Discover Oncology

ANALYSIS Open Access

An explainable machine learning model for predicting the risk of distant metastasis in intrahepatic cholangiocarcinoma: a population-based cohort study

Jinzhe Bi[1] and Yaqun Yu[1*]

*Correspondence:
Yaqun Yu
yyq0129@glmc.edu.cn
[1]Department of Hepatobiliary and Pancreatic Surgery, Affiliated Hospital of Guilin Medical University, Guilin 541001, China

Abstract

Background Distant metastasis (DM) in intrahepatic cholangiocarcinoma (ICC) is associated with poor prognosis and significantly high mortality. Therefore, developing an effective early prediction method for DM risk is crucial for tailoring personalized treatment plans and improving patient outcomes.

Methods This study included data from eligible ICC patients collected from the Surveillance, Epidemiology, and End Results (SEER) database between 2004 and 2021. Feature selection was performed using three methods, including least absolute shrinkage and selection operator (LASSO) regression, the Boruta algorithm, and recursive feature elimination (RFE). Eight machine learning (ML) algorithms were used to develop predictive models. Model performance was evaluated and compared using metrics such as the area under the receiver operating characteristic curve (AUC), area under the precision-recall curve (AUPRC), decision curve analysis (DCA), and calibration curves. The SHapley Additive exPlanations (SHAP) method was applied to rank feature importance and interpret the final model.

Result This study included 8536 ICC patients, including 2816 (33%) with DM. The intersection results of the three feature selection methods identified 10 predictive factors. Among the 8 ML models, the gradient boosting machine (GBM) model achieved the highest AUC (0.802), AUPRC (0.571), and accuracy (0.713), as well as the lowest Brier score (0.177), indicating a comparatively robust overall performance. Calibration curves and DCA indicated that the GBM model has good clinical decision-making capability and predictive performance. SHAP analysis identified the top 10 most relevant features, ranked by relative importance: surgery, N stage, tumor grade, T stage, tumor size, radiotherapy, tumor number, age at diagnosis, chemotherapy, and number of resected lymph nodes (LNs). Additionally, a web-based online calculator was developed to predict the risk of DM in ICC patients, available athttps://bijinzhe.shinyapps.io/icc_dm_shiny/.

Conclusion The GBM model demonstrated considerable potential in predicting the risk of DM in ICC patients. This could assist clinicians in formulating personalized treatment strategies, ultimately improving the overall prognosis of ICC patients.

Keywords Intrahepatic cholangiocarcinoma, Distant metastasis, Machine learning, Interpretable, Prediction model

1 Introduction

Intrahepatic cholangiocarcinoma (ICC) is a highly heterogeneous and aggressively invasive malignant tumor with a poor prognosis, ranking as the second most common primary liver cancer after hepatocellular carcinoma (HCC) [1, 2]. In recent years, the incidence and mortality rates of ICC have shown a significant increasing trend worldwide [3, 4]. Due to the absence of typical clinical symptoms and effective early diagnostic methods, most cases of ICC are detected only after the tumor has advanced to locally advanced or metastatic stages [5, 6]. Currently, radical resection remains a key treatment strategy for ICC. However, despite undergoing curative surgery, approximately 60% of patients experience local recurrence, while nearly 40% develop distant metastases (DM) [7–9]. Additionally, the incidence of occult metastatic lesions in ICC patients is high, with studies showing that over 30% may have occult DM that were not detected by preoperative magnetic resonance imaging [10]. The liver is reported to be the most common site of DM in patients with ICC, followed by the lungs, bones, and brain [11]. Once DM occurs, the overall prognosis worsens significantly, with a median survival of only 12.9 months [12, 13]. Therefore, it is urgent to increase the understanding of the risk factors and mechanisms of DM in ICC and developing accurate and effective predictive methods are crucial for formulating personalized diagnostic, treatment, and follow-up strategies, ultimately improving patient outcomes.

Machine learning (ML), a subset of artificial intelligence (AI), develops predictive models by automatically learning from large datasets to improve prediction algorithms, leading to its widespread application in medical decision-making systems in recent years [14]. ML-based models offer several advantages over traditional models, including the ability to handle complex data, identify nonlinear relationships, and improve prediction accuracy by integrating multiple variables [15]. However, ML models have more complex structures compared to traditional predictive models and often exhibit characteristics of black-box models in statistical analysis, which hinder interpretability [16]. Therefore, to address the challenge of "opening" black-box models, techniques such as SHapley Additive exPlanation (SHAP) values, derived from game theory, have been used to explain the results of complex algorithms, making their decision-making processes more transparent and easier for clinicians to understand [17].

In this study, we aim to utilize the Surveillance, Epidemiology, and End Results (SEER) database and apply ML algorithms to gain a deeper understanding of the risk factors for DM in ICC patients, while developing and validating the optimal predictive model. We use SHAP for both global and local explanations to clarify feature importance and interpret the model's predictions, thereby evaluating the practical significance of the model in predicting DM in ICC patients.

2 Materials and methods

2.1 Patient cohort

In this study, we utilized SEER*Stat (version 8.4.4) to collect detailed data on ICC patients from 17 cancer registries, covering the period from 2004 to 2021, including demographic, clinical, and pathological characteristics. The SEER database, created by the National Cancer Institute (NCI) in 1973, includes survival and cancer incidence data representing approximately 28% of the United States population, collected from reports by national cancer registries [18]. The inclusion criteria for this study were: (1) primary tumor site classified under disease codes C22.0 and C22.1, (2) classification according to the International Classification of Diseases for Oncology, Third Edition (ICD-O-3), with a histological/behavioral code of 8160/3. The exclusion criteria were as follows: patients with non-pathological diagnosis, M stage was not available, unknown tumor size, unknown information regarding surgery, and diagnosed with ICC patients only by autopsy or death certificate. The complete screening process and detailed study design are shown in Fig. 1. Since SEER data is publicly available and does not include any identifiable information or personal details, no additional ethical approval is required.

3 Study variables

The study variables included year of diagnosis, age at diagnosis, sex, race (White, Black, Asia-Pacific, other/unknown), marital status, annual household income, rural-urban classification, T-stage, N-stage, tumor size, tumor number, tumor grade, delayed treatment, surgery, number of resected lymph nodes (LNs), radiotherapy, and chemotherapy. The optimal cutoff values for age and tumor size as continuous variables were determined using X-tile software (version 3.6.1, Yale University, New Haven, CT, USA). X-tile tested various threshold values and selected the value with the smallest p-value as the

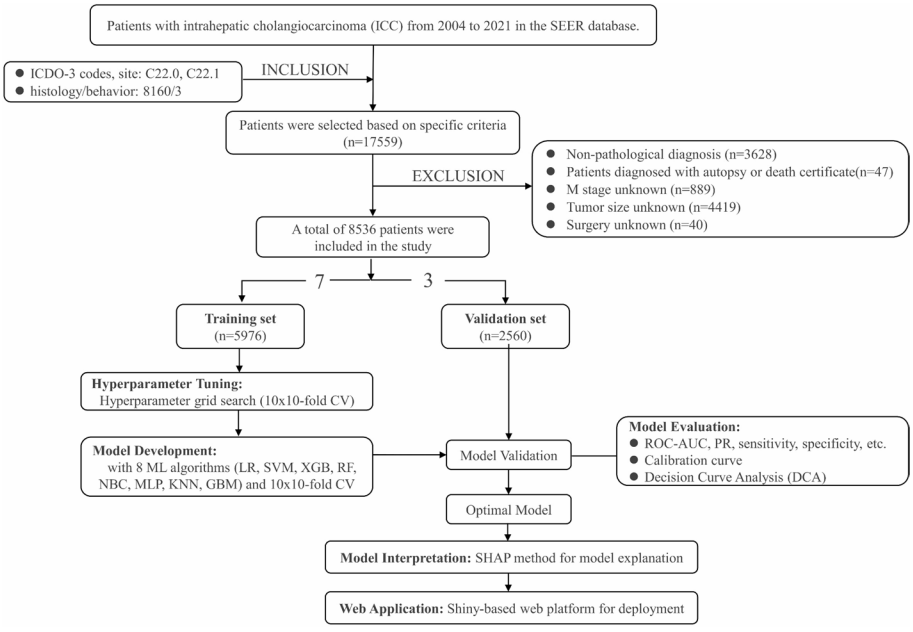

Fig. 1 Flowchart of study design and patient screening. SEER, Surveillance, Epidemiology, and End Results database; ICDO-3, International Classification of Diseases for Oncology, 3rd Edition. *CV*, cross-validation; *LR*, logistic regression; *SVM*, support vector machine; *XGB*, extreme gradient boosting; *RF*, random forest; *NBC*, naive bayes classifier; *MLP*, multi-layer perceptron; *KNN*, k-nearest neighbors; *GBM*, gradient boosting machine; *ROC*, receiver operating characteristic; *AUC*, area under the curve; *PR*, precision-recall; *SHAP*, SHapley Additive exPlanation

best threshold for the analysis [19]. Sensitivity analysis showed that when age and tumor size were converted into categorical variables, the eight ML models still exhibited robust performance (Supplementary Fig. 1). The continuous variable "number of resected LNs" was categorized into three groups: <6, ≥6, and unknown. This categorization adheres to the 8th edition AJCC staging system for ICC, which recommends the removal of at least 6 lymph nodes as the standard for adequate lymphadenectomy. Marital status was simplified into three groups: married, single, and other/unknown. The surgical treatment options were categorized into four groups: none, tumor ablation (surgery codes: 10–17), liver resection (surgery codes: 20–26, 30, 36–38, 50–52, 59, 90), and liver transplantation (surgery codes: 60–61, 75).

4 Selection of variables

In this study, we utilized three independent methods, including Least absolute shrinkage and selection operator (LASSO) regression, the Boruta algorithm and Recursive Feature Elimination (RFE), to screen baseline variables and identify potential predictors in the training set. LASSO is a linear regression method that uses L1 regularization to select key features, simplify the model, and maintain high predictive accuracy [20]. The Boruta algorithm, based on random forests, evaluates feature significance by comparing the importance of original features with that of randomly generated "shadow features" to ensure no potentially significant feature is overlooked [21]. Recursive Feature Elimination (RFE) is a mainstream feature selection method in ML that selects features by recursively training the model and removing the least important ones to achieve optimal performance [22]. RFE utilizes a random forests classifier combined with 10-fold cross-validation (CV), employing the area under the receiver operating characteristic curve (ROCAUC) as the evaluation metric. The overlapping variables identified through the intersection of LASSO regression, Boruta algorithm, and RFE were utilized to establish prediction models, ensuring the robustness of variable selection and improving the model's generalization capability [23–25].

5 Model development and validation

Based on the aforementioned feature selection methods, we developed eight different ML algorithms, including logistic regression (LR), support vector machine (SVM), extreme gradient boosting (XGB), random forest (RF), naive bayes classifier (NBC), multilayer perceptron (MLP), k-nearest neighbor (KNN), and gradient boosting machine (GBM), to predict the risk of DM in ICC patients. Considering the significant impact of class imbalance on model performance in binary classification, we applied the Synthetic Minority Over-sampling Technique (SMOTE) in the training set to address the data imbalance, mitigating overfitting and loss of critical information, and ultimately improving model performance [26]. To optimize the prediction models, we employed a method based on 10 rounds of 10-fold CV combined with the default hyperparameter grid search provided by the "caret" package to determine the final hyperparameters corresponding to the optimal feature subset for each model. Specifically, in each round, the dataset was randomly partitioned into 10 subsets, and a complete 10-fold CV process was executed. Finally, using the optimal feature subset and final hyperparameters obtained from the 10 rounds of 10-fold CV, the model was re-fitted on the training set. This approach effectively reduces the risk of overfitting while enhancing the robustness

of the performance evaluation and the model's generalization ability. We set random seed as "42" in our analysis.

We determined the best model by evaluating multiple performance metrics, including the area under the receiver operating characteristic (ROC) curve (AUC), area under the precision-recall (PR) curve (AUPRC), accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F1 score, and Brier score. For imbalanced datasets, AUPRC is a more reliable metric than AUC for evaluating model performance, so we generated the PR curve and calculated AUPRC as a supplementary metric [27]. The calibration curve was used to assess the alignment between predicted probabilities and actual outcomes. Additionally, the Brier score is a metric used to evaluate the performance of probabilistic prediction models by calculating the difference between estimated and observed risks, ranging from 0 to 1, where a score closer to 0 indicates better calibration [28]. Decision curve analysis (DCA) was employed to evaluate the net benefit of the model at various thresholds. Based on evaluation metrics, we picked the ML model that showed the best predictive results in both the training and validation sets.

6 Model explanation

Accurately interpreting ML prediction models can be challenging. SHAP, a cooperative game theory-based method, is widely employed to interpret black-box models in ML. By calculating each feature's contribution to the prediction, it provides both local and global explanations, thereby aiding in the understanding of the decision-making processes of complex models [29]. Higher SHAP values indicate a greater impact of a particular feature on the model's output. In addition, to facilitate the model's application in clinical settings, we developed an accessible web calculator.

6.1 Statistical analysis

In this study, all data were processed and analyzed using R (version 4.4.1, http://www.r-project.org) and Python (version 3.12, http://www.python.org). Continuous data f ollowing a normal distribution are analyzed using an independent t-test, reported as mean ± standard deviation (SD). Non-normally distributed data are analyzed with the Mann–Whitney *U* test, expressed as median and interquartile range (IQR). Categorical data are presented as frequencies (%) and analyzed using Fisher's exact test or Pearson's chi-square test. All patients were divided into a training set and a validation set in a 7:3 ratio. Feature selection in the training set was performed using a combination of LASSO regression, the Boruta algorithm and RFE methods. Spearman correlation tests and variance inflation factor (VIF) analyses were performed on the selected features to avoid significant covariance or multicollinearity among variables in the model. The SMOTE algorithm was applied to oversample the minority class, effectively resolving the class imbalance in the training set. The validation set remained unchanged to retain the authenticity of the original data distribution. ML models were trained using the R package "caret" (version 6.0.94), which provides a unified interface for various algorithms. Models were built with the train function and appropriate method parameters. The SHAP library in Python (version 0.46.0) was employed for interpretability analysis. A two-sided $P < 0.05$ was considered statistically significant.

7 Results

7.1 Baseline population characteristics

In this study, a total of 8536 ICC patients were included and divided into a training set (n = 5.76) and a validation set (n = 2.60) at a 7:3 ratio. Out of these patients, 2816 cases (33%) had DM, while 5,760 cases (67%) did not. Baseline data comparison between the training and validation sets revealed no significant differences (P > 0.05). Furthermore, significant differences were observed in multiple clinicopathological characteristics between patients with DM and those without DM. For example, patients with DM were more likely to be male, older, and of White ethnicity, with a higher incidence of delayed treatment (over 1 month), elevated tumor grades (II–IV), advanced T stages (II–IV) and N stages (N1), larger tumor diameters, a greater likelihood of having multiple tumors, a lower proportion of undergoing surgery, or more limited lymph node dissection (P < 0.001). Significant differences were also observed between patients with DM and those without DM in terms of receiving radiotherapy and chemotherapy (P < 0.001). The demographic and clinicopathological characteristics of all patients are shown in Table 1.

7.2 Feature selection and collinearity tests

Among all baseline characteristics, we utilized the Boruta algorithm to identify 12 key factors, including year of diagnosis, age at diagnosis, T stage, N stage, tumor size, tumor number, tumor grade, delayed treatment, surgery, the number of resected LNs, radiotherapy, and chemotherapy (Fig. 2A, B). RFE, employing RF as the estimator and performing 10-fold CV, narrowed the predictive factors down to 10 significant variables: age at diagnosis, T stage, N stage, tumor size, tumor number, tumor grade, surgery, number of resected LNs, radiotherapy, and chemotherapy (Fig. 2C). In this study, lambda.1se was determined to be the optimal value, and the 10 key variables selected through LASSO regression included: age at diagnosis, T stage, N stage, tumor size, tumor number, tumor grade, surgery, number of resected LNs, radiotherapy, and chemotherapy (Fig. 2D, E). The intersection of the results from three independent methods was determined to be the optimal feature set, ultimately including 10 clinical features as predictors for developing an ML-based predictive model: age at diagnosis, T stage, N stage, tumor size, tumor number, tumor grade, surgery, number of resected LNs, radiotherapy, and chemotherapy (Fig. 2F). Spearman correlation and VIF tests indicated no strong correlation or multicollinearity between them (Fig. 3A, B).

7.3 Model performance

We conducted 10 rounds of 10-fold CV to construct eight ML models. To determine the optimal predictive model, we evaluated their performance and validated the results on the validation set. As shown in Fig. 4A, B, the GBM model achieved the highest AUC compared to other models, with an AUC of 0.905 (95% CI 0.900–0.910) on the training set and 0.802 (95% CI 0.785–0.819) on the validation set. The forest plot comparing the AUC scores of eight ML models is shown in Fig. 4C. To visually present the model's performance, we generated a plot comparing each participant's predicted values with their actual disease status (Without DM/With DM). Employing a cutoff value of 0.5 for classification, the GBM model accurately identified 81.5% of DM patients in the validation set (Fig. 4D). The PR curve results indicate that the GBM model achieved the highest AUPRC values in both the training set (0.892) and the validation set (0.571) (Fig. 5A, D).

Table 1 Comparison of demographic and clinical features between patients with distant metastasis (DM) and those without DM, and between the training and validation sets

Characteristic	Overall	Without DM	With DM	P-value	Training set	Validation set	P-value
	(N = 8536), n (%)	(N = 5720), n (%)	(N = 2816), n (%)		(N = 5976), n (%)	(N = 2560), n (%)	
Years				<0.001			0.566
2004-2010	1620 (19.0)	1178 (20.6)	442 (15.7)		1160 (19.4)	460 (18.0)	
2011-2015	2397 (28.1)	1593 (27.8)	804 (28.6)		1656 (27.7)	741 (28.9)	
2016-2021	4519 (52.9)	2949 (51.6)	1570 (55.8)		3160 (52.9)	1359 (53.1)	
Age				<0.001			0.814
<60	2360 (27.6)	1457 (25.5)	903 (32.1)		1675 (28.0)	685 (26.8)	
60-79	4987 (58.4)	3411 (59.6)	1576 (56.0)		3468 (58.0)	1519 (59.3)	
≥80	1189 (13.9)	852 (14.9)	337 (12.0)		833 (13.9)	356 (13.9)	
Sex				0.03			0.116
Female	4245 (49.7)	2902 (50.7)	1343 (47.7)		2928 (49.0)	1317 (51.4)	
Male	4291 (50.3)	2818 (49.3)	1473 (52.3)		3048 (51.0)	1243 (48.6)	
Race				0.031			0.408
White	6714 (78.7)	4560 (79.7)	2154 (76.5)		4693 (78.5)	2021 (78.9)	
Black	664 (7.8)	428 (7.5)	236 (8.4)		489 (8.2)	175 (6.8)	
Asia-Pacific	1060 (12.4)	664 (11.6)	396 (14.1)		723 (12.1)	337 (13.2)	
Other/Unknown	98 (1.1)	68 (1.2)	30 (1.1)		71 (1.2)	27 (1.1)	
Marital status				0.451			0.999
Single	1217 (14.3)	803 (14.0)	414 (14.7)		851 (14.2)	366 (14.3)	
Married	5113 (59.9)	3403 (59.5)	1710 (60.7)		3575 (59.8)	1538 (60.1)	
Other/Unknown	2206 (25.8)	1514 (26.5)	692 (24.6)		1550 (25.9)	656 (25.6)	
Delayed treatment				<0.001			0.918
<1 month	2180 (25.5)	1453 (25.4)	727 (25.8)		1537 (25.7)	643 (25.1)	
≥1 month	5843 (68.5)	3806 (66.5)	2037 (72.3)		4073 (68.2)	1770 (69.1)	
None	513 (6.0)	461 (8.1)	52 (1.8)		366 (6.1)	147 (5.7)	
Tumor grade				<0.001			0.998
Grade I	479 (5.6)	405 (7.1)	74 (2.6)		340 (5.7)	139 (5.4)	
Grade II	2444 (28.6)	1898 (33.2)	546 (19.4)		1719 (28.8)	725 (28.3)	
Grade III & IV	2106 (24.7)	1474 (25.8)	632 (22.4)		1470 (24.6)	636 (24.8)	
Unknown	3507 (41.1)	1943 (34.0)	1564 (55.5)		2447 (40.9)	1060 (41.4)	
Surgery				<0.001			0.882
None	5634 (66.0)	3021 (52.8)	2613 (92.8)		3950 (66.1)	1684 (65.8)	
Tumor ablation	222 (2.6)	199 (3.5)	23 (0.8)		159 (2.7)	63 (2.5)	
Liver resection	2410 (28.2)	2248 (39.3)	162 (5.8)		1669 (27.9)	741 (28.9)	
Liver transplantation	270 (3.2)	252 (4.4)	18 (0.6)		198 (3.3)	72 (2.8)	
Number of resected LNs				<0.001			0.546
<6	7858 (92.1)	5175 (90.5)	2683 (95.3)		5514 (92.3)	2344 (91.6)	
≥6	592 (6.9)	489 (8.5)	103 (3.7)		398 (6.7)	194 (7.6)	
Unknown	86 (1.0)	56 (1.0)	30 (1.1)		64 (1.1)	22 (0.9)	
Radiotherapy				<0.001			0.982
No	7099 (83.2)	4692 (82.0)	2407 (85.5)		4973 (83.2)	2126 (83.0)	
Yes	1437 (16.8)	1028 (18.0)	409 (14.5)		1003 (16.8)	434 (17.0)	
Chemotherapy				<0.001			0.944
No	3765 (44.1)	2759 (48.2)	1006 (35.7)		2643 (44.2)	1122 (43.8)	
Yes	4771 (55.9)	2961 (51.8)	1810 (64.3)		3333 (55.8)	1438 (56.2)	
Tumor numbers				<0.001			0.797
Single	6468 (75.8)	4144 (72.4)	2324 (82.5)		4516 (75.6)	1952 (76.3)	
Multiple	2068 (24.2)	1576 (27.6)	492 (17.5)		1460 (24.4)	608 (23.8)	
T stage				<0.001			0.759
T1	3223 (37.8)	2588 (45.2)	635 (22.5)		2232 (37.3)	991 (38.7)	

Table 1 (continued)

Characteristic	Overall	Without DM	With DM	P-value	Training set	Validation set	P-value
	(N = 8536), n (%)	(N = 5720), n (%)	(N = 2816), n (%)		(N = 5976), n (%)	(N = 2560), n (%)	
T2	2194 (25.7)	1438 (25.1)	756 (26.8)		1517 (25.4)	677 (26.4)	
T3	1535 (18.0)	943 (16.5)	592 (21.0)		1095 (18.3)	440 (17.2)	
T4	1092 (12.8)	585 (10.2)	507 (18.0)		786 (13.2)	306 (12.0)	
TX	492 (5.8)	166 (2.9)	326 (11.6)		346 (5.8)	146 (5.7)	
N stage				<0.001			0.994
N0	5649 (66.2)	4352 (76.1)	1297 (46.1)		3954 (66.2)	1695 (66.2)	
N1	2437 (28.5)	1253 (21.9)	1184 (42.0)		1711 (28.6)	726 (28.4)	
NX	450 (5.3)	115 (2.0)	335 (11.9)		311 (5.2)	139 (5.4)	
Tumor size				<0.001			0.793
<4.5	2529 (29.6)	1993 (34.8)	536 (19.0)		1778 (29.8)	751 (29.3)	
4.5-7.9	2869 (33.6)	1972 (34.5)	897 (31.9)		1983 (33.2)	886 (34.6)	
≥8.0	3138 (36.8)	1755 (30.7)	1383 (49.1)		2215 (37.1)	923 (36.1)	
Household income				0.36			0.998
<$45,000	135 (1.6)	100 (1.7)	35 (1.2)		93 (1.6)	42 (1.6)	
$45,000-$60,000	897 (10.5)	615 (10.8)	282 (10.0)		625 (10.5)	272 (10.6)	
>$60,000	7504 (87.9)	5005 (87.5)	2499 (88.7)		5258 (88.0)	2246 (87.7)	
Rural-urban counties				0.825			0.988
Metro < 250,000	618 (7.2)	426 (7.4)	192 (6.8)		428 (7.2)	190 (7.4)	
Metro 250,000-1 million	1930 (22.6)	1284 (22.4)	646 (22.9)		1367 (22.9)	563 (22.0)	
Metro > 1 million	5146 (60.3)	3430 (60.0)	1716 (60.9)		3590 (60.1)	1556 (60.8)	
Non-Metro/Unknown	842 (9.9)	580 (10.1)	262 (9.3)		591 (9.9)	251 (9.8)	

LNs, lymph nodes; *Metro*, metropolitan

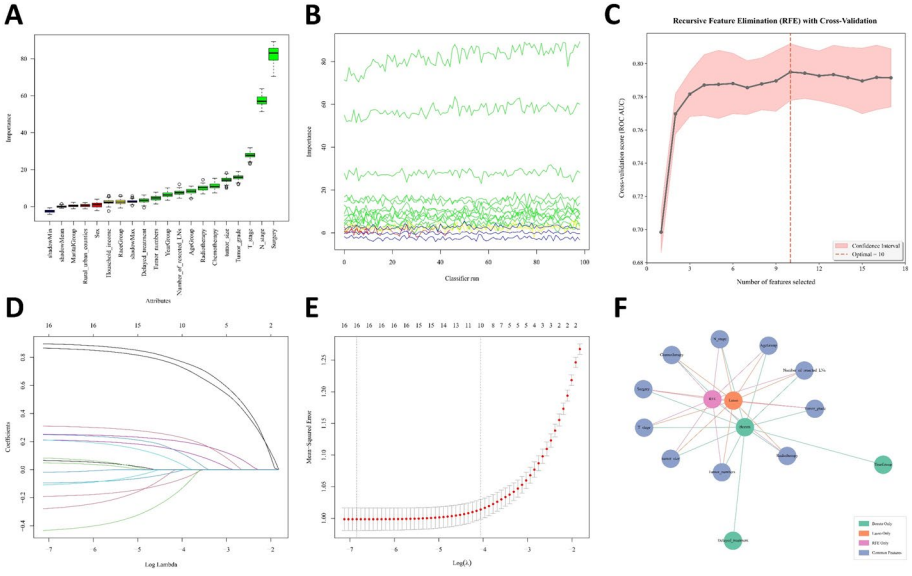

Fig. 2 Feature selection results. **A** Box plots of attributes based on importance values in the Boruta algorithm. **B** Importance values of attributes in each classifier run of the Boruta algorithm. **C** Recursive feature elimination (RFE) based on random forest with 10-fold cross-validation. **D** Least absolute shrinkage and selection operator (LASSO) regression coefficients paths. **E** LASSO regression cross-validation curve. **F** Overlapping features selected by Boruta algorithm, LASSO regression and RFE

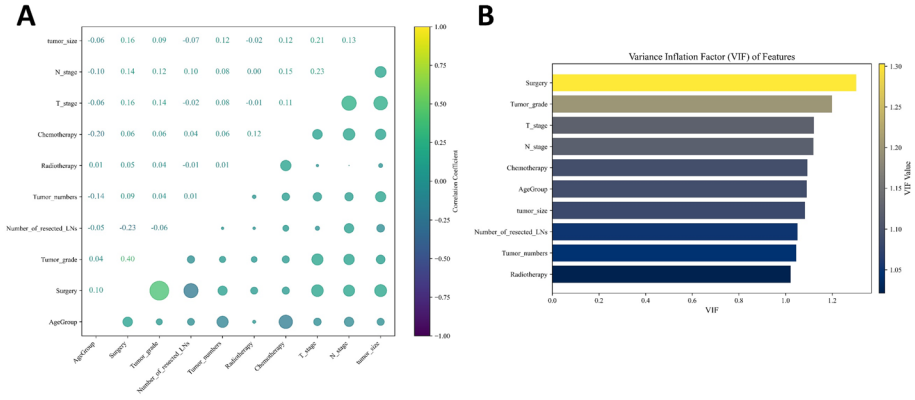

Fig. 3 Spearman's correlation test (**A**) and variance inflation factor test (**B**) among selected features

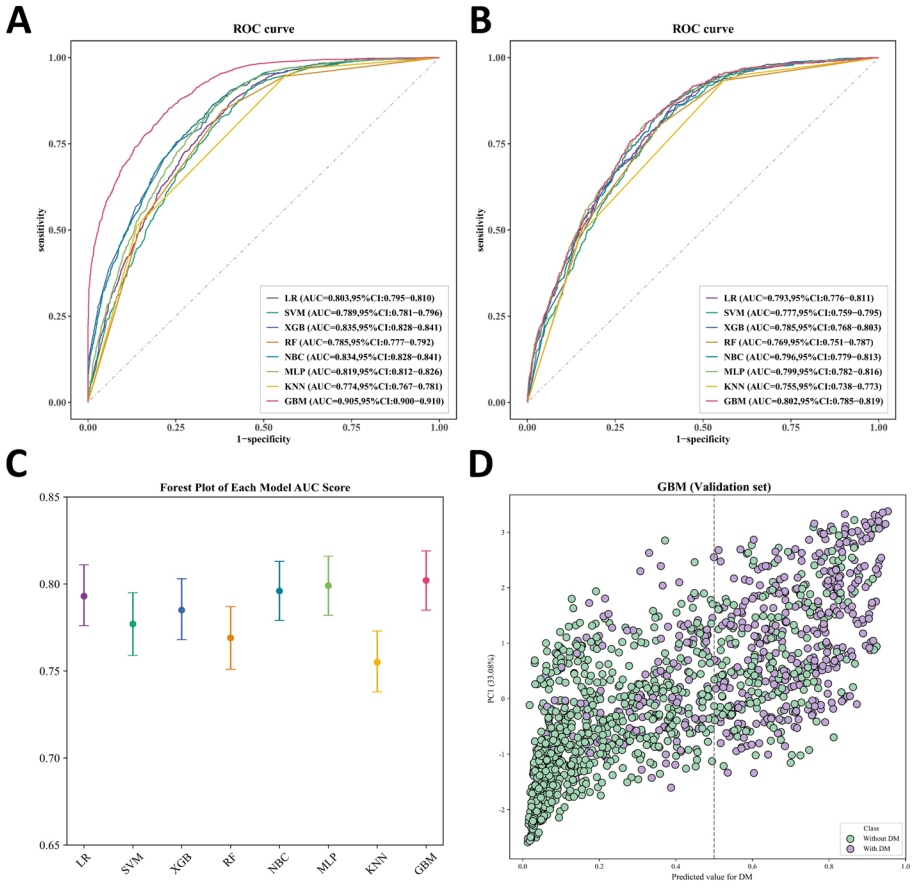

Fig. 4 **A** The training set ROC curve. **B** The validation set ROC curve. **C** Forest plot of AUC values. **D** The predictive performance of the GBM model distinguishes between with distant metastases (DM) (colored in purple) and without DM (colored in green) in the validation set. *LR*, logistic regression; *SVM,* support vector machine; *XGB*, extreme gradient boosting; *RF*, random forest; *NBC*, naive bayes classifier; *MLP*, multi-layer perceptron; *KNN*, k-nearest neighbors; *GBM*, gradient boosting machine

The calibration curves of different ML algorithms indicate that the GBM model exhibits the highest consistency with the ideal prediction curve in both the training and validation sets (Fig. 5B, E). The DCA curves show that, compared to other models, the GBM model achieves higher net benefit across most thresholds in both the training and validation sets (Fig. 5C, F). By comparing eight different ML methods, we generated a table

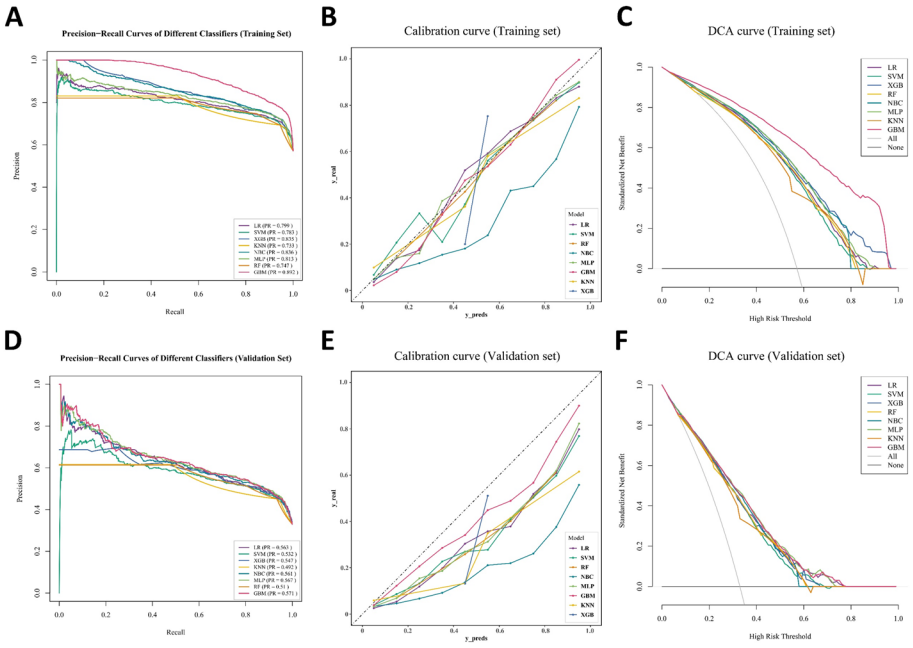

Fig. 5 Performance comparison of different models. PR curve analysis (**A**, **D**), calibration curve analysis (**B**, **E**), and DCA curves for each model (**C**, **F**) assess the performance of eight machine learning (ML) algorithms in predicting distant metastases (DM) in intrahepatic cholangiocarcinoma (ICC) patients across the training and validation sets. *LR*, logistic regression; *SVM*, support vector machine; *XGB*, extreme gradient boosting; *RF*, random forest; *NBC*, naive bayes classifier; *MLP*, multi-layer perceptron; *KNN*, k-nearest neighbors; *GBM*, gradient boosting machine

Table 2 Performance of eight machine learning prediction models in the validation set

Model	Accuracy	Sensitivity	Specificity	PPV	NPV	F1-score	Brier score
LR	0.671	0.87	0.573	0.87	0.573	0.635	0.202
SVM	0.686	0.808	0.625	0.808	0.625	0.629	0.216
XGB	0.678	0.859	0.589	0.859	0.589	0.638	0.249
RF	0.684	0.802	0.626	0.802	0.626	0.626	0.209
NBC	0.683	0.873	0.589	0.873	0.589	0.645	0.301
MLP	0.708	0.815	0.655	0.815	0.655	0.648	0.2
KNN	0.604	0.934	0.441	0.934	0.441	0.608	0.243
GBM	0.713	0.809	0.665	0.809	0.665	0.65	0.177

LR, logistic regression; *SVM*, support vector machine; *XGB*, extreme gradient boosting; *RF*, random forest; *NBC*, naive bayes classifier; *MLP*, multi-layer perceptron; *KNN*, k-nearest neighbors; *GBM*, gradient boosting machine.

displaying evaluation metrics, including accuracy, sensitivity, specificity, PPV, NPV, F1 score, and Brier score. Compared to the other models, the GBM model demonstrated higher accuracy (0.713), specificity (0.665), NPV (0.665), and F1 score (0.65), along with a lower Brier score (0.177). Specific values are provided in Table 2, with the corresponding visualizations shown in Fig. 6A, B. In summary, the GBM model performed the best in both the training and validation sets, and is therefore recommended as the preferred model for predicting DM risk in ICC patients.

7.4 Model explanation

SHAP values provide deeper insights into how the GBM model makes predictions by explaining the contribution of each feature to the model's output. Figure 7A illustrates the direction and strength of each feature's impact on the model's predictions through the SHAP summary plot and presents a bar chart with average SHAP values, showing

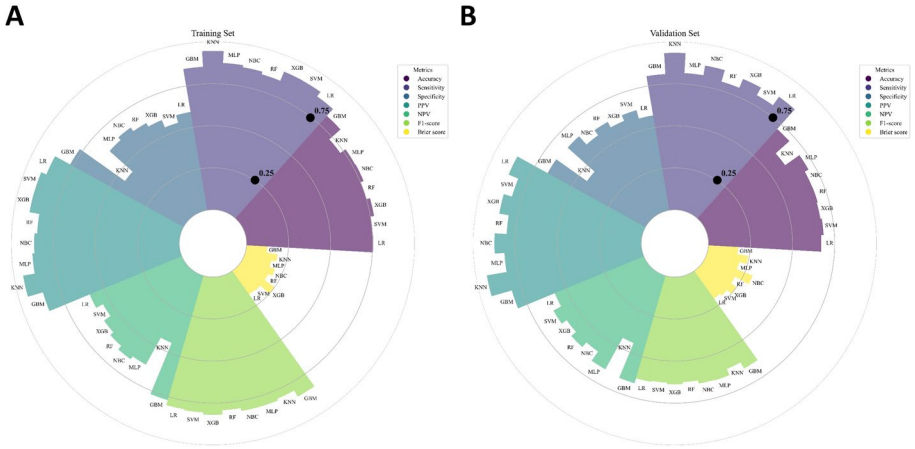

Fig. 6 Evaluate the performance metrics of eight different machine learning methods in the training set (**A**) and validation set (**B**). *PPV*, Positive predictive value; *NPV*, Negative predictive value

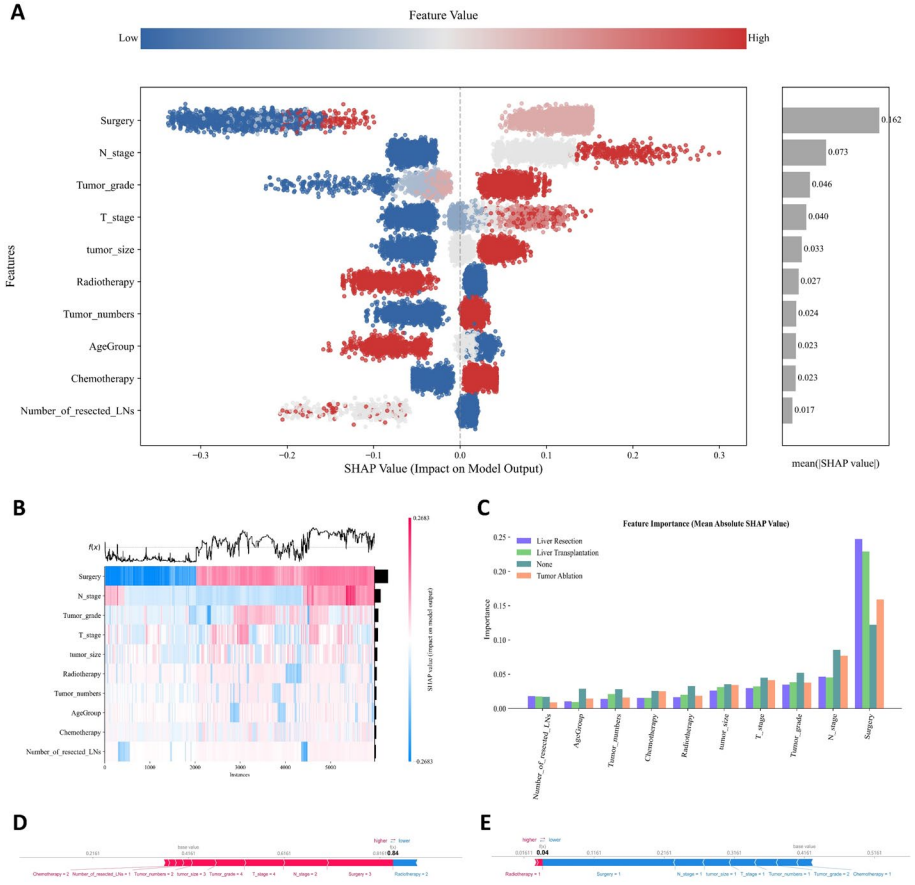

Fig. 7 SHAP model interpretation of feature variables based on the GBM model. **A** The SHAP summary plots displaying the importance ranking of features; **B** The SHAP heatmap clusters hierarchically based on SHAP values. **C** Surgical subgroup-stratified SHAP contribution bar plot. **D** The SHAP force plot for intrahepatic cholangiocarcinoma (ICC) patients with distant metastases (DM). **E** The SHAP force plot for ICC patients without DM. In these plots, red represents variables acting as risk factors for DM, while blue represents variables acting as protective factors

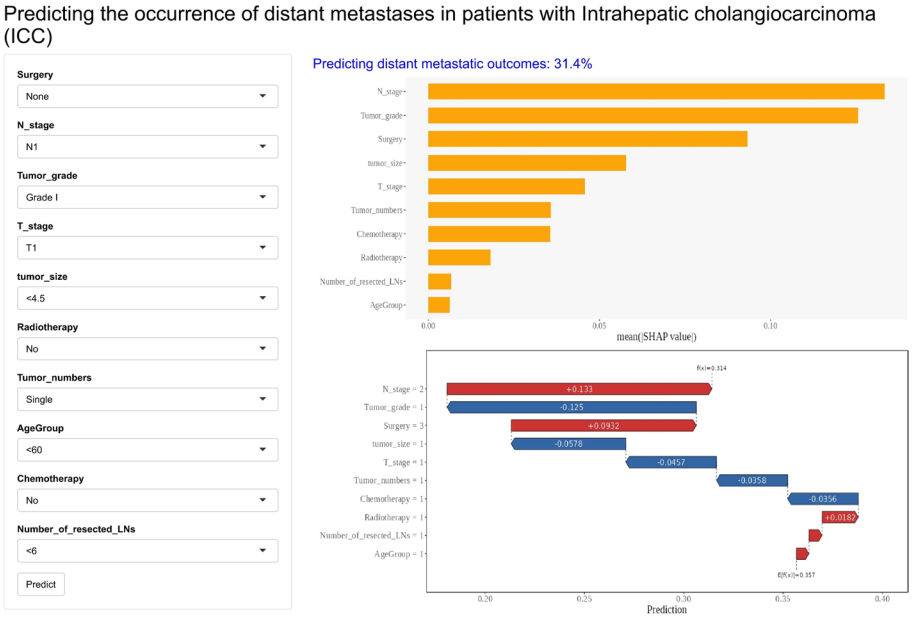

Fig. 8 The web-based calculator for predicting distant metastasis (DM) risk probability in intrahepatic cholangiocarcinoma (ICC)

the contribution of features in descending order: surgery, N stage, tumor grade, T stage, tumor size, radiotherapy, tumor number, age at diagnosis, chemotherapy and number of resected LNs. The SHAP heatmap performs hierarchical clustering of patients based on SHAP values, visually highlighting the distribution of ICC patients with and without DM. Red indicates cases with a high probability of DM, while colorless or blue represents cases with low DM probability or without DM (Fig. 7B). Figure 7C illustrates the SHAP contribution patterns of key features across surgical subgroups. The liver resection and liver transplantation groups showed higher mean SHAP contributions than the tumor ablation and non-surgical groups. In contrast, among non-surgical patients, N stage, tumor grade, T stage, age, and local treatment-related features contributed more to the model predictions. The force plot offers personalized feature attribution through two representative examples, illustrating how SHAP can explain individual model predictions. Figure 7D shows an example of a positive prediction (the predicted probability of DM is 84%), with the absence of surgical treatment, N1 stage, and T4 stage identified as the primary risk factors. Conversely, Fig. 7E shows an example of a negative prediction (the predicted probability of DM is 4%), where liver resection, N0 stage, small tumor size, and T1 stage are key factors associated with a reduced risk of DM. It starts from the base value, the average of all predictions, and each input feature at different levels can increase or decrease the predicted probability. The arrow length reflects the SHAP value of these features. By visualizing the SHAP values, we can assess how each feature impacts the model's predictions for these cases.

7.5 Web calculator

As shown in Fig. 8, we developed a web calculator based on the GBM model to provide real-time, personalized risk estimation of DM in ICC patients for clinical use and dissemination. By entering the actual values of the key features required by the model, the tool automatically predicts the risk of DM for individual ICC patients. The web

calculator is accessible online via the following link: https://bijinzhe.shinyapps.io/icc_dm_shiny/.

8 Discussion

The incidence of ICC is increasing globally. ICC is a highly fatal disease, with 40% of cases resulting in death due to DM [30, 31]. Therefore, timely and accurate identification and prediction of the risk of DM in ICC patients are crucial. Currently, research on predictive models for distant metastases in ICC patients remains notably limited. Moreover, most of these studies rely on a single method (traditional logistic regression) for model construction, which may be insufficient to capture complex relationships, potentially limiting predictive performance [32–34]. With advancements in artificial intelligence, ML offers significant advantages over traditional models in handling multivariable, nonlinear relationships and complex interactions, thereby facilitating the development of clinical predictive models [35, 36]. To the best of our knowledge, this is the first study to utilize multiple ML algorithms for predicting DM in ICC patients, with the goal of enhancing patient outcomes through early intervention.

The lack of standardized guidelines for feature selection leaves the optimal number of features in predictive models uncertain. While more features may enhance information richness, excessive or non-causal features can reduce accuracy and limit clinical use [37]. To assist with feature selection, we adopted a combined strategy of LASSO regression, the Boruta algorithm, and RFE. Based on this, we carefully selected and finalized 10 feature variables, developing and validating 8 ML models. Among these models, the GBM model consistently demonstrated the highest AUC and AUPRC in both the training and validation sets, as well as the lowest Brier score, along with excellent accuracy, calibration, and net benefit. It should also be noted that in this study, the GBM model achieved an AUC of 0.905 on the training set, suggesting potential overfitting by capturing noise in the training data. In contrast, the 10 rounds of 10-fold CV yielded an AUC of 0.802, which, although slightly lower, better reflects the model's generalizability and still demonstrates excellent predictive performance.

The SHAP value analysis revealed that each of the 10 analyzed variables contributed uniquely and meaningfully to the model's performance, including surgery, N stage, tumor grade, T stage, tumor size, radiotherapy, tumor number, age at diagnosis, chemotherapy and number of resected LNs (Fig. 7A). In this study, surgery was recognized as the key variable for predicting DM in ICC patients. Surgical resection remains a vital curative approach for ICC. With advancements in medical technology, an increasing number of patients with primary ICC are undergoing surgical treatment, including those who were previously considered ineligible for surgery [38, 39]. Recent studies have extended the application of surgical resection, including radiofrequency ablation, to advanced intrahepatic cholangiocarcinoma, even in cases of metastatic or multifocal disease [40, 41]. Thus, surgery plays an essential role in the treatment of ICC patients, offering a significant reduction in the risk of DM and contributing to improved patient prognosis. In our study, higher T stage, larger tumor diameter and lymph node metastasis were identified as significant risk factors for DM. The larger the tumor diameter, the higher the T stage, which is associated with increased invasiveness toward surrounding organs and blood vessels, a greater likelihood of positive residual resection margins, and more scattered microscopic lesions, thereby raising the risk of DM. Additionally, ICC can infiltrate

the bloodstream via lymphatic vessels within lymph nodes, leading to the formation of metastatic lesions in distant organs, and this mechanism is well-supported by evidence from studies on other tumor types [42, 43]. Consistent with previous research findings, poorly differentiated tumors are typically associated with higher aggressiveness and an increased risk of metastasis, significantly impacting patient prognosis [44]. Our study revealed that radiotherapy acts as a protective factor against DM in ICC patients. Effective radiotherapy can suppress tumor growth, disrupt angiogenesis, promote tumor cell apoptosis, and modify the immune microenvironment, ultimately reducing the risk of distant metastases [45, 46]. George and Keklikoglou et al. have reported that chemotherapy might increase the probability of distant metastases in malignant tumors, potentially by promoting the expression of metastasis-related genes and enhancing exosome secretion [47, 48]. These studies suggest that while chemotherapy has a positive effect in reducing tumor size, it may also increase the risk of cancer cell metastasis. Our study indicates that tumor multifocality is a significant risk factor for DM. Patients with multifocal tumors typically present with larger primary lesions, higher rates of lymph node involvement, and increased vascular invasion, which significantly increase the risk of DM [49]. Therefore, accurately and effectively assessing the impact of these factors on DM occurrence in ICC patients, along with timely warnings and appropriate interventions, is crucial for improving their long-term survival and prognosis.

ML techniques are often referred to as "black boxes," making their predictive processes difficult to interpret [50]. The lack of transparency may deter clinicians from using these technologies, as they are reluctant to make medical decisions based on unclear or non-interpretable information. To address this issue, a key strength of our study lies in the use of the SHAP method to reasonably elucidate the "black box" nature of ML models. The SHAP method provides both global and local explanations by quantitatively evaluating the specific contribution of each feature variable to the risk of DM in ICC patients, thereby enhancing clinical practitioners' trust in the application of ML models. Specifically, the interpretability provided by our model allows clinicians to intuitively understand how different clinical parameters influence the predicted probability of DM, thereby facilitating personalized and precise risk stratification. Based on the model, clinicians can identify patients with a higher predicted risk, arrange more frequent imaging examinations, or initiate earlier therapeutic interventions to detect and control DM at an early stage. For patients with a lower predicted risk, standard monitoring protocols can be followed, reducing unnecessary medical expenses and alleviating patient anxiety. Additionally, leveraging the convenient tools provided by the Shinyapps platform, we developed an online calculator that allows users to input key parameters and obtain the probability of DM in ICC patients, facilitating practical applications in clinical settings. Another advantage of this study is that all predictive factors incorporated into the model are routinely collected and easily obtainable during hospitalization, which strengthens the feasibility of implementing and promoting the model in clinical practice.

However, certain limitations of this study must be acknowledged. First, the data were derived from a retrospective analysis of the SEER database, which may introduce biases such as selection bias, information bias, and other data-related biases [51]. Second, the limitations of the SEER database made it challenging to obtain additional relevant information, with some critical variables, such as blood biochemical markers and nerve or vascular invasion, being unavailable for timely inclusion, thus constraining further

model optimization. Third, most ICC patients with DM do not undergo surgical intervention due to tumor aggressiveness, which may lead to an overestimation of the predictive significance of the surgical variable in our model. Future studies should incorporate additional objective diagnostic criteria, such as radiomic signatures, biomarkers, or molecular profiles, to further enhance the model's robustness. Furthermore, this study lacks external validation data. In future studies, we aim to collect more comprehensive datasets and conduct in-depth supplementary research, including external validation using independent cohorts from different hospitals to further assess the model's robustness and generalizability.

9 Conclusion

In conclusion, we successfully developed an interpretable ML model to predict the risk of DM in ICC patients. Among the eight algorithms, the GBM model demonstrated the most reliable predictive performance. SHAP analysis provided valuable insights into the model's decision-making process. Additionally, the web-based calculator we developed can assist clinicians in formulating and promptly adjusting personalized clinical decisions.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s12672-025-02952-y.

Supplementary Material 1

Author contributions

Y.Y. oversaw conceptualization and methodology. J.B. wrote the main manuscript text, curated the data and created the visualizations. All authors read and approved the manuscript.

Funding

This study was supported by the Guilin Science Research and Technology Development Project (No. 20230135-1-3) and the Guangxi Medical and Health Key Discipline Construction Project.

Data availability

All data are publicly available, for more information please visit the official website for the Surveillance, Epidemiology, and End Results program (https://seer.cancer.gov).

Declarations

Conflict of interest

All authors declare no conflict of interest.

Ethical approval

Approval was waived by the local ethics committee, as SEER data are publicly available and deidentified.

Received: 26 February 2025 / Accepted: 10 June 2025

Published online: 18 June 2025

References

1. Banales JM, Cardinale V, Carpino G, Marzioni M, Andersen JB, Invernizzi P, Lind GE, Folseraas T, Forbes SJ, Fouassier L, Geier A, Calvisi DF, Mertens JC, Trauner M, Benedetti A, Maroni L, Vaquero J, Macias RI, Raggi C, Perugorria MJ, Gaudio E, Boberg KM, Marin JJ, Alvaro D. Expert consensus document: cholangiocarcinoma: current knowledge and future perspectives consensus statement from the European network for the study of cholangiocarcinoma (ENS-CCA). Nat Rev Gastroenterol Hepatol. 2016;13(5):261–80. https://doi.org/10.1038/nrgastro.2016.51.
2. Banales JM, Marin JJG, Lamarca A, Rodrigues PM, Khan SA, Roberts LR, Cardinale V, Carpino G, Andersen JB, Braconi C, Calvisi DF, Perugorria MJ, Fabris L, Boulter L, Macias RIR, Gaudio E, Alvaro D, Gradilone SA, Strazzabosco M, Marzioni M, Coulouarn C, Fouassier L, Raggi C, Invernizzi P, Mertens JC, Moncsek A, Ilyas SI, Heimbach J, Koerkamp BG, Bruix J, Forner A, Bridgewater J, Valle JW, Gores GJ. Cholangiocarcinoma 2020: the next horizon in mechanisms and management. Nat Rev Gastroenterol Hepatol. 2020;17(9):557–88. https://doi.org/10.1038/s41575-020-0310-z.
3. Sirica AE, Gores GJ, Groopman JD, Selaru FM, Strazzabosco M, Wei Wang X, Zhu AX. Intrahepatic cholangiocarcinoma: continuing challenges and translational advances. Hepatology. 2019;69(4):1803–15. https://doi.org/10.1002/hep.30289.

4. Clements O, Eliahoo J, Kim JU, Taylor-Robinson SD, Khan SA. Risk factors for intrahepatic and extrahepatic cholangiocarcinoma: A systematic review and meta-analysis. J Hepatol. 2020;72(1):95–103. https://doi.org/10.1016/j.jhep.2019.09.007.
5. Esnaola NF, Meyer JE, Karachristos A, Maranki JL, Camp ER, Denlinger CS. Evaluation and management of intrahepatic and extrahepatic cholangiocarcinoma. Cancer. 2016;122(9):1349–69. https://doi.org/10.1002/cncr.29692.
6. Forner A, Vidili G, Rengo M, Bujanda L, Ponz-Sarvisé M, Lamarca A. Clinical presentation, diagnosis and staging of cholangiocarcinoma. Liver Int. 2019;39(Suppl 1):98–107. https://doi.org/10.1111/liv.14086.
7. Mavros MN, Economopoulos KP, Alexiou VG, Pawlik TM. Treatment and prognosis for patients with intrahepatic cholangiocarcinoma: systematic review and Meta-analysis. JAMA Surg. 2014;149(6):565–74. https://doi.org/10.1001/jamasurg.2013.5137.
8. 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(2):198–222. https://doi.org/10.3322/caac.21759.
9. Zhang XF, Beal EW, Bagante F, Chakedis J, Weiss M, Popescu I, Marques HP, Aldrighetti L, Maithel SK, Pulitano C, Bauer TW, Shen F, Poultsides GA, Soubrane O, Martel G, Koerkamp BG, Itaru E, Pawlik TM. Early versus late recurrence of intrahepatic cholangiocarcinoma after resection with curative intent. Br J Surg. 2018;105(7):848–56. https://doi.org/10.1002/bjs.10676.
10. Wiggers JK, Groot Koerkamp B, van Klaveren D, Coelen RJ, Nio CY, Allen PJ, Besselink MG, Busch OR, D'Angelica MI, DeMatteo RP, Kingham TP, van Gulik TM, Jarnagin WR. Preoperative risk score to predict occult metastatic or locally advanced disease in patients with resectable Perihilar cholangiocarcinoma on imaging. J Am Coll Surg. 2018;227(2):238–e2462. https://doi.org/10.1016/j.jamcollsurg.2018.03.041.
11. Yan X, Wang P, Zhu Z, Ning Z, Xu L, Zhuang L, Sheng J, Meng Z. Site-specific metastases of intrahepatic cholangiocarcinoma and its impact on survival: a population-based study. Fut Oncol. 2019;15(18):2125–37. https://doi.org/10.2217/fon-2018-0846.
12. Hewitt DB, Brown ZJ, Pawlik TM. Surgical management of intrahepatic cholangiocarcinoma. Expert Rev Anticancer Ther. 2022;22(1):27–38. https://doi.org/10.1080/14737140.2022.1999809.
13. Yuan C, Zou S, Wang K, Hu Z. Establishment and external validation of prognosis prediction nomogram for patients with distant metastatic intrahepatic cholangiocarcinoma: based on a large population. BMC Cancer. 2024;24(1):227. https://doi.org/10.1186/s12885-024-11976-6.
14. Goecks J, Jalili V, Heiser LM, Gray JW. How machine learning will transform biomedicine. Cell. 2020;181(1):92–101. https://doi.org/10.1016/j.cell.2020.03.022.
15. Zhang X, Ono JP, Song H, Gou L, Ma KL, Ren L. SliceTeller: A data Slice-Driven approach for machine learning model validation. IEEE Trans Vis Comput Graph. 2023;29(1):842–52. https://doi.org/10.1109/TVCG.2022.3209465.
16. Petch J, Di S, Nelson W. Opening the black box: the promise and limitations of explainable machine learning in cardiology. Can J Cardiol. 2022;38(2):204–13. https://doi.org/10.1016/j.cjca.2021.09.004.
17. Lisboa PJG. Open your black box classifier. Healthc Technol Lett. 2023;11(4):210–2. https://doi.org/10.1049/htl2.12050.
18. Cronin KA, Ries LA, Edwards BK. The surveillance, epidemiology, and end results (SEER) program of the National Cancer Institute. Cancer. 2014;120(Suppl 23):3755–7. https://doi.org/10.1002/cncr.29049.
19. Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. Clin Cancer Res. 2004;10(21):7252–9. https://doi.org/10.1158/1078-0432.
20. Frost HR, Amos CI. Gene set selection via LASSO penalized regression (SLPR). Nucl Acids Res. 2017;45(12):e114. https://doi.org/10.1093/nar/gkx291.
21. Wang X, Ren J, Ren H, Song W, Qiao Y, Zhao Y, Linghu L, Cui Y, Zhao Z, Chen L, Qiu L. Diabetes mellitus early warning and factor analysis using ensemble bayesian networks with SMOTE-ENN and Boruta. Sci Rep. 2023;13(1):12718. https://doi.org/10.1038/s41598-023-40036-5.
22. Escanilla NS, Hellerstein L, Kleiman R, Kuang Z, Shull JD, Page D. Recursive Feature Elimination by Sensitivity Testing. Proc Int Conf Mach Learn Appl. 2018; 2018:40–47. https://doi.org/10.1109/ICMLA.2018.00014
23. Swanson K, Wu E, Zhang A, Alizadeh AA, Zou J. From patterns to patients: advances in clinical machine learning for cancer diagnosis, prognosis, and treatment. Cell. 2023;186(8):1772–91. https://doi.org/10.1016/j.cell.2023.01.035.
24. Kaur I, Doja MN, Ahmad T. Data mining and machine learning in cancer survival research: an overview and future recommendations. J Biomed Inf. 2022;128:104026. https://doi.org/10.1016/j.jbi.2022.104026.
25. Elemento O, Leslie C, Lundin J, Tourassi G. Artificial intelligence in cancer research, diagnosis and therapy. Nat Rev Cancer. 2021;21(12):747–52. https://doi.org/10.1038/s41568-021-00399-1.
26. Koivu A, Sairanen M, Airola A, Pahikkala T. Synthetic minority oversampling of vital statistics data with generative adversarial networks. J Am Med Inf Assoc. 2020;27(11):1667–74. https://doi.org/10.1093/jamia/ocaa127.
27. Fu GH, Yi LZ, Pan J. Tuning model parameters in class-imbalanced learning with precision-recall curve. Biom J. 2019;61(3):652–64. https://doi.org/10.1002/bimj.201800148.
28. Yang W, Jiang J, Schnellinger EM, Kimmel SE, Guo W. Modified Brier score for evaluating prediction accuracy for binary outcomes. Stat Methods Med Res. 2022;31(12):2287–96. Epub 2022 Aug 29.
29. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;30:4765–74. https://doi.org/10.48550/arXiv.1705.07874.
30. Abdel-Rahman O. Role of liver-directed local tumor therapy in the management of hepatocellular carcinoma with extrahepatic metastases: a SEER database analysis. Expert Rev Gastroenterol Hepatol. 2017;11(2):183–9. https://doi.org/10.1080/17474124.2017.1259563.
31. Zhang H, Yang T, Wu M, Shen F. Intrahepatic cholangiocarcinoma: epidemiology, risk factors, diagnosis and surgical management. Cancer Lett. 2016;379(2):198–205. https://doi.org/10.1016/j.canlet.2015.09.008.
32. Liu Z, Yi J, Yang J, Zhang X, Wang L, Liu S. Diagnostic and prognostic nomograms for newly diagnosed intrahepatic cholangiocarcinoma with brain metastasis: a population-based analysis. Exp Biol Med (Maywood). 2022;247(18):1657–69. https://doi.org/10.1177/15353702221113828.
33. Zhu SF, Mao BL, Zhuang RY, Huang JY, Wu F, Wang BL, Yan Y. Development and validation of a diagnostic and prognostic model for bone metastasis of intrahepatic cholangiocarcinoma: a population-based analysis. Transl Cancer Res. 2024;13(8):4010–27. https://doi.org/10.21037/tcr-24-567.
34. Coughlin SS, Kapuku G. Commentary: Cancer survivorship and subclinical myocardial damage. Am J Epidemiol. 2022;191(3):367–8. https://doi.org/10.1093/aje/kwz123.

35. Imai S. Data-Driven Clinical Pharmacy Research: Utilizing Machine Learning and Medical Big Data. Biol Pharm Bull. 2024;47(10):1594–1599. https://doi.org/10.1248/bpb.b24-00492. (PMID: 39358238)
36. Benning L, Peintner A, Peintner L. Advances in and the applicability of machine Learning-Based screening and early detection approaches for cancer: a primer. Cancers (Basel). 2022;14(3):623. https://doi.org/10.3390/cancers14030623.
37. Zheng W, Chen S, Fu Z, Zhu F, Yan H, Yang J. Feature selection boosted by unselected features. IEEE Trans Neural Netw Learn Syst. 2022;33(9):4562–74. https://doi.org/10.1109/TNNLS.2021.3058172.
38. Xu C, Li L, Xu W, Du C, Yang L, Tong J, Yi Y. Ultrasound-guided percutaneous microwave ablation versus surgical resection for recurrent intrahepatic cholangiocarcinoma: intermediate-term results. Int J Hyperth. 2019;36(1):351–8. https://doi.org/10.1080/02656736.2019.1571247.
39. Wu L, Tsilimigras DI, Farooq A, Hyer JM, Merath K, Paredes AZ, Mehta R, Sahara K, Shen F, Pawlik TM. Potential survival benefit of radiofrequency ablation for small solitary intrahepatic cholangiocarcinoma in nonsurgically managed patients: a population-based analysis. J Surg Oncol. 2019;120(8):1358–64. https://doi.org/10.1002/jso.25736.
40. Choi WJ, Sapisochin G. Pushing the limits for the surgical treatment of intrahepatic cholangiocarcinoma. Hepatobiliary Surg Nutr. 2023;12(1):99–104. https://doi.org/10.21037/hbsn-22-555.
41. Zhang SJ, Hu P, Wang N, Shen Q, Sun AX, Kuang M, Qian GJ. Thermal ablation versus repeated hepatic resection for recurrent intrahepatic cholangiocarcinoma. Ann Surg Oncol. 2013;20(11):3596–602. https://doi.org/10.1245/s10434-013-3035-1.
42. Pereira ER, Kedrin D, Seano G, Gautier O, Meijer EFJ, Jones D, Chin SM, Kitahara S, Bouta EM, Chang J, Beech E, Jeong HS, Carroll MC, Taghian AG, Padera TP. Lymph node metastases can invade local blood vessels, exit the node, and colonize distant organs in mice. Science. 2018;359(6382):1403–7. https://doi.org/10.1126/science.aal3622.
43. Zhou R, Lu D, Li W, Tan W, Zhu S, Chen X, Min J, Shang C, Chen Y. Is lymph node dissection necessary for resectable intrahepatic cholangiocarcinoma? A systematic review and meta-analysis. HPB (Oxf). 2019;21(7):784–92. https://doi.org/10.1016/j.hpb.2018.12.011.
44. Vijgen S, Terris B, Rubbia-Brandt L. Pathology of intrahepatic cholangiocarcinoma. Hepatobiliary Surg Nutr. 2017;6(1):22–34. https://doi.org/10.21037/hbsn.2016.11.04.
45. Kargiotis O, Geka A, Rao JS, Kyritsis AP. Effects of irradiation on tumor cell survival, invasion and angiogenesis. J Neurooncol. 2010;100(3):323–38. https://doi.org/10.1007/s11060-010-0199-4.
46. Lhuillier C, Rudqvist NP, Elemento O, Formenti SC, Demaria S. Radiation therapy and anti-tumor immunity: exposing Immunogenic mutations to the immune system. Genome Med. 2019;11(1):40. https://doi.org/10.1186/s13073-019-0653-7.
47. Karagiannis GS, Pastoriza JM, Wang Y, Harney AS, Entenberg D, Pignatelli J, Sharma VP, Xue EA, Cheng E, D'Alfonso TM, Jones JG, Anampa J, Rohan TE, Sparano JA, Condeelis JS, Oktay MH. Neoadjuvant chemotherapy induces breast cancer metastasis through a TMEM-mediated mechanism. Sci Transl Med. 2017;9(397):eaan0026. https://doi.org/10.1126/scitranslmed.aan0026.
48. Keklikoglou I, Cianciaruso C, Güç E, Squadrito ML, Spring LM, Tazzyman S, Lambein L, Poissonnier A, Ferraro GB, Baer C, Cassará A, Guichard A, Iruela-Arispe ML, Lewis CE, Coussens LM, Bardia A, Jain RK, Pollard JW, De Palma M. Chemotherapy elicits pro-metastatic extracellular vesicles in breast cancer models. Nat Cell Biol. 2019;21(2):190–202. https://doi.org/10.1038/s41556-018-0256-3.
49. Addeo P, Jedidi I, Locicero A, Faitot F, Oncioiu C, Onea A, Bachellier P. Prognostic impact of tumor multinodularity in intrahepatic cholangiocarcinoma. J Gastrointest Surg. 2019;23(9):1801–9. https://doi.org/10.1007/s11605-018-4052-y.
50. The Lancet Respiratory Medicine. Opening the black box of machine learning. Lancet Respir Med. 2018;6(11):801. https://doi.org/10.1016/S2213-2600(18)30425-9.
51. Park HS, Lloyd S, Decker RH, Wilson LD, Yu JB. Limitations and biases of the surveillance, epidemiology, and end results database. Curr Probl Cancer. 2012;36(4):216–24. https://doi.org/10.1016/j.currproblcancer.2012.03.011.

Publisher's note