molecular pathology features were not integrated into the model, which could have otherwise enhanced its predictive accuracy. Finally, the retrospective nature of data collection from multiple centers resulted in instances of missing data. While this limitation might be offset by strict inclusion and exclusion criteria and a large sample size, prospective international multicenter studies are necessary to further validate the performance of the PAM.

Conclusions

In summary, our study developed the PAM for accurately predicting recurrence in DA patients who underwent radical surgery. PAM demonstrated stable and moderate predictive performance, calibration, and clinical net benefit across three independent validation cohorts. Although moderate in predictive accuracy, the model provides a useful framework for risk stratification in postoperative DA management, aiding clinicians in making more informed, patient-centered decisions.

Abbreviations

DA	Duodenal adenocarcinoma
DC	Duodenal carcinoma
ADC	Ampullary-duodenal carcinoma
NADC	Nonampullary-duodenal carcinoma
PROBAST	Prediction model Risk Of Bias Assessment Tool
TRIPOD	Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis
BMI	Body mass index
WBC	White blood cell
PLT	Platelet
TBIL	Total bilirubin
DBIL	Direct bilirubin
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
ALP	Alkaline phosphatase
AKE	Akritas estimator
GB	Gradient Boosting
GAMB	Boosted Generalized Additive Model
GLMB	Boosted Generalized Linear Model
ST	Survival Tree
CIT	Conditional Inference Tree
RSF	Random Survival Forest
CRF	Conditional Random Forest
AORSF	Accelerated Oblique Random Survival Forest
PR	Penalized Regression
WM	Wrapper method
C-index	Concordance index
ROC	Receiver operating characteristic curves
AUC	Area under the curve
DCA	Decision curve analysis
IQR	Interquartile range
PAM	Penalized Regression + Accelerated Oblique Random Survival Forest model

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12916-025-03912-7.

Additional file 1: TRIPOD Checklist: Prediction Model Development and Validation.

Additional file 2: Table S1—Feature subsets after WM feature selection.

Additional file 3: Figure S1—Evaluating the calibration in 1-, 3-, and 5- year of PAM by calibration curves. (A) Training cohort, (B) Validation cohort 1, (C) Validation cohort 2, (D) Validation cohort 3.

Additional file 4: Figure S2—Evaluating the net benefit of PAM by DCA. (A) 1-year DCA for training cohort, (B) 3-year DCA for training cohort, (C) 5-year DCA for training cohort, (C) 1-year DCA for validation cohort 1, (D) 3-year DCA for validation cohort 1, (V) 5-year DCA for validation cohort 1, (G) 1-year DCA for validation cohort 2, (H) 3-year DCA for validation cohort 2, (I) 5-year DCA for validation cohort 2, (J) 1-year DCA for validation cohort 3, (K) 3-year DCA for validation cohort 3, (L) 5-year DCA for validation cohort 3. DCA, Decision curve analysis.

Additional file 5: Figure S3—Performance of the PAM staging system in the training cohort and three validation cohorts. (A) Training cohort, (B) Validation cohort 1, (C) Validation cohort 2, (D) Validation cohort 3.

Acknowledgements
This research was supported by the CAMS Innovation Fund for Medical Sciences (No. 2022-I2M-1-010).

Authors' contributions
QX, CW, and JZ access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Concept and design: XL, QX, CW, and JZ. Acquisition, analysis, or interpretation of data: XL, QX, CY, ZG, XW, XT1, FM, DW, RL, KG, XT2, YZ, EZ, FC, ZW, JX, YX, CW, JZ. Drafting of the manuscript: XL, QX. Critical revision of the manuscript for important intellectual content: CW, and JZ. Statistical analysis: XL, QX, and JZ. Obtained funding: CW and JZ. Administrative, technical, or material support: CW and JZ. Supervision: CW and JZ. All authors read and approved the final manuscript.

Funding
This work was supported by the CAMS Innovation Fund for Medical Sciences (No. 2022-I2M-1–010).

Data availability
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
This study adheres to ethical standards and was ensured by conducting the study in compliance with the Declaration of Helsinki (revised in 2013), and ethical approval was secured from the Hospital Ethics Committee of the National Cancer Center (No. 22/213–3415). Written informed consent to participate in the study was obtained from all participants prior to their inclusion. No organs or tissues were obtained from vulnerable groups, including prisoners, in this study.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Pancreatic and Gastric Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. [2]National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. [3]General Surgery, Department of Hepatobiliary & Pancreatic Surgery and Minimally Invasive Surgery, Zhejiang Provincial People's Hospital, Affiliated People's Hospital, Hangzhou Medical College, Cancer Center, Hangzhou, Zhejiang, China. [4]Department of General Surgery, First Medical Center of Chinese, PLA General Hospital, Beijing, China. [5]Department of General Surgery, Peking University Third Hospital, Beijing, China. [6]Department of General Surgery, Qilu Hospital of Shandong University, Jinan, Shandong, China. [7]Department of Pancreatic and Biliary Surgery, The First Affiliated Hospital of China Medical University, Shenyang, Liaoning, China. [8]Department of General Surgery, Beijing Friendship Hospital, Capital Medical University, Beijing, China. [9]Department of Hepatobiliary and Pancreaticosplenic Surgery, Beijing Chaoyang Hospital, Capital Medical University, Beijing, China. [10]Hepatobiliary and Pancreatogastric Surgery, Shanxi Province Cancer Hospital/Shanxi Hospital Affiliated to Cancer Hospital, Chinese Academy of Medical Sciences/Cancer Hospital Affiliated to Shanxi Medical University, Taiyuan, Shanxi, China. [11]Department of General Surgery, Peking University First Hospital, Beijing, China. [12]Department of Hepatobiliary Surgery, Shaanxi Provincial People's Hospital, Xi'an, Shaanxi, China. [13]Gastrointestinal Surgery Department, Affiliated Hospital of Chengde Medical University, Chengde, Hebei, China. [14]Department of Oncological Surgery, The Second Hospital of Hebei Medical University, Shijiazhuang, Hebei, China. [15]Department of General Surgery, Beijing Hospital, Beijing, China. [16]General Surgery Department, Capital Medical University Beijing Affiliated Tiantan Hospital, Beijing, China. [17]General Surgery Department, Xuanwu Hospital, Capital Medical University, Beijing, China.

Received: 7 May 2024 Accepted: 26 January 2025
Published online: 21 February 2025

References

1. Severson RK, Schenk M, Gurney JG, Weiss LK, Demers RY. Increasing incidence of adenocarcinomas and carcinoid tumors of the small intestine in adults. Cancer Epidemiol Biomarkers Prev. 1996;5:81–4.
2. Haselkorn T, Whittemore AS, Lilienfeld DE. Incidence of small bowel cancer in the United States and worldwide: geographic, temporal, and racial differences. Cancer Causes Control. 2005;16:781–7.
3. Schottenfeld D, Beebe-Dimmer JL, Vigneau FD. The epidemiology and pathogenesis of neoplasia in the small intestine. Ann Epidemiol. 2009;19:58–69.
4. Benson AB, Venook AP, Al-Hawary MM, Arain MA, Chen YJ, Ciombor KK, et al. Small bowel adenocarcinoma, version 1.2020, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw. 2019;17:1109–33.
5. Jensen KK, Storkholm JH, Chen I, Burgdorf SK, Hansen CP. Long-term results after resection of primary duodenal adenocarcinoma: a retrospective cohort study. Int J Surg. 2022;100:106599.
6. Meijer LL, Alberga AJ, de Bakker JK, van der Vliet HJ, Le Large TYS, van Grieken NCT, et al. Outcomes and treatment options for duodenal adenocarcinoma: a systematic review and meta-analysis. Ann Surg Oncol. 2018;25:2681–92.
7. Solaini L, Jamieson NB, Metcalfe M, Abu Hilal M, Soonawalla Z, Davidson BR, et al. Outcome after surgical resection for duodenal adenocarcinoma in the UK. Br J Surg. 2015;102:676–81.
8. Hirashita T, Ohta M, Tada K, Saga K, Takayama H, Endo Y, et al. Prognostic factors of non-ampullary duodenal adenocarcinoma. Jpn J Clin Oncol. 2018;48:743–7.
9. Poultsides GA, Huang LC, Cameron JL, Tuli R, Lan L, Hruban RH, et al. Duodenal adenocarcinoma: clinicopathologic analysis and implications for treatment. Ann Surg Oncol. 2012;19:1928–35.
10. Mann K, Gilbert T, Cicconi S, Jackson R, Whelan P, Campbell F, et al. Tumour stage and resection margin status are independent survival factors following partial pancreatoduodenectomy for duodenal adenocarcinoma. Langenbecks Arch Surg. 2019;404:439–49.
11. de Bakker JK, Suurmeijer JA, Toennaer JGJ, Bonsing BA, Busch OR, van Eijck CH, et al. Surgical outcome after pancreatoduodenectomy for duodenal adenocarcinoma compared with other periampullary cancers: a nationwide audit study. Ann Surg Oncol. 2023;30:2448–55.
12. Li D, Si X, Wan T, Zhou Y. Outcomes of surgical resection for primary duodenal adenocarcinoma: a systematic review. Asian J Surg. 2019;42:46–52.

13. Cloyd JM, Norton JA, Visser BC, Poultsides GA. Does the extent of resection impact survival for duodenal adenocarcinoma? Analysis of 1,611 cases. Ann Surg Oncol. 2015;22:573–80.
14. Oweira H, Abdel-Rahman O, Mehrabi A, Reissfelder C. Assessment of the external validity of the AJCC 8th staging system for small intestinal adenocarcinoma: a time to reconsider the role of tumor location? J Gastrointest Oncol. 2019;10:421–8.
15. Obermeyer Z, Emanuel EJ. Predicting the future - big data, machine learning, and clinical medicine. N Engl J Med. 2016;375:1216–9.
16. Motwani A, Shukla PK, Pawar M. Ubiquitous and smart healthcare monitoring frameworks based on machine learning: a comprehensive review. Artif Intell Med. 2022;134:102431.
17. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nat Med. 2019;25:24–9.
18. Liu X, Li X, Yu S. CFLAR: a novel diagnostic and prognostic biomarker in soft tissue sarcoma, which positively modulates the immune response in the tumor microenvironment. Oncol Lett. 2024;27:1–14.
19. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med. 2019;170:51–8.
20. Markowetz F. All models are wrong and yours are useless: making clinical prediction models impactful for patients. NPJ Precis Oncol. 2024;8:54.
21. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. 2015;162:W1-73.
22. Bender A, Bischl B, Király FJ, Lang M, Sonabend R. mlr3proba: an R package for machine learning in survival analysis. Bioinformatics. 2020;37(17):2789–91.
23. Dai H, Restaino M, Wang H. A class of nonparametric bivariate survival function estimators for randomly censored and truncated data. Journal of Nonparametric Statistics. 2016;28:736–51.
24. Sk A, Sk M. Credit card fraud detection using ensemble machine learning method - gradient boosting framework. IJRASET. 2023;11:4807–10.
25. Zhou Z, Qiu Z, Niblett B, Johnston A, Schwartzentruber J, Zincir-Heywood N, et al. A boosting approach to constructing an ensemble stack. 2023. p. 133–48.
26. Penny-Dimri JC, Bergmeir C, Reid CM, Williams-Spence J, Perry LA, Smith JA. Tree-based survival analysis improves mortality prediction in cardiac surgery. Front Cardiovasc Med. 2023;10:1211600.
27. Rasmussen NEK, Hansen MF, Stephensen P. Conditional inference trees in dynamic microsimulation - modelling transition probabilities in the SMILE model. DREAM working paper series. 2013.
28. Archetti A, Matteucci M. Federated survival forests. In: 2023 International Joint Conference on Neural Networks (IJCNN). 2023. p. 1–9.
29. Liu J, Yang X, Zhang H, Wang Z, Yan H. Predictive control for unknown dynamics with observation loss: a temporal game-theoretic approach. IEEE Trans Industr Electron. 2024;71:2965–77.
30. Jaeger BC, Welden S, Lenoir K, Speiser JL, Segar MW, Pandey A, et al. Accelerated and interpretable oblique random survival forests. 2022;33(1):192–207.
31. Mbebi AJ, Tong H, Nikoloski Z. L2,1-norm regularized multivariate regression model with applications to genomic prediction. Bioinformatics. 2021;37:2896–904.
32. Xiao Z, Lam H-M. ShinySyn: a Shiny/R application for the interactive visualization and integration of macro- and micro-synteny data. Bioinformatics. 2022;38:4406–8.
33. Fihn SD, Berlin JA, Haneuse SJPA, Rivara FP. Prediction models and clinical outcomes-a call for papers. JAMA Netw Open. 2024;7:e249640.
34. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. Diagn Progn Res. 2019;3:18.
35. Prieto LI, Baker DJ. Cellular senescence and the immune system in cancer. Gerontology. 2019;65:505–12.
36. Maruyama M. Age-associated decline in the immune system. Nihon Rinsho. 2013;71:993–8.
37. Macieira-Coelho A. Putative mechanisms responsible for the decline in cancer prevalence during organism senescence. Biogerontology. 2015;16:559–65.
38. Okada S, Shimomura M, Tsunezuka H, Ishihara S, Ikebe S, Furuya T, et al. High neutrophil count as a negative prognostic factor for relapse in patients with thymic epithelial tumor. Ann Surg Oncol. 2020;27:2438–47.
39. Tanaka H, Ono T, Manabe Y, Kajima M, Fujimoto K, Yuasa Y, et al. Anemia is a prognostic factor for overall survival rate in patients with non-small cell lung cancer treated with stereotactic body radiation therapy. Cancer Manag Res. 2021;13:7447–53.
40. Shibata C, Nakano T, Yasumoto A, Mitamura A, Sawada K, Ogawa H, et al. Comparison of CEA and CA19-9 as a predictive factor for recurrence after curative gastrectomy in gastric cancer. BMC Surg. 2022;22:213.
41. Baqar AR, Wilkins S, Staples M, Angus Lee CH, Oliva K, McMurrick P. The role of preoperative CEA in the management of colorectal cancer: a cohort study from two cancer centres. Int J Surg. 2019;64:10–5.
42. Ma SJ, Yu H, Khan M, Yu B, Santhosh S, Chatterjee U, et al. Defining the optimal threshold and prognostic utility of pre-treatment hemoglobin level as a biomarker for survival outcomes in head and neck cancer patients receiving chemoradiation. Oral Oncol. 2022;133:106054.
43. Weinberg DS, Narayanan AS, Moore TA, Vallier HA. Prolonged resuscitation of metabolic acidosis after trauma is associated with more complications. J Orthop Surg Res. 2015;10:153.
44. Choi KC, Peek-Asa C, Lovell M, Torner JC, Zwerling C, Kealey GP. Complications after therapeutic trauma laparotomy. J Am Coll Surg. 2005;201:546–53.
45. Hee H-Z, Chang K-Y, Huang C-Y, Chang W-K, Tsou M-Y, Lin S-P. Perioperative blood transfusion is dose-dependently associated with cancer recurrence and mortality after head and neck cancer surgery. Cancers (Basel). 2022;15:99.
46. Linder BJ, Frank I, Cheville JC, Tollefson MK, Thompson RH, Tarrell RF, et al. The impact of perioperative blood transfusion on cancer recurrence and survival following radical cystectomy. Eur Urol. 2013;63:839–45.
47. Zhang J, Chen S, Yan Y, Zhu X, Qi Q, Zhang Y, et al. Extracellular ubiquitin is the causal link between stored blood transfusion therapy and tumor progression in a melanoma mouse model. J Cancer. 2019;10:2822–35.
48. Vamvakas EC, Blajchman MA. Transfusion-related immunomodulation (TRIM): an update. Blood Rev. 2007;21:327–48.
49. Nakagawa K, Sho M, Okada K-I, Akahori T, Aoyama T, Eguchi H, et al. Surgical results of non-ampullary duodenal cancer: a nationwide survey in Japan. J Gastroenterol. 2022;57:70–81.
50. Zhou Y-M, Sui C-J, Li B, Xu F, Kan T, Yang J-M. Results of en bloc resection for hepatocellular carcinoma extending to adjacent organs. Can J Surg. 2012;55:222–6.
51. Christophi C. Hepatocellular carcinoma: implications of microvascular invasion and tumour thrombi. ANZ J Surg. 2007;77:99–100.
52. Hirozane T, Nakayama R, Yamaguchi S, Mori T, Asano N, Asakura K, et al. Recurrent malignant peripheral nerve sheath tumor presenting as an asymptomatic intravenous thrombus extending to the heart: a case report. World J Surg Oncol. 2022;20:8.
53. Acher AW, Abbott DE. Rethinking resection and transplant candidacy for HCC: should tumor biology replace size-based criteria? Ann Surg Oncol. 2020;27:1309–11.
54. Jin X, Shen C, Yang X, Yu Y, Wang J, Che X. Association of tumor size with myometrial invasion, lymphovascular space invasion, lymph node metastasis, and recurrence in endometrial cancer: a meta-analysis of 40 studies with 53,276 patients. Front Oncol. 2022;12:881850.
55. Hatta W, Koike T, Takahashi S, Shimada T, Hikichi T, Toya Y, et al. Risk of metastatic recurrence after endoscopic resection for esophageal squamous cell carcinoma invading into the muscularis mucosa or submucosa: a multicenter retrospective study. J Gastroenterol. 2021;56:620–32.
56. Berger NG, Ridolfi TJ, Ludwig KA. Delayed gastrointestinal recovery after abdominal operation - role of alvimopan. Clin Exp Gastroenterol. 2015;8:231–5.
57. Tahiri M, Sikder T, Maimon G, Teasdale D, Hamadani F, Sourial N, et al. The impact of postoperative complications on the recovery of elderly surgical patients. Surg Endosc. 2016;30:1762–70.
58. Huang C-W, Wu T-H, Hsu H-Y, Pan K-T, Lee C-W, Chong S-W, et al. Reappraisal of the role of alkaline phosphatase in hepatocellular carcinoma. J Pers Med. 2022;12:518.
59. Kim SH, Shin K-H, Moon S-H, Jang J, Kim HS, Suh J-S, et al. Reassessment of alkaline phosphatase as serum tumor marker with high specificity in osteosarcoma. Cancer Med. 2017;6:1311–22.

Publisher's Note