

RESEARCH

Open Access



Preoperative prediction of intrahepatic cholangiocarcinoma lymph node metastasis by means of machine learning: a multicenter study in China

Tingfeng Huang^{1†}, Hongzhi Liu^{1†}, Zhaowang Lin^{2†}, Jie Kong^{3†}, Kongying Lin¹, Zhipeng Lin¹, Yifan Chen¹, Qizhu Lin¹, Weiping Zhou⁴, Jingdong Li⁵, Jiang-Tao Li⁶ and Yongyi Zeng^{1*}

Abstract

Background: Hepatectomy is currently the most effective modality for the treatment of intrahepatic cholangiocarcinoma (ICC). The status of the lymph nodes directly affects the choice of surgical method and the formulation of postoperative treatment plans. Therefore, a preoperative judgment of lymph node status is of great significance for patients diagnosed with this condition. Previous prediction models mostly adopted logistic regression modeling, and few relevant studies applied random forests in the prediction of ICC lymph node metastasis (LNM).

Methods: A total of 149 ICC patients who met clinical conditions were enrolled in the training group. Taking into account preoperative clinical data and imaging features, 21 indicators were included for analysis and modeling. Logistic regression was used to filter variables through multivariate analysis, and random forest regression was used to rank the importance of these variables through the use of algorithms. The model's prediction accuracy was assessed by the concordance index (C-index) and calibration curve and validated with external data.

Result: Multivariate analysis shows that Carcinoembryonic antigen (CEA), Carbohydrate antigen19-9 (CA19-9), and lymphadenopathy on imaging are independent risk factors for lymph node metastasis. The random forest algorithm identifies the top four risk factors as CEA, CA19-9, and lymphadenopathy on imaging and Aspartate Transaminase (AST). The predictive power of random forest is significantly better than the nomogram established by logistic regression in both the validation group and the training group (Area Under Curve reached 0.758 in the validation group).

Conclusions: We constructed a random forest model for predicting lymph node metastasis that, compared with the traditional nomogram, has higher prediction accuracy and simultaneously plays an auxiliary role in imaging examinations.

Keywords: Intrahepatic cholangiocarcinoma, Lymph node metastasis, Machine learning

[†]Tingfeng Huang, Hongzhi Liu, Zhaowang Lin and Jie Kong contributed equally as first authors.

*Correspondence: lamp197311@126.com; lamp197311@fjmu.edu.cn

¹ Department of Hepatobiliary Surgery, Mengchao Hepatobiliary Hospital of Fujian Medical University, Xihong Road 312, Fuzhou, Fujian 350025, PR, China

Full list of author information is available at the end of the article

Introduction

Intrahepatic cholangiocarcinoma (ICC) is an uncommon but lethal disease that originates from bile duct epithelial cells above the secondary bile duct branch [1]. This condition highly malignant and has a poor prognosis [2] and its global incidence has been reported to be on the rise [3–5]. About 35% of the patients with ICC



already had lymph node metastasis (LNM) at the time of diagnosis [6]. An international multicenter study shows that LNM in cases of ICC is associated with an adverse effect on survival rates [7]. Therefore, exploring a new and effective method to evaluate the status of lymph node metastasis is of great significance for the diagnosis and treatment of patients with ICC [8–10].

Clinicians who rely solely on the imaging data to judge whether there is lymph node metastasis or not are susceptible to missed or inaccurate diagnosis [11]. Therefore, it is particularly important to comprehensively evaluate the status of the lymph nodes to better understand a patient's condition. In the era of rapid development of medical technology and diversified treatment options, the demand for methods to accurately judge the conditions of lymph nodes is increasing every year. Traditionally, logistic regression analysis and nomogram have been widely used in building prognostic predictive models [13, 14]. Some variables such as CA19-9, CEA, lymph node size on imaging (CT/MRI) (≥ 1 cm) are commonly extracted in the modeling process. However, the linearity assumption cannot model the complex, multidimensional and nonlinear relationship between variables, so these approaches have several limitations. Recently, machine learning has become one of the noteworthy directions of prediction models [15]. It uses nonlinear functions and takes into account the interaction between variables, thus improving prediction accuracy [16]. Currently, machine-learning techniques are rarely utilized in predicting LNM of ICC and many studies that applied this method were limited with relatively small sample sizes and lack of external validation dataset. Improved predictive ability could be achieved by applying traditional statistical and machine learning methods. An increasing number of studies have been paying attention to the limitations and advantages of traditional methods and machine learning and how they interfere with each other.

There are many machine learning methods available in the research community and random forest is one of the most used ones in clinical classification issues [17]. Random forests operate by constructing a multitude of decision trees. In the training process, the interaction between each variable can be detected and their importance can be sorted. In order to use preoperative indicators, including many complex variables, we used machine learning and logical regression in the modeling process. Relevant literature shows that patients with hepatitis B virus (HBV) infection tend to have a lower rate of lymph node metastasis [18, 19]. However, the P value of HBV was less than 0.05 in the multivariate analysis, so we did not include it in our modeling.

The purpose of this study is to add hepatitis as a predictor and to develop a model to accurately predict ICC LNM before surgery. The differences between machine learning and line diagram were also compared, and the credibility of the model was verified by external data.

Materials and methods

Patients of the training cohort

This study was approved by the Ethics Committee of the Mengchao Hepatobiliary Hospital of Fujian Medical University and exempts the requirement of written informed consent (IRB No 2019_049_01). All procedures were performed in accordance with the World Medical Association Declaration of Helsinki. The database was retrospectively derived from patients with pathologically confirmed ICC who underwent hepatic resection and lymph node dissection at the Eastern Hepatobiliary Surgery Hospital, Second Military Medical University (EHS) ($n=149$, from Jan. 2013 to Sept. 2018). We validated our model with an external data set consisting of three cohorts of patients from Mengchao Hepatobiliary Hospital of Fujian Medical University, The Second Affiliated Hospital Zhejiang University School of Medicine, and the Affiliated Hospital of North Sichuan Medical College ($n=62$).

The inclusion criteria were as follows: (1) pathological confirmation of ICC, (2) no intrahepatic and extrahepatic metastases, (3) Child–Pugh A/B before surgery, (4) underwent lymph node dissection. We excluded patients who: (1) did not undergo lymph node dissection or palliative surgery, (2) died within 30 days after surgery, (3) had incomplete preoperative imaging or serological data. (4) underwent neoadjuvant chemotherapy. Qualified patients from Eastern Hepatobiliary Surgery Hospital between 2013 and 2018 were placed in the training group, whereas patients who met the conditions in the other three hospitals between 2016 and 2020 were used for external verification.

Clinicopathologic variables

The serological data we used were the results of the most recent test conducted two weeks before the operation. The preoperative imaging data were obtained from contrast-enhanced Computed Tomography or contrast-enhanced Magnetic Resonance Imaging (CT/MRI) prior to surgery. Based on our clinical experience, the variables included sex, age, hepatitis B, tumor size on imaging (MRI/CT), number of tumors on imaging (single or multiple), liver cirrhosis on imaging, enlargement of the lymph nodes on CT/MRI imaging, total bilirubin, ALT, GGT, AST/ALT ($\leq 1, > 1$), ALP, PT, white blood cell count (WBC), platelets (PLT), red blood cell (RBC), alpha-fetoprotein (AFP) (≤ 20 ng/ml, > 20 ng/ml), carcinoembryonic antigen (CEA) (≤ 10 ng/ml, > 10 ng/ml).

ml), and carbohydrate antigen 19–9 (CA19-9) (≤ 39 ng/ml, > 200 u/ml). Furthermore, the maximum tumor diameter on imaging was defined as the biggest diameter. Solitary tumor was defined as only one tumor lesion in the liver, and two or more cancerous lesions were defined as multifocality. Enlargement of the lymph nodes was defined as > 1 cm on CT/MRI imaging.

Statistical analysis

All data analyses were performed by R software (version 4.1.0, <http://www.r-project.org>). The Mann–Whitney U test was used to compare continuous variables, and chi-square or Fisher’s exact test were used to compare categorical variables. C-Index was used to evaluate the forecasting effect of the model. All *P* values were based on a two-sided statistical analysis. $P < 0.05$ was considered statistically significant in single factor and multifactor analyses, and the variables with $p \leq 0.05$ were used to establish the prediction model. A small amount of the missing sample was supplemented by the mean imputation. C-Index is mainly used to reflect the discrimination ability of various forecasting models, and to investigate whether the model is correct in forecasting. In this research, The C-Index was calculated using the Hmisc (4.7) package of R language (4.1.3).

Surgical strategy

Hepatectomy was performed in patients who met the following criteria: 1) diagnosed with a technically resectable tumor with no evidence of extrahepatic metastases, and 2) generally good condition with healthy liver functions and adequate remaining liver volume. LND (lymph nodes dissection) was performed if 1) the patient was most likely to have LNM as determined by the preoperative multidisciplinary team, and 2) had enlarged lymph nodes that were manually detected by the surgeon during the procedure. The methods included subtotal hepatectomy and small hepatectomy. According to Couinaud’s classification, major hepatectomy is defined as the removal of three or more liver segments, and resection of less than three liver segments is defined as minor hepatectomy. The procedure of LND included skeletonization of the hepatoduodenal ligament and resection of para hepatic artery lymph nodes at least to the second station, which was a little different from each center.

Results

Baseline characteristics of patients

The database was retrospectively derived from patients with pathologically confirmed ICC who underwent

Table 1 Characteristics of patients in the primary and validation cohorts

	LNM level	Primary cohort			Validation cohort		
		0	1	p	0	1	p
n		82	67		31	31	
Sex (%)	Female	37 (45.12)	35 (52.24)	0.4839	14 (45.16)	17 (54.84)	0.6115
	Male	45 (54.88)	32 (47.76)		17 (54.84)	14 (45.16)	
Age (mean (SD))		55.793 (9.037)	53.284 (10.766)	0.1241	55.742 (10.395)	58.774 (10.191)	0.2507
num (%)	Single	75 (91.64)	59 (1.94)	0.6794	29 (93.55)	25 (80.65)	0.2557
	Multiple	7 (8.36)	7 (8.06)		2 (6.45)	6 (19.35)	
hepatitis (%)	negative	43 (52.44)	49 (73.13)	0.0157	27 (87.10)	23 (74.19)	0.3349
	positive	39 (47.56)	18 (26.87)		4 (12.90)	8 (25.81)	
size (mean (SD))		6.498 (2.343)	8 (11.94)	0.7527	5.830 (2.589)	6.405 (2.490)	0.377
cirrhosis (%)	negative	72 (87.80)	62 (92.54)	0.4956	28 (90.32)	30 (96.77)	0.6052
	positive	10 (12.20)	5 (7.46)		3 (9.68)	1 (3.23)	
rhm (%)	negative	71 (86.59)	44 (65.67)	0.0047	15 (48.39)	10 (32.26)	0.3004
	positive	11 (13.41)	23 (34.33)		16 (51.61)	21 (67.74)	
TBL (mean (SD))		14.583 (10.476)	16.791 (34.271)	0.5811	53.084 (112.683)	39.371 (86.214)	0.5925
ALB (mean (SD))		42.341 (3.378)	42.893 (4.656)	0.4045	39.274 (5.043)	38.535 (4.234)	0.5346
ALT (mean (SD))		32.822 (26.540)	33.976 (35.924)	0.822	37.871 (39.604)	45.161 (77.202)	0.6416
GGT (mean (SD))		127.151 (163.976)	145.381 (227.913)	0.5717	216.935 (115.562)	160.613 (208.391)	0.5025
ALP (mean (SD))		127.622 (96.705)	123.164 (71.726)	0.7545	191.968 (142.448)	207.161 (144.561)	0.6783
PT (mean (SD))		15.463 (4.301)	14.349 (3.784)	0.0992	15.484 (2.091)	15.829 (2.349)	0.5435
WBC (mean (SD))		6.639 (1.924)	6.923 (2.192)	0.401	6.659 (1.836)	7.438 (2.247)	0.1401
RBC (mean (SD))		4.357 (0.375)	4.407 (0.441)	0.4488	4.431 (0.488)	10.307 (32.612)	0.3199
PLT (mean (SD))		206.244 (69.260)	230.910 (81.970)	0.0483	209.971 (74.339)	265.626 (100.379)	0.0159
AFP,ng/ml (%)	≤ 20	70 (85.37)	58 (86.57)	1	30 (96.77)	28 (90.32)	0.6052
	> 20	12 (14.63)	9 (13.43)		1 (3.23)	3 (9.68)	
SL (%)	≤ 1	26 (31.71)	21 (31.34)	1	14 (45.16)	13 (41.94)	1
	> 1	56 (68.29)	46 (68.66)		17 (54.84)	18 (58.06)	
CEA ,ng/ml (%)	≤ 10	71 (86.59)	45 (67.16)	0.0082	27 (87.10)	20 (64.52)	0.0752
	> 10	11 (13.41)	22 (32.84)		4 (12.90)	11 (35.48)	
CA199,ng/ml (%)	≤ 39	32 (39.02)	16 (23.88)	0.016	14 (45.16)	6 (19.35)	0.0361
	39–200	29 (35.37)	19 (28.36)		5 (16.13)	3 (9.68)	
	> 200	21 (25.61)	32 (47.76)		12 (38.71)	22 (70.97)	

0 negative, 1 Positive, Sex: 0 female 1 male, LNM Pathology confirmed lymphnode metastasis, num Number of tumors on imaging (single or multiple), hepatitis Hepatitis B or not, TBL Total Bilirubin, ALB album in, ALT alanine amino transferase, PT prothrombintime, WBC leukocyte, RBC erythrocyte, PLT platelet, AFP Alpha fetoprotein, SL AST/ALT, CEA Carcinoembryonic antigen, CA199 Carbohydrate antigen 199, ALP Alpha fetoprotein

Table 2 Univariate and multivariate logistic regression analysis

	Characteristics	Univariate			Multivariate		
		OR	CI	P	OR2	CI2	P2
1	AFP	0.91	0.36–2.3	0.834	NA	NA	NA
2	Age	0.97	0.94–1.01	0.125	NA	NA	NA
3	ALB	1.04	0.95–1.13	0.408	NA	NA	NA
4	ALP	1	1–1	0.753	NA	NA	NA
5	ALT	1	0.99–1.01	0.821	NA	NA	NA
6	AST	1	1–1	0.56	NA	NA	NA
7	CA19–9	1.76	1.17–2.65	0.007	1.69	1.07–2.67	0.025
8	CEA	3.16	1.4–7.13	0.006	2.97	1.2–7.37	0.019
9	GGT	1	1–1	0.572	NA	NA	NA
10	hepatitis	0.41	0.2–0.81	0.011	0.49	0.23–1.05	0.066
11	PLT	1	1–1.01	0.051	NA	NA	NA
12	PT	0.93	0.86–1.01	0.1	NA	NA	NA
13	RBC	1.37	0.61–3.05	0.446	NA	NA	NA
14	rcirrhosis	0.58	0.19–1.79	0.344	NA	NA	NA
15	rhm	3.37	1.5–7.59	0.003	3.74	1.53–9.16	0.004
16	mm	1.45	0.5–4.24	0.494	NA	NA	NA
17	rsize	1.02	0.9–1.15	0.751	NA	NA	NA
18	sex	0.75	0.39–1.44	0.388	NA	NA	NA
19	SL	1.02	0.51–2.04	0.962	NA	NA	NA
20	TBIL	1	0.99–1.02	0.593	NA	NA	NA
21	WBC	1.07	0.91–1.25	0.399	NA	NA	NA

Abbreviations: AFP Alpha fetoprotein, ALB albumin, ALP Alpha fetoprotein, ALT Alanine aminotransferase, AST Aspartate Transaminase, CA19-9 Carbohydrate antigen 19–9, CEA Carcinoembryonic antigen, GGT γ-glutamyl transferase, hepatitis Hepatitis B positive, PLT platelet, PT Prothrombin time, RBC erythrocyte, rcirrhosis Imaging hepatic cirrhosis, rlm lymph node metastasis on imaging, mm multiple tumors on imaging, rsize Imaging tumor size, SL AST/ALT, TBIL Total bilirubin, WBC leukocyte

hepatic resection and lymph node dissection at Eastern Hepatobiliary Surgery Hospital, Second Military Medical University (EHSU) (*n* = 149). We validated our model with an external data set consisting of three cohorts of patients from Mengchao Hepatobiliary Hospital of Fujian Medical University, The Second Affiliated Hospital Zhejiang University School of Medicine, Affiliated Hospital of North Sichuan Medical College (*n* = 62).

Patient baseline demographic and clinical data are shown in Table 1. There were no significant differences in the percentage of LNM between the training group and the validation group (*P* = 0.944). The median patient age was 56.0 years (IQR 47.0–62.0) in the training cohort, and the median patient age was 60.0 years (IQR 52.0–64.0) in the validation cohort. Most patients were male (*n* = 77, 51.7%), and the majority of patients had a single lesion (*n* = 134, 90%) with a median tumor size of 6.2 cm (IQR 4.8–8.0) in the training group. Similar results were also observed in validation groups, and no statistical difference was detected.

Construction of the model

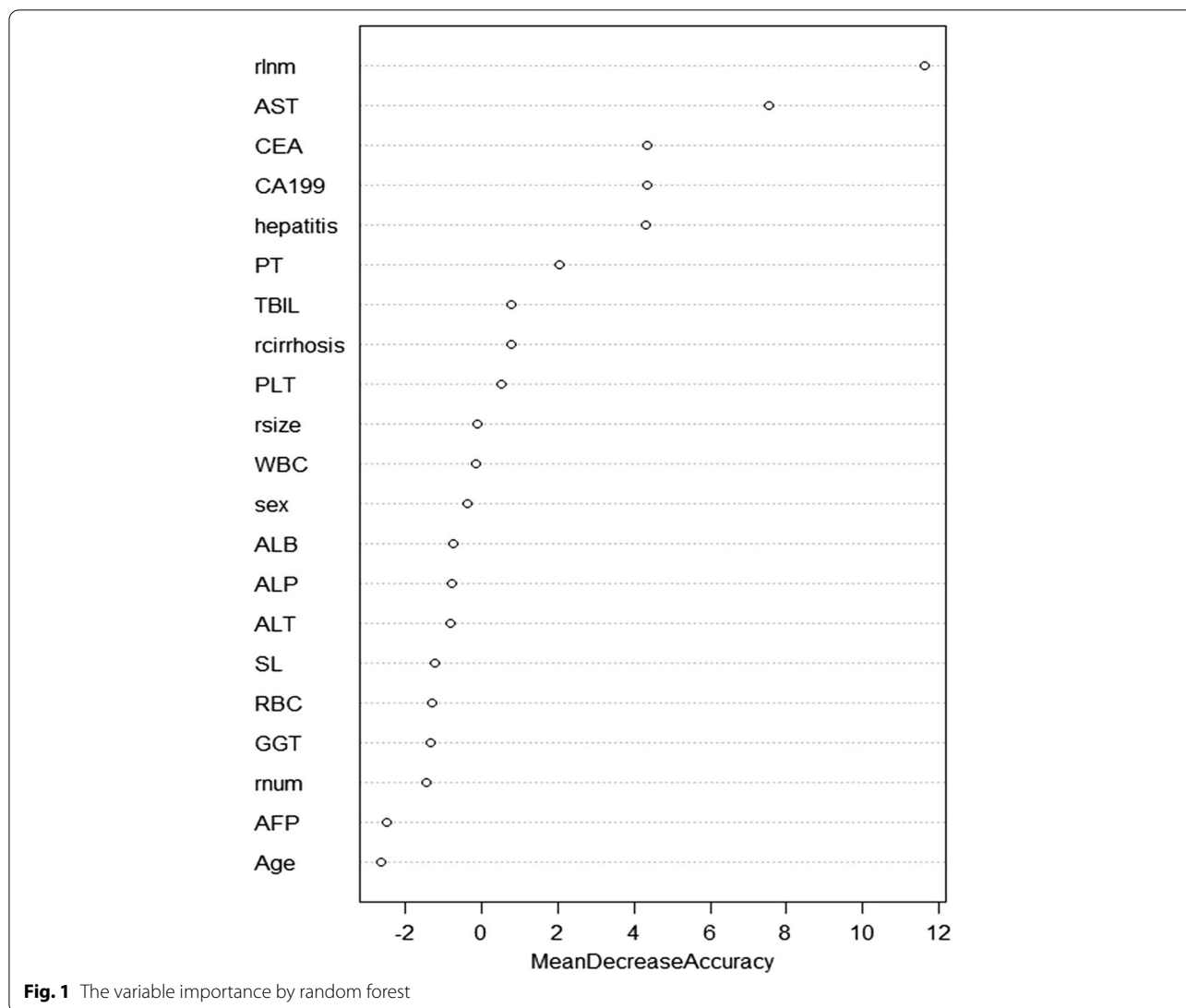
Variables

The following univariate and multivariate logistic regression analysis showed potential independent risk factors,

including CEA, CA19-9, and enlargement of the lymph nodes on CT/MRI imaging (Table 2). The importance ranking of factors calculated by random forest is shown in Fig. 1. The importance ranking of variables related to lymph node metastasis was obtained according to the Gini index algorithm of the importance of random forest variables. The top four variables were: lymph node metastases on imaging, AST, CEA > 10 ng/ml, and CA19-9 > 39 ng/ml.

Establishment and verification of the predictive nomogram

The results of the aforementioned multivariate logistic regression analysis showed that elevated CA19-9 levels (>200u/ml, *P* = 0.019), elevated CEA (>10 ng/ml, *P* = 0.025) and enlargement of the lymph nodes (>0.1 cm, *P* = 0.004) were independent postoperative risk factors to construct the nomogram (Fig. 2). In the nomogram, the value for each patient is located on each variable axis, and a line is drawn upward to determine the number of points received for each variable value. The sum of these numbers is located on the total score axis, and a line is drawn down to the outcome axis to determine the likelihood of developing lymph node metastases. The C-index was 0.733 (95% CI:0.654–0.813) for the prediction nomogram in the training

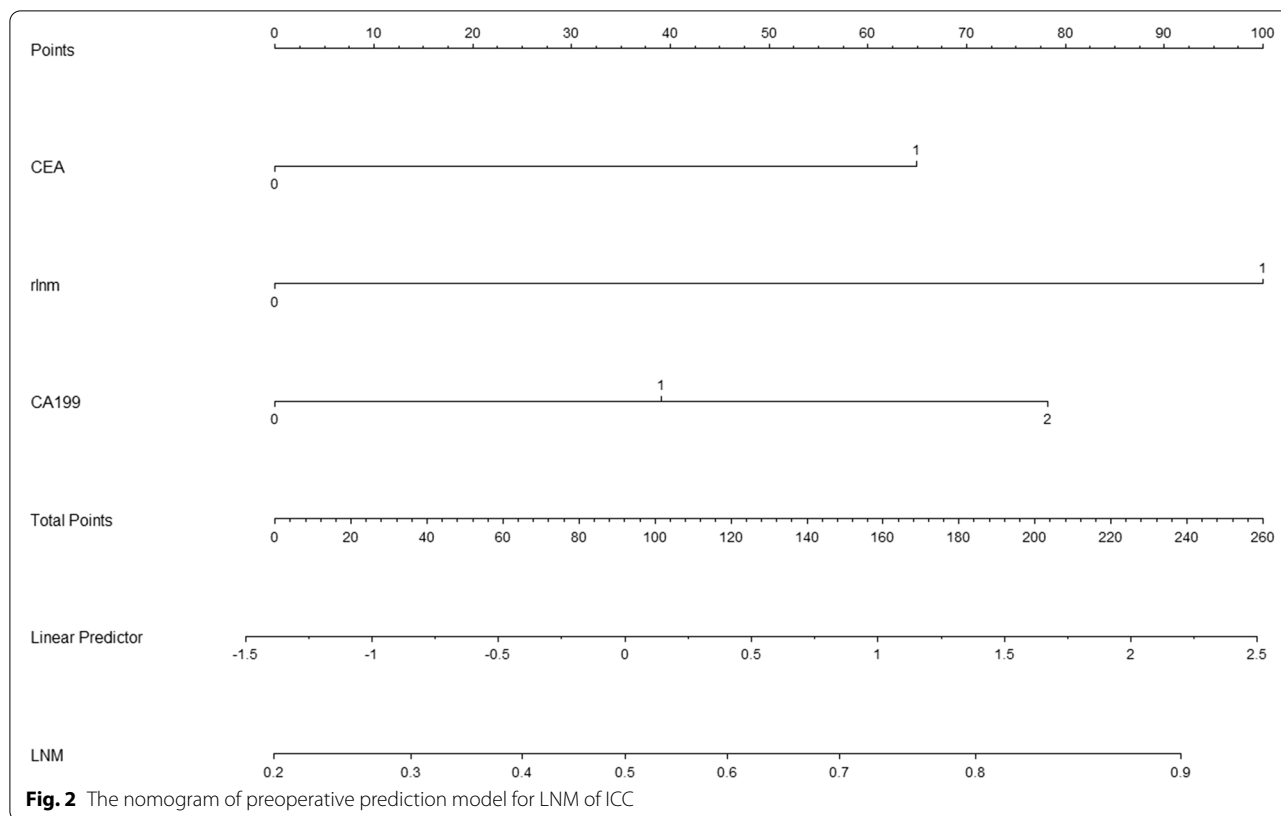


group (Fig. 3a). In the validation group, the C-index was 0.707 (95% CI: 0.576–0.837) for the prediction of LNM (Fig. 3b). In the logistic regression test group, the Sensitivity is 0.72, the Specificity is 0.59, the positive predictive value is 0.42, the negative predictive value is 0.84, the false positive is 0.60, the false negative is 0.28, the Accuracy is 0.63 (95%CI:0.50–0.75). In the logistic regression train group, the Sensitivity is 0.69, the Specificity is 0.68, the positive predictive value is 0.78, the negative predictive value is 0.57, the false positive is 0.32, the false negative is 0.31, the Accuracy is 0.68 (95%CI:0.60–0.76). The calibration curve is illustrated in Fig. 4A/B. The curve shows that the occurrence of lymph node metastasis predicted before the operation

was in good agreement with the actual situation in both the prediction and the validation groups.

Development and validation of the random forest

We constructed random forest classifiers consisting of 500 classification and regression trees (CART), using the implementation in the Random forest R package. The prediction factors were incorporated into the training by the integrated algorithm, the importance of variables was ranked according to the algorithm, and then a stochastic forest random model was constructed. The C-index was 0.837(95%CI: 0.777–0.896) for the random forest in the training group (Fig. 5a). In the validation cohort, the C-index was 0.758 (95% CI: 0.656–0.860) for the prediction of LNM (Fig. 5b). In the Random forest train group, the Sensitivity is 0.82, the Specificity is 0.90, the positive



predictive value is 0.93, the negative predictive value is 0.75, the false positive is 0.41, the false negative is 0.28, the Accuracy is 0.85 (95%CI:0.78–0.90). In the Random forest test group, the Sensitivity is 0.86, the Specificity is 0.70, the positive predictive value is 0.61, the negative predictive value is 0.90, the false positive is 0.11, the false negative is 0.19, the Accuracy is 0.76 (95%CI:0.63–0.86).

Discussion

Intrahepatic cholangiocarcinoma (ICC) is a rare and fatal disease [2, 5], and lymph node metastasis associated with this condition is an independent risk factor that affects the prognosis of ICC patients [20], and the choice of surgical methods for treatment [21]. The vast majority of related studies in this realm used logistic regression to screen risk factors and then established nomograms to try to predict lymph node metastasis [22, 23]. However, most studies lacked validation with multicenter data, and only a few researchers have applied random forest to predict lymph node metastasis in patients with ICC. To the best of our knowledge, we are the first group to use random forest to predict ICC lymph node metastases, compare the method with traditional techniques and validate the prediction accuracy with external data.

Multivariate logistic regression is widely used to construct nomograms. It removes confounding variables

and helps better visualize the results. However, logistic regression can only establish a linear model, and too many variables can easily affect the model fitting and equation stability [23]. Machine learning is a versatile tool for data analysis that has been widely used in predicting other carcinomas [25–27]. The calculation process of random forest is a black box, but its accuracy is high, its operational speed is fast, and the calculations do not easily overfit [24]. However, it is important to acknowledge that this method may have specific limitations in the case of small sample sizes [28]. In this paper, a logistic regression model and a random forest model were established simultaneously, and the accuracy of the two methods was further compared.

Our nomogram was established based on important risk factors for lymph node metastasis screened through logistic regression, including elevated CA19-9 (greater than 200u/ml $P=0.019$), elevated CEA (greater than 10 ng/ml $P=0.025$), and lymphadenopathy on imaging (largest lymph node diameter greater than 0.1 cm $P=0.004$). Previous studies have confirmed that these are important risk factors for lymph node metastasis in ICC [29–31]. It has also been reported that D-dimer combined with CA19-9 can predict lymph node metastasis in ICC [32]. Other scholars have also created nomograms established by logistic regression, and their results

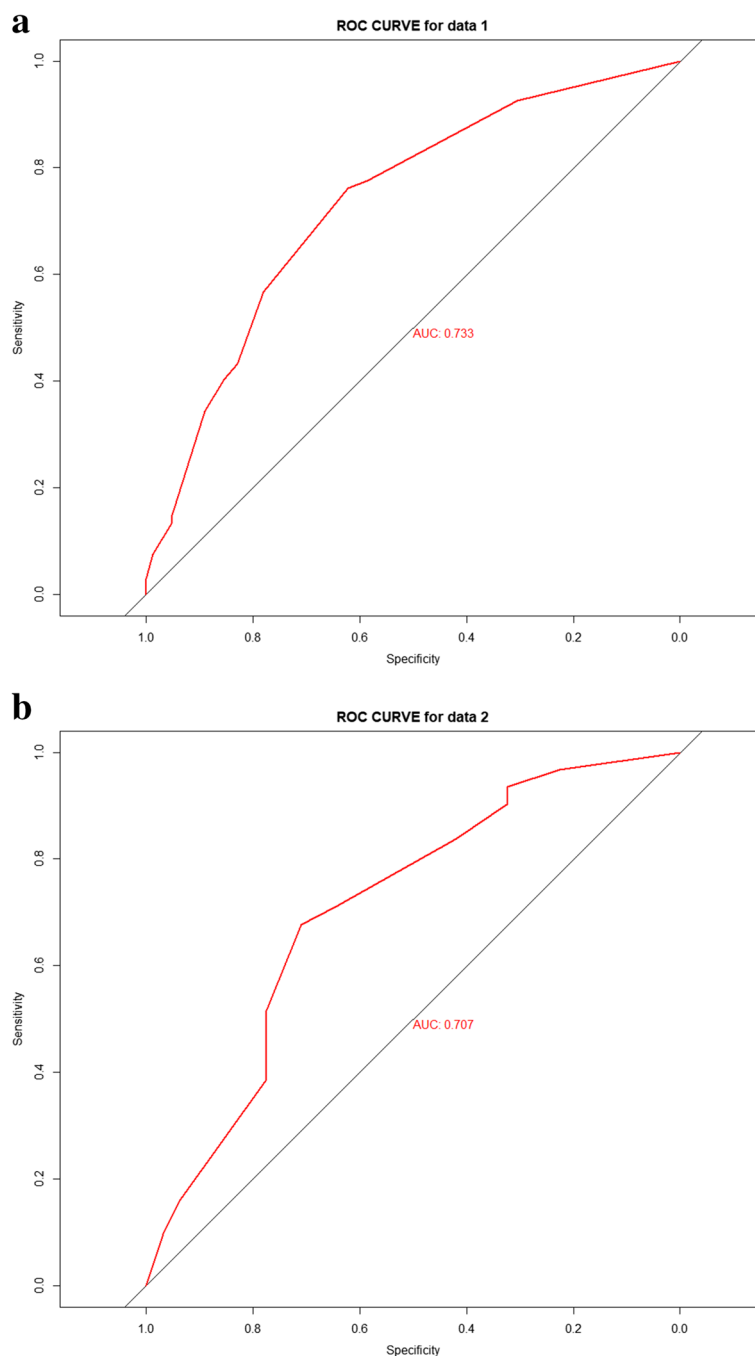


Fig. 3 a The C-index of nomogram in training group. b The C-index of nomogram in validation group

showed that they had good predictive power [12–14]. However, logistic regression is considered a limited method because it is difficult to further improve its prediction accuracy.

Recently, machine learning has been widely used in different research fields. Therefore, we innovatively adopted Random forest algorithm for LNM prediction in ICC.

In our study, the Random forest showed a huge advantage and better prediction accuracy over traditional nomograms. Through random forest, we also acquired preoperative serological indicators that are highly correlated with lymph node metastasis, such as AST, CEA, and CA19-9. Taking CA19-9 as an example, some scholars believe that para-aortic lymph node metastases are

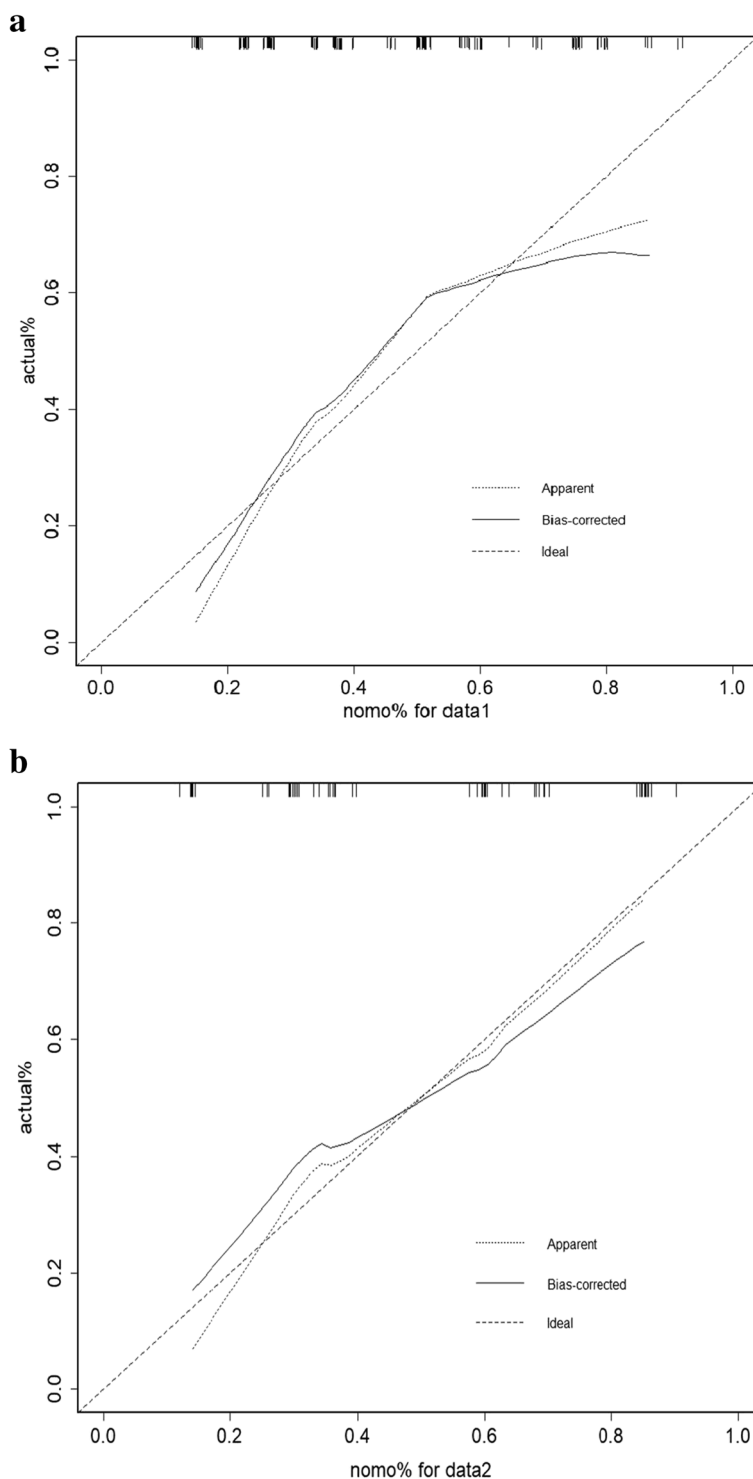


Fig. 4 **a** The calibration curve of train group. **b** The calibration curve of test group

already systemic metastases [33]. Preoperative CA19-9 levels have the potential to predict para-aortic lymph node metastasis, especially when CA19-9 levels are

higher than 200 U/ml [31]. Furthermore, it has been reported that elevated AST in other tumors may be associated with lymph node metastasis [34]. This correlation

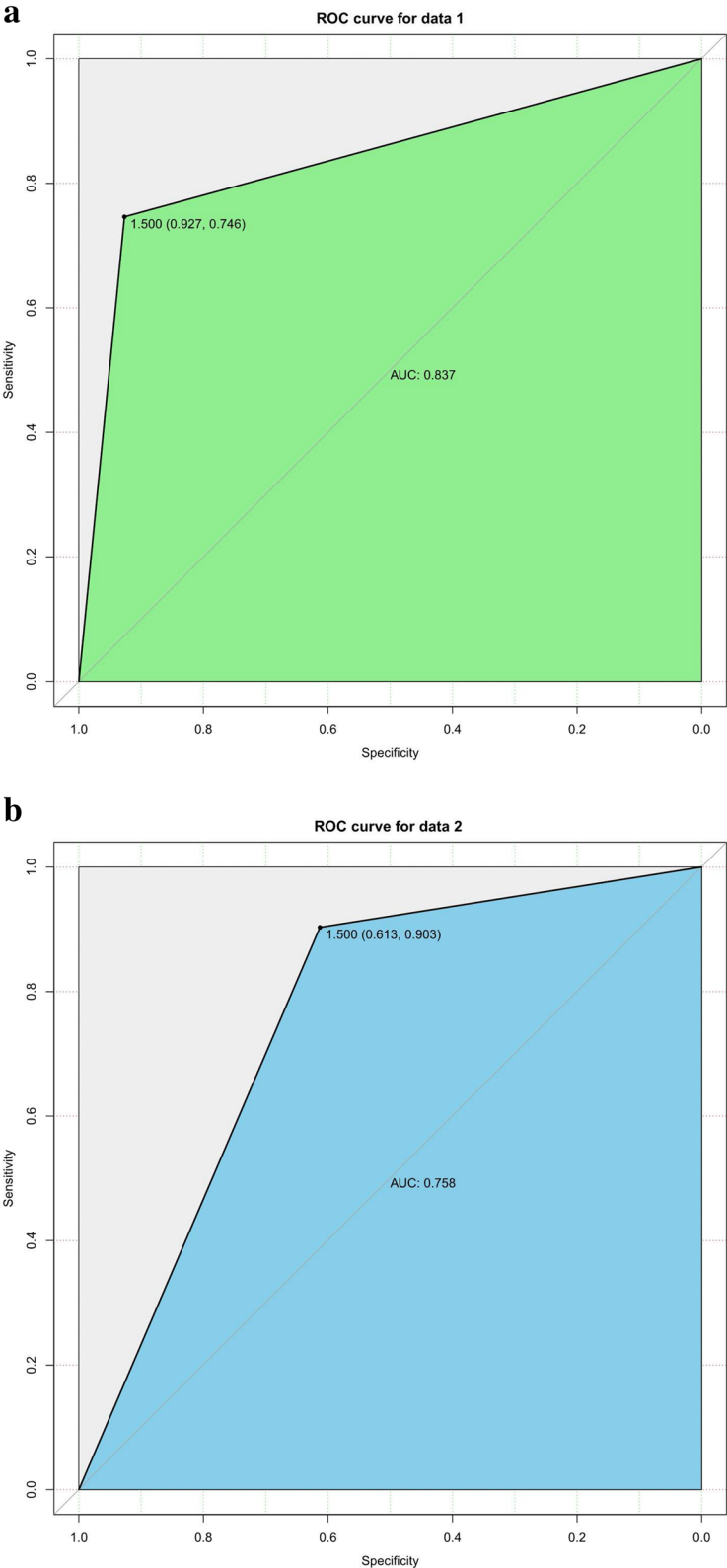


Fig. 5 **a** The C-index of random forest in training group. **b** The C-index of random forest in validation group

may be linked with elevated AST occurring under conditions that lead to a higher proliferative state, tissue damage, and increased tumor cell renewal [35]. So, serological indicators combined with preoperative imaging can help us better determine the status of lymph nodes.

Nevertheless, it is important to acknowledge the limitation of our research. Considering that the accuracy of random forest, if the number of cases is increased further, random forest regression would be a method that is more persuasive and superior in quality. Also, if the sample size increases, the model will reduce the risk of overfitting. Furthermore, although each center involved in this study is a large Grade 3A hospital (3A hospitals are top-class hospitals in China), there may be some differences in the diagnosis and treatment of ICC patients among different medical institutions. However, the multicenter data is more representative and increases the applicability of the prediction model. At last, although relevant data from different centers in China were included for analysis, we still lacked data from international centers. Therefore, in the future, if the sample size can be expanded and data from large scale international centers can be included, our machine learning method can be improved to make more accurate predictions. Previous studies have shown that ICC patients with lymph node metastases tend to have poor survival rates [7, 20]. However, due to a lack of follow-up data, a detailed survival analysis of predicted patients could not be performed in this study. This informational gap can be solved if prospective studies are carried out in the future. At the same time, patients with hilar invasion tend to have lymph node metastasis [36], but due to the limitations of our data, this part of the content has not been studied in depth. We hope that we can further improve the relevant content in the future.

In summary, this study proposes a preoperative model for predicting LNM of ICC patients using random forest, a commonly used machine learning method. Our research design proved that this method is effective in distinguishing lymph node status. This research presents a novel approach, as this is the first time the random forest method is used to predict ICC lymph node metastasis. Compared with traditional methods, the random forest technique can trace more accurate predictions, which is why it is considered advantageous for complex medical studies like ours. The use of multicenter data made our research more representative and our results more reliable and applicable to other contexts. Through the double verification process of the nomogram and the random forest, patients can accurately judge the status of their lymph nodes before the operation and have access to essential data to guide their surgical treatments.

Acknowledgements

Not Applicable.

Authors' contributions

Yongyi Zeng, Tingfeng Huang, Hongzhi Liu contributed to the study conception and design. Hongzhi Liu, Jie Kong, Zhaowang Lin, Kongying Lin, Zhipeng Lin, Yifan Chen, Qizhu Lin contributed to analysis and interpretation of data. Yongyi Zeng, Jingdong Li, Weiping Zhou, Jiang-Tao Li contributed to offering the data. Tingfeng Huang, Hongzhi Liu drafted the manuscript. Yongyi Zeng contributed to revising the article, critical revision, and final approval. The author(s) read and approved the final manuscript.

Funding

This study was supported by the Science and Technology project of Fuzhou (Grant number: 2020-WS-106), Startup Fund for Scientific Research, Fujian Medical University (2019QH1294), Startup Fund for Scientific Research, Fujian Medical University (2019QH1297), Health research talent training project of Fujian province (2019-1-85), Fuzhou Science and Technology Bureau project (2020-WS-57), Fuzhou Science and Technology Bureau project (2020-WS-92), Provincial Clinical Research Center for Hepatobiliary and Pancreatic Tumors (2020Y2013), Key Clinical Specialty Discipline Construction Program of Fuzhou (201912002) and the Scientific Foundation of Fuzhou Municipal Health Commission (2021-S-wp1). The funding bodies had no role in the study design, data collection and analysis, interpretation, or in writing the manuscript.

Availability of data and materials

All data included in this study are available upon request by contact with the corresponding author.

Declarations

Ethics approval and consent to participate

The study was approved by the Ethics Committee of the Mengchao Hepatobiliary Hospital of Fujian Medical University, and exempts the requirement of written informed consent (IRB No 2020_003_01). All procedures were performed in accordance with World Medical Association Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Hepatobiliary Surgery, Mengchao Hepatobiliary Hospital of Fujian Medical University, Xihong Road 312, Fuzhou, Fujian 350025, PR, China. ²Department of Radiology, Mengchao Hepatobiliary Hospital of Fujian Medical University, Fuzhou, China. ³Department of Hepatobiliary, Heze Municipality Hospital, Heze, China. ⁴The Third Department of Hepatic Surgery, Eastern Hepatobiliary Surgery Hospital, Second Military Medical University, Shanghai, China. ⁵Department of General Surgery, Institute of Hepatobiliary-Pancreatic-Intestinal Diseases, Affiliated Hospital of North Sichuan Medical College, Nanchong, China. ⁶Department of Hepatobiliary and Pancreatic Surgery, The Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou, China.

Received: 3 June 2022 Accepted: 22 August 2022

Published online: 29 August 2022

References

1. Liver Cancer Study Group of Japan. General rules for the clinical and pathological study of primary liver cancer. Fris. English. Tokyo: Kanehara & Co Ltd; 1997.
2. Meng ZW, Han SH, Zhu JH, Zhou LY, Chen YL. Risk Factors for Cholangiocarcinoma After Initial Hepatectomy for Intrahepatic Stones. *World J Surg.* 2017;41(3):835–43.
3. Nathan H, Pawlik TM, Wolfgang CL, Choti MA, Cameron JL, Schulick RD. Trends in survival after surgery for cholangiocarcinoma: a

- 30-year population-based SEER database analysis. *J Gastrointest Surg.* 2007;11(11):1488–96 discussion 1496–7.
4. Njei B. Changing pattern of epidemiology in intrahepatic cholangiocarcinoma. *Hepatology.* 2014;60(3):1107–8.
 5. Saha SK, Zhu AX, Fuchs CS, Brooks GA. Forty-Year Trends in Cholangiocarcinoma Incidence in the U.S. Intrahepatic Disease on the Rise. *Oncologist.* 2016;21(5):594–9.
 6. 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.
 7. de Jong MC, Nathan H, Sotiropoulos GC, Paul A, Alexandrescu S, Marques H, Pulitano C, et al. Intrahepatic cholangiocarcinoma: an international multi-institutional analysis of prognostic factors and lymph node assessment. *J Clin Oncol.* 2011;29(23):3140–5.
 8. Adachi T, Eguchi S. Lymph node dissection for intrahepatic cholangiocarcinoma: a critical review of the literature to date. *J Hepatobiliary Pancreat Sci.* 2014;21(3):162–8.
 9. Shirabe K, Mano Y, Taketomi A, Soejima Y, Uchiyama H, Aishima S, et al. Clinicopathological prognostic factors after hepatectomy for patients with mass-forming type intrahepatic cholangiocarcinoma: relevance of the lymphatic invasion index. *Ann Surg Oncol.* 2010;17(7):1816–22.
 10. Yoh T, Cauchy F, Le Roy B, Seo S, Taura K, Hobeika C, et al. Prognostic value of lymphadenectomy for long-term outcomes in node-negative intrahepatic cholangiocarcinoma: A multicenter study. *Surgery.* 2019;166(6):975–82.
 11. Adachi T, Eguchi S, Beppu T, Ueno S, Shiraiishi M, Okuda K, et al. Prognostic Impact of Preoperative Lymph Node Enlargement in Intrahepatic Cholangiocarcinoma: A Multi-Institutional Study by the Kyushu Study Group of Liver Surgery. *Ann Surg Oncol.* 2015;22(7):2269–78.
 12. Wang Y, Li J, Xia Y, Gong R, Wang K, Yan Z, et al. Prognostic nomogram for intrahepatic cholangiocarcinoma after partial hepatectomy. *J Clin Oncol.* 2013;31(9):1188–95.
 13. Tsilimigras DI, Sahara K, Paredes AZ, Moro A, Mehta R, Moris D, et al. Predicting Lymph Node Metastasis in Intrahepatic Cholangiocarcinoma. *J Gastrointest Surg.* 2021;25(5):1156–63.
 14. Meng ZW, Lin XQ, Zhu JH, Han SH, Chen YL. A nomogram to predict lymph node metastasis before resection in intrahepatic cholangiocarcinoma. *J Surg Res.* 2018;226:56–63.
 15. Chen G, Wang R, Zhang C, Gui L, Xue Y, Ren X, et al. Integration of pre-surgical blood test results predict microvascular invasion risk in hepatocellular carcinoma. *Comput Struct Biotechnol J.* 2021;16(19):826–34.
 16. Werner S, Schmidt L, Marchand V, Kemmer T, Falschlunger C, Sednev MV, et al. Machine learning of reverse transcription signatures of variegated polymerases allows mapping and discrimination of methylated purines in limited transcriptomes. *Nucleic Acids Res.* 2020;48(7):3734–46.
 17. Zhang B, Chang K, Ramkissoon S, Tanguturi S, Bi WL, Reardon DA, et al. Multimodal MRI features predict isocitrate dehydrogenase genotype in high-grade gliomas. *Neuro Oncol.* 2017;19(1):109–17.
 18. Zhang L, Cai JQ, Zhao JJ, Bi XY, Tan XG, Yan T, et al. Impact of hepatitis B virus infection on outcome following resection for intrahepatic cholangiocarcinoma. *J Surg Oncol.* 2010;101(3):233–8.
 19. Zhou HB, Wang H, Li YQ, Li SX, Wang H, Zhou DX, et al. Hepatitis B virus infection: a favorable prognostic factor for intrahepatic cholangiocarcinoma after resection. *World J Gastroenterol.* 2011;17(10):1292–303.
 20. Hu H, Xu G, Du S, Luo Z, Zhao H, Cai J. The role of lymph node dissection in intrahepatic cholangiocarcinoma: a multicenter retrospective study. *BMC Surg.* 2021;21(1):359.
 21. Bagante F, Gani F, Spolverato G, Xu L, Alexandrescu S, Marques HP, et al. Intrahepatic Cholangiocarcinoma: Prognosis of Patients Who Did Not Undergo Lymphadenectomy. *J Am Coll Surg.* 2015;221(6):1031–40 e1–4.
 22. Cheng Z, Lei Z, Shen F. Coming of a precision era of the staging systems for intrahepatic cholangiocarcinoma? *Cancer Lett.* 2019;28(460):10–7.
 23. Yoh T, Hatano E, Seo S, Terajima H, Uchida Y, Taura K, et al. Preoperative criterion identifying a low-risk group for lymph node metastasis in intrahepatic cholangiocarcinoma. *J Hepatobiliary Pancreat Sci.* 2018;25(6):299–307.
 24. Wang N, Cao Y, Song W, He K, Li T, Wang J, et al. Serum peptide pattern that differentially diagnoses hepatitis B virus-related hepatocellular carcinoma from liver cirrhosis. *J Gastroenterol Hepatol.* 2014;29(7):1544–50.
 25. Saria S, Butte A, Sheikh A. Better medicine through machine learning: What's real, and what's artificial? *PLoS Med.* 2018;15(12):e1002721.
 26. Kamerzell TJ, Middaugh CR. Prediction Machines: Applied Machine Learning for Therapeutic Protein Design and Development. *J Pharm Sci.* 2021;110(2):665–81.
 27. Song D, Wang Y, Wang W, Wang Y, Cai J, Zhu K, et al. Using deep learning to predict microvascular invasion in hepatocellular carcinoma based on dynamic contrast-enhanced MRI combined with clinical parameters. *J Cancer Res Clin Oncol.* 2021;147(12):3757–67.
 28. Kanwal F, Taylor TJ, Kramer JR, Cao Y, Smith D, Gifford AL, et al. Development, Validation, and Evaluation of a Simple Machine Learning Model to Predict Cirrhosis Mortality. *JAMA Netw Open.* 2020;3(11):e2023780.
 29. Kraus C, Klöbl M, Tik M, Auer B, Vanicek T, Geissberger N, et al. The pulvinar nucleus and antidepressant treatment: dynamic modelling of antidepressant response and remission with ultra-high field functional MRI. *Mol Psychiatry.* 2019;24(5):746–56.
 30. Yamada T, Nakanishi Y, Okamura K, Tsuchikawa T, Nakamura T, Noji T, et al. Impact of Serum Carbohydrate Antigen 19-9 Level on Prognosis and Prediction of Lymph Node Metastasis in Patients With Intrahepatic Cholangiocarcinoma. *J Gastroenterol Hepatol.* 2018;33(9):1626–33.
 31. Asaoka T, Kobayashi S, Hanaki T, Iwagami Y, Tomimaru Y, Akita H, et al. Clinical significance of preoperative CA19-9 and lymph node metastasis in intrahepatic cholangiocarcinoma. *Surg Today.* 2020;50(10):1176–86.
 32. Chen Q, Zheng Y, Zhao H, Cai J, Wang L, Zhao J, et al. The combination of preoperative D-dimer and CA19-9 predicts lymph node metastasis and survival in intrahepatic cholangiocarcinoma patients after curative resection. *Ann Transl Med.* 2020;8(5):192.
 33. Schwarz L, Lupinacci RM, Svrcek M, Lesurtel M, Bubenheim M, Vuarnesson H, et al. Para-aortic lymph node sampling in pancreatic head adenocarcinoma. *Br J Surg.* 2014;101(5):530–8.
 34. Mori K, Janisch F, Mostafaei H, Lysenko I, Kimura S, Egawa S, et al. Prognostic value of preoperative blood-based biomarkers in upper tract urothelial carcinoma treated with nephroureterectomy: A systematic review and meta-analysis. *Urol Oncol.* 2020;38(5):315–33.
 35. Warburg O. On the origin of cancer cells. *Science.* 1956;123:309–14.
 36. Umeda Y, Mitsuhashi T, Kojima T, Satoh D, Sui K, Endo Y, et al. Impact of lymph node dissection on clinical outcomes of intrahepatic cholangiocarcinoma: Inverse probability of treatment weighting with survival analysis. *J Hepatobiliary Pancreat Sci.* 2022;29(2):217–29.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

