Original Article

DOI: 10.5582/bst.2024.01312

Machine learning-based prognostic prediction and surgical guidance for intrahepatic cholangiocarcinoma

Long Huang[1,2,§], Jianbo Li[1,2,§], Shuncang Zhu[1,2,§], Liang Wang[3], Ge Li[4], Junyong Pan[5], Chun Zhang[6], Jianlin Lai[1,2], Yifeng Tian[1,2,*], Shi Chen[1,2,*]

[1] Shengli Clinical Medical College of Fujian Medical University, Fuzhou, Fujian, China;
[2] Department of Hepatobiliary Pancreatic surgery, Fujian Provincial Hospital, Fuzhou University Affiliated Provincial Hospital, Fuzhou, Fujian, China;
[3] Department of Hepatopancreatobiliary Surgery, First Affiliated Hospital of Fujian Medical University, Fuzhou, Fujian, China;
[4] Department of Hepatobiliary Surgery, Fujian Medical University Union Hospital, Fuzhou, Fujian, China;
[5] Department of Hepatobiliary and Pancreatic Surgery, the Second Affiliated Hospital, Fujian Medical University, Quanzhou, Fujian, China;
[6] Department of General Surgery, Mindong Hospital Affiliated to Fujian Medical University, Ningde, Fujian, China.

SUMMARY The prognosis following radical surgery for intrahepatic cholangiocarcinoma (ICC) is poor, and optimal follow-up strategies remain unclear, with ongoing debates regarding anatomic resection (AR) versus non-anatomic resection (NAR). This study included 680 patients from five hospitals, comparing a combination of eight feature screening methods and 11 machine learning algorithms to predict prognosis and construct integrated models. These models were assessed using nested cross-validation and various datasets, benchmarked against TNM stage and performance status. Evaluation metrics such as area under the curve (AUC) were applied. Prognostic models incorporating screened features showed superior performance compared to unselected models, with AR emerging as a key variable. Treatment recommendation models for surgical approaches, including DeepSurv, neural network multitask logistic regression (N-MTLR), and Kernel support vector machine (SVM), indicated that N-MTLR's recommendations were associated with survival benefits. Additionally, some patients identified as suitable for NAR were within groups previously considered for AR. In conclusion, three robust clinical models were developed to predict ICC prognosis and optimize surgical decisions, improving patient outcomes and supporting shared decision-making for patients and surgeons.

Keywords intrahepatic cholangiocarcinoma, individualized treatment, machine learning, prediction tool, shared decision-making

1. Introduction

Intrahepatic cholangiocarcinoma (ICC) ranks as the second most prevalent primary liver cancer, following hepatocellular carcinoma (HCC). For individuals with resectable ICC, the prognosis post-resection is discouraging, with a 5-year survival rate of only 25-35%. Notably, tumor recurrence accounts for the majority of deaths, contributing to 60-70% of cases (*1-3*). Consequently, precise prognostic assessment is of significant importance to guide personalized treatment strategies and improve the overall prognosis for ICC patients. The majority of clinical investigations concerning ICC rely on radiomic features to predict prognosis. However, it is a challenge to acquire radiomic features, and determining the region of interest (ROI) introduces subjectivity. As a result, these models are inherently intricate and hard to interpret (*4,5*). Consequently, these factors pose significant obstacles to the practical clinical application of such models.

Despite significant research advancements such as chemotherapy, targeted therapy, and immunotherapy, which have provided valuable scientific and clinical insights into the treatment of ICC (*6-8*), surgical resection remains the main potentially curative treatment. In the case of HCC, there has been frequent discussion about the difference in long-term prognosis between anatomic resection (AR) and non-anatomic resection (NAR) (*9-11*). However, in the context of ICC, the advantages of AR versus NAR remain uncertain (*12-14*). It is worth noting that the number of patients with ICC combined with cholelithiasis is higher in Eastern countries compared with that in Western countries, and the specific surgical approach in such cases remains undetermined (*15*). In conventional clinical studies, conclusions are often drawn at the population level, but these conclusions may not necessarily benefit patients in real-world scenarios (*16*).

Artificial intelligence, particularly machine learning, exhibits undeniable advantages in addressing these issues. Machine learning has the capacity to enhance population-level evidence and facilitate the development of personalized treatment strategies for patients. However, only a few studies have compared various machine learning methods to construct high-performance predictive models for predicting recurrence and survival rates in ICC patients following radical surgery. Furthermore, there is a notable gap in the literature concerning personalized predictions for selecting surgical approaches in cases of ICC.

In the present study, multiple machine learning algorithms, dimensionality reduction algorithms, and integrated learning methods were employed to investigate models capable of predicting post-radical surgery prognosis for ICC. Additionally, these models were compared with those developed by the American Joint Committee on Cancer (AJCC) 8th edition staging system. Notably, in the model interpretation, AR significantly reduces the risk of recurrence and mortality. Therefore, multiple models were developed, including deep learning, to explore personalized surgical recommendation models to enhance patient prognosis. We encapsulate the algorithm as a program and upload it to GitHub. The decision-making procedure in these models was analyzed to gain valuable insights into the factors influencing the prognosis of ICC.

2. Patients and Methods

2.1. Patients

Data were gathered from five hospitals (Fujian provincial hospital 218 patients, First affiliated hospital of Fujian medical university 163 patients, Fujian medical university union hospital 117 patients, The second affiliated hospital, Fujian medical university 133 patients, Mindong hospital affiliated to Fujian medical university 49 patients). To gather a comprehensive dataset, three hospitals in Fuzhou were used as the training-validation set, while the remaining data were used as the external test set. To ensure appropriate patient follow-up for at least 2 years, the collected data encompasses a period beginning from January 2021 to January 2023. Within this period, an event-free outcome was defined as no death in 2 years and no recurrence in 1 year.

The inclusion criteria for this study were as follows: *i*) confirmation of ICC through postoperative histopathology; *ii*) initial treatment was surgical resection (involving either AR or NAR); *iii*) patients with R0 margins. Conversely, the exclusion criteria were as follows: *i*) patients with severe underlying diseases; *ii*) patients with pre-resection metastases; *iii*) patients who passed away within 30 days of surgery; *iv*) patients who died to causes other than disease under investigation. An overview of the study workflow is depicted in Figure 1.

2.2. Definition of anatomic resection

AR was defined as the complete removal of the Couinaud segment, which included procedures like segmental hepatectomy, lobectomy, or hemihepatectomy. On the other hand, NAR was defined as the partial removal of portal tributaries associated with the affected segment. This classification includes procedures involving partial resection and tumor enucleation (*17,18*).

2.3. Development and validation of models

Following data preprocessing, 11 machine learning algorithms were applied to each of the 8 feature selections to predict recurrence and mortality in ICC. Subsequently, the top three models with the highest Area Under the Curve (AUC) for each feature were selected to explore the integrated model. The single model with the highest AUC, the integrated model, and the TNM-based model were evaluated through cross-validation and their performance on the external validation dataset. In this regard, the receiver-operator characteristic (ROC) curve, AUC, and decision curve analysis (DCA) were employed as indicators. The integrated model was selected when it outperformed other models in various aspects. Conversely, when a single algorithm exhibited superior performance and high AUC, it was selected as the ultimate model. Furthermore, for each feature set, the model with the highest AUC was selected for model interpretation and variable importance ranking. In the present study, AR was selected as an important parameter in all features for predicting recurrence and mortality. The analysis revealed that AR correlates with the probability of recurrence and mortality in ICC patients. To address the risk of overfitting in deep learning, the prior training-validation set was utilized for the training set, and the external validation set was utilized for the validation set in the surgical recommendation model. After excluding intraoperative and postoperative variables, the variables jointly selected by 8 feature selections were incorporated into the prediction models for surgical modality. The models were achieved using DeepSurv (*19*), neural network multitask logistic regression (N-MTLR) (*20*), and Kernel support vector model (SVM) as the base models. Hazard ratio (HR), median overall survival (OS), and significance were determined through log-rank tests for eligible recommended treatments. Subsequently, the appropriate recommended treatment model was employed for individual predictions, and the respective eligible populations for AR and NAR were summarized for personalized forecasts. The calculation of HR was modified utilizing the inverse probability of treatment weighting (IPTW) method to balance potential selections between AR and NAR for patients (*21*).

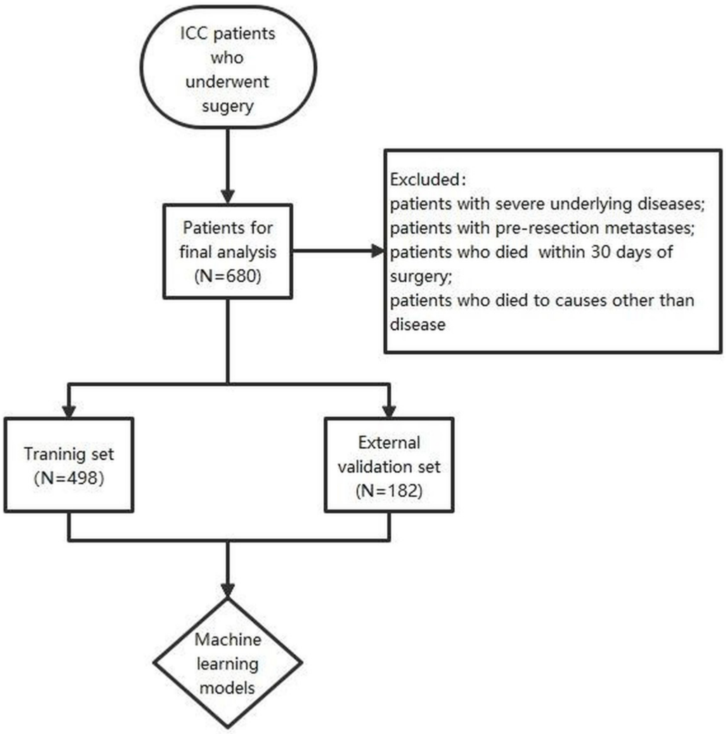

Figure 1. The overall flowchart of the study.

2.4. Statistics analysis

All analyses were carried out using Python 3.7 and R 4.1.3. $P < 0.05$ was considered Statistically significant. Details of data preprocessing, modeling, and validation approaches are presented in Supplemental Data (*https://www.biosciencetrends.com/action/getSupplementalData.php?ID=227*).

3. Results

3.1. Patient characteristics

The study involved 680 patients with a median follow-up of 932 days. The training-validation dataset consisted of data from 498 patients, and the data for the remaining 182 patients were used as the external test dataset. Patient demographic and clinical parameters are summarized in Table 1, indicating the external test dataset had higher percentages of patients with hepatolithiasis (56.0% *vs.* 53.8%) and TNM8 N1-stage (40.1% *vs.* 34.1%). On the other hand, some indicators were lower in the external test dataset in comparison with the training dataset, including TNM8 T1a-stage (11.5% *vs.* 15.1%) and AR (42.8% *vs.* 62.9%).

3.2. Model construction, validation, and interpretation for predicting prognosis

Following data preprocessing, 19 continuous and 7 discrete features were used in machine learning. Since adjuvant chemotherapy after surgery for ICC has become a standard treatment, 672 patients (99%) in this study cohort received standard adjuvant chemotherapy, with only 8 patients not receiving chemotherapy. Given the high consistency of adjuvant chemotherapy in this study, chemotherapy was not included as an independent variable in the analysis. The features retained after each feature selection method are shown in Supplemental Table S1 (*https://www.biosciencetrends.com/action/getSupplementalData.php?ID=227*). The results obtained from incorporating machine learning algorithms and feature sets are shown in Figure 2. The presented heatmaps illustrate AUC for various combinations of machine learning algorithms and feature selection methods. Meanwhile, the nested cross-validation approach was utilized to optimize model hyperparameters and evaluate models. Evaluations of the benchmark model, single models, and integrated model on the training-validation and external test datasets are presented in Figure 3, Supplemental Figure S1 and Table S2 (*https://www.biosciencetrends.com/action/getSupplementalData.php?ID=227*). In the context of the recurrence and mortality prediction table, despite some overlap in the confidence intervals of the baseline model, the proposed model demonstrated superior performance (Table 2). More specifically, Figure 4 and Supplemental Figure S6-S7 (*https://www.biosciencetrends.com/action/getSupplementalData.php?ID=227*) indicate that the integrated model exhibited enhanced consistency in both the calibration curve and DCA for both recurrence and mortality. While there were negligible deviations in the AUC for the recurrence model between the integrated and single models, the integrated model outperformed

Table 1. Demographic and clinical parameters for combined training-validation and test datasets (before imputation)

Parameter	Combined Training & Validation Sets (*n* = 498)	External Test Set (*n* = 182)	*p* value
Age, median (IQR), year	61 (54.0-67.0)	64.5 (55.0-70.0)	< 0.001
Sex, *n* (%)			0.17
Female	234 (47.0)	97 (53.2)	
Male	264 (53.0)	85 (46.7)	
BMI, median (IQR), Kg/m²	22.890 (20.9,24.4)	23.1800 (21.5,24.9)	0.1
Missing, *n* (%)	7 (1.4)	4 (2.1)	
Hepatolithiasis, *n* (%) (I)			0.66
Yes	268 (53.8)	102 (56.0)	
No	230 (46.2)	80 (44.0)	
Vascular invasion, *n* (%) (I)			0.6
Yes	236 (47.4)	90 (49.5)	
No	263 (52.6)	92 (50.5)	
Acute cholangitis, *n* (%)			< 0.001
Yes	208 (41.8)	113 (62.1)	
No	290 (58.2)	69 (37.9)	
TNM8 T stage, *n* (%)			0.19
T1a	75 (15.1)	21 (11.5)	
T1b	60 (12.0)	27 (14.8)	
T2	48 (9.6)	27 (14.8)	
T3	252 (50.6)	82 (45.1)	
T4	63 (12.7)	25 (13.7)	
TNM8 N stage, *n* (%)			0.18
N1	170 (34.1)	73 (40.1)	
N0	328 (67.2)	109 (59.9)	
Tumor distribution, *n* (%)			0.19
Left hemiliver	257 (51.6)	83 (45.6)	
No	241 (48.3)	99 (54.4)	
Maximum tumor diameter, median (IQR), cm	5.0 (3.5,7.0)	4.85 (3.4, 7.0)	0.7
Anatomic resection, *n* (%)			< 0.001
Yes	313 (62.9)	78 (42.8)	
No	185 (37.1)	104 (57.1)	
Operative blood loss, median (IQR), mL	400 (200, 600)	450 (200, 700)	0.81
Missing, *n* (%)	15 (3.0)	5 (2.7)	
Number of lymphatic dissection, *n* (%)	5 (3.7)	5 (4.8)	0.23
Neutrophil ratio, median (IQR), %	87.8 (84.2, 90.6)	88.6 (84.9, 91.3)	0.12
Missing, *n* (%)	17 (3.4)	0 (0.0)	
Lymphocyte ratio, median (IQR), %	6.3 (4.6, 8.5)	6.60 (4.9, 8.8)	0.47
*Lymphocyte count, median (IQR), 10^9/L	0.80 (0.6, 1.1)	0.8950 (0.6, 1.2)	0.14
Missing, *n* (%)	17 (3.4)	5 (2.7)	
HB, mean(IQR), g/L	112 (97.0, 127.0)	108.0 (96.0 , 124.5)	0.13
Missing, *n* (%)	33 (6.6)	19 (10.4)	
*WBC count, median (IQR), 10^9/L	12.77 (10.1, 15.1)	13.350 (10.4, 15.8)	0.2
Missing, *n* (%)	9 (1.8)	0 (0.0)	
*PLT count, median (IQR), 10^9/L	186 (139.0, 232.0)	198. 0 (145.5, 240.5)	0.15
Missing, *n* (%)	3 (0.6)	7 (3.0)	
CA199, median (IQR), U/mL	80.4 (12.9,449.9)	83.1 (14.9, 611.3)	0.61
CA125, median (IQR), U/mL	10.4 (4.1,28.2)	9.9 (3.8, 27.1)	0.76
Missing, *n* (%)	78 (15.7)	44 (24.2)	
CEA, median (IQR), ng/mL	2.9000 (1.4,5.4)	3.0450 (1.5, 5.7)	0.67
ALB, median (IQR), g/L	29.9 (26.0,33.1)	29.8 (26.0, 32.9)	0.63
Missing, *n* (%)	1 (0.2)	1 (0.5)	
DBIL, median (IQR), umol/L	9.450 (5.4,19.9)	9.40 (5.7, 23.0)	0.77
IBIL, median (IQR), umol/L	11.500 (7.3,19.4)	11.450 (7.4, 18.0)	0.8
ALP, median (IQR), U/L	102.0 (70.0,212.5)	115.0 (73.0, 228.5)	0.32
Missing, *n* (%)	49 (9.8)	10 (5.5)	
GGT, median (IQR), U/L	114 (76,201)	120.9550 (73.0, 228.5)	0.27
Missing, *n* (%)	29 (5.8)	0 (0.0)	
Recurrence at 1 year, *n* (%)	274 (55.0)	114 (62.6)	0.081
Death at 2 years, *n* (%)	249 (50.0)	96 (52.7)	0.55
Median length of OS	739.50 (432.25, 1125.25)	831.5 (656.0, 1201.0)	< 0.001

Only features used in modeling are presented. Categorical data are summarized with median, percentages, and p-values pertaining to Fisher's exact test. Continuous data are summarized with median and IQR, and p-values pertain to the Wilcoxon rank sum test. Variables marked with (I) were based on preoperative imaging studies, while all other tumor-related variables were based on histopathological examination. *Abbreviations*: IQR, interquartile range; BMI, body mass index; CA199, carbohydrate antigen 199; CA125, carbohydrate antigen 125; CEA, carcinoembryonic antigen; HB, hemoglobin; WBC, white blood cell; PLT, platelet; ALB, albumin; DBIL, direct bilirubin; IBIL, indirect bilirubin; ALP, alkaline phosphatase; GGT, γ-glutamyl transpeptidase.

Table 2. AUC with 95% confidence intervals for each prediction model's validation and external test dataset

Outcome	Validation set		External test set	
	AUC	95% CI	AUC	95% CI
OS				
Single model	0.949	0.912-0.974	0.848	0.791-0.907
Integrated model	0.923	0.913-0.942	0.917	0.887-0.951
TNM based model	0.841	0.813-0.878	0.857	0.804-0.917
Recurrence				
Single model	0.893	0.861-0.924	0.946	0.918-0.981
Integrated model	0.918	0.903-0.937	0.877	0.838-0.919
TNM based model	0.807	0.773-0.842	0.825	0.764-0.887

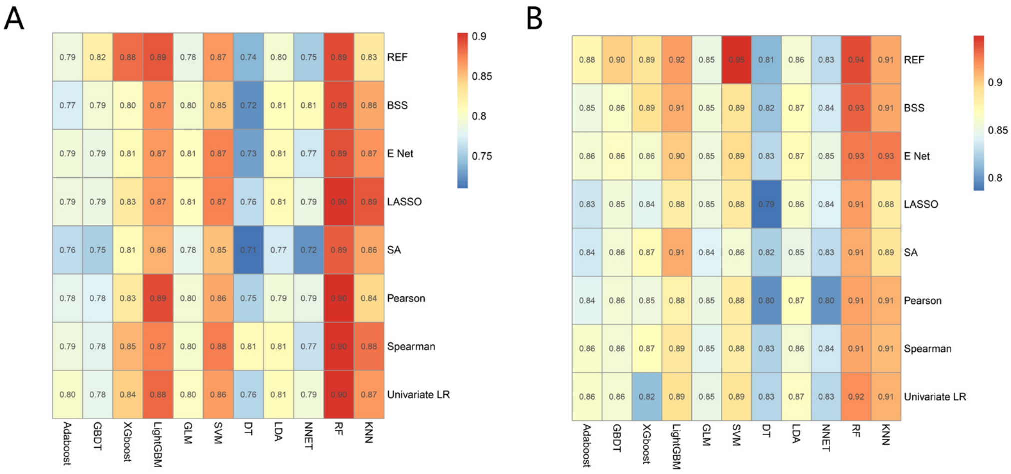

Figure 2. Heatmaps illustrating the performance of each machine learning algorithm (columns) with each feature reduction method (rows). (A) Heatmap for predicting recurrence; **(B)** Heatmap for predicting overall survival. *Abbreviations*: RFE, recursive feature elimination; BSS, best subset selection; E Net, elastic net; LASSO, least absolute shrinkage and selection operator; SA, simulated annealing; Univariate LR, univariate logistic regression; AdaBoost, adaptive boosting machine; GBDT, gradient boosting decision tree; XGboost, extreme gradient boosting machine; LightGBM, light gradient boosting machin; GLM, generalised linear model; SVM, support vector machine; DT, decision tree; LDA, linear discriminant analysis; NNET, neural network; RF, random forest; KNN, K nearest neighbours.

single models in terms of DCA and AUC for mortality. Considering the complexity and efficiency of the model implementation, random forest (RF) was used for recurrence, while the integrated model of SVM, RF, and K-nearest neighbors (KNN) was used for OS. When ranking the importance of both recurrence and mortality variables, AR held a more critical position. The models were further explained through the Shapley additive explanations (SHAP) analysis (Supplemental Figure S2 and S5, *https://www.biosciencetrends.com/action/getSupplementalData.php?ID=227*). The analysis indicates that the presence of vascular invasion and hepatolithiasis in patients increases the mortality rate, while AR reduces the mortality rate. In SHAP analysis of recurrence and mortality, operative blood loss exhibited unstable patterns across various models. This parameter can either increase or decrease the outcome variable.

3.3. Construction of a surgical prediction model

Given the significant importance of AR in SHAP analysis and the results of previous feature screening, data on BMI, CA199, presence of vascular infiltrates and hepatolithiasis on imaging, and AR were included in the surgical approach recommendation models. Hyperparametric search results for each model are shown in Supplemental Table S3-S4, *https://www.biosciencetrends.com/action/getSupplementalData.php?ID=227*). The N-MTLR model recommended treatments that were associated with significantly higher survival in both the training and validation datasets (Table 3 and Figure 4), with HR of 0.333 (95% CI: 0.262-0.424; p < 0.001) in the training dataset and 0.561 (95% CI: 0.357-0.882; p = 0.012) in the validation dataset. To consider potential patient selection differences between AR and NAR, comparisons were conducted using IPTW, with higher weight assigned to underrepresented patients in each treatment group. IPTW results showed a performance similar to that of conventional HR. In the DeepSurv, N-MTLR model, and Kernel SVM models, AR was recommended for 571 (84.0%), 493 (72.5%), and 304 (44.7%) patients, respectively. Among patients with hepatolithiasis, AR was recommended for 199 (53.8%) patients. Notably, in the subgroup of patients with both hepatolithiasis and vascular invasion, surgical recommendations based on the N-MTLR model also

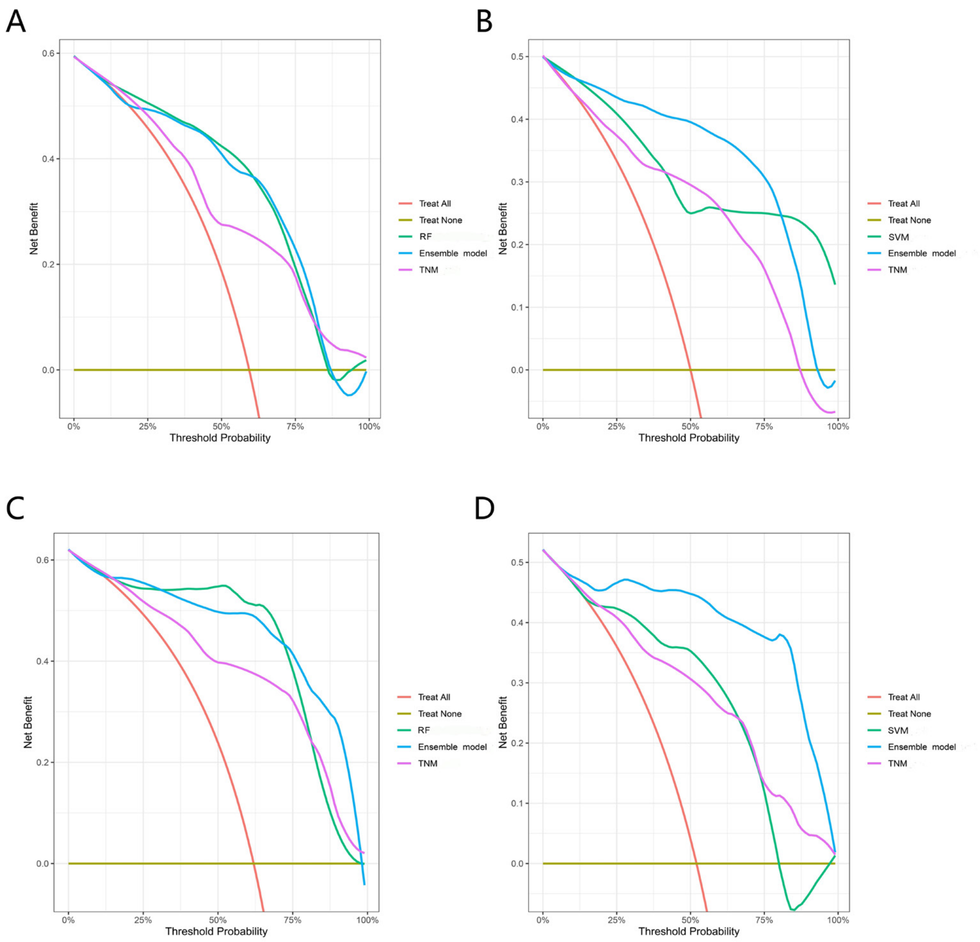

Figure 3. The DCA curves for single algorithms with the highest AUC, the ensemble model, and the baseline model in the nested cross-validation (A-B) and external validation dataset (C-D). (A) The DCA curves for predicting recurrence in the nested cross-validation; **(B)** The DCA curves for predicting overall survival in the nested cross-validation; **(C)** The DCA curves for predicting recurrence in the external validation dataset; **(D)** The DCA curves for predicting overall survival in the external validation dataset.

Table 3. Survival predictions for treatment according to model recommendations

Model	Validation set: Patients receiving recommended treatment	Validation set: Patients not receiving recommended treatment	HR (95% CI)	*p* value	HR, IPTW (95% CI)	*p* value
N-MTLR						
Development Set	980.0 (812.0)	567.0 (478.0)	0.333 (0.262, 0.424)	< 0.001	0.409 (0.316, 0.528)	< 0.001
Validation Set	858.0 (259.0)	769.0 (666.0)	0.561 (0.357, 0.882)	0.011	0.597 (0.386, 0.925)	0.021
DeepSurv						
Development Set	815.0 (99.0)	862.0 (603.0)	0.919 (0.780, 1.251)	0.91	2.269 (1.590, 3.238)	< 0.001
Validation Set	792.0 (745.0)	701.0 (673.0)	0.8602 (0.490, 1.510)	0.60	1.205 (0.605, 2.402)	0.59
Kernel SVM						
Development Set	637.0 (709.0)	740.0 (694.0)	3.662 (2.000, 6.709)	0.053	6.703 (3.478, 12.918)	< 0.001
Validation Set	821.0 (52.0)	832.0 (559.0)	1.722 (0.987, 3.005)	0.39	2.236 (1.335, 3.746)	0.0022

Abbreviations: HR, hazard ratio; IPTW, inverse probability of treatment weighting; IQR, interquartile range; N-MTLR, neural multitask logistic regression; OS, overall survival; SVM, Support Vector Machine. HRs are given for the patients who received the recommended treatment compared with those who did not.

demonstrated benefits for patients (Supplemental Figure S8, *https://www.biosciencetrends.com/action/getSupplementalData.php?ID=227*). In cases of confirmed hepatolithiasis without vascular invasion on imaging, AR was recommended for 12 (8.5%) patients. The procedure for integrating the recommendation model is available at *https://github.com/haizhili/Prognostic_Prediction_and_Surgical_Guidance_for_ICC*.

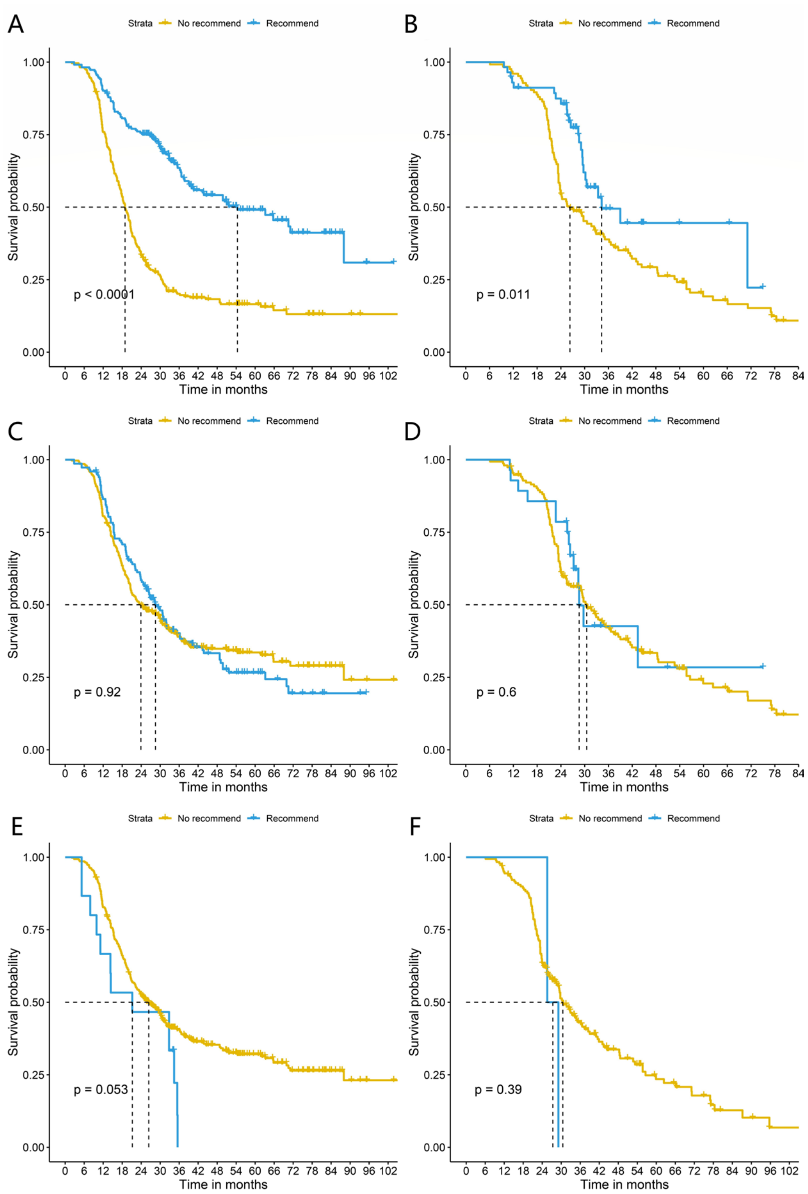

Figure 4. Results for (A-B) N-MLTR, (C-D) DeepSurv, and (E-F) Kernel SVM models. (A) The Kaplan-Meier curves for the N-MLTR model in the training dataset; **(B)** The Kaplan-Meier curves for the N-MLTR model in the validation dataset; **(C)** The Kaplan-Meier curves for the DeepSurv model in the training dataset; **(D)** The Kaplan-Meier curves for the DeepSurv model in the validation dataset; **(E)** The Kaplan-Meier curves for the Kernel SVM model in the training dataset; **(F)** The Kaplan-Meier curves for the Kernel SVM model in the validation dataset.

4. Discussion

In the present multicenter study, an incremental analysis was conducted. In the first step, multiple machine learning algorithms and various dimensionality reduction techniques were compared using routine medical data to develop and validate predictive models. These models can effectively and precisely predict recurrence and OS. AR emerged as a variable, consistently appearing in all feature selections. It is worth noting that AR is a variable with strong correlations with both the recurrence and OS. Furthermore, considering the importance of explaining medical decisions to patients, the models with the highest AUC for each feature selection were interpreted. These interpretations consistently highlighted the risk-reducing effects of AR on both recurrence and mortality, reflecting its overall benefit in the population. However, it should be indicated that the model interpretation focuses on the overall benefit of AR in the population but fails to analyze the advantages of NAR, which remains valuable in real-world clinical practice (*22,23*). Therefore, in the second step, surgical modality recommendation models were developed for both AR and NAR, including deep learning techniques, to enable individual-level predictions. The findings revealed that the majority of patients were suitable candidates for AR. Meanwhile, it was found that individuals who could be ideal candidates for NAR were also considered suitable candidates for AR. This refinement in population characteristics provides valuable insights for clinical practice.

Accurate prediction of postoperative recurrence and survival among ICC patients holds critical importance (*24,25*). Although AR has demonstrated improved outcomes in HCC, revealing its benefits in ICC requires further investigations (*10,11*). Conventional treatment decisions typically encounter some shortcomings, including poor personalization, and dependence on physician preference and group-level data (*12-14*). To resolve these shortcomings and accurately predict ICC recurrence and survival, numerous predictive models using routine clinical data have been developed. Notably, a model based on the N-MTLR model was introduced, providing personalized surgical recommendations. This advancement benefits patients and assists physicians in making treatment decisions, thereby improving ICC care. The AUC values of various machine learning models for predicting recurrence and OS remain consistent across combined training-validation and external validation datasets. Minor performance variations were observed in OS models during cross-validation and validation on the external validation datasets. These variations were especially more pronounced through recursive feature selection. However, these models consistently outperformed the TNM-based prognostic model. An additional advantage of the developed models lies in the use of integrated modeling. Integrated models can enhance the final predictive performance beyond individual predictive models. This enhancement is achieved by combining diverse predictive models that have been trained using distinct architectures and hyperparameters. The integration of individual classifiers in a parallel manner increased consistency across various datasets. Notably, integrated models are not sensitive to the challenges imposed by the "curse of dimensionality", where the predictive or discriminative efficacy of a model rapidly declines as the data dimensionality increases (*26-28*). It is worth noting that classical models were employed in the present study to predict recurrence, which can provide clear explanations for their predictions. In the medical field, model explanation facilitates understanding reasons for making particular decisions.

This article employs several classical feature selection methods such as annealing, recursive feature elimination, optimal subset, and correlation coefficient to select variables. These methods are used to determine variables from different perspectives, which can improve the accuracy and generalization ability of prognostic models and reduce overfitting (*29*). Moreover, AR was screened out in different feature selection methods, indicating that it is statistically significant. Meanwhile, the model interpretation showed that AR affects the survival of patients with ICC. Therefore, several variables were used after removing intraoperative and postoperative variables, most often screened out by feature engineering as inputs to the surgery recommendation model.

In the model interpretation, it was observed that AR and the absence of hepatolithiasis and vascular invasion may have positive effects on the prognosis of ICC patients. However, other variables such as BMI may also affect the outcome differently, suggesting that the effect of the same variable on the result is not unique across models (*22,23*). Accordingly, it was inferred that AR does not benefit all patients and an individual-level analysis of individuals who were recommended AR to gain a deeper understanding of its applicability.

In this study, the surgical recommendation model based on the N-MTLR model indicated that 493 patients (72.5%) might be suitable candidates for AR. Unlike previous population-based studies that relied on a single standard, the proposed model comprehensively considered multiple crucial preoperative variables, providing more detailed suggestions for individualized decision-making. Overall, AR demonstrated significant advantages in terms of recurrence and survival rates for most patients, particularly for those with larger tumors and without liver dysfunction, making it the more appropriate surgical option. However, the model also showed that non-anatomic resection (NAR) might be a better option for certain patient groups. NAR offers advantages such as being less invasive, preserving more liver tissue, and promoting faster postoperative recovery. Specifically, for patients with smaller tumors, NAR and AR showed minimal differences in recurrence and survival rates, and NAR could reduce unnecessary liver tissue removal, preserving more liver function. Additionally, NAR proved beneficial for patients with limited liver function (*e.g.*, those with cirrhosis or other chronic liver diseases), significantly reducing the risk of postoperative liver failure while maintaining a high survival rate. Among 42 patients (13.0%) with vascular invasion detected through imaging, NAR may also be the more appropriate choice. It should be noted that hepatolithiasis is classified as a poor prognostic factor for ICC, though further studies are needed to determine the best surgical approach for ICC patients with hepatolithiasis. In clinical practice, AR is typically used to resect biliary lesions associated with stones. However, the study showed that 46.2% of ICC patients with hepatolithiasis (171 patients) might be more suitable candidates for NAR, highlighting the importance of individualized treatment decisions.

The current approach to clinical decision-making often relies on physician preference and group-level evidence-based clinical practices to advise patients. However, this method of decision-making may not always offer the most suitable treatment options for individuals and may not effectively incorporate the unique characteristics of each patient. However, clinicians can employ various algorithms to provide individualized treatment recommendations for patients (*16,30,31*). Researchers typically use clinical data to investigate the benefits of AR and NAR. For instance, investigations reveal that patients with ICC combined with hepatolithiasis benefit more from AR than NAR (*32*). In the developed model, only 12 patients (8.5%) with hepatolithiasis and no vascular invasion were considered suitable candidates for AR. In contrast to conventional machine learning research that primarily focuses on predicting prognosis, this article focuses on developing a personalized surgical modality recommendation algorithm. This algorithm not only enhanced patient outcomes but also can be encapsulated in a compact executable file on computers. This feature simplifies its clinical application for healthcare providers. The proposed recommendation model used routine clinical data, facilitating its application and disseminating the research. Meanwhile, instead of traditional nomogram scores, the executable file directly provides an appropriate surgical approach for the patient, which makes the output more concise and easier to use.

In addition to remarkable advantages, this study also has some shortcomings. First, the data used in this study were retrospective, potentially introducing regression bias. The interpolation method used to address missing data may affect the integrity of the clinical data, emphasizing the need for future international prospective clinical trials to validate these findings. Secondly, the data used in this article did not incorporate information from radiomics. As a result, more advanced imaging-based models may outperform the proposed model. However, this model utilized routine preoperative and postoperative examination results as input data, which effectively minimized the additional time and costs typically associated with data preparation and processing. This approach also simplifies replication in primary care hospitals. Finally, while the absence of specific data regarding postoperative adjuvant treatments in the raw data might affect the results, it is important to note that more than 99% of the patients in this study received standard postoperative adjuvant chemotherapy. The high consistency of adjuvant therapy within the patient population significantly reduced the potential impact of such treatments on the comparison between surgical approaches (AR and NAR). This consistency enhanced the model's applicability and reliability in this standardized treatment population. The present study provides guidance for developing models focusing on surgical procedure data.

In conclusion, this article compares various machine learning algorithms and feature selection methods to develop two predictive models for recurrence and OS following radical resection in ICC patients. The results demonstrate that the developed model outperforms conventional approaches. Additionally, an advanced preoperative surgical recommendation system based on clinical data was introduced. This model enhances patient-centered decision-making and suggests personalized treatments. The recommended surgical approach exhibited significant improvements in patient prognosis. This study offers fresh insights into the clinical application of surgical procedures for ICC, emphasizing the potential for more effective treatment strategies.

Acknowledgements

We thank all patients and institutions involved in this study. We are deeply saddened by the unfortunate passing of Jianbo Li. We express our heartfelt gratitude and respect for his dedication and contributions to this research.

Funding: This work was supported by grants from Fujian Research and Training Grants for Young and Middle-aged Leader in Healthcare (2021060 to Shi Chen); Natural Science Foundation of Fujian Province, Youth Program (2022J05212 to Long Huang); and Fujian Provincial Health Science and Technology Program for Cultivating Young and Middle-aged Key Talents (2021GGA006 to Long Huang).

Conflict of Interest: The authors have no conflicts of interest to disclose.

References

1. Banales JM, Marin JJG, Lamarca A, *et al*. Cholangiocarcinoma 2020: The next horizon in mechanisms and management. Nat Rev Gastroenterol Hepatol. 2020; 17:557-588.
2. Sposito C, Ratti F, Cucchetti A, Ardito F, Ruzzenente A, Di Sandro S, Maspero M, Ercolani G, Di Benedetto F, Guglielmi A, Giuliante F, Aldrighetti L, Mazzaferro V. Survival benefit of adequate lymphadenectomy in patients undergoing li ver resection for clinically node-negative intrahepatic cholangiocarcinoma. J Hepatol. 2023; 78:356-363.
3. Wang Y, Li J, Xia Y, Gong R, Wang K, Yan Z, Wan X, Liu G, Wu D, Shi L, Lau W, Wu M, Shen F. Prognostic nomogram for intrahepatic cholangiocarcinoma after partial hepatectomy. J Clin Oncol. 2013; 31:1188-1195.
4. Leech M, Osman S, Jain S, Marignol L. Mini review: Personalization of the radiation therapy management of prostate cancer using MRI-based radiomics. Cancer Lett. 2021; 498:210-216.
5. Vial A, Stirling D, Field M, Ros M, Miller AA. The role of deep learning and radiomic feature extraction in cancer-specific predictive modelling: A review. Transl Cancer Res. 2018; 7:803-816.
6. Moris D, Palta M, Kim C, Allen PJ, Morse MA, Lidsky ME. Advances in the treatment of intrahepatic cholangiocarcinoma: An overview of the current and future therapeutic landscape for clinicians. CA Cancer J Clin. 2023; 73:198-222.
7. Vogel A, Bridgewater J, Edeline J, Kelley RK, Klümpen HJ, Malka D, Primrose JN, Rimassa L, Stenzinger A, Valle JW, Ducreux M; ESMO Guidelines Committee. Biliary tract cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. Ann Oncol. 2023; 34:127-140.
8. Coffman-D'Annibale KL, Greten TF. Top advances of the year: Hepatobiliary cancers. Cancer. 2023; 129:1149-1155.
9. Jiao S, Li G, Zhang D, Xu Y, Liu J, Li G. Anatomic versus non-anatomic resection for hepatocellular carcinoma, do we have an answer? A meta-analysis. Int J Surg. 2020; 80:243-255.
10. Minagawa M, Mise Y, Omichi K, Ichida H, Mizuno T, Yoshioka R, Imamura H, Yanagisawa N, Inoue Y, Takahashi Y, Saiura A. Anatomic resection for

hepatocellular carcinoma: Prognostic impact ass essed from recurrence treatment. Ann Surg Oncol. 2022; 29:913-921.
11. Shindoh J, Kobayashi Y, Umino R, Kojima K, Okubo S, Hashimoto M. Successful anatomic resection of tumor-bearing portal territory delays long-term stage progression of hepatocellular carcinoma. Ann Surg Oncol. 2021; 28:844-853.
12. Li B, Song JL, Aierken Y, Chen Y, Yang JY. Nonanatomic resection is not inferior to anatomic resection for primary intrahepatic cholangiocarcinoma: A propensity score analysis. Sci Rep. 2018; 8:17799.
13. Si A, Li J, Yang Z, Xia Y, Yang T, Lei Z, Cheng Z, Pawlik TM, Lau WY, Shen F. Impact of anatomical versus non-anatomical liver resection on short- and long-term outcomes for patients with intrahepatic cholangiocarcinoma. Ann Surg Oncol. 2019; 26:1841-1850.
14. Wang WQ, Li J, Liang BY, Lv X, Zhu RH, Wang JL, Huang ZY, Yang SH, Zhang EL. Anatomical liver resection improves surgical outcomes for combined hepatocellular-cholangiocarcinoma: A propensity score matched study. Front Oncol. 2022; 12:980736.
15. Suzuki Y, Mori T, Momose H, Matsuki R, Kogure M, Abe N, Isayama H, Tazuma S, Tanaka A, Takikawa H, Sakamoto Y. Predictive factors for subsequent intrahepatic cholangiocarcinoma associated with hepatolithiasis: Japanese National Cohort Study for 18 years. J Gastroenterol. 2022; 57:387-395.
16. Pfob A, Mehrara BJ, Nelson JA, Wilkins EG, Pusic AL, Sidey-Gibbons C. Towards patient-centered decision-making in breast cancer surgery: Machine learning to predict individual patient-reported outcomes at 1-year follow-up. Ann Surg. 2023; 277:e144-e152.
17. Liu J, Zhuang G, Bai S, Hu Z, Xia Y, Lu C, Wang J, Wang C, Liu L, Li F, Wu Y, Shen F, Wang K. The comparison of surgical margins and type of hepatic resection for hepatocellular carcinoma with microvascular invasion. Oncologist. 2023; 28:e1043-e1051.
18. Zhang XP, Xu S, Lin ZY, Gao QL, Wang K, Chen ZL, Yan ML, Zhang F, Tang YF, Zhao ZM, Li CG, Lau WY, Cheng SQ, Hu MG, Liu R. Significance of anatomical resection and resection margin status in pa tients with HBV-related hepatocellular carcinoma and microvascular inv asion: A multicenter propensity score-matched study. Int J Surg. 2023; 109:679-688.
19. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: Personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Med Res. 2018; 18:24.
20. Fotso S. Deep neural networks for survival analysis based on a multi-task framework. *https://ui.adsabs.harvard.edu/abs/2018arXiv180105512F* (accessed September 1, 2024).
21. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Stat Med. 2015; 34:3661-3679.
22. Lundberg S, Lee SI. A unified approach to interpreting model predictions. *https://arxiv.org/pdf/1705.07874* (accessed September 1, 2024).
23. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?": Explaining the predictions of any classifier. ACM. *https://ui.adsabs.harvard.edu/abs/2016arXiv160204938T* (accessed September 1, 2024).
24. Ji GW, Jiao CY, Xu ZG, Li XC, Wang K, Wang XH. Development and validation of a gradient boosting machine to predict p rognosis after liver resection for intrahepatic cholangiocarcinoma. BMC Cancer. 2022; 22:258.
25. Shen X, Zhao H, Jin X, Chen J, Yu Z, Ramen K, Zheng X, Wu X, Shan Y, Bai J, Zhang Q, Zeng Q. Development and validation of a machine learning-based nomogram for prediction of intrahepatic cholangiocarcinoma in patients with intrahepatic lithiasis. Hepatobiliary Surg Nutr. 2021; 10:749-765.
26. Bayoudh K, Knani R, Hamdaoui F, Mtibaa A. A survey on deep multimodal learning for computer vision: Advances, trends, applications, and datasets. Vis Comput. 2022; 38:2939-2970.
27. Dong A, Li Z, Wang M, Shen D, Liu M. High-order laplacian regularized low-rank representation for multimodal dementia diagnosis. Front Neurosci. 2021; 15:634124.
28. Guggenmos M, Schmack K, Veer IM, Lett T, Sekutowicz M, Sebold M, Garbusow M, Sommer C, Wittchen HU, Zimmermann US, Smolka MN, Walter H, Heinz A, Sterzer P. A multimodal neuroimaging classifier for alcohol dependence. Sci Rep. 2020; 10:298.
29. Hindocha S, Charlton TG, Linton-Reid K, Hunter B, Chan C, Ahmed M, Robinson EJ, Orton M, Ahmad S, McDonald F, Locke I, Power D, Blackledge M, Lee RW, Aboagye EO. A comparison of machine learning methods for predicting recurrence and death after curative-intent radiotherapy for non-small cell lung cancer: Development and validation of multivariable clinical prediction models. EBioMedicine. 2022; 77:103911.
30. Kalimouttou A, Lerner I, Cheurfa C, Jannot AS, Pirracchio R. Machine-learning-derived sepsis bundle of care. Intensive Care Med. 2023; 49:26-36.
31. Liu P, Li S, Zheng T, *et al.* Subphenotyping heterogeneous patients with chronic critical illness to guide individualised fluid balance treatment using machine learning: A retrospective cohort study. EClinicalMedicine. 2023; 59:101970.
32. Wu JY, Huang WT, He WB, Dai GF, Lv JH, Qiu FN. Long-term outcomes of anatomic *vs.* non-anatomic resection in intrahepatic cholangiocarcinoma with hepatolithiasis: A multicenter retrospective study. Front Med (Lausanne). 2023; 10:1130692.

Received October 1, 2024; Revised October 14, 2024; Accepted November 24, 2024.

[§]These authors contributed equally to this work.
**Address correspondence to:*
Shi Chen and Yifeng Tian, Shengli Clinical Medical College of Fujian Medical University, Fujian Medical University, Department of Hepatobiliary Pancreatic Surgery, Fujian Provincial Hospital, Fuzhou University Affiliated Provincial Hospital, Fuzhou 350001, Fujian, China.
E-mail: wawljwalj@163.com (SC); tianyifeng@fjmu.edu.cn (YT)

Released online in J-STAGE as advance publication December 8, 2024.