Title page

Title: Insights to obstructive jaundice: comprehensive analysis and machine learning-based diagnostics in over 5000 individuals

Running Head: Machine learning-based diagnostics of obstructive jaundice

Authors: Ningyuan Wen[#1,2], Yaoqun Wang[#1,2], Xianze Xiong[1,2], Jianrong Xu[1,2], Shaofeng Wang[1,2], Yuan Tian[1,2], Di Zeng[1,2], Xingyu Pu[1,2], Geng Liu*[1,2], Bei Li*[1,2], Jiong Lu*[1,2], Nansheng Cheng*[1,2]

[#]Joint first author

*Joint last author

[1]Division of Biliary Surgery, Department of General Surgery, West China Hospital, Sichuan University, Chengdu 610041, Sichuan, China

[2]Research Center for Biliary Diseases, West China Hospital, Sichuan University, Chengdu 610041, Sichuan, China

Correspondence:

Professor Nansheng Cheng

Postal address: Division of Biliary Surgery, Department of General Surgery, West China Hospital, Sichuan University, Chengdu 610041, Sichuan, China

Email: nanshengcheng@yeah.net

Tel: +86-18980601366

Fax: +86 028-85422465

or

Professor Jiong Lu

Postal address: Division of Biliary Surgery, Department of General Surgery, West China Hospital, Sichuan University, Chengdu 610041, Sichuan, China

Email: lujiong@scu.edu.cn

Tel: +86-18602806495

or

Professor Bei Li

Postal address: Division of Biliary Surgery, Department of General Surgery, West China Hospital, Sichuan University, Chengdu 610041, Sichuan, China

Email: libei@scu.edu.cn

Tel: +86-18980600913

or

Dr. Geng Liu

Postal address: Research Center for Biliary Diseases, West China Hospital, Sichuan University, Chengdu 610041, Sichuan, China

Email: liu2geng6@163.com

Tel: +86- 15120090397

Declaration of interests:

Authors disclose no conflict of interest for this study.

Financial support statement:

This work was supported by Sichuan Provincial Commission of Health Science Project (20PJ059); Sichuan Science and Technology Program (Grant No.2022YSF0060, Grant No.2022YSF0114, Grant No.2022NSFSC0680, Grant No. 2023YFS0094); 1·3·5 project for disciplines of excellence–Clinical Research Incubation Project, West China Hospital, Sichuan University (20HXFH021); 1·3·5 project for disciplines of excellence, West China Hospital, Sichuan University (ZYJC21049); The Key Research and Development Program sponsored by the Ministry of Science and Technology of Chengdu (Grant No. 2021-YF05- 00065-SN).

Author contributions:

Study concept and design: NW, YW, BL, JL, GL and NC; Data acquisition: NW, YW, YT, BL, JL and NC; Data analysis and interpretation: NW, YW, GL, JX and DZ; Implementation of machine learning: NW, YW, XP and GL; Drafting of the manuscript: NW, SW, GL, BL, JL and CN; Funding: XX, BL, JL and CN. All authors have read and critically revised the manuscript and agreed to the published version.

Electronic word count: 5774

Electronic word count of the abstract: 267

Number of figures: 5

Number of tables: 2

Abstract

Background: Obstructive jaundice is a common problem associated with diverse etiologies which has not been thoroughly investigated in large-scale cohorts. Our study involved the largest retrospective cohort of obstructive jaundice to date, exploring the spectrum of diseases while establishing a diagnostic system with machine learning (ML) methods based on routine laboratory tests.

Methods: This study involves two retrospective observational cohorts from China. The biliary surgery cohort (BS cohort, n=349) served for initial data exploration and external validation of ML models, while the large general cohort (LG cohort, n=5726) enabled comprehensive data analysis and ML model construction. Interpretable ML techniques were employed to derive insights from the models.

Results: The LG cohort exhibited a more diverse disease spectrum compared to the BS cohort, with pancreatic adenocarcinoma, common bile duct stones, distal cholangiocarcinoma, perihilar cholangiocarcinoma, and acute pancreatitis (non-calculous) identified as the top five causes of obstructive jaundice. Traditional serum markers such as CA 19-9 and CEA did not emerge as standalone diagnostic markers for obstructive jaundice. Leveraging ML techniques, we developed two models collectively named as the MOLT model: one effectively distinguishes between benign and malignant causes (AUROC=0.862), while the other provides nuanced insights by further categorizing malignancies into three tiers and benign diseases into two (ACC=0.777). Interpretable ML tools revealed key features contributing to the decision-making process of each model.

Conclusions: Through our study, we uncovered the diagnostic potential of routine laboratory

tests in obstructive jaundice, enabling the development of a practical diagnostic tool based on interpretable ML models. These findings may pave the way for personalized and user-friendly diagnosis of obstructive jaundice, thereby aiding clinical decision-making.

Keywords:

Biliary Tract; Bile Duct Obstruction; Biliary Atresia; Biliary Tract Neoplasms; Gallstones; Pancreatitis; Pancreatic Ductal Carcinoma

Introduction

Obstructive jaundice is a common problem associated with various hepato-pancreato-biliary (HPB) diseases [1]. Its etiology includes both benign and malignant conditions. Benign causes of obstructive jaundice encompass a wide range of diseases, such as common bile duct stones, benign biliary strictures, Mirizzi syndrome, etc. Malignant causes typically involve tumors originating from the pancreas, bile ducts, or ampulla of Vater, including pancreatic adenocarcinoma, cholangiocarcinoma, ampullary carcinoma, and more [2-4]. Additionally, intrahepatic mass involving the hepatic hilus and metastatic tumors from other sites may also lead to obstructive jaundice [5, 6]. These diverse etiologies require careful evaluation and management to determine the appropriate course of treatment.

Over the past few decades, significant advancements have been made in the diagnostic approaches for obstructive jaundice. Serum-based diagnostics and imaging techniques have played pivotal roles in this regard. Serum-based diagnostics, including liver function tests (LFTs) and tumor markers, have greatly facilitated the initial assessment of obstructive jaundice. LFTs provide valuable information about liver enzymes, bilirubin levels, and other parameters indicative of hepatobiliary dysfunction. Tumor markers, such as carbohydrate antigen 19-9 (CA 19-9) and carcinoembryonic antigen (CEA), aid in the detection and monitoring of malignancies causing obstructive jaundice. On the other hand, imaging modalities have revolutionized the diagnosis and characterization of obstructive jaundice. Techniques such as ultrasonography (US), computed tomography (CT), magnetic resonance imaging (MRI) and endoscopic retrograde cholangiopancreatography (ERCP) offer detailed anatomical visualization of the hepatobiliary

system, enabling the identification of the underlying cause of obstruction [7]. Moreover, advancements in imaging technology, such as contrast-enhanced imaging and three-dimensional reconstruction, have further enhanced diagnostic accuracy of obstructive jaundice-associated HPB diseases [8, 9].

Despite these advancements, challenges persist in the diagnostic approach to obstructive jaundice. Serum-based markers may lack specificity and sensitivity, limiting their utility as standalone diagnostic tools [10]. Meanwhile, imaging modalities may encounter limitations in differentiating benign from malignant etiologies or accurately characterizing the extent of disease involvement, especially when the lesions are small [11]. Moreover, access to advanced imaging techniques may be limited under certain healthcare settings, hampering timely diagnosis and management. Additionally, in the field of obstructive jaundice research, there are still many gaps to be addressed. One of the most glaring issues is the lack of large-scale cohort studies, both retrospective and prospective. This has created numerous challenges, particularly in understanding the proportions of different diseases contributing to obstructive jaundice. This is one of the significant aspects our research endeavored to address, as we embarked on analyzing the specific causes to obstructive jaundice in a large retrospective cohort of over 5000 individuals. During this process, we discovered that much of the information carried by various clinical markers remained underutilized.

Hence, harnessing the power of state-of-the-art machine learning (ML) methods, renowned for their adaptability and ability to extract intricate patterns from complex datasets, we aimed to develop a robust diagnostic tool for obstructive jaundice [12, 13]. This tool, enabling the integration of diverse clinical laboratory tests to enhance accuracy and efficiency of diagnosis,

can be easily implemented due to its straightforward and user-friendly nature. Our objective was for it to not only differentiate between benign and malignant obstructive jaundice, but also to classify more specific etiologies based on this foundation. We validated its efficacy in real-world settings using an external surgical cohort, while employing interpretable ML techniques to offer transparency and insights into the decision-making process. By uncovering the untapped potential of routine clinical markers and leveraging ML techniques, we aimed to enhance the diagnosis and management of obstructive jaundice and improve patient outcomes.

Patients and Methods

Participants

The study protocol was approved by the Ethics Committee Biomedical Research, West China Hospital of Sichuan University, involving two retrospective observational cohorts from a single center (West China Hospital, Chengdu, China), and was preregistered in Open Science Framework (registration DOI: https://doi.org/10.17605/OSF.IO/DC4B8). The study workflow is visualized with a graphical abstract. This study involved two retrospective observational cohorts from a single center: the biliary surgery cohort (BS cohort) served as the dataset for initial data exploration and external validation of ML models, while the large general cohort (LG cohort) was utilized for comprehensive data analysis and ML model construction. The BS cohort consisted of patients diagnosed with obstructive jaundice who were admitted to the department of biliary surgery between February 2022 and September 2023. A total of 349

patients were eventually included in the BS cohort after screening with predefined criteria (Fig.S1). For the LG cohort, data was reviewed from all hospitalized patients diagnosed with obstructive jaundice in the general hospital between January 2008 and January 2022. A total of 5726 patients were included in the LG cohort after screening according to our predefined inclusion/exclusion criteria (Fig.1).

In short, these are the principles for patients to be included: (1) obstructive jaundice as a documented diagnosis; (2) reconfirmation of the diagnosis based on elevated cholestatic parameters (bilirubin, alkaline phosphatase and γ-glutamyltransferase); (3) etiology pathologically confirmed via ERCP, PTCD, FNB or surgical intervention; (4) age over 18 years. Additionally, patients with obstructive jaundice secondary to HPB surgery were excluded, ruling out iatrogenic factors. The aforementioned criteria delineated the spectrum of diseases covered by this study, as summarized in Figure 3.

Collection of clinical data

The clinical data of each patient was retrieved from the medical record archive of our institution and underwent de-identification. The following information underwent further investigation: age, sex, clinical diagnosis, pathological report and clinical laboratory test results.

The following test results were included: α-fetoprotein (AFP, ng/mL), carcinoembryonic antigen (CEA, ng/mL), cancer antigen 125 (CA 125, U/mL), cancer antigen 19-9 (CA 19-9, U/mL), red blood cell count (RBC, $\times 10^{12}$/L), hemoglobin (HGB, g/L), hematocrit (HCT, L/L), mean corpuscular volume (MCV, fL), mean corpuscular hemoglobin concentration (MCHC, g/L),

mean corpuscular hemoglobin (MCH, pg), red cell distribution width-coefficient of variation (RDW-CV, %), red cell distribution width-standard deviation (RDW-SD, fL), platelet count (PLT, $\times 10^9$/L), mean platelet volume (MPV, fL), platelet distribution width (PDW, %), large platelet ratio (P-LCR%, %), white blood cell count (WBC, $\times 10^9$/L), neutrophil percentage (NEUT%, %), lymphocyte percentage (LYM%, %), monocyte percentage (MONO%, %), eosinophil percentage (EO%, %), basophil percentage (BASO%, %), neutrophil count (NEUT#, $\times 10^9$/L), lymphocyte count (LYM#, $\times 10^9$/L), monocyte count (MONO#, $\times 10^9$/L), eosinophil count (EO#, $\times 10^9$/L), basophil count (BASO#, $\times 10^9$/L), total bilirubin (TBIL, μmol/L), direct bilirubin (DBIL, μmol/L), indirect bilirubin (IBIL, μmol/L), alanine aminotransferase (ALT, IU/L), aspartate aminotransferase (AST, IU/L), alkaline phosphatase (ALP, IU/L), γ-glutamyl transferase (γ-GT, IU/L), albumin (ALB, g/L), globulin (Glo, g/L), albumin/globulin ratio (A/G), glucose (GLU, mmol/L), urea (UREA, mmol/L), creatinine (CREA, μmol/L), cystatin C (CysC, mg/L), uric acid (UA, μmol/L), triglycerides (TG, mmol/L), cholesterol (CHOL, mmol/L), high-density lipoprotein cholesterol (HDL, mmol/L), low-density lipoprotein cholesterol (LDL, mmol/L), creatine kinase (CK, IU/L), lactate dehydrogenase (LDH, IU/L), hydroxybutyrate dehydrogenase (HBDH, IU/L), prothrombin time (PT, s), international normalized ratio (INR), activated partial thromboplastin time (APTT, s), fibrinogen (Fbg, g/L), thrombin time (TT, s), and C-reactive protein (CRP, mg/L).

Of note, one patient may undergo multiple examinations for the same item during the course of treatment. Only the laboratory results during the initial diagnosis of obstructive jaundice were utilized. If multiple test results still persist, median value was taken for further analysis.

Development and validation of ML models

Based on these clinical data, we explored a diverse array of prediction models in two different tasks. Firstly, we delved into binary classification models, which focused on distinguishing between benign and malignant diseases. We utilized various ML algorithms to achieve best predictive performance. These algorithms included logistic regression (LR), decision tree models, K-nearest neighbors (KNN), random forest (RF), support vector machine (SVM), eXtreme Gradient Boosting (XGBoost), and lightGBM. To evaluate model performance, we employed a comprehensive set of evaluation metrics, including area under the receiver operating characteristic curve (AUROC) with 95% confidence intervals, accuracy (ACC), area under the precision-recall curve (AUPR), F1 score, sensitivity and specificity.

Furthermore, beyond binary classification models, we extended our analysis to include multi-class classification models which further categorized diseases into five detailed categories based on their characteristics. These multi-class models were constructed using decision tree, XGBoost, RF, SVM, lightGBM, and KNN algorithms. Evaluation metrics of multi-class models included accuracy (ACC), area under the receiver operating characteristic curve weighted by prevalence (AUNU), Macro F1 score, precision score and recall score.

To assess the robustness and generalizability of the constructed models, internal validation and external validation were both conducted in binary classification models. Internal validation but not external validation was carried out for multi-class classification models, as the spectrum of diseases was limited in the BS cohort. All models were developed by R 4.1.3 (R Foundation for Statistical Computing, Vienna, Austria) using the mlr3 machine learning framework.

In addition, interpretability analysis was conducted on the optimized models to gain insights into the decision-making processes. Various tools from the iml package and mlr3verse package in R were employed for this purpose, including feature importance analysis and SHapley Additive exPlanations (SHAP) values.

Statistical analysis

Statistical analyses were conducted using R software. Shapiro-Wilk test and QQ plot were employed to assess the normality of data distribution. Continuous variables with a normal distribution were expressed as mean with standard deviation (SD), while those not following a normal distribution were expressed as median with interquartile range (IQR). Categorical data were presented as frequencies and percentages. For comparisons between groups, independent samples t-tests and Mann-Whitney U tests were performed for continuous variables with and without normal distribution, respectively, while chi-square tests were conducted for categorical variables. The DeLong test was employed for the assessment of model performance. A significance level of $P < 0.05$ was considered statistically significant.

Results

Patient profiles and disease spectrums of the study cohorts

We initially analyzed data from the biliary surgery cohort (BS cohort) which included 349

consecutive surgical inpatients diagnosed with obstructive jaundice (Fig. S1). Demographically, there were 216 (62%) male and 133 (38%) female, with 60% of them aged over 60. In terms of the disease spectrum, there were 204 cases (58.5%) with malignant obstructive jaundice and 145 cases (41.5%) with benign etiologies (Fig. S2A). As the BS cohort exclusively included patients from a biliary surgery center, it can be inferred that the majority of these patients presented with biliary-related issues rather than hepato-pancreatic diseases. Our analysis revealed that patients in the BS cohort were predominantly associated with biliary malignancies, accounting for 49.0% of all types of diseases and 86.8% of all malignancies (Fig. S2C). In terms of benign etiologies, calculous diseases, including common bile duct stones (CBDS) and hepatolithiasis (HL), were identified as the main cause to obstructive jaundice, accounting for 67.6% of all benign causes.

Undoubtedly, these preliminary results may not accurately reflect the true distribution of diseases in the general population, given that patients were selectively admitted to the surgical ward. Therefore, we proceeded to analyze obstructive jaundice in a large general cohort (LG cohort), comprising 5726 patients from a comprehensive medical center over a span of 14 years (Fig. 1). We first observed similarities between the LG cohort and BS cohort in terms of sex, age, and the relative proportion of benign and malignant diseases, underscoring the representativeness of our previous observations (Fig. S2B). Still, statistical analysis revealed significant differences in the disease spectrum as well as other baseline characteristics between the LG and BS cohorts (Table. S1). Based on a larger sample size, the LG cohort was able to unveil some previously undisclosed insights into obstructive jaundice, summarized in Fig. 2. To sum up, biliary malignancies (1657 cases, 28.94%), pancreatic malignancies (1106 cases, 19.32%), ampullary malignancies (252 cases, 4.40%), hepatic malignancies (190 cases, 3.32%), metastatic cancers

(360 cases, 6.29%) and other rare malignancies (155 cases, 2.71%) built up the malignant side of the disease spectrum; while calculous diseases (1257 cases, 21.95%), inflammatory diseases (566 case, 9.88%) and other benign causes (171 cases, 2.99%) composed the benign counterpart. The composition of a more detailed category of diseases is also depicted in Fig. 2. The top five leading causes of obstructive jaundice were revealed as pancreatic adenocarcinoma (1094 cases, 19.11%), CBDS (1046 cases, 18.27%), distal cholangiocarcinoma (dCCA) (595 cases, 10.39%), perihilar cholangiocarcinoma (pCCA) (573 cases, 10.01%) and acute pancreatitis (non-calculous) (328 cases, 5.73%).

To summarize, before we went further into the development of diagnostic models for obstructive jaundice, comprehensive analysis was conducted of our data and obtained valuable insights into the spectrum of diseases underlying this condition.

Requirement for machine learning-based diagnostics in obstructive jaundice

Our next objective was to evaluate the effectiveness of current diagnostic markers for obstructive jaundice, aiming to ascertain whether a more effective diagnostic approach is warranted. In the exploratory phase of this study, we conducted a comparative analysis of the baseline characteristics of patients with benign and malignant obstructive jaundice in the BS cohort. The results suggested that there was a significant difference in multiple clinical indices between benign and malignant groups, particularly evident in tumor markers and LFTs (Table. S2 and Figure. S3A). However, despite the evident statistical difference, the diagnostic performance of single laboratory markers for distinguishing benign from malignant conditions was suboptimal.

The top five diagnostic markers were CA 19-9 (AUROC=0.768), DBIL (AUROC=0.736), TBIL (AUROC=0.730), CEA (AUROC=0.697), and DBIL/TBIL ratio (AUROC=0.696) (Figure. S3B). We then continued our investigation in the LG cohort to further validate these results based on a larger sample size. Accordingly, a comparative analysis was conducted to unveil statistical variances in the baseline characteristics between benign and malignant groups (Table. 1).

Table. 1. Comparative analysis of baseline characteristics between benign and malignant groups in the LG cohort.

Variables	LG cohort (borderline disease excluded, n=5714)		
	Benign obstruction (n=1994)	Malignant obstruction (n=3720)	pVal
Sex			0.36
Male	1151 (57.7%)	2099 (56.4%)	
Female	843 (42.3%)	1621 (43.6%)	
Age (years)	53.9 (16.1)	59.9 (11.7)	<0.01
CA 19-9 (U/ml)	31.43 [13.63, 84.61]	132.64 [34.76, 306.48]	<0.01
CEA (ng/ml)	2.12 [1.35, 3.35]	3.41 [2.07, 6.39]	<0.01
CA 125 (U/ml)	19.33 [11.61, 41.29]	24.35 [14.15, 51.65]	<0.01
AFP (ng/ml)	2.64 [1.90, 3.75]	2.92 [2.14, 4.23]	<0.01
TBIL (μmol/L)	53.0 [42.4, 81.7]	54.8 [44.3, 119.4]	<0.01
DBIL (μmol/L)	21.3 [13.3, 51.1]	24.7 [16.5, 88.1]	<0.01
IBIL (μmol/L)	9.2 [6.6, 13.2]	8.0 [5.6, 11.9]	<0.01
ALT (IU/L)	42 [24, 73]	39 [26, 63]	0.20
AST (IU/L)	41 [27, 75]	33 [24, 57]	0.02
ALP (IU/L)	139 [87, 226]	131 [82, 229]	0.42
γ-GT (IU/L)	119 [37, 288]	81 [27, 195]	<0.01
ALB (g/L)	38.5 (5.3)	37.5 (5.8)	<0.01
Glo (g/L)	28.5 (6.0)	27.2 (5.2)	<0.01
A/G	1.37 [1.17, 1.56]	1.41 [1.23, 1.61]	0.07
GLU (mmol/L)	5.74 [5.17, 6.68]	5.83 [5.23, 6.86]	0.36
UREA (mmol/L)	4.8 [3.9, 5.8]	4.7 [3.9, 5.6]	0.59
CREA (μmol/L)	55.48 [52.00, 63.98]	53.27 [48.11, 61.00]	0.04
CysC (mg/L)	0.95 [0.84, 1.08]	0.94 [0.83, 1.06]	0.24
UA (μmol/L)	228 (79)	207 (89)	0.03
TG (mmol/L)	1.33 [1.02, 1.72]	1.37 [1.03, 1.80]	0.38
CHOL (mmol/L)	4.20 [3.45, 5.01]	4.26 [3.42, 5.30]	0.25

HDL (mmol/L)	1.00 [0.61, 1.32]	0.87 [0.41, 1.29]	<0.01
LDL (mmol/L)	2.16 (0.87)	2.06 (0.87)	0.06
CK (IU/L)	65 [44, 104]	71 [47, 101]	0.31
LDH (IU/L)	177 [155, 203]	179 [157, 201]	0.79
HBDH (IU/L)	137 [124, 159]	138 [123, 156]	0.78
RBC (×10^12/L)	3.84 (0.79)	3.58 (0.65)	<0.01
HGB (g/L)	115 (22)	109 (19)	<0.01
HCT (L/L)	0.35 (0.07)	0.33 (0.06)	<0.01
MCV (fL)	93.0 [89.5, 96.3]	93.7 [90.2, 97.0]	<0.01
MCHC (g/L)	328 (11)	327 (11)	0.26
MCH (pg)	30.6 [29.3, 31.7]	30.8 [29.6, 32.0]	<0.01
RDW-CV (%)	14.6 [13.5, 16.0]	15.2 [14.2, 16.6]	<0.01
RDW-SD (fL)	49.1 [45.1, 53.5]	51.6 [47.9, 55.9]	<0.01
PLT (×10^9/L)	187 [138, 241]	194 [148, 245]	<0.01
MPV (fL)	11.9 (1.2)	11.8 (1.2)	0.35
PDW (%)	15.3 (3.2)	15.1 (3.1)	0.08
P-LCR% (%)	39.9 (9.8)	39.5 (9.9)	0.27
WBC (×10^9/L)	6.97 [5.54, 9.26]	7.36 [5.74, 9.52]	<0.01
NEUT% (%)	72.0 [65.0, 78.9]	73.6 [67.3, 79.3]	<0.01
LYM% (%)	17.3 [11.6, 23.9]	15.0 [10.7, 20.9]	<0.01
MONO% (%)	7.0 [5.9, 8.3]	7.2 [6.2, 8.5]	<0.01
EO% (%)	1.4 [0.8, 2.3]	1.5 [0.9, 2.4]	0.04
BASO% (%)	0.4 [0.3, 0.5]	0.4 [0.3, 0.5]	0.58
NEUT# (×10^9/L)	4.81 [3.58, 6.95]	5.27 [3.87, 7.44]	<0.01
LYM# (×10^9/L)	1.17 [0.85, 1.53]	1.09 [0.83, 1.41]	<0.01
MONO# (×10^9/L)	0.48 [0.38, 0.62]	0.52 [0.41, 0.66]	<0.01
EO# (×10^9/L)	0.09 [0.05, 0.16]	0.10 [0.06, 0.16]	<0.01
BASO# (×10^9/L)	0.03 [0.02, 0.04]	0.03 [0.02, 0.04]	0.26
PT (s)	11.7 [10.9, 12.8]	11.6 [11.0, 12.6]	0.38

INR	1.04 [0.97, 1.14]	1.03 [0.97, 1.12]	0.06
APTT (s)	28.1 [26.1, 30.8]	28.0 [26.1, 30.8]	0.85
Fbg (g/L)	3.45 (1.24)	3.66 (1.08)	<0.01
TT (s)	18.4 [17.3, 19.5]	18.2 [17.1, 19.4]	0.03
CRP (mg/L)	14.8 [4.7, 42.4]	19.8 [6.8, 72.5]	<0.01

Continuous variables with normal distribution are presented as mean value (SD) while others are presented as median (IQR). Categorical variables are presented as frequency (percentage).

Interestingly, a greater number of laboratory indices were identified to show statistically significant distinctions in the LG cohort as opposed to the BS cohort, most likely due to the considerably larger sample size. In contrast, in the LG cohort, most laboratory indices demonstrated a notable decrease in diagnostic potency. The diagnostic efficacy of tumor markers was generally less satisfactory in the LG cohort (Figure. 3A). Among the top five indicators in terms of diagnostic efficacy, only CA 19-9 had an AUROC exceeding 0.7 (AUROC=0.712), followed by CEA (AUROC=0.685), Age (AUROC=0.617), RDW-SD (AUROC=0.616) and DBIL (AUROC=0.613) (Figure. 3B). The DeLong test indicated a significant decrease in diagnostic efficacy of indicators including CA 19-9 and DBIL in the LG cohort compared with the BS cohort (Figure. 3C). These results suggested that in a surgical cohort with a simpler diagnostic environment, common laboratory indices may have acceptable power as standalone diagnostic markers; but in a larger, more comprehensive patient cohort with a more complex diagnostic environment, individual serum markers are no longer robust diagnostic tools for obstructive jaundice. Therefore, we propose constructing a diagnostic model for obstructive jaundice that can integrate multiple test indicators.

We first attempted a traditional linear regression model to distinguish benign and malignant etiologies. The diagnostic model construction process, employing stepwise logistic regression, systematically incorporated a selection procedure involving the inclusion and exclusion of variables from an initial pool of 57 indicators (Figure. S4A). The diagnostic efficacy of the stepwise logistic regression model was found to be moderate, achieving AUROC values of 0.784 and 0.791 in the internal and external validation sets, respectively. However, when compared to the subsequently established ML model, it presented significantly lower AUROC values, along

with inadequate sensitivity and specificity (Figure. S4B & S4C). These results highlighted the inherent advantage of ML techniques in this particular task. Similarly, in subsequently established ML models, we observed that the number of features included in the model significantly influenced its diagnostic efficacy (Figure. 3D). These results indicate that constructing high-efficiency diagnostic models based on common laboratory markers is feasible, but it requires multiple parameters or features to jointly exert their utility. Therefore, in the final version of our ML models, all 57 features were included. These features delineate common dimensions of disease characteristics, including demographic features, tumor markers, complete blood count, comprehensive metabolic panel, clotting screen and inflammatory markers (Figure. 3E).

Establishment, validation and interpretation of diagnostic models to distinguish benign and malignant obstructions

After confirming the feasibility of constructing a ML diagnostic model for obstructive jaundice based on 57 common clinical features, we proceeded to optimize the diagnostic performance of this binary diagnostic model. A series of mainstream ML methods were employed to construct the model, with their performance compared to select the optimal model. The RF model ultimately stood out for its remarkable performance both in the internal and external validation sets (Figure. 4A, Figure. 4B and Table. 2).

Table. 2. Comparison of performance among various ML methods employed for construction of the binary MOLT model.

Cohort	Model	Index					
		AUROC (95%CI)	ACC	AUPR	F1 score	Sensitivity	Specificity
LG cohort (training)	Random Forest	1.000 (1.000-1.000)	0.996	1.000	0.997	0.987	1.000
	LightGBM	0.999 (0.999-1.000)	0.983	0.999	0.988	0.956	0.996
	XGBoost	0.891 (0.878-0.904)	0.836	0.842	0.888	0.573	0.961
	Decision Tree	0.836 (0.820-0.853)	0.825	0.744	0.874	0.681	0.894
	SVM	0.920 (0.907-0.932)	0.872	0.892	0.911	0.680	0.963
	Logistic Regression	0.811 (0.793-0.829)	0.787	0.707	0.855	0.501	0.923
	KNN	0.992 (0.990-0.995)	0.907	0.982	0.935	0.731	0.990
LG cohort (internal validation)	Random Forest	0.875 (0.855-0.896)	0.824	0.816	0.878	0.566	0.955
	LightGBM	0.907 (0.891-0.923)	0.842	0.851	0.887	0.650	0.938
	XGBoost	0.852 (0.832-0.873)	0.798	0.774	0.862	0.504	0.947
	Decision Tree	0.826 (0.803-0.849)	0.791	0.748	0.848	0.613	0.881
	SVM	0.817 (0.793-0.842)	0.772	0.719	0.841	0.500	0.910
	Logistic Regression	0.781 (0.754-0.809)	0.757	0.672	0.835	0.431	0.923
	KNN	0.713 (0.684-0.743)	0.705	0.582	0.799	0.354	0.883
BS cohort (external validation)	Random Forest	0.862 (0.819-0.904)	0.828	0.865	0.864	0.676	0.936
	LightGBM	0.822 (0.774-0.870)	0.802	0.827	0.841	0.669	0.897
	XGBoost	0.725 (0.667-0.782)	0.699	0.709	0.768	0.483	0.853
	Decision Tree	0.678 (0.615-0.741)	0.731	0.689	0.783	0.586	0.833
	SVM	0.856 (0.812-0.900)	0.808	0.844	0.828	0.834	0.789
	Logistic Regression	0.800 (0.754-0.846)	0.722	0.734	0.771	0.614	0.799
	KNN	0.816 (0.770-0.862)	0.751	0.781	0.806	0.559	0.887

Cells labeled red/yellow represent the first/second-best performance in each cohort.

Of note, the lightGBM model and the SVM model also exhibited impressive performance. DeLong's test between ROC curves was conducted to finalize our choice (Figure. S5). Although the lightGBM model excelled in internal prediction (AUROC 0.907 vs. 0.875, DeLong's test $p<0.01$), the RF model demonstrated greater stability in the external validation cohort (AUROC 0.862 vs. 0.822, DeLong's test $p=0.03$). This robust performance across different diagnostic contexts is essential for ensuring the model's effectiveness. Subsequently, the RF model was selected and designated as the MOLT model, standing for **M**achine learning of **O**bstructive jaundice based on common **L**aboratory **T**ests.

To present the decision-making process of the MOLT model in a more transparent manner, we employed several methods for model interpretation. Feature importance scores highlighted the top-ranking features distinguishing benign from malignant etiologies. The top 10 features were identified as age, CA 19-9, CEA, CHOL, ALB, DBIL, A/G, RDW-SD, Fbg and AST (Figure. 4D and Figure. S6A). While feature importance scores provide a global view of feature importance across the dataset, SHAP values offer a more nuanced understanding of how each feature influences individual predictions, taking interactions and dependencies between features into account [14, 15]. Top 10 features ranked by SHAP value were CA 19-9, CEA, age, ALB, PLT, CHOL, RDW-CV, RDW-SD, Fbg and DBIL (Figure. 4E and Figure. S6B). By analyzing the overall interaction strength, we also provided insights into the complexity of relationships between predictor variables in the MOLT model (Figure. 4F). Partial Dependence Plots (PDP) and Individual Conditional Expectation (ICE) Plots were generated for the top three features identified by SHAP values (age, CA 19-9 and CEA), illustrating how individual feature values impact the model's predictions (Figure. 4G). With the SHAP-interpreted ML model,

individualized decision-making processes were elucidated, allowing for a comprehensive understanding of how the model arrives at predictions for each specific case (Figure. 4H). Decision boundaries pertaining to key features were also visualized (Figure. S7).

Establishment, validation and interpretation of multi-class models for further classification of obstructive jaundice

In clinical practice, merely obtaining information about the benign or malignant nature of obstructive jaundice is insufficient to support subsequent treatment options. In benign diseases, clinicians wish to differentiate patients with calculous disease from those with non-calculous disease, while in malignant cases, it is important for the specific degree of tumor malignancy to be assessed. Therefore, building upon the initial model, we converted the binary classification target into a multi-classification target for the construction of a more complex model, namely the multi-class MOLT model, allowing the differentiation between calculous benign lesions, non-calculous benign lesions, metastatic malignancies, pancreato-biliary malignancies and other types of malignancies (Figure. 5A). Similar to the original binary MOLT model, a series of ML methods were employed, with the best-performing one selected to optimize model performance. The outcomes of model construction and internal validation were summarized (Figure. 5B). Notably, external validation was unable to be carried out for the multi-class MOLT model as there were no metastatic patient in the BS cohort. Outperforming the others, the XGBoost model showcased impressive diagnostic efficiency, boasting an ACC of 0.777 and an AUNU of 0.882, a notable achievement given the complexity of the task encompassing five classes.

ML interpretability tools were also utilized to explain the multi-class MOLT model. Feature importance scores highlighted the top-ranking features contributing to this model, namely CA 19-9, age, CEA, ALB, CHOL, AFP, UA, A/G, RBC and CA125 (Figure. 5C). SHAP values were utilized to assess how individual features influenced the model's decisions within each disease category (Figure. 5D).ALB, A/G, CA 19-9, CEA and AST were top 5 features contributing to the diagnosis of benign calculous disease; CA19-9, PDW, ALB, CEA and P-LCR% were top 5 features contributing to the diagnosis of benign non-calculous disease; CEA, age, TBIL, CA 19-9 and RBC were top 5 features contributing to the diagnosis of metastatic malignancies; ALB, CA 19-9, CEA, A/G and RBC were top 5 features contributing to the diagnosis of pancreato-biliary malignancies; while AFP, PLT, ALB, A/G and HDL were top 5 features contributing to the diagnosis of other malignancies. These results offer valuable insights into obstructive jaundice and provide a practical diagnostic tool.

Discussion

Until now, large-scale cohorts regarding obstructive jaundice remain scarce, which has led to a limited understanding of the proportions of specific types of diseases contributing to this condition, as well as limited insight into the efficacy of existing diagnostic approaches. Our study included the largest retrospective cohort of obstructive jaundice to-date, in order to delineate the spectrum of diseases associated with this condition. During the analysis of the patients' baseline, we discovered the untapped diagnostic potential in various clinical diagnostic tests. Consequently, we developed ML models based on common clinical laboratory tests, which

not only distinguishes between benign and malignant obstructions, but also further differentiates between calculous benign lesions, non-calculous benign lesions, metastatic malignancies, pancreato-biliary malignancies and other types of malignancies. To ensure transparency in the decision-making process, interpretable ML tools were utilized to decipher these models.

Undoubtedly, the spectrum of diseases contributing to obstructive jaundice may vary across different countries and regions. However, the dearth of knowledge in this field has restricted our understanding of this condition. Our study has emerged as a significant contribution to this field, offering a comprehensive analysis of over 5000 patients with obstructive jaundice in a single Chinese center over a period of 14 years. Smaller retrospective cohort studies conducted across Europe, Australia, Central Asia and South Asia have provided valuable insights into obstructive jaundice [16-22]. Notably, Garcea et al.'s retrospective analysis of over 1000 cases in the United Kingdom found similar disease patterns to ours, with CBD stones and pancreatic ductal adenocarcinoma as primary benign and malignant etiologies, respectively [19]. Similarly, Björnsson et al.'s analysis of 241 patients in Sweden revealed a slightly higher incidence of malignant obstruction (63.9%) compared to benign cases, with cholangiocarcinoma accounting for one-third of malignant obstructions, mirroring our findings [16]. These findings suggest that the disease spectrum of obstructive jaundice may be more consistent across different regions than previously believed. Moreover, our study supplemented these insights by revealing additional dimensions of this condition. Firstly, we observed that non-calculous benign etiologies might have been underreported, as most of the previous studies identified CBD stones as the predominant benign cause. Our findings indicated that CBD stones only accounted for approximately half of the benign cases, while around one-third were associated with diverse non-

calculous factors. In addition, there may have been an underestimation of metastatic causes of obstructive jaundice, considering the lower likelihood of obtaining a pathological diagnosis in these patients. Furthermore, intrahepatic lesions involving the hepatic hilus, mainly intrahepatic cholangiocarcinoma (iCCA) and hepatocellular carcinoma (HCC), constitute a significant proportion (approximately 10%) of all cases.

Regarding the diagnosis of obstructive jaundice, our study also yielded valuable insights. We not only presented a practical diagnostic tool for distinguishing between different causes, but also enhanced comprehension to obstructive jaundice by providing transparency into the decision-making process. According to previous studies, calculous diseases in benign obstructive jaundice can be accurately distinguished through various imaging modalities [23-25]. However, distinguishing non-calculous benign diseases from malignant diseases can be considerably challenging [26-29]. As our study revealed that over one-third of the benign cases were associated with non-calculous etiologies, there is a need to place greater emphasis on addressing this issue. Meanwhile, in the realm of malignancy, clinicians seek to stratify cancers according to their level of aggressiveness. To this end, we employed ML techniques to develop two models: one effectively distinguishes between benign and malignant causes, while the other offers nuanced insights by further classifying malignancies into three tiers and benign diseases into two. Subsequently, these models may facilitate the application of appropriate diagnostic and therapeutic interventions.

It is noteworthy that the concept of integrating diverse laboratory test results into a unified model did not arise arbitrarily. There has long been ample evidence pointing towards this direction. For instance, multiple studies have observed that patients with malignant obstruction tends to be

older than the benign group [16, 20]. Similarly, benign obstructive jaundice is observed to be associated with lower bilirubin levels, as biliary obstructions caused by calculous disease tend to be intermittent [1, 19]. Furthermore, there is documented evidence suggesting an association between obstructive jaundice and renal injury, with the severity of renal dysfunction potentially reflecting the nature of the disease [30-32]. In this study, by leveraging interpretable ML models, we gained further insight about how these previous findings contribute to the diagnostic model. While traditional markers such as CA 19-9, CEA, age and bilirubin levels remained significant, factors like albumin levels, cholesterol levels and red cell distribution width emerged as noteworthy contributors. These findings warrant further investigation to provide physiological and pathological evidence elucidating their mechanisms.

Several limitations of this study should be noted. Firstly, as this study exclusively enrolled patients with a confirmed pathological diagnosis, the findings regarding the proportions of specific types of diseases contributing to obstructive jaundice may be subject to bias due to variations in the likelihood of different diseases to be biopsied. Secondly, this is a single-center study, primarily involving patients from China. As a result, the potential impact of geographic variations among patients and differences in detection methods across various clinical laboratories on the study outcomes was not addressed, even though we observed similarities in our findings with previous researches from other regions. Furthermore, our MOLT model exhibited exceptional specificity; however, its sensitivity fell short of expectations. This underscores the need for future studies to improve both the specificity and sensitivity of diagnostics. Despite these limitations, our study stands as the largest cohort study in the field of obstructive jaundice to date, with robust diagnostic tools developed through the utilization of

state-of-the-art techniques.

To conclude, our study delineated the disease spectrum of obstructive jaundice and developed a series of ML models, collectively termed the MOLT model, to differentiate between benign and malignant obstructions and further categorize diseases into five distinct categories. These models underwent meticulous interpretation, providing transparency into the decision-making process. These findings illuminate the diagnostic potential of routine laboratory tests in obstructive jaundice, highlighting the role of ML models in enhancing diagnostic accuracy in complex clinical conditions like obstructive jaundice. These insights may facilitate personalized and user-friendly diagnosis of obstructive jaundice, thereby aiding clinical decision-making.

Abbreviations

ML, machine learning; ERCP, endoscopic retrograde cholangiopancreatography; PTCD, percutaneous transhepatic cholangiography drainage; iCCA, intrahepatic cholangiocarcinoma; pCCA, perihilar cholangiocarcinoma; dCCA, distal cholangiocarcinoma; AFP, α-fetoprotein; CEA, carcinoembryonic antigen; CA 125, cancer antigen 125; CA 19-9, cancer antigen 19-9; RBC, red blood cell count; HGB, hemoglobin; HCT, hematocrit; MCV, mean corpuscular volume; MCHC, mean corpuscular hemoglobin concentration; MCH, mean corpuscular hemoglobin; RDW-CV, red cell distribution width-coefficient of variation; RDW-SD, red cell distribution width-standard deviation; PLT, platelet count; MPV, mean platelet volume; PDW, platelet distribution width; P-LCR%, large platelet ratio; WBC, white blood cell count; NEUT%, neutrophil percentage; LYM%, lymphocyte percentage; MONO%, monocyte percentage; EO%, eosinophil percentage; BASO%, basophil percentage; NEUT#, neutrophil count; LYM#,

lymphocyte count; MONO#, monocyte count; EO#, eosinophil count; BASO#, basophil count; TBIL, total bilirubin; DBIL, direct bilirubin; IBIL, indirect bilirubin; ALT, alanine aminotransferase; AST, aspartate aminotransferase; ALP, alkaline phosphatase; γ-GT, γ-glutamyl transferase; ALB, albumin; Glo, globulin; A/G, albumin/globulin ratio; GLU, glucose; UREA, urea; CREA, creatinine; CysC, cystatin C; UA, uric acid; TG, triglycerides; CHOL, cholesterol; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; CK, creatine kinase; LDH, lactate dehydrogenase; HBDH, hydroxybutyrate dehydrogenase; PT, prothrombin time; INR, international normalized ratio; APTT, activated partial thromboplastin time; Fbg, fibrinogen; TT, thrombin time; CRP, C-reactive protein.

Data sharing

We have made the source code and datasets used in this study publicly available on GitHub. The project, titled "Machine Learning of Obstructive Jaundice based on Common Laboratory Tests (the MOLT model)", can be accessed at https://github.com/re5yho/Machine-learning-of-Obstructive-jaundice-based-on-common-Laboratory-Tests-the-MOLT-model-.git. Researchers interested in replicating or extending our work are encouraged to explore the repository.

Acknowledgements

We would like to express our gratitude to Professor Jingxin Zhang from Harbin University of Commerce for imparting ML methods based on the mlr3 framework in R. We would also like to

express our appreciation to Smart Server Medical Art (https://smart.servier.com/) and Scidraw (https://scidraw.io/) for providing free medical illustrations.

References

[1] Jarnagin WR. Blumgart's Surgery of the Liver, Biliary Tract and Pancreas, 2-Volume Set: Elsevier Health Sciences; 2022.

[2] Pereira SP, Goodchild G, Webster GJM. The endoscopist and malignant and non-malignant biliary obstruction. Biochim Biophys Acta Mol Basis Dis 2018;1864:1478-1483.

[3] Kapoor BS, Mauri G, Lorenz JM. Management of Biliary Strictures: State-of-the-Art Review. Radiology 2018;289:590-603.

[4] Fry DE. Obstructive jaundice. Causes and surgical interventions. Postgrad Med 1988;84:217-222, 227, 230.

[5] Okamoto T. Malignant biliary obstruction due to metastatic non-hepato-pancreato-biliary cancer. World J Gastroenterol 2022;28:985-1008.

[6] Lu J, Li B, Li FY, Ye H, Xiong XZ, Cheng NS. Long-term outcome and prognostic factors of intrahepatic cholangiocarcinoma involving the hepatic hilus versus hilar cholangiocarcinoma after curative-intent resection: Should they be recognized as perihilar cholangiocarcinoma or differentiated? Eur J Surg Oncol 2019;45:2173-2179.

[7] She YM, Ge N. The value of endoscopic ultrasonography for differential diagnosis in obstructive jaundice of the distal common bile duct. Expert Rev Gastroenterol Hepatol 2022;16:653-664.

[8] Heilmaier C, Lutz AM, Bolog N, Weishaupt D, Seifert B, Willmann JK. Focal liver lesions:

detection and characterization at double-contrast liver MR Imaging with ferucarbotran and gadobutrol versus single-contrast liver MR imaging. Radiology 2009;253:724-733.

[9] Lv WJ, Zhao XY, Hu DD, Xin XH, Qin LL, Hu CH. Insight into Bile Duct Reaction to Obstruction from a Three-dimensional Perspective Using ex Vivo Phase-Contrast CT. Radiology 2021;299:597-610.

[10] Marrelli D, Caruso S, Pedrazzani C, Neri A, Fernandes E, Marini M, et al. CA19-9 serum levels in obstructive jaundice: clinical value in benign and malignant conditions. Am J Surg 2009;198:333-339.

[11] Joo I, Lee JM, Yoon JH. Imaging Diagnosis of Intrahepatic and Perihilar Cholangiocarcinoma: Recent Advances and Challenges. Radiology 2018;288:7-13.

[12] 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;186:1772-1791.

[13] Haug CJ, Drazen JM. Artificial Intelligence and Machine Learning in Clinical Medicine, 2023. N Engl J Med 2023;388:1201-1208.

[14] Marcílio WE, Eler DM. From explanations to feature selection: assessing SHAP values as feature selection mechanism. 2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI); 2020 7-10 Nov. 2020; 2020. p. 340-347.

[15] Molnar C, Casalicchio G, Bischl B. iml: An R package for interpretable machine learning. Journal of Open Source Software 2018;3:786.

[16] Björnsson E, Gustafsson J, Borkman J, Kilander A. Fate of patients with obstructive jaundice. Journal of hospital medicine 2008;3:117-123.

[17] Siddique K, Ali Q, Mirza S, Jamil A, Ehsan A, Latif S, et al. Evaluation of the aetiological

spectrum of obstructive jaundice. J Ayub Med Coll Abbottabad 2008;20:62-66.

[18] Patel VB, Musa RK, Patel N, Patel SD. Role of MRCP to determine the etiological spectrum, level and degree of biliary obstruction in obstructive jaundice. Journal of family medicine and primary care 2022;11:3436-3441.

[19] Garcea G, Ngu W, Neal CP, Dennison AR, Berry DP. Bilirubin levels predict malignancy in patients with obstructive jaundice. HPB (Oxford) 2011;13:426-430.

[20] Chalya PL, Kanumba ES, McHembe M. Etiological spectrum and treatment outcome of Obstructive jaundice at a University teaching Hospital in northwestern Tanzania: A diagnostic and therapeutic challenges. BMC Res Notes 2011;4:147.

[21] Little JM, Cunningham P. Obstructive jaundice in a referral unit: surgical practice and risk factors. Aust N Z J Surg 1985;55:427-432.

[22] Malchow-Møller A, Matzen P, Bjerregaard B, Hilden J, Holst-Christensen J, Staehr Johansen T, et al. Causes and characteristics of 500 consecutive cases of jaundice. Scand J Gastroenterol 1981;16:1-6.

[23] Soto JA, Alvarez O, Lopera JE, Múnera F, Restrepo JC, Correa G. Biliary obstruction: findings at MR cholangiography and cross-sectional MR imaging. Radiographics 2000;20:353-366.

[24] Katabathina VS, Dasyam AK, Dasyam N, Hosseinzadeh K. Adult bile duct strictures: role of MR imaging and MR cholangiopancreatography in characterization. Radiographics 2014;34:565-586.

[25] Sharma M, Pathak A, Sharma Y. Endoscopic ultrasound in CBD stone. Gastroenterology 2009;137:e7-8.

[26] Lanzillotta M, Mancuso G, Della-Torre E. Advances in the diagnosis and management of IgG4 related disease. Bmj 2020;369:m1067.

[27] Okazaki K, Uchida K, Koyabu M, Miyoshi H, Ikeura T, Takaoka M. IgG4 cholangiopathy: current concept, diagnosis, and pathogenesis. J Hepatol 2014;61:690-695.

[28] Dyson JK, Beuers U, Jones DEJ, Lohse AW, Hudson M. Primary sclerosing cholangitis. Lancet (London, England) 2018;391:2547-2559.

[29] Lohse AW, Mieli-Vergani G. Autoimmune hepatitis. J Hepatol 2011;55:171-182.

[30] Liu J, Qu J, Chen H, Ge P, Jiang Y, Xu C, et al. The pathogenesis of renal injury in obstructive jaundice: A review of underlying mechanisms, inducible agents and therapeutic strategies. Pharmacological research 2021;163:105311.

[31] Martínez-Cecilia D, Reyes-Díaz M, Ruiz-Rabelo J, Gomez-Alvarez M, Villanueva CM, Álamo J, et al. Oxidative stress influence on renal dysfunction in patients with obstructive jaundice: A case and control prospective study. Redox biology 2016;8:160-164.

[32] Green J, Better OS. Systemic hypotension and renal failure in obstructive jaundice-mechanistic and therapeutic aspects. Journal of the American Society of Nephrology : JASN 1995;5:1853-1871.

Figure legends

Figure. 1. Flowchart illustrating the patient selection process from an initial pool of 20,545 patients to establish the LG cohort.

Figure. 2. Summary chart depicting the disease spectrum of obstructive jaundice observed in the LG cohort, highlighting the prevalence of various benign and malignant etiologies.

Figure. 3. The necessity and feasiblility of contructing a ML model for the clinical diagnosis of obstructive jaundice. Traditional serum makers, including **(A)** tumor makers and **(B)** top 5 biomarkers ranked by AUROC, were found to exhibit suboptimal diagnostic efficacy in the LG cohort. **(C)** In the LG cohort, the diagnostic scenario becomes more intricate, causing markers like CA 19-9 and DBIL to exhibit diminished efficacy compared to the BS cohort, necessitating the development of combined ML models. **(D)** The number of features included in the ML model significantly influenced its diagnostic efficacy, as ML model with 57 features outperformed the others (DeLong's test $p<0.01$). **(E)** The included features encompass a wide spectrum of clinical characteristics including demographic features, tumor markers, complete blood count, comprehensive metabolic panel, clotting screen and inflammatory markers.

Figure. 4. The establishment, validation and interpretation of binary MOLT model. (A) In the internal validation set, the lightGBM model showcased best performance measured by AUROC, while **(B)** the RF model showcased best performance in the external validation set. **(C)**

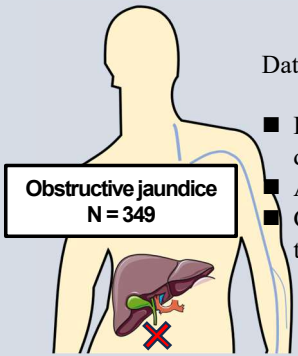
The RF model was subsequently designated as the MOLT model, which exhibited better performance compared with traditional CA 19-9 in the external validation set. Interpretation of the MOLT model revealed **(D)** features with top-ranked feature importance score, **(E)** features with top-ranked SHAP values and **(F)** features with top-ranked interaction score. **(G)** PDP and ICE plots were created for the top three features (age, CA 19-9, and CEA) identified by SHAP values, elucidating their impact on the model's predictions. **(H)** The SHAP-interpreted ML model clarified individualized decision-making processes, offering understanding into prediction rationale for each case.

Figure. 5. The establishment, validation and interpretation of multi-class MOLT model. (A) Extending the binary MOLT model, a five-class ML task was formulated. **(B)** The performance of diverse multi-class models was gauged using metrics like ACC, AUNU, macro F1 score, precision and recall scores. Similarly, the decision-making process of the multi-class MOLT model was elucidated with **(C)** feature importance score and **(D)** SHAP value.

Biliary surgery cohort

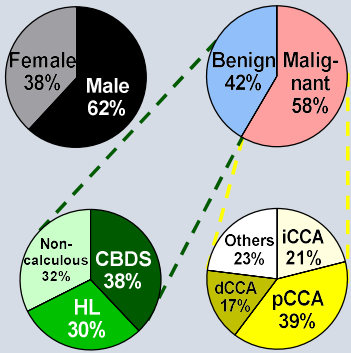
Pilot exploration & External validation

Patients from a surgical center



Data acquired:

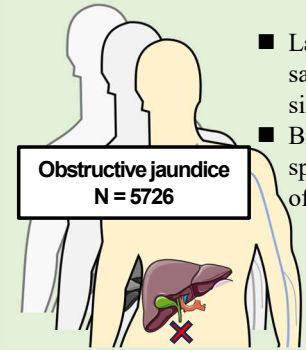
- Pathological diagnosis
- Age, sex
- Common lab tests



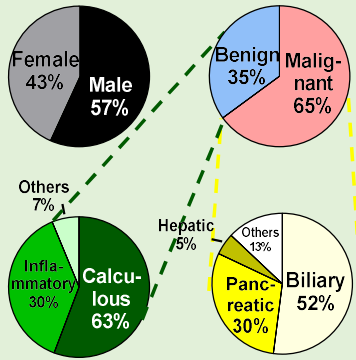
Large general cohort

Comprehensive analysis & Model construction

Patients from a general hospital



- Larger sample size
- Broader spectrum of diseases

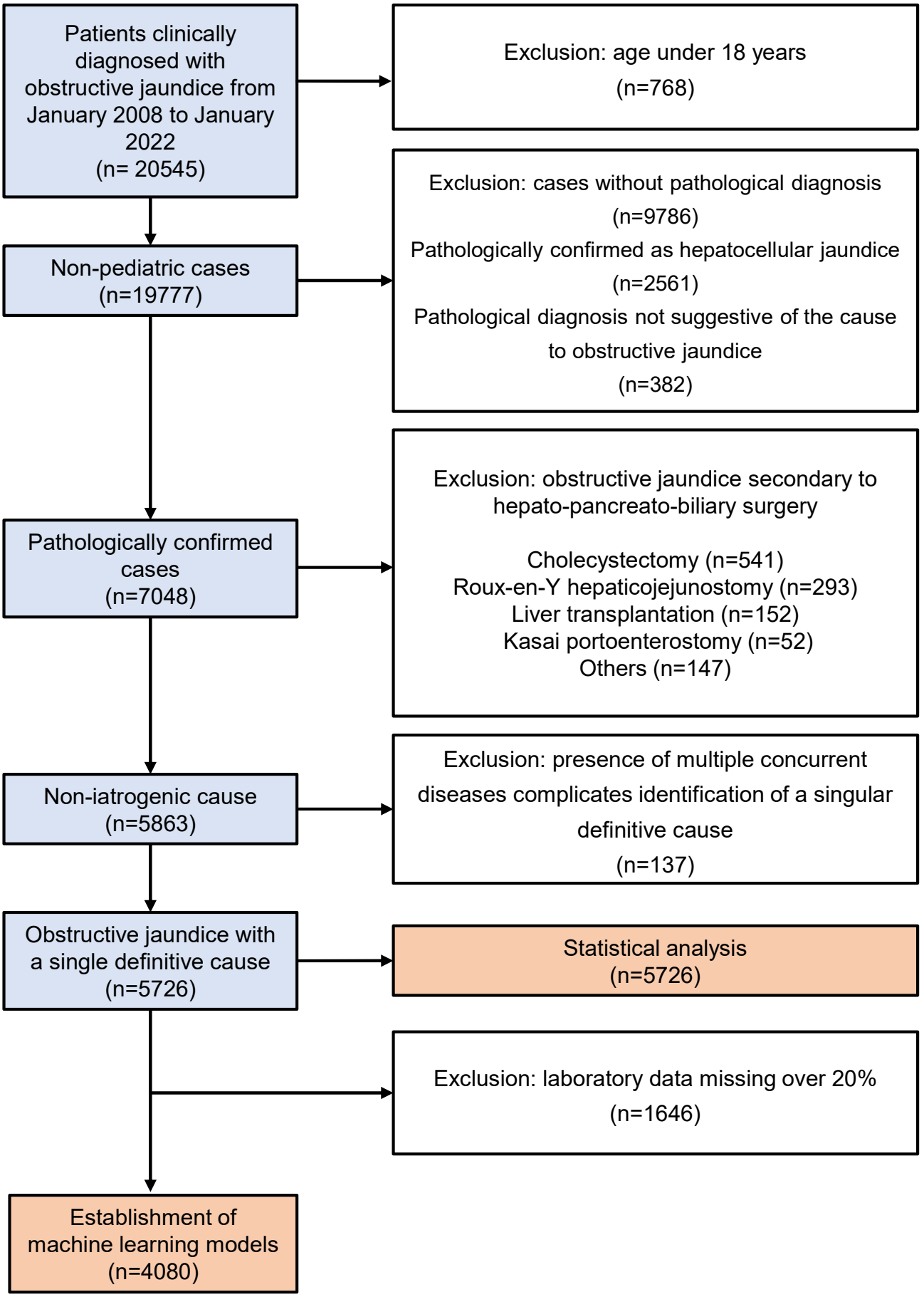


Statistical analysis

- Comprehensive analysis of causes to obstructive jaundice
- Impaired diagnostic accuracy of traditional serum markers in large general cohort

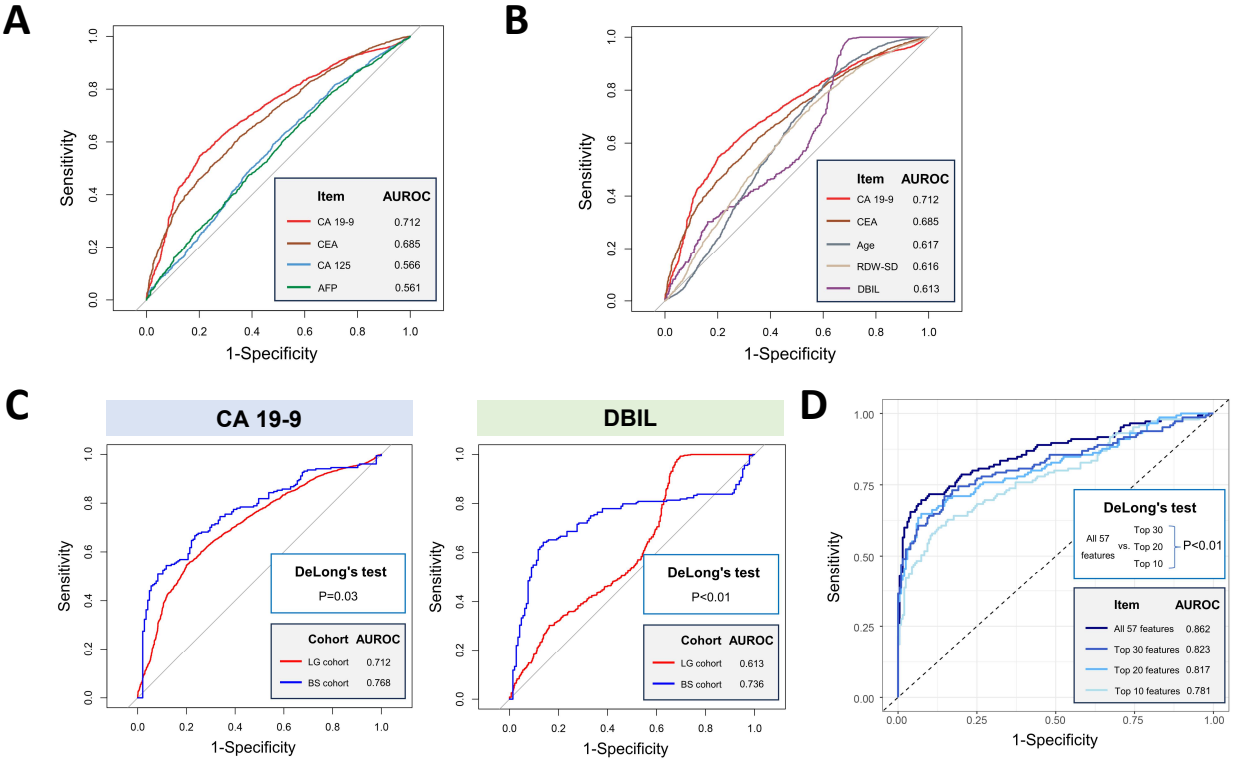
Machine learning models

- Diagnosis of benign/malignant obstructive jaundice (external validation AUROC = 0.862)
- Multiclass model for further classification (internal validation ACC = 0.777)
- Clinical interpretation of the models


Patients clinically diagnosed with obstructive jaundice from January 2008 to January 2022
(n= 20545)
Exclusion: age under 18 years
(n=768)
Non-pediatric cases
(n=19777)
Exclusion: cases without pathological diagnosis
(n=9786)
Pathologically confirmed as hepatocellular jaundice
(n=2561)
Pathological diagnosis not suggestive of the cause to obstructive jaundice
(n=382)
Pathologically confirmed cases
(n=7048)
Exclusion: obstructive jaundice secondary to hepato-pancreato-biliary surgery
Cholecystectomy (n=541)
Roux-en-Y hepaticojejunostomy (n=293)
Liver transplantation (n=152)
Kasai portoenterostomy (n=52)
Others (n=147)
Non-iatrogenic cause
(n=5863)
Exclusion: presence of multiple concurrent diseases complicates identification of a singular definitive cause
(n=137)
Obstructive jaundice with a single definitive cause
(n=5726)
Statistical analysis
(n=5726)
Exclusion: laboratory data missing over 20%
(n=1646)
Establishment of machine learning models
(n=4080)

Category	Number of Cases (proportion)	Type of Disease		Number of Cases (proportion)	Disease	Number of Cases (proportion)
Malignant obstruction	3720 (64.97%)	Biliary		1657 (28.94%)	Intrahepatic cholangiocarcinoma	320 (5.59%)
					Perihilar cholangiocarcinoma	573 (10.01%)
					Distal cholangiocarcinoma*	595 (10.39%)
					Gallbladder carcinoma	169 (2.95%)
		Pancreatic		1106 (19.32%)	Pancreatic adenocarcinoma	1094 (19.11%)
					Others	12 (0.21%)
		Ampullary		252 (4.40%)	Ampullary carcinoma*	252 (4.40%)
		Hepatic		190 (3.32%)	Hepatocellular carcinoma	178 (3.11%)
					Others	12 (0.21%)
		Metastatic		360 (6.29%)	Colorectal cancer metastasis	163 (2.85%)
					Gastric cancer metastasis	121 (2.11%)
					Lung cancer metastasis	26 (0.45%)
					Breast cancer metastasis	17 (0.30%)
					Other origins	33 (0.58%)
		Others		155 (2.71%)		
Borderline disease	12 (0.21%)					
Benign obstruction	1994 (34.82%)	Calculus		1257(21.95%)	Common bile duct stone	1046 (18.27%)
					Intrahepatic stone	137 (2.39%)
					Pancreatic duct stone**	37 (0.65%)
					Gallstone (Mirizzi syndrome)	37 (0.65%)
		Inflammatory		566 (9.88%)	Acute pancreatitis**	328 (5.73%)
					Chronic pancreatitis**	44 (0.75%)
					IgG4-related disese	74 (1.29%)
					Primary sclerosing cholangitis	48 (0.84%)
					Autoimmune hepatitis	59 (1.03%)
					Biliary tree infection (non-calculus)	13 (0.23%)
		Others		171 (2.99%)		

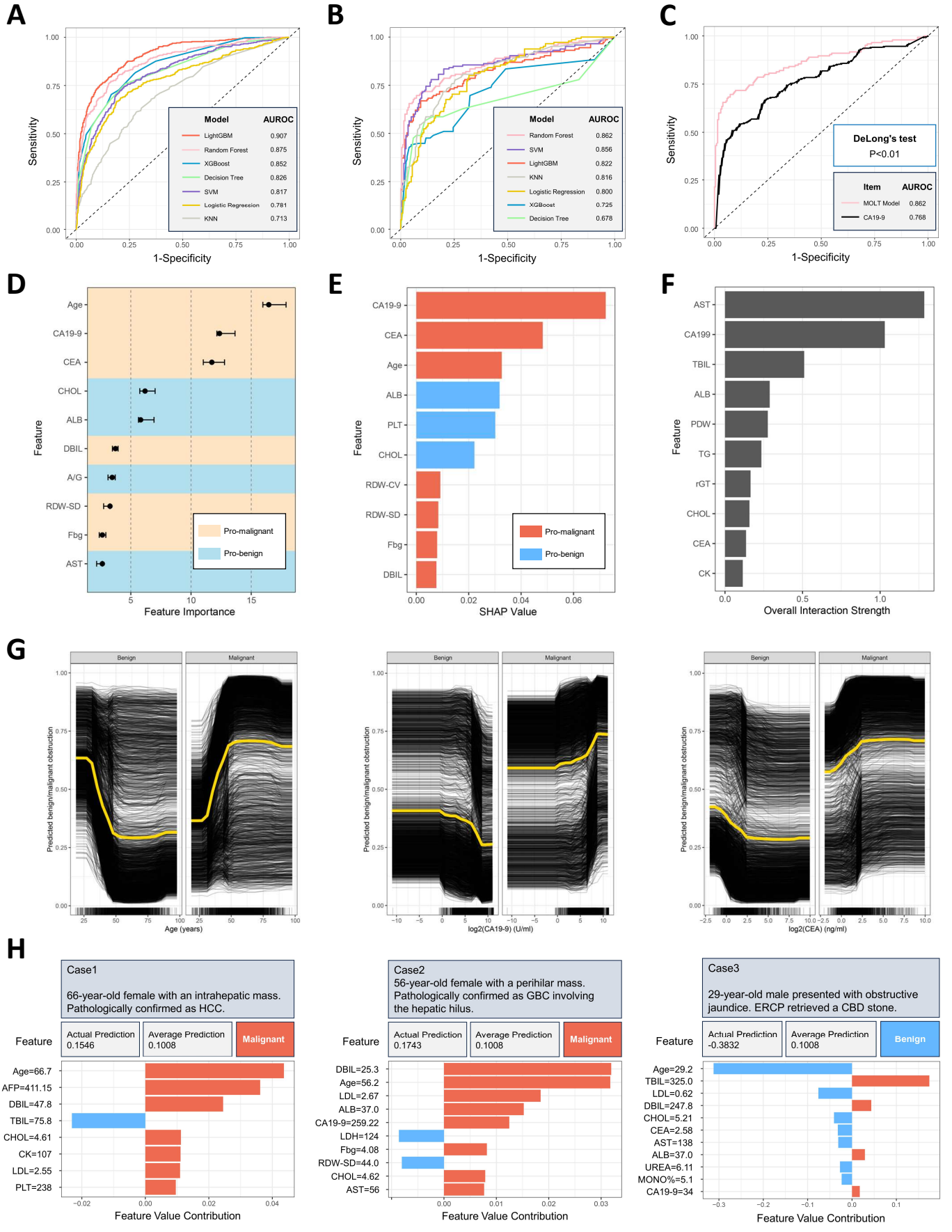
* Distal cholangiocarcinoma and ampullary carcinoma can be hard to completely differentiate due to their anatomical proximity. Pathological reports were referred to for the most likely site of origin.
**A considerable proportion of pancreatitis are caused by calculus disease. In this study, patients were categorized into “calculous” only when there was definite evidence of stones.



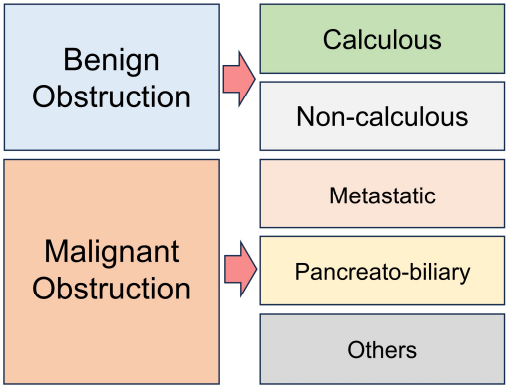
E

Demographic features	Tumor markers	Complete blood count
Age, Sex	CA 19-9, CEA CA 125, AFP	RBC, HGB, HCT, MCV, MCHC, MCH, RDW-CV, RDW-SD, PLT, MPV, PDW, P-LCR%, WBC, NEUT%, LYM%, MONO%, EO%, BASO%, NEUT#, LYM#, MONO#, EO#, BASO#

Comprehensive metabolic panel	Clotting screen	Inflammatory marker
TBIL, DBIL, IBIL, ALT, AST, ALP, γ-GT, ALB, Glo, A/G, GLU, UREA, CREA, CysC, UA, TG, CHOL, HDL, LDL, CK, LDH, HBDH	PT, INR, APTT, Fbg, TT	CRP


A
Sensitivity
1-Specificity
Model AUROC
LightGBM 0.907
Random Forest 0.875
XGBoost 0.852
Decision Tree 0.826
SVM 0.817
Logistic Regression 0.781
KNN 0.713
B
Sensitivity
1-Specificity
Model AUROC
Random Forest 0.862
SVM 0.856
LightGBM 0.822
KNN 0.816
Logistic Regression 0.800
XGBoost 0.725
Decision Tree 0.678
C
Sensitivity
1-Specificity
DeLong's test
P<0.01
Item AUROC
MOLT Model 0.862
CA19-9 0.768
D
Feature
Age
CA19-9
CEA
CHOL
ALB
DBIL
A/G
RDW-SD
Fbg
AST
Pro-malignant
Pro-benign
Feature Importance
E
Feature
CA19-9
CEA
Age
ALB
PLT
CHOL
RDW-CV
RDW-SD
Fbg
DBIL
Pro-malignant
Pro-benign
SHAP Value
F
Feature
AST
CA199
TBIL
ALB
PDW
TG
rGT
CHOL
CEA
CK
Overall Interaction Strength
G
Benign
Malignant
Predicted benign/malignant obstruction
Age (years)
log2(CA19-9) (U/ml)
log2(CEA) (ng/ml)
H
Case1
66-year-old female with an intrahepatic mass. Pathologically confirmed as HCC.
Feature
Actual Prediction 0.1546
Average Prediction 0.1008
Malignant
Age=66.7
AFP=411.15
DBIL=47.8
TBIL=75.8
CHOL=4.61
CK=107
LDL=2.55
PLT=238
Feature Value Contribution
Case2
56-year-old female with a perihilar mass. Pathologically confirmed as GBC involving the hepatic hilus.
Feature
Actual Prediction 0.1743
Average Prediction 0.1008
Malignant
DBIL=25.3
Age=56.2
LDL=2.67
ALB=37.0
CA19-9=259.22
LDH=124
Fbg=4.08
RDW-SD=44.0
CHOL=4.62
AST=56
Feature Value Contribution
Case3
29-year-old male presented with obstructive jaundice. ERCP retrieved a CBD stone.
Feature
Actual Prediction -0.3832
Average Prediction 0.1008
Benign
Age=29.2
TBIL=325.0
LDL=0.62
DBIL=247.8
CHOL=5.21
CEA=2.58
AST=138
ALB=37.0
UREA=6.11
MONO%=5.1
CA19-9=34
Feature Value Contribution

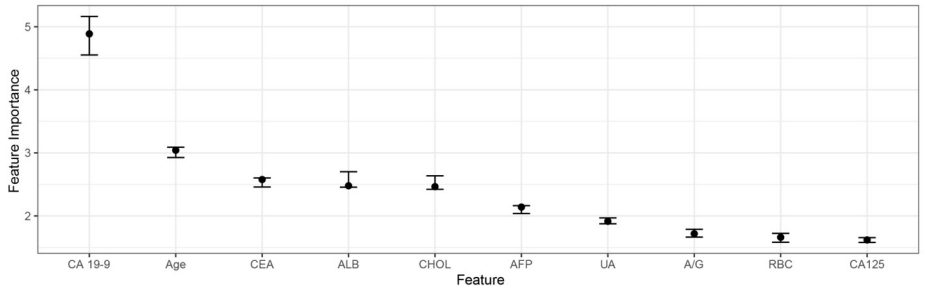
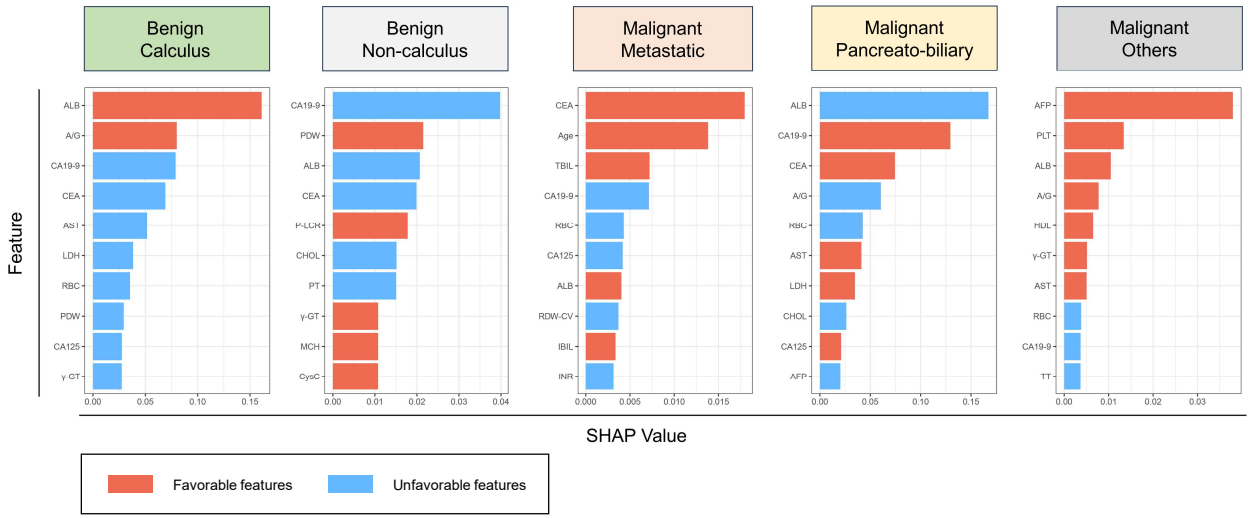
A



B

Cohort	Model	Index				
		ACC	AUROC (AUNU)	Macro F1 Score	Precision Score	Recall Score
Cohort LG (training)	XGBoost	1.000	1.000	1.000	1.000	1.000
	Random Forest	0.974	1.000	0.961	0.991	0.935
	LightGBM	0.928	0.994	0.912	0.968	0.869
	KNN	0.869	0.995	0.819	0.943	0.743
	SVM	0.745	0.897	0.562	0.857	0.493
	Decision Tree	0.658	0.724	-	0.386	0.348
Cohort LG (internal validation)	XGBoost	0.777	0.882	0.637	0.736	0.598
	Random Forest	0.758	0.889	0.572	0.814	0.531
	LightGBM	0.747	0.850	0.473	0.555	0.442
	KNN	0.664	0.759	0.573	0.719	0.532
	SVM	0.626	0.787	0.378	0.697	0.357
	Decision Tree	0.615	0.695	-	0.334	0.330

C



D