

RESEARCH

Open Access



Interpretable machine learning-based clinical prediction model for predicting lymph node metastasis in patients with intrahepatic cholangiocarcinoma

Hui Xie^{1†}, Tao Hong^{2†}, Wencai Liu^{3†}, Xiaodong Jia^{4†}, Le Wang⁵, Huan Zhang⁶, Chan Xu⁷, Xiaoke Zhang⁶, Wen-Le Li^{7*}, Quan Wang^{8*}, Chengliang Yin^{9*} and Xu Lv^{10*}

Abstract

Objective Prediction of lymph node metastasis (LNM) for intrahepatic cholangiocarcinoma (ICC) is critical for the treatment regimen and prognosis. We aim to develop and validate machine learning (ML)-based predictive models for LNM in patients with ICC.

Methods A total of 345 patients with clinicopathological characteristics confirmed ICC from Jan 2007 to Jan 2019 were enrolled. The predictors of LNM were identified by the least absolute shrinkage and selection operator (LASSO) and logistic analysis. The selected variables were used for developing prediction models for LNM by six ML algorithms, including Logistic regression (LR), Gradient boosting machine (GBM), Extreme gradient boosting (XGB), Random Forest (RF), Decision tree (DT), Multilayer perceptron (MLP). We applied 10-fold cross validation as internal validation and calculated the average of the areas under the receiver operating characteristic (ROC) curve to measure the performance of all models. A feature selection approach was applied to identify importance of predictors in each model. The heat map was used to investigate the correlation of features. Finally, we established a web calculator using the best-performing model.

Results In multivariate logistic regression analysis, factors including alcoholic liver disease (ALD), smoking, boundary, diameter, and white blood cell (WBC) were identified as independent predictors for LNM in patients with ICC. In internal validation, the average values of AUC of six models ranged from 0.820 to 0.908. The XGB model was identified as the best model, the average AUC was 0.908. Finally, we established a web calculator by XGB model, which was useful for clinicians to calculate the likelihood of LNM.

[†]Hui Xie, Tao Hong, Wencai Liu, and Xiaodong Jia contributed equally to this article.

*Correspondence:

Wen-Le Li
Drlee0910@163.com
Quan Wang
quanwang1987@163.com
Chengliang Yin
chengliangyin@163.com
Xu Lv
yxlx0732024@163.com

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



Conclusion The proposed ML-based predicted models had a good performance to predict LNM of patients with ICC. XGB performed best. A web calculator based on the ML algorithm showed promise in assisting clinicians to predict LNM and developed individualized medical plans.

Keywords Intrahepatic cholangiocarcinoma, Machine learning algorithms, Lymph node metastasis, Web calculator

Intrahepatic cholangiocarcinoma (ICC) is the second most common pathological type of primary liver cancer, after hepatocellular carcinoma (HCC) [1], accounting for approximately 10%~20% of all cases [2, 3]. The incidence rate of ICC has increased during the last several decades [1, 4, 5]. ICC has an extremely poor prognosis and also is a highly invasive malignant tumor [1, 2], the 5-year overall survival rate has been reported in the range of 22–44% [1, 6]. In the progress of invasion, lymph node metastasis (LNM) is commonly observed, the rate of lymph node metastasis is about 25%~50% [7]. Median survival times in ICC patients with no lymph node metastasis is 19.0~37.6 months, whereas those with LNM had only 9.0~22.9 months [8]. Surgery serves as the major method of treatment for ICC patients [3], lymphadenectomy is crucial to accurately stage the disease and guide decisions around adjuvant chemotherapy [9]. However, no international consensus has been reached on management of the lymph nodes during the operation. Based on the essential impact of lymph node metastasis on staging and treatment in ICC patients, the identification of the probability of LNM has great effective clinical significance [10, 11].

Usually, radiological image is a main method to judge lymph node status, however the limitations can't be ignored. The sensitivity and specificity of CT diagnosis is 40%~50% and 77%~77%, respectively, and MRI is lower than CT scan [12], although the positron emission tomography (PET/CT) has higher accuracy in the assessment of LNM in patients with ICC [13], due to the high cost of PET/CT, it is not possible to routinely monitor all patients with this method. In clinic practice, pathology serves as the gold standard for LNM, but detailed information is unknown until after surgery [10]. Thus, reliable prediction models of LNM through clinical factors are urgent required. Various prediction models [3, 7, 14–18] have been constructed to predict the prognostic of ICC patients. As for the prediction model of LNM, although previous studies [7–9, 16, 18–20] have integrated potential risk factors to construct several predictive models, we don't found that current studies have developed and validated a model to predict LNM using ML algorithms.

Recently, Machine learning (ML) algorithm, as an emerging and popular type of artificial intelligence (AI), has attracted more and more attention due to the ability to predict events occurrence and outcome and was widely applied to health-care data analysis, aid in clinical

decision-making [21], especially in predicting possibility of metastatic diseases in malignant tumor patients [22, 23].

Herein, we developed and validated ML-based models using clinical characteristics to predict the probability of LNM in ICC patients. And a machine learning algorithm with the strongest predictive power is visualized by using a web calculator. This study will be helpful for surgical planning and clinical management.

Methods

Patient population

The Ethics Commission of the Fifth Medical Center of PLA General Hospital approved this present retrospective study (2019002D). All patients signed informed consent before surgery. Between Jan 2007 and Jan 2019, 345 patients who underwent surgical resection and regional lymphadenectomy for ICC at the Fifth Medical Center of PLA General Hospital were enrolled in this study.

Included patients had ICC proven by histopathology. The exclusion criteria were as follows:(1) history of other malignant tumors; (2) undergoing anticancer therapy (radiotherapy or chemotherapy) for liver malignancy before surgery; (3) primary liver cancer with mixed types and metastatic liver tumors; (4) incomplete clinical records.

Feature selection for modeling

The collected clinical features were conducted dimension reduction and screened by LASSO analysis, which was utilized to select optimal features with non-zero coefficients as risk factors from the development cohort and minimize the risk of overfitting [24]. The results of backward step-wise regression analysis in the optimal features datasets were included in univariate and multivariate logistics regression analyses. Then, the clinical variables in the univariate regression independently related to LNM were further analyzed by multivariate regression analysis, the LNM independently related variables with p -values < 0.05 in multivariate regression analysis were presented to generate predictive models for patients with ICC.

Development of the predictive models

Machine learning algorithms outperform traditional regression methods when predicting the outcomes [25].

In this study, we implemented six ML algorithms to develop predictive models as follows: Random Forest (RF), Logistic regression (LR), Extreme gradient boosting (XGB), Gradient boosting machine (GBM), Multi-layer perceptron (MLP), and Decision tree (DT) [26, 27]. Afterward we employed 10-fold cross-validation in the model development and calculated the average value of AUC of the receiver operating characteristic curve to compare prediction power of illustrated models. Using the Permutation Importance analysis to assess the importance of predictors in each ML-based model predicting LNM. We calculated Pearson's correlation coefficients to assess collinearity among the variables and plotted the correlation heat map. Finally, based on the best-performing model, we designed a web calculator as a predictive tool easily and accurately accessible to clinicians, making it possible to quantitatively calculate the individual probability of LNM.

Statistical analysis

We applied the mean \pm standard deviation (SD) to described the continuous variables and compared using the student's t tests, while categorical variables were expressed as percentages or frequencies and determined the significant difference using the chi-square test. Statistical analysis was performed with R software (version 4.05), including logistics regression analysis, baseline tables. Machine learning models and web calculator were built using Python (version 3.8). Statistical significance levels were set at .05.

Results

Baseline characteristics

The baseline characteristics between ICC patients with LNM and without LNM are detailed shown in Table 1. According to the inclusion and exclusion criteria, a total of 345 ICC patients have been enrolled. The median survival time was 20.49 months in patients without LNM, which was significantly different from patients with LNM (the median survival time = 7.83 months). Patients with LNM had higher mortality and shorter survival time than those without LNM ($p < 0.001$). This revealed that lymph node metastasis has a huge negative effect on survival of ICC patients. Patients with tumor diameter > 5 cm were more susceptible to metastases in lymph node. In addition, smoking, ALD (alcohol liver disease), white blood cell (WBC), boundary and diameter were all significantly associated with LNM (P -value < 0.05). However, there were no significant differences in NAFLD (non-alcoholic fatty liver disease), hyperlipidemia, image number, and Mg between the two groups (Table 1, $p > 0.05$).

Table 1 Baseline data of lymphatic metastasis in patients with intrahepatic cholangiocarcinoma

Characteristics	level	No(N=117)	Yes(N=228)	<i>p</i>
Status (%)	alive	48 (41.0)	27 (11.8)	<0.001
	dead	69 (59.0)	201 (88.2)	
Times (mean (SD))	NA	20.49 (19.05)	7.83 (5.88)	<0.001
NAFLD (%)	no	114 (97.4)	217 (95.2)	0.472
	yes	3 (2.6)	11 (4.8)	
ALD (%)	no	82 (70.1)	194 (85.1)	0.002
	yes	35 (29.9)	34 (14.9)	
Smoking (%)	no	74 (63.2)	76 (33.3)	<0.001
	yes	43 (36.8)	152 (66.7)	
Hyperlipidemia (%)	no	113 (96.6)	226 (99.1)	0.202
	yes	4 (3.4)	2 (0.9)	
Image number (%)	double	29 (24.8)	50 (21.9)	0.293
	more	5 (4.3)	4 (1.8)	
	single	83 (70.9)	174 (76.3)	
Boundary (%)	no	56 (47.9)	195 (85.5)	<0.001
	yes	61 (52.1)	33 (14.5)	
Diameter G (%)	<5 cm	54 (46.2)	23 (10.1)	<0.001
	>10 cm	7 (6.0)	29 (12.7)	
	5-10cm	56 (47.9)	176 (77.2)	
WBC (mean (SD))	NA	6.05 (2.40)	7.18 (1.96)	<0.001
Mg (mean (SD))	NA	0.87 (0.07)	0.90 (0.29)	0.269

LASSO and logistic regression for models development feature selection

Of all clinical features, 93 features were reduced to 12 potential predictors with nonzero coefficients in the LASSO logistic regression analysis (Fig. 1). LASSO analysis, a method suitable for data dimension reduction and feature selection of high-dimensional data, makes the relatively unimportant features coefficients zero by the regularization technique [28]. By backward stepwise regression, we selected 9 variables to univariate and multivariable logistics regression. The univariate logistics regression analysis found that 5 factors related to LNM, then according to the results in multivariable logistics regression analysis, ALD (yes, OR = 0.25, 95% CI = 0.13-0.5, $p < 0.001$), smoking (yes, OR = 3.83, 95% CI = 2.13-6.88, $p < 0.001$), boundary (yes, OR = 0.31, 95% CI = 0.17-0.55, $p < 0.001$), Diameter (5-10 cm, OR = 3.14, 95% CI = 1.63-6.06, $p = 0.000$; >10 cm, OR = 5.89, 95% CI = 2.06-16.85, $p = 0.001$), and WBC (the serum level $>7180/\mu\text{L}$, OR = 1.18, 95% CI = 1.03-1.35, $p = 0.016$) were identified as independent factors associated with LNM in patients with ICC (Table 2), among five variables, a distinct boundary and ALD were independent protective factors. Therefore, machine learning models were developed based on above five independent predictive factors (Table 2).

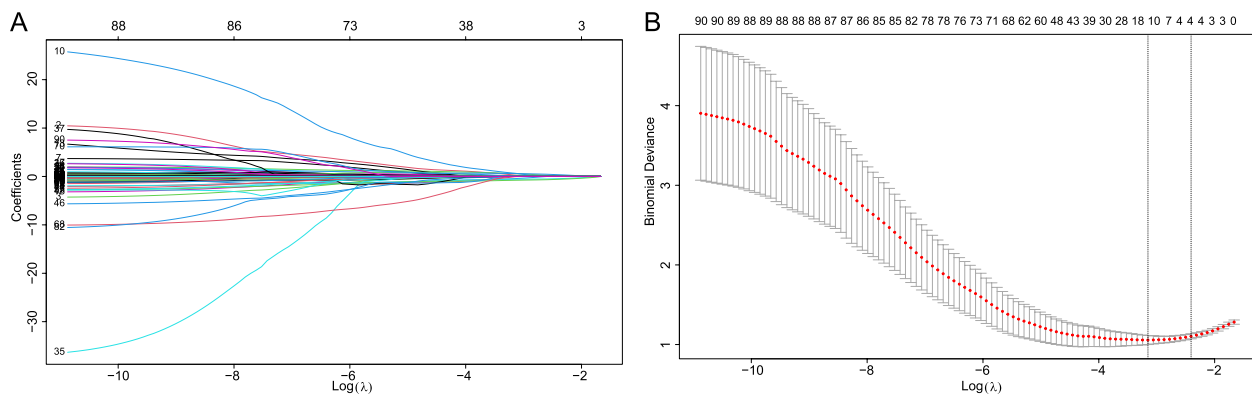


Fig. 1 The results of LASSO regression

Table 2 Univariate and multivariable logistics regression

Characteristics	OR	CI	P	OR	CI	P
ALD						
no	Ref	Ref	Ref	Ref	Ref	Ref
yes	0.41	0.24-0.7	0.001	0.25	0.13-0.5	<0.001
Boundary						
no	Ref	Ref	Ref	Ref	Ref	Ref
yes	0.16	0.09-0.26	<0.001	0.31	0.17-0.55	<0.001
Diameter G						
<5cm	Ref	Ref	Ref	Ref	Ref	Ref
>10cm	9.73	3.73-25.37	<0.001	5.89	2.06-16.85	0.001
5-10cm	7.38	4.16-13.09	<0.001	3.14	1.63-6.06	0.001
Hyperlipidemia						
no	Ref	Ref	Ref	Ref	Ref	Ref
yes	0.25	0.05-1.39	0.113	NA	NA	NA
Image number						
double	Ref	Ref	Ref	Ref	Ref	Ref
more	0.46	0.12-1.87	0.28	NA	NA	NA
single	1.22	0.72-2.06	0.467	NA	NA	NA
Mg	26.77	0.55-1299.91	0.097	NA	NA	NA
NAFLD						
no	Ref	Ref	Ref	Ref	Ref	Ref
yes	1.93	0.53-7.04	0.322	NA	NA	NA
Smoking						
no	Ref	Ref	Ref	Ref	Ref	Ref
yes	3.44	2.16-5.48	<0.001	3.83	2.13-6.88	<0.001
WBC	1.34	1.18-1.53	<0.001	1.18	1.03-1.35	0.016

Performance of developed models

Six machine learning algorithm models based on the five variables were constructed. Internal validation was assessed, the performance of each machine learning algorithm was evaluated by 10-fold cross-validation, the average AUC values for evaluate performance in

each model were calculated as follows: XGB: Average AUC=0.908; LR: Average AUC=0.820; MLP: Average AUC=0.840; DT: Average AUC=0.831; RF: Average AUC=0.876; GBM: Average AUC=0.864. As illustrated in Fig. 2. XGB algorithm had better accuracy in predicting LNM than the other five models.



Fig. 2 10-fold cross validation of machine learning algorithms

Variable importance and Pearson correlation of variables

Permutational importance quantified the variable importance in each ML algorithms (Fig. 3), WBC ranked first in five algorithms, the importance of variables in the XGB model is arranged in the following order: WBC, boundary, diameter G, smoking, ALD. In Fig. 4, we evaluated the correlation of the variables using Pearson’s correlation, and visualized the relationship of them via a heat map, indicating that no significant correlation and no

collinearity among the variables for LNM, indicating that the variables are independent of each other and no collinearity among the variables. WBC, followed by boundary, were the most important features in XGB, a significant negative correlation had been found between them.

Establishment of a web calculator

Based on the XGB model, we built an easy-to-use web calculator based on the XGB algorithm for clinicians to

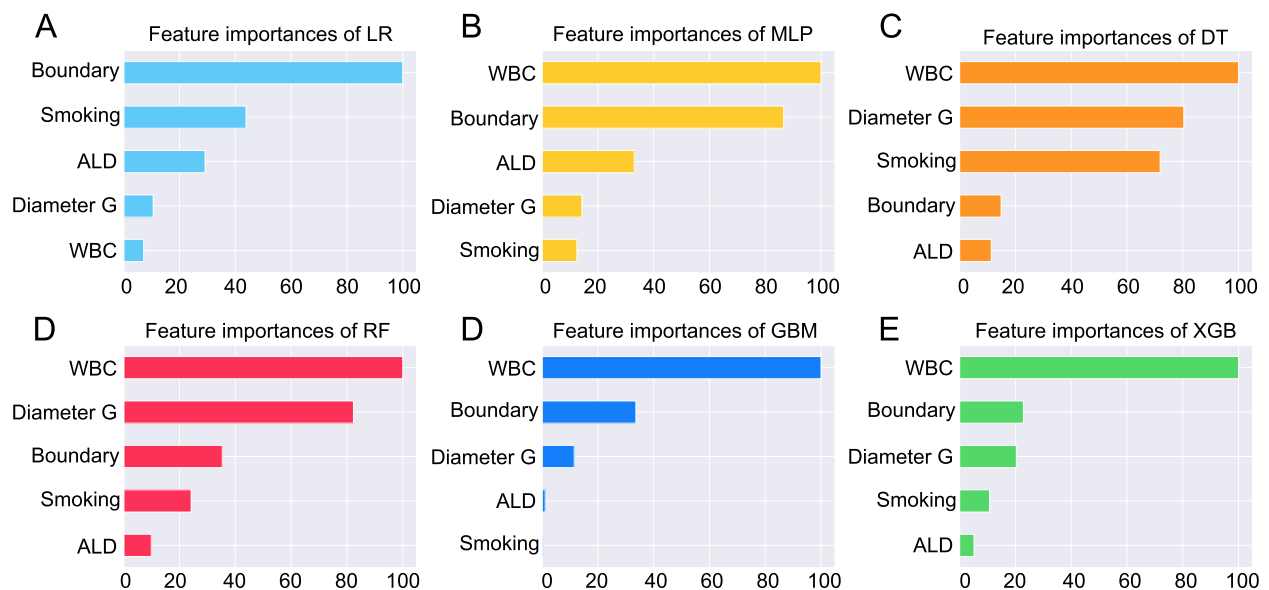


Fig. 3 Importance of the independent variables

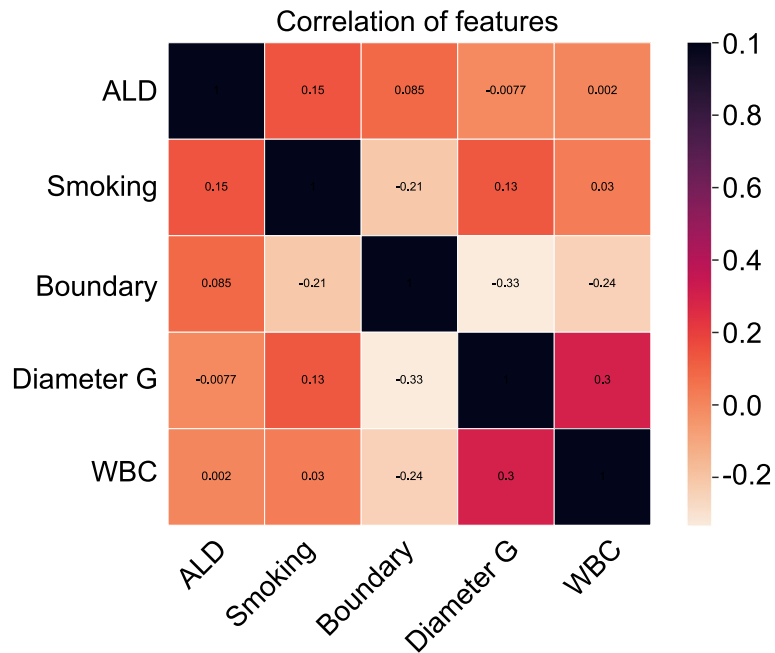


Fig. 4 Correlations between the independent variables

×

Variables

WBC

0.00

 20.00

Diameter.G

<5cm ▾

Boundary

No ▾

Smoking

No ▾

ALD

No ▾

A machine learning-based predictive model for predicting lymph node metastasis in patients with bile duct cancer

Predict

Risk grouping for LNM: Low Risk

Probability of LNM: 6.5%

☰

Fig. 5 The web-based calculator for predicting LNM in patients with ICC

calculate the individualized likelihood of LNM in ICC patients with a simple input of easily accessible clinical variables (Fig. 5).

Discussion

Intrahepatic cholangiocarcinoma originates from the malignant transformation of the bile ducts epithelium, and represents more aggressive compared to HCC [1], with the 5-year overall survival ranging from 15% to 40% [1, 6]. The incidence of LNM in ICC is much higher than that in HCC [29]. Indeed lymph node status is critical for therapy selection and has been identified as one of the most important factors for prognosis [6]. A few of studies demonstrated that lymphadenectomy (LND) improved long-term survival outcome of ICC patients [30, 31], thus, LND should be a routine method for radical resection in ICC [32, 33]. Whereas other studies reported that LND didn't improve survival outcome of ICC patients, with associated surgery-related complications [34, 35]. It's reported that approximately 50% of the patients did not dissect lymph node dissection [36], which may result in mis- or under-staging and further compromised their outcomes [32, 36]. For ICC patients, accurate prediction of LNM will facilitate clinical treatment decision-making for the appropriate diagnosis and surgical planning.

Accordingly, we used a novel type of AI-machine learning to predict LNM in ICC patients. Using ML algorithms, we developed and validated six models to predict LNM in 345 patients with ICC. We found that XGB model (average AUC=0.908) had greatest predictive performance in internal validation. Unlike some nomogram models [14, 19], we further provided dynamic construction. Consequently, based on the XGB model, a web calculator has been established to estimate visually individual probability of LNM and improved the applicability of the model.

In our study, multivariate logistic regression analysis founded that ALD, smoking, boundary, diameter, and WBC were independent predictive factors of LNM in patients with ICC (Table 2). As an independent risk factor, the influence of WBC on prognosis has been reported. Shirono et al [37] found that the serum WBC level was negatively associated with survival time in ICC patients, furthermore illustrated that patients with the WBC level was more than 6800/ μ L had a short survival time. In this study, we demonstrated that WBC was an independent predictor for the presentation of LNM in ICC patients. We also revealed that the risk of LNM was significantly increased when serum WBC level was more than 7180/ μ L. According to the permutation importance of variables in Fig. 3, WBC ranks first among the five prediction models and deserves the most attention when predicting LNM. WBCs include monocytes,

lymphocytes and neutrophils. Monocytes have roles in promoting tumor invasion and angiogenesis [38]. In addition, tumor-associated macrophages developed from monocytes, can promote tumor lymphangiogenesis by the secretion of pro-lymphangiogenic factors and trans-differentiation into lymphatic endothelial cells [39]. Submerb et al. reported that the monocyte in patients with Cholangiocarcinoma is correlated with a poor prognosis [40]. On the other hand, lymphocytes play an essential role in immune response, low counts may result in an insufficient immunological reaction against tumor progression and metastasis [38]. Previous research has revealed that lymphocyte to monocyte ratio (LMR) was associated with N stage and distant metastasis [41]. Peng et al. reported that the pre-LMR served as a predictor for early recurrence of Cholangiocarcinoma [42]. Meanwhile, a high neutrophil count was associated with poor prognosis and recurrence in ICC [43]. Stefan et al. reported that neutrophil to lymphocyte ratio was independently associated with worse overall survival among ICC patients [44]. In the present study, a high WBC level maybe reflect increasing in monocytes or neutrophil. The effects of monocytes, lymphocytes and neutrophils on lymph node metastasis should be further studied.

In addition, we concluded that tumors with diameter less than 5cm were less likely to occur LNM, which is similar to previous conclusion [20]. What's more, we performed more detailed studies for tumor (diameter>5cm), according to multivariate logistics regression analysis results, compared to tumor with 5-10cm, larger tumor (diameter more than 10cm) had a higher metastatic risk to lymph nodes (OR:5.89 VS 3.14). Due to the biological growth behavior of ICC, larger tumor volume means that the tumor has a longer growth cycle and further increases the possibility of lymph node invasive risk.

In addition, the present study found that the type of ICC boundary on radiological image was closely related to LNM, a distinct boundary played a protective role in reducing the likelihood of LNM occurrence, similar result has been reported previously [20]. Microinvasion may reveal a possible mechanism of tumor aggressiveness to lymph nodules [45]. As showed in Fig. 4, boundary served as the second important feature after WBC. Two other independent predictive factors were ALD and smoking. A meta-analysis of eight studies [46] reported that alcohol was major risk factors for ICC. Drinking alcohol causes alcoholic liver disease, which is greatly associated with increased ICC risk [47], as smoking dose [48]. Nonetheless, the relationship between ALD, smoking and LNM in ICC patients was comprehended poorly. Interestingly, we found that ALD was a protective factor for LNM. This finding seems to contradict the

existing literature identifying ALD as a risk factor for various cancers, including ICC [46, 47]. To reconcile this apparent paradox, we propose several hypotheses. First, ALD-induced immunosuppression may alter the host's immune landscape, reducing the attack of immune cells on cancer cells and thus reducing the spread of lymphoid tumors (Gao & Bataller, 2011). Second, liver pathology associated with ALD, particularly cirrhosis, may adversely alter the hepatic microenvironment, impeding tumor cell migration and invasion due to tissue reorganization and vascular changes [49]. Third, there may be a potential selection for survival bias, whereby ALD patients who die prematurely due to liver disease complications do not have sufficient time to develop LNM, leading to an underestimation of the risk factors associated with LNM in long-lived populations. Finally, the chronic inflammatory state associated with ALD may inhibit tumor spread, contrary to the generally accepted view that inflammation promotes cancer progression [50, 51]. These considerations highlight the complexity and individual variability of tumor biology and underscore the need for further research to elucidate the mechanisms by which ALD affects ICC metastatic behavior, thereby providing new insights into therapeutic approaches and patient management. Smoking was significantly associated with LNM and was an independent risk factor for LNM. Therefore, in people with a preliminary diagnosis of ICC, we recommend smoking cessation. However, whether quitting smoking can reduce the risk of LNM in patients with a history of smoking needs to be further verified.

To our knowledge, this paper is the first study to develop and validate a predictive models for predicting LNM in ICC applying machine learning algorithms. The model distinguishes from linear models adopted by previous studies, which can maximize clinical parameters and improve the diagnosis accuracy.

The XGB model initially proposed by Chen et al. in 2016 possessed the best prediction performance [22], it has a high accuracy and fast processing time and has been regarded as a more reliable algorithm when the sample size is limited [52]. Therefore, XGB is suitable for our study which is a small sample from a single medical center.

Finally, we established a concise, visualizable and dynamic online application based on XGB model, the real-time risk of LNM can be calculated and more rational and specific treatment regimens for patients can be tailored according to the personal information. For example, when an ICC patient presented with the following clinical characteristics: tumor diameter less

than 5 cm, no boundary, no smoking, ALD and serum WBC count is 5000/ μ L. We inputted above data into the web calculator, then the application integrated each factor and calculated automatically total probability of LNM, the output result was approximately 6.5% (Fig. 5), indicating that the patient had a low risk to lymph node metastasis. Therefore, we do not recommend further PET/CT monitoring and lymph node dissection.

Conclusions

To sum up, we constructed a machine learning-based predictive model with a good performance to predict LNM in patient with ICC based on independent factors, including ALD, smoking, boundary, tumor diameter and WBC level. In addition, we did an attempt to translate research outputs into clinical practices by building an online calculator, and the real-time predictive tool may aid in decision-making and management of ICC patients.

Limitations

Some limitations in our study can't be ignored. Firstly, as a retrospective study, selection bias was inevitable. In addition, the present study is small sample size from a single institution, our study is the lack of validation in an external dataset. In the future, external validation and large-scale multicenter studies will be required to validate our results. Thirdly, the inclusion of variables may affect the accuracy of the prediction model due to the highly subjective of the discrimination of tumor boundary and measure of diameter. Finally, there is a lack of analysis of LNM by neutrophils, lymphocytes and monocytes. According to previous studies [41], preoperative lymphocyte/monocyte ratio is associated with metastasis, and studying the subsets of WBC may improve the accuracy of prediction.

Authors' contributions

CLY, QW, XL and WLL designed the study. WCL, QW, HX and TH performed the study and analyzed the data. TH, HZ, WLL and CX and wrote the manuscript. HX and XL provided the expert consultations and clinical suggestions. XKZ, CX, QW and XDJ conceived of the study, CLY participated in its design and coordination, ALL authors helped to draft the manuscript.

Funding

This study was supported by grants from Natural Science Foundation of Beijing Municipality (7212099), National Natural Science Foundation of China (81902495), Medical Big Data and AI R & D Project of General Hospital (2019MBD-001 and 2019MBD-025), National Natural Science Foundation of China (NO. 81774304).

Availability of data and materials

The data used in this study can be obtained from the corresponding author on reasonable grounds.

Declarations

Ethics approval and consent to participate

The Ethics Commission of the Fifth Medical Center of PLA General Hospital approved this present retrospective study (2019002D).

Consent for publication

Not Applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of General Surgery, Yan 'an People's Hospital, Yan 'an, China. ²Department of Cardiac Surgery, Fuwai Hospital Chinese Academy of Medical Sciences, Shenzhen, China. ³Department of Orthopaedic Surgery, the First Affiliated Hospital of Nanchang University, Nanchang, China. ⁴Senior Department of Oncology, Fifth Medical Center of PLA General Hospital, Beijing, China. ⁵Department of thoracic surgery, the first affiliated hospital of Dalian Medical University, Dalian, China. ⁶Graduate School of Shaanxi University of Chinese Medicine, Xianyang 712046, China. ⁷State Key Laboratory of Molecular Vaccinology and Molecular Diagnostics & Center for Molecular Imaging and Translational Medicine, School of Public Health, Xiamen University, Xiamen 361102, China. ⁸Radiation Oncology Department, Fifth Medical Center of PLA General Hospital, Beijing, China. ⁹Faculty of Medicine, Macau University of Science and Technology, Macau, China. ¹⁰Department of General Surgery, Yixing Cancer Hospital, Yixing, Jiangsu 214200, China.

Received: 15 August 2023 Accepted: 5 April 2024

Published online: 19 April 2024

References

- Bridgewater J, Galle PR, Khan SA, Llovet JM, Park JW, Patel T, Pawlik TM, Gores GJ. Guidelines for the diagnosis and management of intrahepatic cholangiocarcinoma. *J Hepatol*. 2014;60(6):1268–89.
- Wu L, Tsilimigras DI, Paredes AZ, Mehta R, Hyer JM, Merath K, Sahara K, Bagante F, Beal EW, Shen F, et al. Trends in the incidence, treatment and outcomes of patients with intrahepatic cholangiocarcinoma in the usa: facility type is associated with margin status, use of lymphadenectomy and overall survival. *World J Surg*. 2019;43(7):1777–87.
- Tsilimigras DI, Sahara K, Paredes AZ, Moro A, Mehta R, Moris D, Guglielmi A, Aldrighetti L, Weiss M, Bauer TW, et al. Predicting lymph node metastasis in intrahepatic cholangiocarcinoma. *J Gastrointest Surg*. 2021;25(5):1156–63.
- Yeesoonsang S, McNeil E, Virani S, Bilheem S, Pittayawonganon C, Jiraphongsa C, Sriplung H. Trends in incidence of two major subtypes of liver and bile duct cancer: hepatocellular carcinoma and cholangiocarcinoma in songkhla, southern thailand, 1989–2030. *J Cancer Epidemiol*. 2018;2018:8267059.
- West J, Wood H, Logan RF, Quinn M, Aithal GP. Trends in the incidence of primary liver and biliary tract cancers in England and Wales 1971–2001. *Br J Cancer*. 2006;94(11):1751–8.
- Amini N, Spolverato G, Kim Y, Gupta R, Margonis GA, Ejaz A, Pawlik TM. Lymph node status after resection for gallbladder adenocarcinoma: prognostic implications of different nodal staging/scoring systems. *J Surg Oncol*. 2015;111(3):299–305.
- Yamaoka R, Ishii T, Kawai T, Yasuchika K, Miyauchi Y, Kojima H, Katayama H, Ogiso S, Fukumitsu K, Uemoto S. CD90 expression in human intrahepatic cholangiocarcinoma is associated with lymph node metastasis and poor prognosis. *J Surg Oncol*. 2018;118(4):664–74.
- Yoh T, Hatano E, Seo S, Terajima H, Uchida Y, Taura K, Yasuchika K, Uemoto S. Preoperative criterion identifying a low-risk group for lymph node metastasis in intrahepatic cholangiocarcinoma. *J Hepatobiliary Pancreat Sci*. 2018;25(6):299–307.
- Ji GW, Zhu FP, Zhang YD, Liu XS, Wu FY, Wang K, Xia YX, Zhang YD, Jiang WJ, Li XC, et al. A radiomics approach to predict lymph node metastasis and clinical outcome of intrahepatic cholangiocarcinoma. *Eur Radiol*. 2019;29(7):3725–35.
- 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.
- Ribero D, Pinna AD, Guglielmi A, Ponti A, Nuzzo G, Giulini SM, Aldrighetti L, Calise F, Gerunda GE, Tomatis M, et al. Surgical approach for long-term survival of patients with intrahepatic cholangiocarcinoma: a multi-institutional analysis of 434 patients. *Arch Surg*. 2012;147(12):1107–13.
- Sainani NI, Catalano OA, Holalkere NS, Zhu AX, Hahn PF, Sahani DV. Cholangiocarcinoma: Current and Novel Imaging Techniques. *Radiographics*. 2008;28(5):1263–87.
- Hu JH, Tang JH, Lin CH, Chu YY, Liu NJ. Preoperative staging of cholangiocarcinoma and biliary carcinoma using 18F-fluorodeoxyglucose positron emission tomography: a meta-analysis. *J Investig Med*. 2018;66(1):52–61.
- Nam K, Hwang DW, Shim JH, Song TJ, Lee SS, Seo DW, Lee SK, Kim MH, Kim KH, Hwang S, et al. Novel preoperative nomogram for prediction of futile resection in patients undergoing exploration for potentially resectable intrahepatic cholangiocarcinoma. *Sci Rep*. 2017;7:42954.
- Hyder O, Marques H, Pulitano C, Marsh JW, Alexandrescu S, Bauer TW, Gambin TC, Sotiropoulos GC, Paul A, Barroso E, et al. A nomogram to predict long-term survival after resection for intrahepatic cholangiocarcinoma: an Eastern and Western experience. *JAMA Surg*. 2014;149(5):432–8.
- Navarro JG, Lee JH, Kang I, Rho SY, Choi GH, Han DH, Kim KS, Choi JS. Prognostic significance of and risk prediction model for lymph node metastasis in resectable intrahepatic cholangiocarcinoma: do all require lymph node dissection? *HPB (Oxford)*. 2020;22(10):1411–9.
- Zhu Y, Mao Y, Chen J, Qiu Y, Wang Z, He J. Preoperative computed tomography features of intrahepatic cholangiocarcinoma for predicting lymph node metastasis and overall survival. *J Comput Assist Tomogr*. 2019;43(5):729–35.
- Yamada T, Nakanishi Y, Okamura K, Tsuchikawa T, Nakamura T, Noji T, Asano T, Tanaka K, Kurashima Y, Ebihara Y 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.
- 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.
- Chen YX, Zeng ZC, Tang ZY, Fan J, Zhou J, Jiang W, Zeng MS, Tan YS. Prediction of the lymph node status in patients with intrahepatic cholangiocarcinoma: analysis of 320 surgical cases. *Front Oncol*. 2011;1:42.
- Zhu J, Zheng J, Li L, Huang R, Ren H, Wang D, Dai Z, Su X. Application of machine learning algorithms to predict central lymph node metastasis in T1–T2, non-invasive, and clinically node negative papillary thyroid carcinoma. *Front Med (Lausanne)*. 2021;8:635771.
- Kwak MS, Eun YG, Lee JW, Lee YC. Development of a machine learning model for the prediction of nodal metastasis in early T classification oral squamous cell carcinoma: SEER-based population study. *Head Neck*. 2021;43(8):2316–24.
- Liu WC, Li ZQ, Luo ZW, Liao WJ, Liu ZL, Liu JM. Machine learning for the prediction of bone metastasis in patients with newly diagnosed thyroid cancer. *Cancer Med*. 2021;10(8):2802–11.
- Wen J, Pan T, Yuan YC, Huang QS, Shen J. Nomogram to predict post-operative infectious complications after surgery for colorectal cancer: a retrospective cohort study in China. *World J Surg Oncol*. 2021;19(1):204.
- Cai J, Luo J, Wang S, Yang S. Feature selection in machine learning: a new perspective. *Neurocomputing*. 2018;300:70–9.
- Gonzalez GH, Tahsin T, Goodale BC, Greene AC, Greene CS. Recent advances and emerging applications in text and data mining for biomedical discovery. *Brief Bioinform*. 2016;17(1):33–42.
- Ngiam KY, Khor IW. Big data and machine learning algorithms for healthcare delivery. *The Lancet Oncology*. 2019;20(5):e262–73.
- Wu TT, Yang DL, Li H, Guo YS. Development and validation of a nomogram to predict in-hospital cardiac arrest among patients admitted with acute coronary syndrome. *Am J Emerg Med*. 2021;49:240–8.
- Ohtsuka M, Ito H, Kimura F, Togawa A, Yoshidome H, Miyazaki M. Results of surgical treatment for intrahepatic cholangiocarcinoma and clinicopathological factors influencing survival. *Br J Surg*. 2002;89(12):1525–31.

30. Vitale A, Moustafa M, Spolverato G, Gani F, Cillo U, Pawlik TM. Defining the possible therapeutic benefit of lymphadenectomy among patients undergoing hepatic resection for intrahepatic cholangiocarcinoma. *J Surg Oncol*. 2016;113(6):685–91.
31. Yamashita Y, Taketomi A, Morita K, Fukuhara T, Ueda S, Sanefuji K, Iguchi T, Kayashima H, Sugimachi K, Maehara Y. The impact of surgical treatment and poor prognostic factors for patients with intrahepatic cholangiocarcinoma: retrospective analysis of 60 patients. *Anticancer Res*. 2008;28(4C):2353–9.
32. Weber SM, Ribero D, O'Reilly EM, Kokudo N, Miyazaki M, Pawlik TM. Intrahepatic cholangiocarcinoma: expert consensus statement. *HPB (Oxford)*. 2015;17(8):669–80.
33. Ruzzenente A, Conci S, Viganò L, Ercolani G, Manfreda S, Bagante F, Ciangherotti A, Pedrazzani C, Pinna AD, Iacono C, et al. Role of lymph node dissection in small (≤ 3 cm) intrahepatic cholangiocarcinoma. *J Gastrointest Surg*. 2019;23(6):1122–9.
34. Uchiyama K, Yamamoto M, Yamaue H, Ariizumi S, Aoki T, Kokudo N, Ebata T, Nagino M, Ohtsuka M, Miyazaki M, et al. Impact of nodal involvement on surgical outcomes of intrahepatic cholangiocarcinoma: a multicenter analysis by the Study Group for Hepatic Surgery of the Japanese Society of Hepato-Biliary-Pancreatic Surgery. *J Hepatobiliary Pancreat Sci*. 2011;18(3):443–52.
35. Chang M-E, Lei H-J, Chen M-H, Yeh Y-C, Li C-P, Hung Y-P, Hsia C-Y, Liu C-A, Chau G-Y, Chao Y. Evaluation of prognostic factors and implication of lymph node dissection in intrahepatic cholangiocarcinoma: 10-year experience at a tertiary referral center. *J Chin Med Assoc*. 2017;80(3):140–6.
36. Bagante F, Gani F, Spolverato G, Xu L, Alexandrescu S, Marques HP, Lame-las J, Aldrighetti L, Gamblin TC, Maitzel SK, et al. Intrahepatic cholangiocarcinoma: prognosis of patients who did not undergo lymphadenectomy. *J Am Coll Surg*. 2015;221(6):1031–1040.e1–4.
37. Shirono T, Niizeki T, Iwamoto H, Shimose S, Suzuki H, Kawaguchi T, Kamachi N, Noda Y, Okamura S, Nakano M, et al. Therapeutic outcomes and prognostic factors of unresectable intrahepatic cholangiocarcinoma: a data mining analysis. *J Clin Med*. 2021;10(5):987.
38. Kawai M, Hirono S, Okada KI, Miyazawa M, Shimizu A, Kitahata Y, Kobayashi R, Ueno M, Hayami S, Tanioka K, et al. Low lymphocyte monocyte ratio after neoadjuvant therapy predicts poor survival after pancreatectomy in patients with borderline resectable pancreatic cancer. *Surgery*. 2019;165(6):1151–60.
39. Riabov V, Gudima A, Wang N, Mickley A, Orekhov A, Kzhyshkowska J. Role of tumor associated macrophages in tumor angiogenesis and lymphangiogenesis. *Front Physiol*. 2014;5:75.
40. Subimerb C, Pinlaor S, Lulitanond V, Khuntikeo N, Okada S, McGrath MS, Wongkham S. Circulating CD14(+) CD16(+) monocyte levels predict tissue invasive character of cholangiocarcinoma. *Clin Exp Immunol*. 2010;161(3):471–9.
41. Hu RJ, Ma JY, Hu G. Lymphocyte-to-monocyte ratio in pancreatic cancer: prognostic significance and meta-analysis. *Clin Chim Acta*. 2018;481:142–6.
42. Peng D, Lu J, Hu H, Li B, Ye X, Cheng N. Lymphocyte to Monocyte Ratio Predicts Resectability and Early Recurrence of Bismuth-Corlette Type IV Hilar Cholangiocarcinoma. *J Gastrointest Surg*. 2020;24(2):330–40.
43. Watanabe A, Harimoto N, Araki K, Kubo N, Igarashi T, Tsukagoshi M, Ishii N, Yamanaka T, Yoshizumi T, Shirabe K. Absolute neutrophil count predicts postoperative prognosis in mass-forming intrahepatic cholangiocarcinoma. *Anticancer Res*. 2019;39(2):941–7.
44. Buettner S, Spolverato G, Kimbrough CW, Alexandrescu S, Marques HP, Lame-las J, Aldrighetti L, Gamblin TC, Maitzel SK, Pulitano C, et al. The impact of neutrophil-to-lymphocyte ratio and platelet-to-lymphocyte ratio among patients with intrahepatic cholangiocarcinoma. *Surgery*. 2018;164(3):411–8.
45. Bi AH, Zeng ZC, Ji Y, Zeng HY, Xu C, Tang ZY, Fan J, Zhou J, Zeng MS, Tan YS. Impact factors for microinvasion in intrahepatic cholangiocarcinoma: a possible system for defining clinical target volume. *Int J Radiat Oncol Biol Phys*. 2010;78(5):1427–36.
46. Palmer WC, Patel T. Are common factors involved in the pathogenesis of primary liver cancers? A meta-analysis of risk factors for intrahepatic cholangiocarcinoma. *J Hepatol*. 2012;57(1):69–76.
47. Xiong J, Yin Z, Xu W, Shen Z, Li Y, Lu X. Alcoholic liver disease and risk of cholangiocarcinoma: a systematic review and meta-analysis. *OncoTargets and therapy*. 2018;11:8211–9.
48. Petrick JL, Campbell PT, Koshiol J, Thistle JE, Andreotti G, Beane-Freeman LE, Buring JE, Chan AT, Chong DQ, Doody MM, et al. Tobacco, alcohol use and risk of hepatocellular carcinoma and intrahepatic cholangiocarcinoma: The Liver Cancer Pooling Project. *Br J Cancer*. 2018;118(7):1005–12.
49. El-Serag HB. Hepatocellular carcinoma. *N Engl J Med*. 2011;365(12):1118–27.
50. El-Serag HB. Hepatocellular carcinoma: recent trends in the United States. *Gastroenterology*. 2004;127(5 Suppl 1):S27–34.
51. Shiels MS, O'Brien TR. Recent decline in hepatocellular carcinoma rates in the United States. *Gastroenterology*. 2020;158(5):1503–1505.e1502.
52. Ye J, Hua M, Zhu F. Machine learning algorithms are superior to conventional regression models in predicting risk stratification of COVID-19 patients. *Risk Manag Healthc Policy*. 2021;14:3159–66.

Publisher's Note

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