

# Predicting Intrahepatic Cholestasis of Pregnancy: A Retrospective Cohort Study of a Comprehensive Clinical Prediction Model

Huan Liang<sup>1</sup>, Ye Tian<sup>2</sup>, Jie Gao<sup>3</sup>, Jing Teng<sup>4</sup>

<sup>1</sup>Department of Gynecology, The Central Hospital of Enshi Tujia and Miao Autonomous Prefecture, Enshi, Hubei, 445000, People's Republic of China; <sup>2</sup>School of Medicine, Hubei Minzu University, Enshi, Hubei, 445000, People's Republic of China; <sup>3</sup>Department of Ultrasonic Imaging, The Central Hospital of Enshi Tujia and Miao Autonomous Prefecture, Enshi, Hubei, 445000, People's Republic of China; <sup>4</sup>Department of Gynecology, The Affiliated Hospital of Hubei University of Chinese Medicine, Hubei Provincial Hospital of Traditional Chinese Medicine, Hubei Key Laboratory of Theory and Application Research of Liver and Kidney in Traditional Chinese Medicine, Wuhan, Hubei, 430061, People's Republic of China

Correspondence: Jing Teng, Email [tengjingtj@126.com](mailto:tengjingtj@126.com)

**Objective:** This study aimed to develop a comprehensive machine learning (ML)-based prediction model for intrahepatic cholestasis of pregnancy (IHCP) by integrating multi-modal data including demographic characteristics, laboratory biochemical indicators, and ultrasonic echocardiographic parameters. The model was designed to stratify ICP severity and remain applicable in settings lacking total bile acid (TBA) testing, which addresses current diagnostic gaps and may support the reduction of adverse perinatal outcomes.

**Methods:** A retrospective cohort of 750 pregnant women (525 in training, 225 in testing) between July 2020 and October 2023 from the Central Hospital of Enshi Tujia and Miao Autonomous Prefecture was recruited for the study. Seven ML algorithms (Logistic regression, Decision Tree, Random Forest [RF], Extreme Gradient Boosting [XGBoost], Regularized Support Vector Machine [RSVM], Multilayer Perceptron [MLP], and Elastic Net [ENET]).

**Results:** The RF model exhibited superior performance, achieving ROC-AUC of 0.90 (training) and 0.86 (testing), with sensitivity and specificity both  $\geq 0.93$  in the testing cohort. Key predictors included pruritus, TBA, glycocholic acid, alkaline phosphatase, and ultrasonic indicators (ventricular wall mean thickness, myocardial echogenicity). Notably, the model retained efficacy without TBA, maintaining precision  $\geq 0.75$  across recall values of 0.6–0.9.

**Conclusion:** The multi-modal RF model effectively predicts IHCP, enables severity stratification, and enhances accessibility in resource-limited settings, providing valuable support for targeted clinical interventions and may support the reduction of adverse perinatal outcomes.

**Keywords:** intrahepatic cholestasis of pregnancy, machine learning, random forest, prediction model, ultrasonic radiomics, biochemical indicators

## Introduction

Intrahepatic cholestasis of pregnancy (IHCP) is a pregnancy-specific liver disorder that mainly occurs in the second and third trimesters. Its incidence varies greatly by region. China's Yangtze River Basin, including Sichuan, Chongqing, Shanghai and Jiaying, reports an incidence of 3%–7%.<sup>1,2</sup> South American countries have rates as high as 9.2%–15.6% and Scandinavian nations record 1.5%.<sup>3</sup> In contrast, North America and parts of Europe have lower prevalence, with rates of 0.2%–0.3% and 0.1%–0.2% respectively.<sup>4</sup> Clinically, IHCP is marked by skin pruritus (usually without skin lesions) and elevated serum total bile acid (TBA), often with mild increases in liver function markers. Though benign for mothers, it endangers fetuses severely.<sup>5</sup> Bile acids can cross the placental barrier to accumulate in fetal tissues and amniotic fluid, causing preterm birth, meconium-stained amniotic fluid, fetal distress or stillbirth.<sup>6,7</sup> Notably, maternal bile acid over 100  $\mu\text{mol/L}$  raises fetal death risk tenfold, emphasizing the need for timely diagnosis.

Despite research efforts, IHCP's causes and mechanisms remain unclear. It may be associated with pregnancy hormones, genetic factors, environmental influences and immune dysregulation.<sup>8,9</sup> Previous studies have explored risk factors such as prior IHCP, multiple pregnancies and gestational diabetes, but no consensus has been reached due to ethnic and environmental differences across cohorts. Diagnostic attempts have included the Z allele variant of the SERPINA1 gene, apolipoprotein A2 (APOA2) and specific microRNAs, yet these methods are costly and less accessible in many clinical settings.<sup>10–12</sup> Traditional fetal monitoring tools, such as fetal heart rate monitoring and ultrasound, fail to predict IHCP-related risks effectively.<sup>13,14</sup> Most existing studies on IHCP also rely on univariate analysis, which provides limited clinical value for practical decision-making.

Current diagnostic gaps highlight the necessity of this study. Lack of TBA testing in resource-limited settings delays IHCP diagnosis and increases the risk of adverse fetal outcomes. Currently, few studies have integrated biochemical indicators such as GOT (Glutamic-Oxaloacetic Transaminase) and GPT (Glutamic-Pyruvic Transaminase) with fetal ultrasound radiomics. Fetal ultrasound radiomics is a technique that captures subtle changes in the fetus and placenta, and integrating it with biochemical indicators could help build more comprehensive predictive models for IHCP.<sup>15,16</sup> Machine learning (ML) excels at analyzing high-dimensional data but is rarely applied in IHCP research. To date, no studies have used ML to classify pregnancies into normal, mild IHCP and severe IHCP categories.

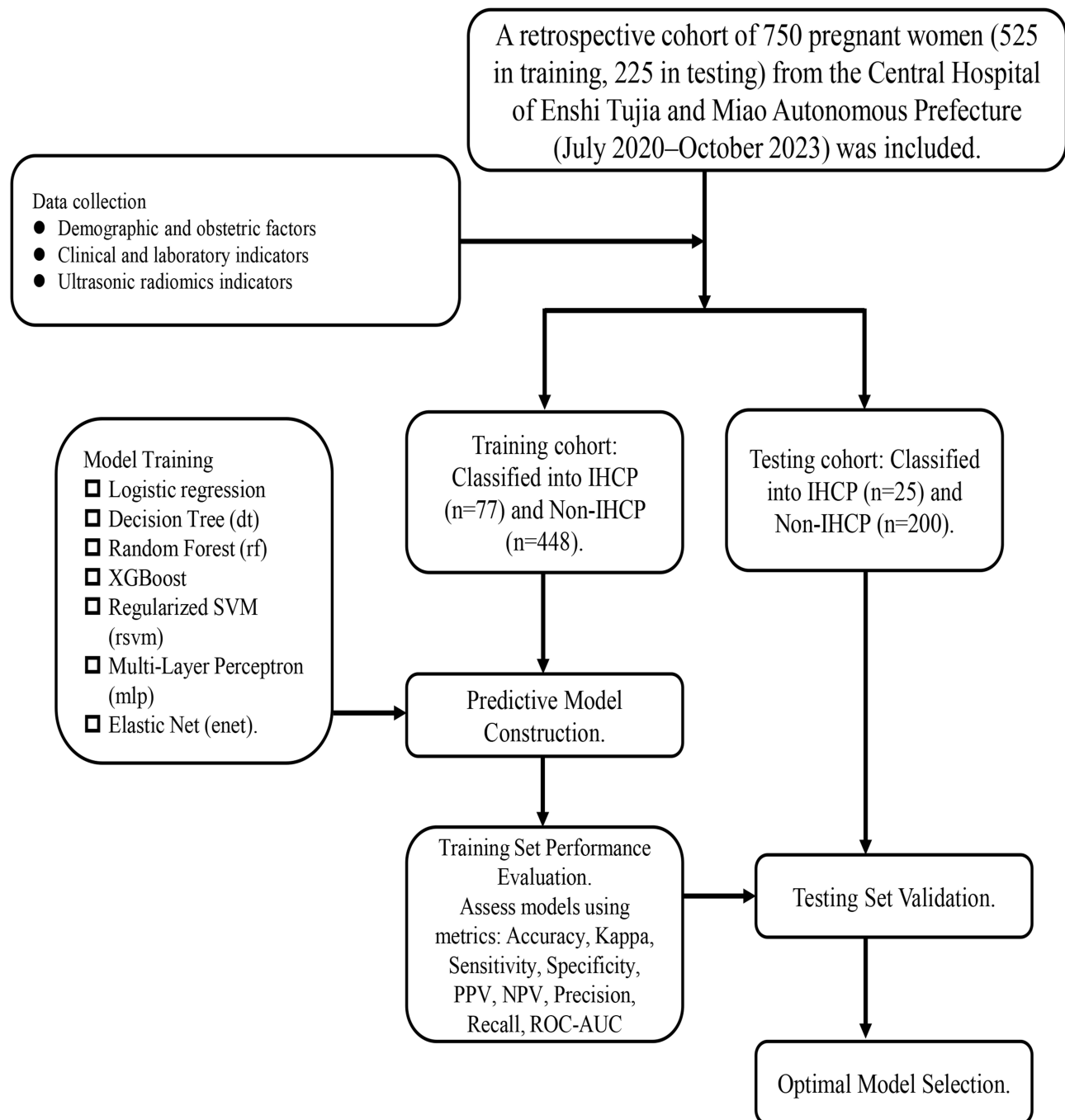
This study aims to develop an ML-based IHCP prediction model. Its innovation lies in three points. The first is integrating multiple types of data to construct a multi-modal model. The second is stratifying IHCP severity to support targeted clinical interventions. The third is designing the model to be usable even without TBA testing. This work seeks to improve the accuracy of IHCP prediction, optimize clinical management strategies and reduce adverse perinatal outcomes associated with the disorder.

## Materials and Methods

### Study Participants and Data Collection

This retrospective study was conducted at the Central Hospital of Enshi Tujia and Miao Autonomous Prefecture. Pregnant women were enrolled between July 2020 and October 2023, including those diagnosed with intrahepatic cholestasis of pregnancy (IHCP) and those with normal pregnancies. Cases with incomplete clinical, laboratory, or ultrasonic data were excluded to ensure data quality. Inclusion criteria as follows: (1) Pregnant women who received regular prenatal care and underwent ultrasonic examinations; (2) Those with a clearly defined gestational age, confirmed by early ultrasound (before 14 weeks of gestation) or a reliable menstrual history; (3) Cases diagnosed with IHCP in accordance with the 2024 ICP Diagnosis and Treatment Guide; for normal pregnancies, participants were required to have no symptoms of skin pruritus, normal liver function, and serum TBA levels  $<10 \mu\text{mol/L}$ ; (4) Availability of complete clinical, laboratory, and ultrasonic data for analysis. Exclusion Criteria: (1) Pregnant women with concurrent liver diseases other than IHCP, such as viral hepatitis or autoimmune liver disease; (2) Those with obstetric complications unrelated to IHCP, including fetal chromosomal abnormalities or structural malformations; (3) Presence of severe systemic diseases, such as cardiovascular disorders or renal insufficiency; (4) History of receiving treatments that could affect liver function or bile acid metabolism prior to enrollment; (5) Incomplete clinical data, loss to follow-up, or refusal to provide relevant information.

After exclusion, 750 participants were included in the final cohort. This cohort was randomly divided into a training cohort (70%,  $n=525$ ) and a testing cohort (30%,  $n=225$ ). The training cohort comprised 77 participants with IHCP and 448 with normal pregnancies, while the testing cohort included 25 participants with IHCP and 200 with normal pregnancies. Data collected for each participant covered three categories: baseline demographic characteristics (maternal age, body mass index [BMI], gravidity), laboratory biochemical indicators (total bile acids [TBA], glycocholic acid [GCA], alkaline phosphatase [ALP], alanine aminotransferase [ALT], aspartate aminotransferase [AST], total bilirubin [TBIL], cholesterol), and ultrasonic radiomics indicators (ventricular wall mean, myocardial echogenicity, tricuspid regurgitation velocity, ventricular cavity area, interventricular septum, aortic annulus diameter, pulmonary valve velocity, mitral E/A ratio, cardiac output, left ventricular [Lv] ejection fraction). The flowchart for patient inclusion and predictive model construction is shown in [Figure 1](#).



**Figure 1** Flowchart for Patient Inclusion and Predictive Model Construction.

## Patient Data Confidentiality Statement

The retrospective cohort study was approved by the Ethics Committee of the Central Hospital of Enshi Tujia and Miao Autonomous Prefecture (IRB-2020-089). In accordance with the ethical review opinion, the requirement for individual informed consent for accessing patient medical records was waived due to the retrospective nature of the study.

All patient data used in this research, including demographic characteristics, laboratory biochemical indicators, ultrasonic radiomic parameters and clinical medical records, were collected and managed in strict compliance with the ethical principles of the Declaration of Helsinki and its subsequent amendments, as well as relevant national and institutional regulations on medical data privacy and security. Identifiable personal information of all participants

(such as name, ID number, contact information, etc.) has been completely anonymized and de-identified during data collection and processing to ensure that individual privacy is not disclosed. Data storage adopts encrypted management mode, and only the research team members who have obtained the ethical approval and signed the confidentiality agreement have the right to access the research data. All data will only be used for the purpose of this research and will not be disclosed to any third party or used for other research or commercial purposes without the approval of the Ethics Committee.

## Diagnostic Criteria of IHCP

Diagnostic criteria for IHCP were based on the ICP Diagnosis and Treatment Guide (2024) released by the Obstetrics Group of the Obstetrics and Gynecology Branch of the Chinese Medical Association. Two criteria were required for diagnosis. The first was unexplained skin pruritus during pregnancy. The second was either unexplained liver function abnormalities in pregnant women or serum TBA  $\geq 10$   $\mu\text{mol/L}$ . Most participants showed resolution or alleviation of these symptoms after childbirth, which supported diagnostic confirmation.

IHCP severity was also classified per the guide. Mild ICP was defined by serum total bile acid (TBA) levels ranging from 10  $\mu\text{mol/L}$  to 99  $\mu\text{mol/L}$ , with skin pruritus as the only main symptom and no other obvious clinical manifestations. Severe ICP was defined by serum TBA levels  $\geq 100$   $\mu\text{mol/L}$ , accompanied by severe pruritus and at least one additional risk factor, including multiple pregnancies, hypertensive disorders of pregnancy (HDP), recurrent ICP, ICP-related perinatal death, and early-onset ICP (onset before 28 weeks of gestation).

This severity classification is consistent with the current international clinical guidelines for ICP, in which the TBA level of  $\geq 100$   $\mu\text{mol/L}$  is identified as the critical threshold for severe ICP due to its tenfold increased risk of adverse fetal and neonatal outcomes including stillbirth.

## Candidate Predictors

Candidate predictors were selected via a systematic evidence-based process. First, a comprehensive literature search was performed across PubMed, Web of Science, Cochrane Library, Scopus, Medline, and Embase. Search terms included “intrahepatic cholestasis of pregnancy”, “IHCP”, “risk factors”, and “predictors” to capture relevant studies.

Each retrieved study was reviewed to extract potential predictors, and systematic reviews/meta-analyses were consulted to avoid omissions. The research team held discussions to evaluate each candidate based on two key factors: frequency of mention in high-quality literature (reflecting evidence strength) and clinical accessibility (ensuring routine measurability).

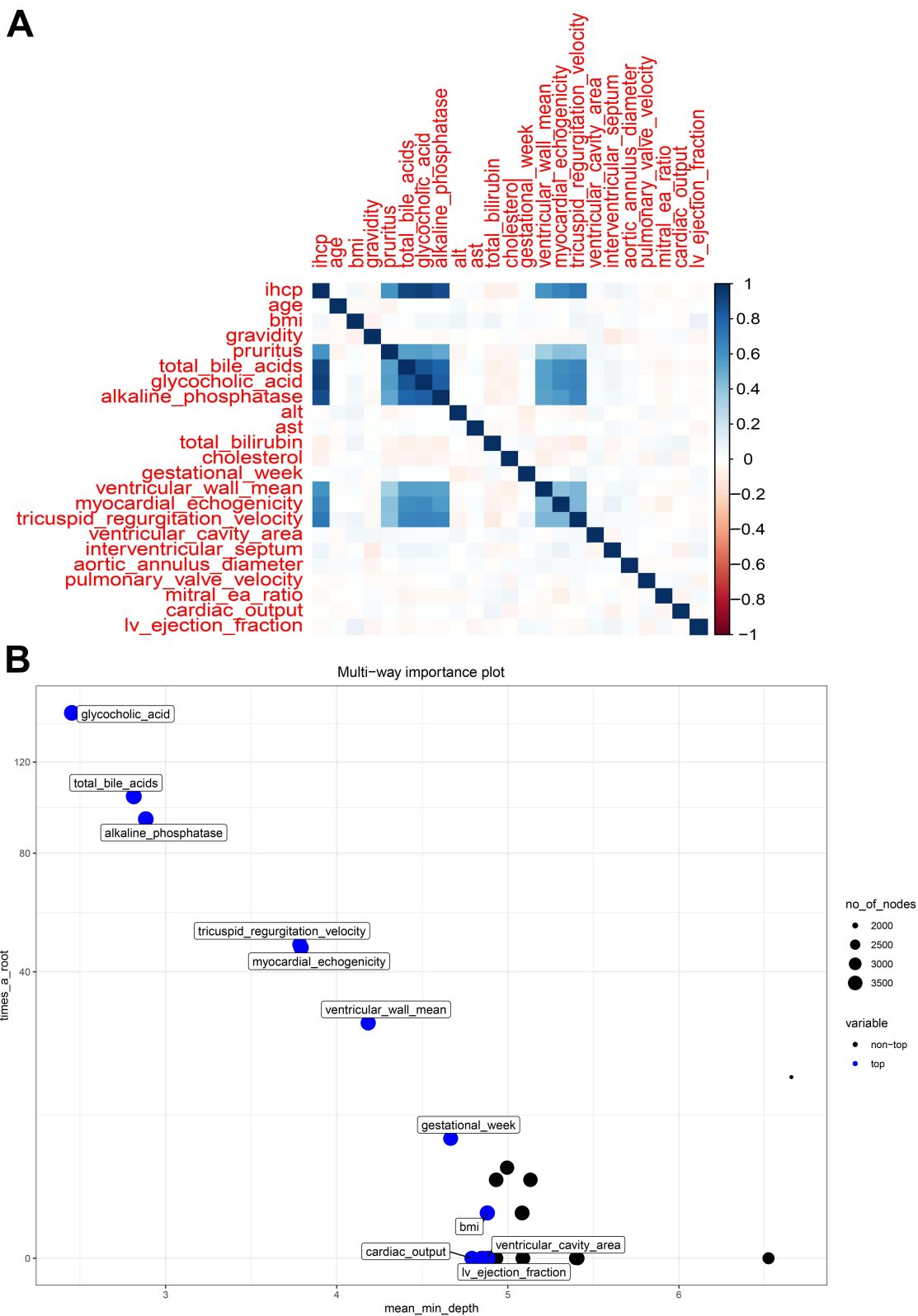
A total of 21 candidate predictors were finalized, grouped into three categories. Demographic and obstetric factors included maternal age, BMI, and gravidity. Clinical and laboratory indicators included pruritus, TBA, GCA, ALP, ALT, AST, TBIL, cholesterol, and gestational week. Ultrasonic radiomics indicators included ventricular wall mean, myocardial echogenicity, tricuspid regurgitation velocity, ventricular cavity area, interventricular septum, aortic annulus diameter, pulmonary valve velocity, mitral E/A ratio, cardiac output, and Lv ejection fraction.

## Model Development and Evaluation

Seven machine learning algorithms were selected as candidates: Logistic regression, Decision Tree (DT), Random Forest (RF), Extreme Gradient Boosting (XGBoost), Regularized Support Vector Machine (RSVM), Multilayer Perceptron (MLP), and Elastic Net (ENET).

The cohort was split into training (70%) and testing (30%) sets using stratified random sampling to maintain the ratio of IHCP to normal cases. Each model was trained on the training set and validated via five repeats of tenfold cross-validation. Mean and median values from cross-validation were used to score models, and box plots were generated to visualize performance (Figure 2B for variable importance reference). The top-performing models were optimized via grid search for hyperparameter tuning.

Model performance on training and testing sets was evaluated using ten metrics: accuracy, Kappa coefficient, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), precision, recall, F1-score, and area under the receiver operating characteristic curve (ROC-AUC). Heatmaps of these metrics were



**Figure 2** Pearson Correlation Heatmap and SHAP-based Importance Ranking for IHCP-Related Variables. **(A)** Pearson Correlation Heatmap: Quantitatively presenting the linear correlation coefficients between all 21 candidate predictors and the outcome variable IHCP, with the color gradient reflecting the direction and strength of the correlation and numerical values marking the specific correlation coefficient. **(B)** Multi-way Importance Plot based on SHAP: Ranking all candidate variables according to their SHAP weights to quantify the relative contribution of each predictor to IHCP prediction, where higher weights represent greater predictive importance of the variable in the model.

generated to compare model performance across datasets. Precision-recall curves were plotted to assess performance in imbalanced data, and curves of performance versus high-risk thresholds and cost-benefit ratios were generated to explore clinical applicability. LASSO regression results were visualized to show feature distribution and coefficient changes with log lambda. ROC-AUC values of selected models were further compared to confirm discriminative ability. Pearson correlation analysis was conducted to examine linear relationships between predictors and IHCP.

## Statistical Analysis

Data processing and analyses were conducted using Python 3.9, SIMCA-P 13.0 (Umetrics, Sweden), and SPSS 21.0 (IBM Corporation, USA). Data distribution was tested via the Kolmogorov–Smirnov test. Most continuous variables showed non-normal distribution and were described as median [interquartile range (IQR)]. Categorical variables were presented as frequency and percentage. Kruskal–Wallis H rank sum test was used for continuous variable comparisons across multiple groups, with pairwise comparisons via Mann–Whitney *U*-test (Bonferroni-corrected). For categorical variables, Pearson chi-square test, continuity-corrected chi-square test, or Fisher’s exact probability test was selected based on data characteristics. The significance level was set at  $\alpha=0.05$ , with Bonferroni correction for multiple pairwise comparisons.

Feature selection was performed in two steps to reduce redundancy. First, Spearman correlation analysis was used to assess redundancy; for any pair of predictors with correlation coefficient  $>0.8$ , one was retained based on clinical relevance. Second, Least Absolute Shrinkage and Selection Operator (LASSO) regression (via scikit-learn in Python 3.9) was applied to further select variables with optimal diagnostic performance by shrinking irrelevant predictor coefficients to zero. SIMCA-P 13.0 was used for partial least squares discriminant analysis (PLS-DA) of bile acid spectra, with results visualized in two-dimensional and three-dimensional score plots.

The statistical power of this study was calculated based on a sample size of 750 participants, an expected effect size of 0.3, and a significance level of  $\alpha=0.05$ , with the calculated power reaching 0.92, which ensured the study’s ability to detect meaningful differences between groups. Model validation included internal validation via five repeats of tenfold cross-validation and stratified random sampling to maintain the ratio of ICP to normal cases in the training and testing cohorts, which guaranteed the robustness and generalizability of the model.

## Results

### Baseline Characteristics, Clinical Indicators and Ultrasonic Radiomics Indicators

After excluding cases with incomplete data, 750 pregnant women were included in the final analysis, split into a training cohort ( $n=525$ ) and a testing cohort ( $n=225$ ). The training cohort had 77 (14.7%) participants with IHCP and 448 (85.3%) with normal pregnancies, while the testing cohort comprised 25 (11.1%) IHCP cases and 200 (88.9%) normal pregnancies (Table 1). In both cohorts, IHCP groups differed significantly from non-IHCP groups in several key indicators. Skin pruritus was markedly more prevalent in IHCP groups (training: 92.2%; testing: 92.0%) than non-IHCP groups (training: 15.4%; testing: 14.0%, both  $p<0.001$ ). Serum total bile acid (TBA), glycocholic acid (GCA), and alkaline phosphatase (ALP) were significantly higher in IHCP groups across both cohorts (all  $p<0.001$ ). Ultrasonically, IHCP groups showed higher ventricular wall mean thickness, myocardial echogenicity, and tricuspid regurgitation velocity (all  $p<0.001$ ) in both cohorts. The IHCP group had a higher median body mass index (BMI) in the training cohort (24.89 vs 23.96  $\text{kg/m}^2$ ,  $p=0.028$ ) but not in the testing cohort ( $p=0.336$ ). No significant differences were found in age, gravidity, alanine aminotransferase (ALT), aspartate aminotransferase (AST), total bilirubin (TBIL), or gestational week between groups (all  $p>0.05$ ), and cholesterol differed only in the training cohort ( $p=0.028$ ).

Collectively, these data from Table 1 confirm that pruritus, TBA, GCA, ALP, ventricular wall mean thickness, myocardial echogenicity, and tricuspid regurgitation velocity are consistent distinguishing features between IHCP and non-IHCP pregnancies, laying a foundation for subsequent predictor selection in the IHCP prediction model.

**Table 1** Baseline Characteristics and Clinical/Ultrasonic Radiomics Indicators of Pregnant Women in Training and Testing Cohorts Stratified by IHCP Status

Variables	Training Cohort			P-value	Testing Cohort			P-value
	Overall (N=525)	IHCP (N=77)	Non-IHCP (N=448)		Overall (N=225)	IHCP (N=25)	Non-IHCP (N=200)	
Age (median [IQR]),years	28.00 [26.00, 31.00]	29.00 [26.00, 31.00]	28.00 [25.00, 31.00]	0.153	28.00 [25.00, 31.00]	28.00 [26.00, 31.00]	28.00 [25.00, 31.00]	0.677
BMI (median [IQR]),kg/m <sup>2</sup>	24.19 [21.98, 26.46]	24.89 [22.97, 26.78]	23.96 [21.84, 26.26]	0.028	24.25 [21.92, 26.34]	25.12 [23.16, 26.78]	24.06 [21.84, 26.22]	0.336
Gravidity (%)								
1	136 (25.9)	21 (27.3)	115 (25.7)	0.665	58 (25.8)	7 (28.0)	51 (25.5)	0.669
2	222 (42.3)	29 (37.7)	193 (43.1)		92 (40.9)	10 (40.0)	82 (41.0)	
3	143 (27.2)	25 (32.5)	118 (26.3)		64 (28.4)	8 (32.0)	56 (28.0)	
4	21 (4.0)	2 (2.6)	19 (4.2)		11 (4.9)	0 (0.0)	11 (5.5)	
5	3 (0.6)	0 (0.0)	3 (0.7)		0 (0.0)	0 (0.0)	0 (0.0)	
Pruritus (%)								
0	385 (73.3)	6 (7.8)	379 (84.6)	<0.001	174 (77.3)	2 (8.0)	172 (86.0)	<0.001
1	140 (26.7)	71 (92.2)	69 (15.4)		51 (22.7)	23 (92.0)	28 (14.0)	
Total bile acids (median [IQR]), $\mu$ mol/L	7.49 [6.04, 9.37]	33.95 [26.36, 38.86]	7.08 [5.75, 8.54]	<0.001	7.00 [5.73, 8.85]	34.52 [26.99, 37.12]	6.66 [5.54, 8.31]	<0.001
Glycocholic acid (median [IQR]),mg/L	1.53 [1.10, 2.16]	18.48 [14.67, 21.81]	1.40 [1.05, 1.82]	<0.001	1.62 [1.10, 2.17]	15.98 [13.45, 22.93]	1.48 [1.06, 1.96]	<0.001
Alkaline phosphatase (median [IQR]),U/L	133.54 [112.71, 157.60]	277.13 [255.01, 304.35]	127.94 [109.55, 145.52]	<0.001	134.35 [113.42, 157.12]	271.77 [260.34, 301.56]	130.48 [110.68, 149.03]	<0.001
ALT (median [IQR]),U/L	31.50 [26.48, 38.23]	31.53 [26.77, 38.54]	31.48 [26.37, 38.21]	0.754	31.71 [27.36, 38.96]	31.29 [25.75, 43.11]	32.21 [27.63, 38.57]	0.689
AST (median [IQR]),U/L	26.57 [22.38, 31.77]	27.75 [24.41, 31.77]	26.13 [22.08, 31.77]	0.133	25.64 [21.06, 31.49]	26.57 [22.97, 30.66]	25.58 [21.05, 31.66]	0.62
Total bilirubin (median [IQR]), $\mu$ mol/L	8.93 [7.46, 10.61]	8.84 [7.16, 10.26]	8.96 [7.58, 10.68]	0.224	8.82 [7.49, 10.63]	8.46 [6.45, 9.96]	8.93 [7.56, 10.72]	0.187
Cholesterol (median [IQR]), mmol/L	5.24 [4.73, 5.73]	5.04 [4.66, 5.47]	5.27 [4.74, 5.76]	0.028	5.20 [4.66, 5.70]	5.29 [4.75, 5.62]	5.18 [4.66, 5.70]	0.674
Gestational week (median [IQR])	31.69 [29.52, 33.91]	31.79 [29.08, 34.08]	31.67 [29.61, 33.80]	0.762	31.75 [29.75, 34.26]	32.00 [29.86, 34.08]	31.75 [29.74, 34.27]	0.956
Ventricular wall mean (median [IQR]),mm	3.61 [3.30, 3.95]	4.45 [4.11, 4.70]	3.52 [3.26, 3.79]	<0.001	3.55 [3.26, 3.77]	4.50 [4.18, 4.73]	3.49 [3.20, 3.72]	<0.001
Myocardial echogenicity (median [IQR])	52.49 [47.77, 57.85]	66.40 [61.42, 70.82]	51.30 [46.75, 55.35]	<0.001	52.13 [47.10, 56.84]	67.70 [65.54, 72.51]	51.33 [46.36, 54.84]	<0.001
Tricuspid regurgitation velocity (median [IQR]),m/s	0.84 [0.77, 0.93]	1.13 [1.02, 1.23]	0.82 [0.77, 0.87]	<0.001	0.83 [0.78, 0.90]	1.15 [1.02, 1.23]	0.81 [0.77, 0.87]	<0.001
Ventricular cavity area (median [IQR]),mm <sup>2</sup>	16.67 [15.54, 18.16]	16.83 [15.84, 18.22]	16.65 [15.43, 18.13]	0.258	16.63 [15.44, 18.18]	17.33 [16.10, 18.18]	16.60 [15.40, 18.17]	0.298
Interventricular septum (median [IQR]),mm	3.10 [2.82, 3.33]	3.10 [2.89, 3.27]	3.09 [2.81, 3.34]	0.571	3.13 [2.88, 3.36]	3.23 [2.93, 3.48]	3.12 [2.86, 3.35]	0.128
Aortic annulus diameter (median [IQR]),mm	6.65 [6.25, 7.09]	6.75 [6.19, 7.11]	6.64 [6.25, 7.08]	0.651	6.68 [6.26, 7.14]	6.50 [5.98, 6.78]	6.70 [6.27, 7.20]	0.084
Pulmonary valve velocity (median [IQR]),m/s	0.94 [0.86, 1.01]	0.93 [0.86, 1.03]	0.94 [0.86, 1.01]	0.774	0.95 [0.87, 1.03]	0.94 [0.87, 1.03]	0.95 [0.87, 1.02]	0.988
Mitral ea ratio (median [IQR])	1.07 [0.98, 1.17]	1.07 [0.97, 1.16]	1.07 [0.98, 1.17]	0.885	1.07 [0.99, 1.16]	1.07 [0.99, 1.15]	1.07 [0.99, 1.16]	0.916
Cardiac output (median [IQR]),mL/min	238.75 [217.55, 262.13]	233.98 [215.76, 263.50]	239.30 [217.95, 261.49]	0.95	241.21 [215.64, 265.22]	237.36 [209.85, 254.27]	241.51 [216.82, 265.78]	0.35
Lv ejection fraction (median [IQR]),%	64.54 [61.75, 67.67]	63.62 [61.73, 67.03]	64.70 [61.78, 67.74]	0.33	64.58 [61.01, 67.20]	61.62 [59.84, 65.99]	64.83 [61.63, 67.33]	0.05

**Abbreviations:** IHCP, Intrahepatic Cholestasis of Pregnancy; BMI, Body Mass Index; ALT, Alanine Aminotransferase; AST, Aspartate Aminotransferase; IQR, Interquartile Range; Lv ejection fraction, Left Ventricular Ejection Fraction.

## Screening of Candidate Variables for the IHCP Prediction Model

To identify clinically meaningful and statistically robust predictors for the IHCP prediction model, three analytical approaches (Pearson correlation, SHAP-based importance ranking, and LASSO regression) were applied, with results presented in [Figure 2](#) and [Supplementary Figure 1](#). Pearson correlation analysis showed pruritus had the strongest positive correlation with IHCP, followed by TBA, GCA, and ALP. Three ultrasonic indicators (ventricular wall mean thickness, myocardial echogenicity, tricuspid regurgitation velocity) had weak negative correlations, while age and ALT had near-zero correlations.

In the SHAP-based importance ranking, GCA, TBA, and ALP stood out as the top predictors for IHCP. The three ultrasonic indicators mentioned earlier also ranked highly in terms of predictive value, whereas gestational week and BMI fell into the “non-top” category due to low SHAP weights, indicating they had little influence on IHCP prediction. LASSO regression included pre-regularization feature distribution and coefficient path. At optimal  $\text{Log } \lambda$ , only 7 variables (pruritus, TBA, GCA, ALP, and the three ultrasonic indicators) retained non-zero coefficients. These analyses consistently identified 7 core variables, excluding redundancies and providing a robust basis for IHCP prediction model development.

## Construction of the IHCP Prediction Model

Seven machine learning models were constructed for IHCP prediction, and their performance is characterized in [Table 2](#) and [Figure 3](#). In the training cohort, the RF model exhibited the most balanced and high-level performance among all models. It achieved an accuracy of 0.79, a Kappa coefficient of 0.45, with sensitivity and specificity both reaching 0.95, a positive predictive value (PPV) of 0.97, a negative predictive value (NPV) of 0.41, an F-Measure of 0.91, and a ROC-AUC of 0.90. The DT model obtained a high accuracy of 0.86 but showed poor discriminative ability. Its ROC-AUC was only 0.55, and its F-Measure was also low at 0.55. XGBoost, RSVM, MLP, and ENET displayed moderate performance in the training cohort. Their ROC-AUC values ranged from 0.47 to 0.72, and their F-Measure values were between 0.47 and 0.73.

In the testing cohort, the RF model maintained robust performance. Its sensitivity and specificity remained at 0.93, ROC-AUC stayed at 0.86, and F-Measure was 0.87, while accuracy decreased slightly to 0.64. The superiority of the RF model is further verified in [Supplementary Figure 2](#), in the testing scenario, the ROC-AUC of the RF model (0.86) was notably higher than that of the DT model (0.53), XGBoost (0.64), RSVM (0.45), and ENET (0.62). Among all constructed models, the RF model consistently demonstrated the most stable and effective predictive ability across the training and testing cohorts. This finding confirms that the RF model is the optimal choice for IHCP prediction.

## Evaluating Machine Learning Models for IHCP Prediction and Optimal RF Model Identification

To identify reliable tools for the clinical risk assessment of IHCP, this study evaluated the predictive efficacy of multiple machine learning models. Key performance data of these models were presented in [Figures 4](#) and [5](#). Overall, seven tested machine learning models and two control groups (All/None) exhibited significant performance variability across different parameters as shown in [Figure 4](#). Most models failed to maintain stable performance. The DT and RSVM models showed sharp performance declines when parameters deviated from moderate ranges. The XGBoost, ENET, and Logistic Regression models only achieved moderate stability, while the All/None control groups consistently delivered low performance. In contrast, the RF model exhibited notably robust adaptability, showing significant advantages over other models. Further validation was conducted using Precision-Recall (PR) curves in [Figure 5](#), which covered the seven machine learning models. This validation confirmed the superiority of the RF model. When the recall (the ability to detect positive IHCP cases) ranged from 0.6 to 0.9, the precision (the accuracy of positive predictions) of the RF model remained above 0.75. This performance far exceeded that of other models, for which the precision was generally below 0.65 within the same recall range.

Collectively, the RF model emerged as the optimal choice for IHCP prediction. Its stable performance across parameter variations and excellent balance between precision and recall enabled the accurate identification of high-risk

**Table 2** Performance Metrics of Different Machine Learning Models on Train and Test Datasets

Dataset	Model	Accuracy	Kappa Coefficient	Sensitivity	Specificity	PPV	NPV	Precision	Recall	F-Measure	Roc_Auc
Training cohort	Logistic	0.66	0.17	0.82	0.83	0.90	0.24	0.66	0.67	0.66	0.66
	DT	0.86	0.15	0.77	0.77	0.86	1.00	0.55	0.56	0.55	0.55
	RF	0.79	0.45	0.95	0.95	0.97	0.41	0.91	0.91	0.91	0.90
	Xgboost	0.55	0.16	0.85	0.86	0.95	0.23	0.72	0.73	0.72	0.72
	RSVM	0.70	0.06	0.73	0.73	0.86	0.19	0.47	0.48	0.47	0.47
	MLP	0.65	0.16	0.82	0.82	0.90	0.24	0.65	0.66	0.65	0.65
	ENET	0.65	0.17	0.82	0.82	0.91	0.24	0.65	0.65	0.65	0.64
Testing cohort	Logistic	0.75	0.22	0.83	0.84	0.89	0.30	0.68	0.68	0.68	0.67
	DT	0.86	0.11	0.76	0.77	0.86	1.00	0.54	0.54	0.54	0.53
	RF	0.64	0.03	0.93	0.93	0.85	0.17	0.87	0.87	0.87	0.86
	Xgboost	0.55	0.12	0.82	0.82	0.91	0.21	0.65	0.65	0.65	0.64
	RSVM	0.68	-0.05	0.72	0.73	0.84	0.12	0.46	0.46	0.46	0.45
	MLP	0.73	0.25	0.80	0.81	0.91	0.30	0.62	0.62	0.62	0.61
	ENET	0.73	0.25	0.81	0.81	0.91	0.30	0.63	0.63	0.63	0.62

**Abbreviations:** PPV, Positive Predictive Value; NPV, Negative Predictive Value; Roc\_Auc, ROC Area Under the Curve; DT, Decision Tree; RF, Random Forest; RSVM, Regularized Support Vector Machine; MLP, Multilayer Perceptron; ENET, Elastic Net.

IHCP cases. Meanwhile, this model also minimized false positives, thereby providing critical support for clinical risk stratification and the planning of targeted interventions.

## Discussion

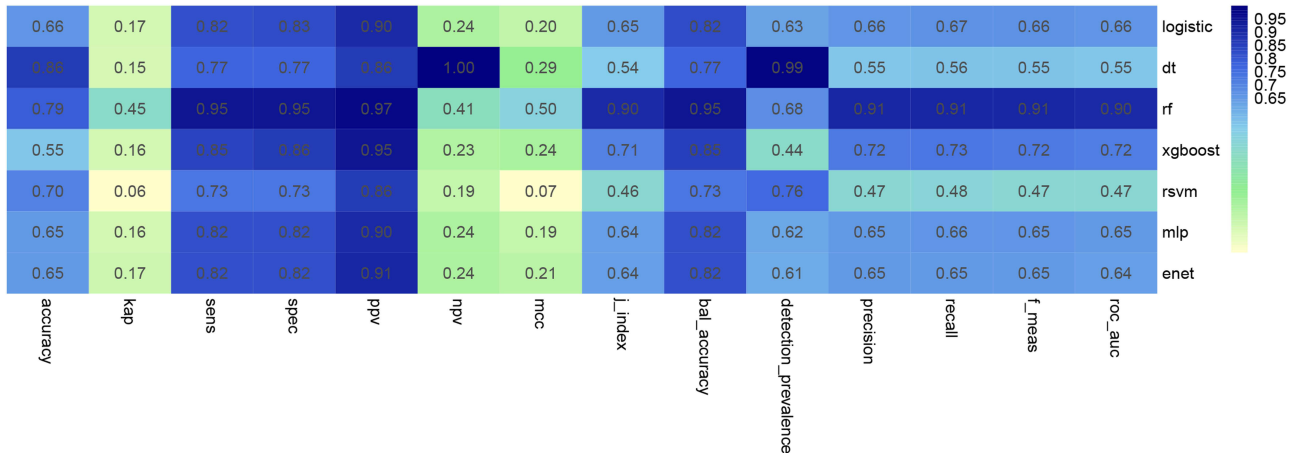
IHCP is a pregnancy-specific liver disorder that primarily manifests in the second and third trimesters, with incidence rates differing significantly across regions. Clinically, it is defined by skin pruritus and elevated serum total bile acid, and although it is generally benign for mothers, it poses serious threats to fetuses.<sup>17,18</sup> These threats include preterm birth, meconium-stained amniotic fluid, fetal distress, and even stillbirth, with maternal bile acid levels over 100  $\mu\text{mol/L}$  increasing the risk of fetal death tenfold. While its pathogenesis is linked to genetic, hormonal, immunological, and environmental factors, the exact mechanisms remain unclear.<sup>19–21</sup> Existing research often relies on univariate analysis, which offers limited clinical utility, and diagnostic methods such as specific genetic variants or microRNAs are costly and inaccessible in many settings. Traditional fetal monitoring tools also fail to effectively predict IHCP-related risks.

Nevertheless, this retrospective cohort study yields novel insights. By incorporating baseline demographic characteristics, laboratory biochemical parameters, and ultrasound radiomic features, along with a comparative evaluation of seven machine learning algorithms, the Random Forest model was identified as the optimal predictive tool. This model exhibited robust and stable performance in both the training and validation cohorts, with corresponding ROC-AUC values of 0.90 and 0.86, respectively, while maintaining a favorable balance between precision and recall. Notably, this study advances the field by establishing a multi-modal predictive model for stratifying the severity of intrahepatic cholestasis of pregnancy (IHCP), facilitating the implementation of targeted interventions. Furthermore, the model remains functional in the absence of total bile acid testing, thereby addressing critical clinical gaps in resource-limited settings. Compared to prior studies, this approach enhances the accuracy and accessibility of IHCP prediction, thereby improving clinical management and reducing adverse perinatal outcomes.<sup>22–24</sup>

The superiority of the RF model in this study is not merely a statistical outcome but a reflection of its inherent capacity to manage the complexity and multi-modality of IHCP-related data. Unlike logistic regression, which presupposes linear correlations between predictors, or decision trees that tend to overfit, the RF algorithm integrates multiple decision trees to reduce bias and variance. This characteristic holds particular value considering the interconnectedness of demographic, biochemical, and ultrasonic factors in IHCP pathogenesis. Its steady performance across both training and testing cohorts, with ROC-AUC values surpassing 0.85, highlights its strength in identifying subtle patterns. For instance, the faint yet significant connections between ventricular wall mean thickness and IHCP severity might go unnoticed by simpler models.

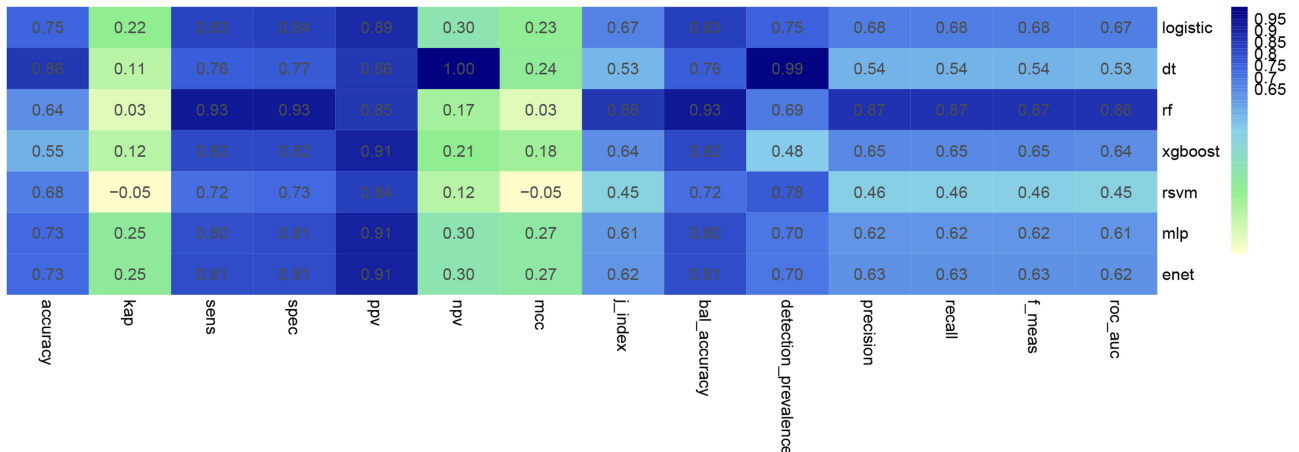
**A**

**Model Evaluation Metrics Heatmap(Train Datasets)**



**B**

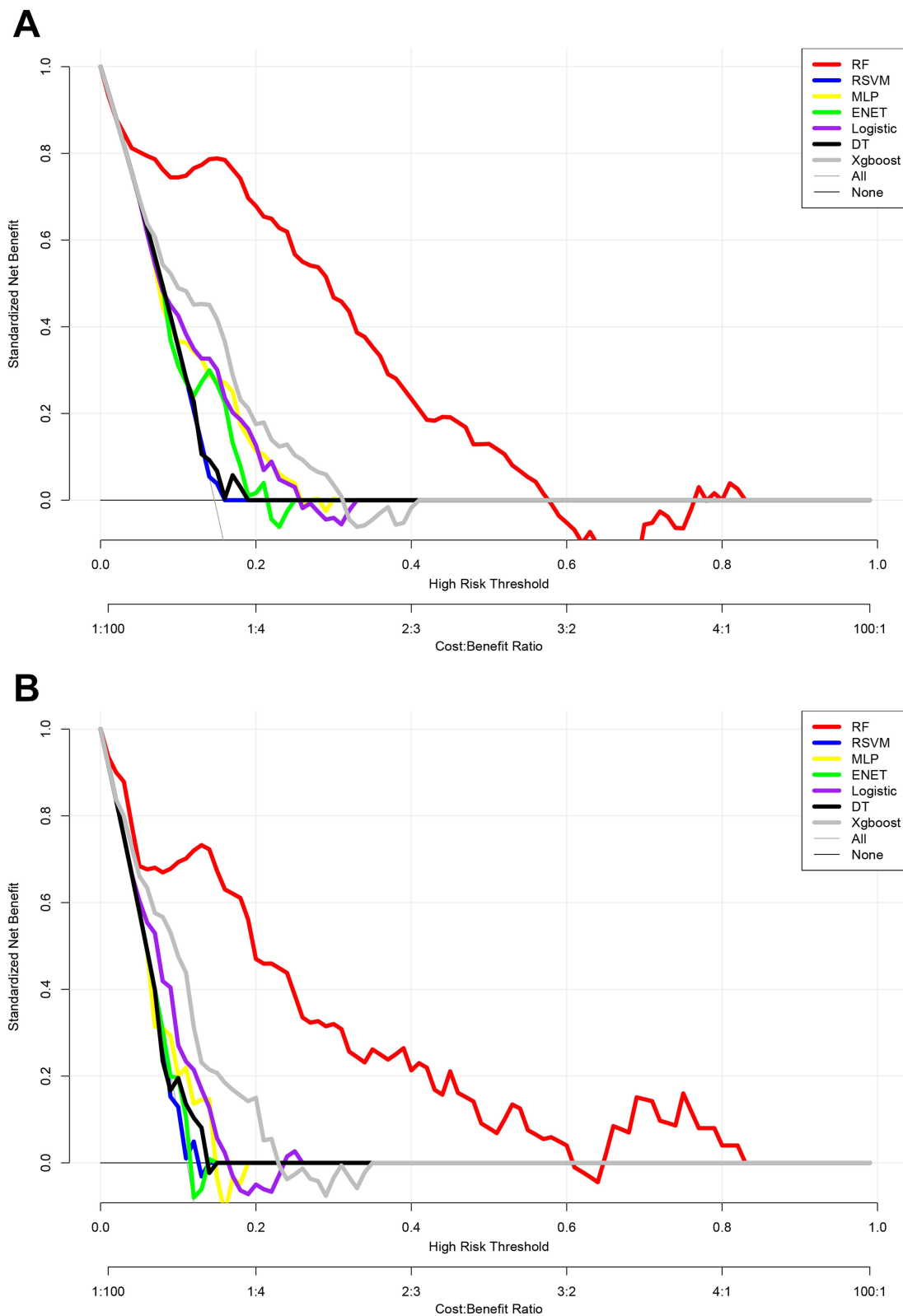
**Model Evaluation Metrics Heatmap(Test Datasets)**



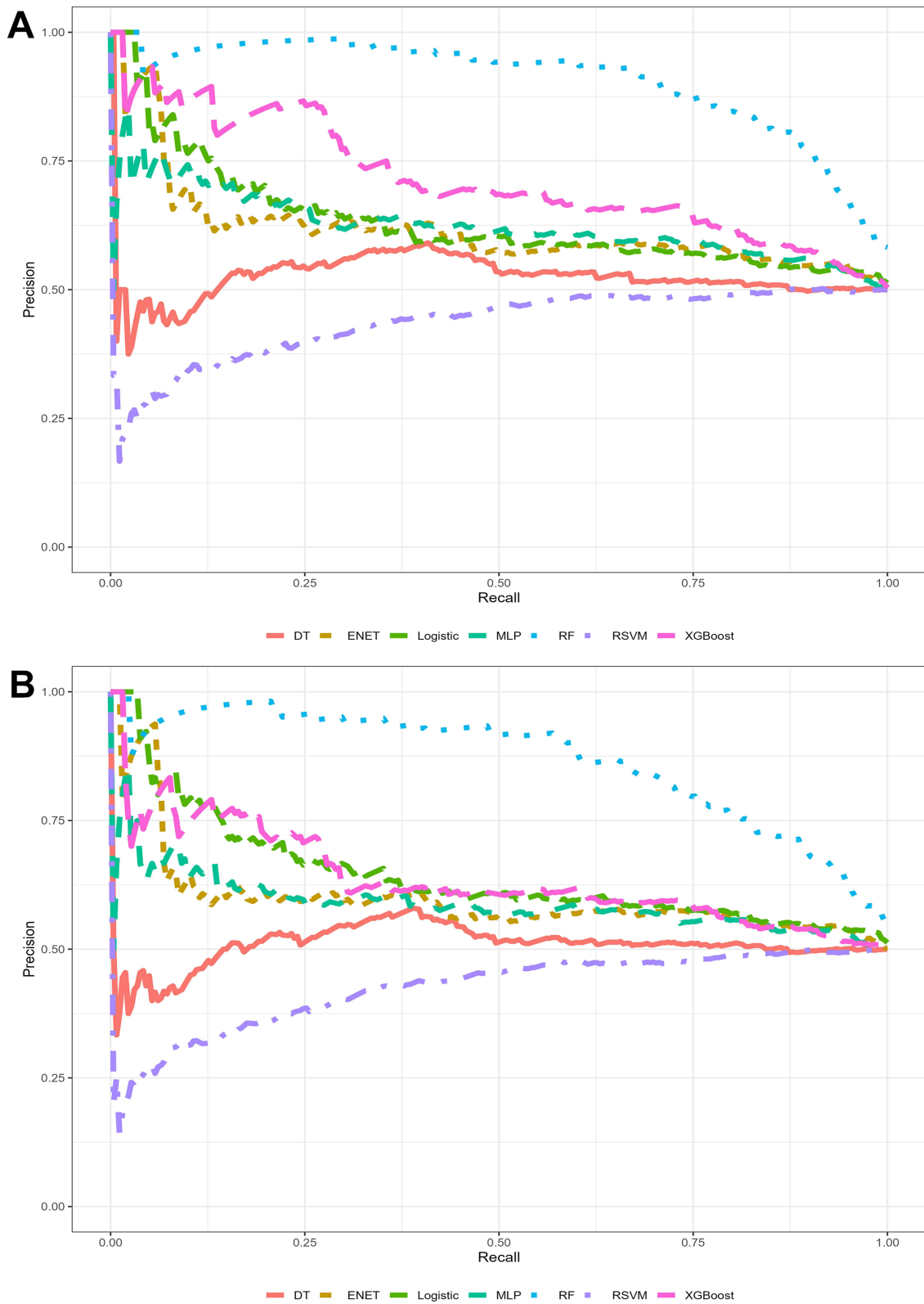
**Figure 3** Heatmaps of Model Evaluation Metrics for Training and Test Datasets. **(A)** Training Dataset Metric Heatmap: Visualizing the performance of seven machine learning models (Logistic regression, DT, RF, XGBoost, RSVM, MLP, ENET) on the training cohort (n=525) across ten key evaluation metrics (accuracy, Kappa coefficient, sensitivity, specificity, PPV, NPV, precision, recall, F1-score, ROC-AUC), with color intensity indicating the magnitude of metric values for intuitive comparison of model performance differences. **(B)** Test Dataset Metric Heatmap: Displaying the performance of the same seven machine learning models on the independent test cohort (n=225) using the identical ten evaluation metrics, which directly reflects the robustness and generalizability of each model by comparing the metric changes between the training and test datasets.

A key innovation of this research is the incorporation of ultrasonic radiomics into predictive modeling. Traditional IHCP studies have mainly centered on serum markers such as TBA and GCA, but our results indicate that fetal cardiac parameters, including myocardial echogenicity and tricuspid regurgitation velocity, independently contribute to predictive accuracy.<sup>25–27</sup> This aligns with growing evidence suggesting that intrauterine bile acid buildup may cause subclinical cardiac dysfunction in fetuses, a condition detectable through advanced ultrasonography.<sup>28–30</sup> By including these radiomic features, the RF model moves beyond the constraints of approaches relying solely on biomarkers, offering a more comprehensive evaluation of fetal susceptibility to IHCP-related complications.

Notably, the model maintains its predictive capability even without TBA data, addressing a long-standing issue in resource-limited settings. In areas where TBA testing is unavailable or delayed, relying on pruritus, ALP levels, and ultrasonic indicators can still facilitate timely risk stratification.<sup>21,31</sup> This is clinically important because untreated IHCP in such settings significantly raises the risk of fetal mortality. The SHAP analysis, which ranked GCA and ALP highly among non-TBA predictors, further supports the viability of this alternative approach and provides an evidence-based basis for adapting the model to low-resource environments.



**Figure 4** Curves of Model Performance vs High Risk Threshold and Cost-Benefit Ratio. **(A)** Model Performance vs High Risk Threshold Curve: Depicting the dynamic changes in predictive performance of seven machine learning models and two control groups (All/None) with the adjustment of high-risk threshold values (0.0–1.0), which evaluates the adaptability of each model to different clinical risk stratification criteria. **(B)** Model Performance vs Cost-Benefit Ratio Curve: Demonstrating the variation in clinical predictive performance of the above models and control groups with the change of cost-benefit ratios (1:100 to 100:1), which provides an evidence-based basis for the clinical applicability and cost-effectiveness of each model in different practical application scenarios.



**Figure 5** Precision-Recall Curves of Different Machine Learning Models. **(A)** Full-Range Precision-Recall Curve: Showing the complete relationship between precision and recall of seven machine learning models for IHCP prediction across the entire recall range, with the area under the curve reflecting the overall predictive performance of each model for the imbalanced IHCP cohort data. **(B)** Clinical Focus Range Precision-Recall Curve: Highlighting the precision-recall performance of the seven models within the clinically critical recall range (0.6–0.9) for identifying high-risk IHCP cases, which directly compares the ability of each model to maintain high prediction precision while ensuring sufficient recall rate in clinical practice.

When viewed in the context of existing literature, our findings both confirm and expand on previous knowledge. Consistent with earlier research, pruritus and elevated bile acid metabolites emerged as strong predictors, reinforcing their role as fundamental clinical markers.<sup>17,32,33</sup> However, the addition of ultrasonic radiomics introduces a new aspect. While prior studies have linked fetal ultrasound to IHCP outcomes, this is the first to quantify its role in predictive modeling using machine learning. The RF model's ability to differentiate between mild and severe IHCP, based on TBA thresholds and associated risk factors like multiple pregnancies, also advances clinical practice by enabling targeted interventions. These include closer monitoring for severe cases and delayed delivery for milder ones.

It is important to recognize the limitations of this study to guide future research. Firstly, the retrospective design, although effective in capturing real-world clinical data, may introduce selection bias. This is especially true for the inclusion of ultrasonic parameters, which were not consistently recorded across all participants. Secondly, the cohort was sourced from a single center in the Yangtze River Basin, a region with a relatively high IHCP incidence. External validation in populations with lower prevalence, such as those in North America or Northern Europe, is needed to confirm its generalizability. Thirdly, despite its strong performance, the "black-box" nature of the RF model limits its interpretability. This issue could be resolved by integrating explainable AI techniques to clarify the decision-making process.

Future research could focus on three areas. Expanding the dataset to include multi-center, multi-ethnic cohorts would enhance the model's external validity, ensuring it can be applied across diverse populations. Incorporating longitudinal data, such as serial measurements of bile acids and ultrasonic parameters throughout pregnancy, might improve the model's ability to predict the onset and progression of IHCP over time. Additionally, exploring epigenetic or metabolomic markers alongside current predictors could refine risk stratification, particularly for cases with unclear clinical symptoms.

## Conclusion

In summary, this study establishes a multi-modal RF model as a reliable tool for IHCP prediction, bridging gaps in accessibility and accuracy. By combining biochemical and ultrasonic data, it not only improves clinical decision-making but also facilitates equitable healthcare delivery, even in settings with limited diagnostic resources. As IHCP continues to pose a major threat to perinatal health, such advancements in predictive modeling are crucial for reducing adverse outcomes and enhancing maternal-fetal care globally.

## Funding

Project Name: 2023-2024 Annual TCM Scientific Research Project of Hubei Provincial Administration of Traditional Chinese Medicine(Project No.: ZY2023Q029).

## Disclosure

The authors report no conflicts of interest in this work.

## References

- Huang L, Li X, Liu T, et al. Effect of intrahepatic cholestasis of pregnancy on infantile food allergy: a retrospective longitudinal study cohort in Southwest China. *Eur J Obstet Gynecol Reprod Biol.* 2022;272:110–115. doi:10.1016/j.ejogrb.2022.03.026
- Zhan Y, Xu T, Chen T, Wang X. Intrahepatic cholestasis of pregnancy and maternal dyslipidemia: a systematic review and meta-analysis. *Acta obstetrica et Gynecologica Scandinavica.* 2022;101(7):719–727. doi:10.1111/aogs.14380
- Biocca MJ, Sperling JD, Chauhan SP. Intrahepatic cholestasis of pregnancy: review of six national and regional guidelines. *Eur J Obstet Gynecol Reprod Biol.* 2018;231:180–187. doi:10.1016/j.ejogrb.2018.10.041
- García-Romero CS, Guzman C, Cervantes A, Cerbón M. Liver disease in pregnancy: medical aspects and their implications for mother and child. *Ann. Hepatol.* 2019;18(4):553–562. doi:10.1016/j.aohep.2019.04.009
- Saad AF, Pacheco LD, Chappell L, Saade GR. Intrahepatic cholestasis of pregnancy: toward improving perinatal outcome. *Reprod Sci.* 2022;29(11):3100–3105. doi:10.1007/s43032-021-00740-x
- Smith DD, Rood KM. Intrahepatic Cholestasis of Pregnancy. *Clini Obstetrics and Gynecol.* 2020;63(1):134–151. doi:10.1097/GRF.0000000000000495
- Palmer KR, Xiaohua L, Mol BW. Management of intrahepatic cholestasis in pregnancy. *Lancet.* 2019;393(10174):853–854. doi:10.1016/S0140-6736(18)32323-7
- Xiao J, Li Z, Song Y, et al. Molecular pathogenesis of intrahepatic cholestasis of pregnancy. *Can J Gastroenterol Hepatol.* 2021;2021:6679322. doi:10.1155/2021/6679322

9. Niemyjska-Dmoch W, Kosiński P, Węgrzyn P, Luterek K, Jezela-Stanek A. Intrahepatic cholestasis of pregnancy and theory of inheritance of the disease. Literature review. *J Matern Fetal Neonatal Med.* 2023;36(2):2279020. doi:10.1080/14767058.2023.2279020
10. Zeng W, Hou Y, Gu W, Chen Z. Proteomic biomarkers of intrahepatic cholestasis of pregnancy. *Reprod Sci.* 2024;31(6):1573–1585. doi:10.1007/s43032-023-01437-z
11. Zu Y, Guo S, Li G, et al. Serum microRNAs as non-invasive diagnostic biomarkers for intrahepatic cholestasis of pregnancy. *Am J Transl Res.* 2022;14(9):6763–6773.
12. Deniz CD, Ozler S, Sayin FK. Association of adverse outcomes of intrahepatic cholestasis of pregnancy with zonulin levels. *J Obstetrics Gynaecol.* 2021;41(6):904–909. doi:10.1080/01443615.2020.1820463
13. Mathur D, Morgan M, McKenzie J, Wakefield D, Janicki MB, Figueroa R. Intrahepatic cholestasis of pregnancy: dilemma in diagnosis and management. *J Matern Fetal Neonatal Med.* 2022;35(25):8975–8981. doi:10.1080/14767058.2021.2008896
14. Ozel A, Alici Davutoglu E, Eric Ozdemir M, Oztunc F, Madazli R. Assessment of fetal left ventricular modified myocardial performance index and its prognostic significance for adverse perinatal outcome in intrahepatic cholestasis of pregnancy. *J Matern Fetal Neonatal Med.* 2020;33(12):2000–2005. doi:10.1080/14767058.2018.1535588
15. Donet A, Girault A, Pinton A, Lepercq J. Intrahepatic cholestasis of pregnancy: is a screening for differential diagnoses necessary? *J Gynecol Obstetrics and Human Reproduction.* 2020;49:101907. doi:10.1016/j.jogoh.2020.101907
16. Rodriguez M, Bombin M, Ahumada H, Bachmann M, Egaña-Ugrinovic G, Sepúlveda-Martínez A. Fetal cardiac dysfunction in pregnancies affected by intrahepatic cholestasis of pregnancy: a cohort study. *J Obstetrics and Gynaecol Res.* 2022;48(7):1658–1667. doi:10.1111/jog.15283
17. Ovadia C, Seed PT, Sklavounos A, et al. Association of adverse perinatal outcomes of intrahepatic cholestasis of pregnancy with biochemical markers: results of aggregate and individual patient data meta-analyses. *Lancet.* 2019;393(10174):899–909. doi:10.1016/S0140-6736(18)31877-4
18. Ovadia C, Williamson C. Intrahepatic cholestasis of pregnancy: recent advances. *Clin Dermatol.* 2016;34(3):327–334. doi:10.1016/j.clindermatol.2016.02.004
19. Azzaroli F, Turco L, Lisotti A, Calvanese C, Mazzella G. The pharmacological management of intrahepatic cholestasis of pregnancy. *Current Clin Pharmacol.* 2011;6(1):12–17. doi:10.2174/157488411794941313
20. Dixon PH, Levine AP, Cebola I, et al. GWAS meta-analysis of intrahepatic cholestasis of pregnancy implicates multiple hepatic genes and regulatory elements. *Nat Commun.* 2022;13(1):4840. doi:10.1038/s41467-022-29931-z
21. Tang M, Xiong L, Cai J, et al. Intrahepatic cholestasis of pregnancy: insights into pathogenesis and advances in omics studies. *Hepatology Internat.* 2024;18(1):50–62. doi:10.1007/s12072-023-10604-y
22. Zhang X, Chen Y, Salerno S, et al. Prediction of intrahepatic cholestasis of pregnancy in the first 20 weeks of pregnancy. *J Matern Fetal Neonatal Med.* 2022;35(25):6329–6335. doi:10.1080/14767058.2021.1911996
23. Ren Y, Shan X, Ding G, et al. Risk factors and machine learning prediction models for intrahepatic cholestasis of pregnancy. *BMC Pregnancy Childbirth.* 2025;25(1):89. doi:10.1186/s12884-025-07180-4
24. Asali A, Ravid D, Shalev H, et al. Intrahepatic cholestasis of pregnancy: machine-learning algorithm to predict elevated bile acid based on clinical and laboratory data. *Arch Gynecol Obstetrics.* 2021;304(3):641–647. doi:10.1007/s00404-021-05994-z
25. He J, Zhu X, Yang X, Wang H. Predictive efficacy of machine-learning algorithms on intrahepatic cholestasis of pregnancy based on clinical and laboratory indicators. *J Matern Fetal Neonatal Med.* 2025;38(1):2413854. doi:10.1080/14767058.2024.2413854
26. Manzotti C, Casazza G, Stimac T, Nikolova D, Gluud C. Total serum bile acids or serum bile acid profile, or both, for the diagnosis of intrahepatic cholestasis of pregnancy. *Cochrane Database Syst Rev.* 2019;7(7):Cd012546. doi:10.1002/14651858.CD012546.pub2
27. Obiegbusi CN, Dong XJ, Obiegbusi SC, Jin X, Okoene IK. Predictors of adverse fetal outcomes in intrahepatic cholestasis of pregnancy (ICP): a narrative review. *Reprod Sci.* 2024;31(2):341–351. doi:10.1007/s43032-023-01329-2
28. Lin J, Gu W, Hou Y. Diagnosis and prognosis of early-onset intrahepatic cholestasis of pregnancy: a prospective study. *J Matern Fetal Neonatal Med.* 2019;32(6):997–1003. doi:10.1080/14767058.2017.1397124
29. Kowalska-Kańska A, Maciejewski T, Niemiec KT. The concentrations of bile acids and erythropoietin in pregnant women with intrahepatic cholestasis and the state of the fetus and newborn. *Medycyna wieku rozwojowego.* 2013;17(3):232–245.
30. Glantz A, Marschall HU, Mattsson LA. Intrahepatic cholestasis of pregnancy: relationships between bile acid levels and fetal complication rates. *Hepatology.* 2004;40(2):467–474. doi:10.1002/hep.20336
31. Zhou Q, Yuan Y, Wang Y, et al. The severity of intrahepatic cholestasis during pregnancy increases risks of adverse outcomes beyond stillbirth: evidence from 15,826 patients. *BMC Pregnancy Childbirth.* 2024;24(1):476. doi:10.1186/s12884-024-06645-2
32. Ovadia C, Sajous J, Seed PT, et al. Ursodeoxycholic acid in intrahepatic cholestasis of pregnancy: a systematic review and individual participant data meta-analysis. *Lancet Gastroenterol Hepatol.* 2021;6(7):547–558. doi:10.1016/S2468-1253(21)00074-1
33. Majsterek M, Wierzchowska-Opoka M, Makosz I, Kreczyńska L, Kimber-Trojnar Ż, Leszczyńska-Gorzela B. Bile acids in intrahepatic cholestasis of pregnancy. *Diagnostics.* 2022;12(11). doi:10.3390/diagnostics12112746

International Journal of General Medicine

Publish your work in this journal

The International Journal of General Medicine is an international, peer-reviewed open-access journal that focuses on general and internal medicine, pathogenesis, epidemiology, diagnosis, monitoring and treatment protocols. The journal is characterized by the rapid reporting of reviews, original research and clinical studies across all disease areas. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/international-journal-of-general-medicine-journal>

**Dovepress**  
Taylor & Francis Group