Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A High-Performance Transfer Learning Model for Early Screening of Biliary Atresia in the Neonatal Period

Dingding Wang
Beijing Children's Hospital
Zonghan Li
Beijing Children's Hospital
Zhaozhou liu
Beijing Children's Hospital
Dayan Sun
Beijing Children's Hospital
Yuyan Jin
Beijing Children's Hospital
Jinshi Huang
Beijing Children's Hospital
Guantong Li
15065532698@163.com

Beijing Children's Hospital

Research Article

Keywords: biliary atresia, neonate, diagnosis, machine learning, transfer learning

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10753339/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Biliary atresia (BA) is a life-threatening cause of neonatal cholestasis, and early recognition is essential to enable timely Kasai portoenterostomy. However, clinical diagnosis during the neonatal period remains challenging due to the overlap of symptoms with other cholestatic disorders.

Methods

We retrospectively analyzed 268 cholestatic neonates (≤ 30 days old) evaluated at Beijing Children's Hospital between 2017 and 2025, including 86 BA and 182 non-BA infants. Clinical features and routine biochemical indicators were collected. Traditional machine learning (ML) models and a transfer learning (TL) model were developed; model performance was assessed by receiver operating characteristic curves and standard classification metrics. Model interpretability was evaluated using SHAP. A graphical user interface (GUI) was created for clinical deployment.

Results

Multivariate analysis identified acholic stools, gamma-glutamyl transferase (GGT), total bilirubin (TBIL), and total bile acids (TBA) as independent predictors of BA. LASSO selected six key variables (sex, acholic stools, ALB, GGT, TBIL, TBA) for ML model construction. Among traditional algorithms, Random Forest showed the best performance. The TL model demonstrated improved discrimination, achieving an AUC of 0.987 in the validation cohort. SHAP analysis confirmed that acholic stools, GGT, and TBIL contributed most to BA prediction. A GUI prototype was developed to facilitate convenient clinical use.

Conclusion

This study suggests that a TL-based model integrating common clinical features and routine biochemical markers may assist in early identification of BA among neonates. Although the internal performance was promising, external validation is needed to assess its broader applicability and clinical utility.

1. Introduction

Biliary atresia (BA) is a progressive fibro-obliterative cholangiopathy of infancy involving both intrahepatic and extrahepatic bile ducts. It leads to obstructive cholestasis and, if untreated, progresses to irreversible cirrhosis. The incidence of BA ranges from 1 in 5,000 to 20,000 live births [1]. Kasai portoenterostomy (KPE) remains the primary surgical treatment. Previous studies have demonstrated that the timing of surgery is a critical determinant of prognosis, and infants who undergo KPE within 30

days of life achieve significantly better bile drainage and native liver survival [2, 3]. However, delayed diagnosis and referral remain common in clinical practice, resulting in many infants receiving surgery beyond the optimal window, thereby compromising long-term outcomes [4]. Thus, the development of a rapid, accurate, and widely applicable tool for the early identification of BA is essential for improving patient prognosis.

Intraoperative cholangiography is considered the gold standard for diagnosing BA, but it requires anesthesia and surgical conditions, making it unsuitable for early screening [5]. Imaging modalities such as abdominal ultrasound, hepatobiliary scintigraphy, and magnetic resonance cholangiopancreatography have been applied in clinical practice; however, their diagnostic performance largely depends on operator expertise, and their accuracy is often limited in the early stages of the disease[6, 7]. In recent years, several novel biomarkers, such as matrix metalloproteinase-7 (MMP-7), have shown promising potential in the early diagnosis of BA [8, 9]. Nevertheless, these tests remain costly, technically complex, and less accessible, which limits their widespread clinical application. In contrast, routine serum biochemical markers are more advantageous due to their low cost, simplicity, and wide availability. Previous studies have demonstrated that commonly used laboratory parameters, such as γ-glutamyl transferase, bilirubin, and total bile acids, can help distinguish BA from non-BA cholestasis to some extent [10, 11]. Therefore, when these biochemical markers are combined with clinical features, they may provide a reliable foundation for developing a simple, cost-effective, and widely applicable model for early diagnosis and screening of BA.

In recent years, machine learning (ML) has shown significant advantages in medical diagnostics, enabling the integration of multidimensional information and capturing nonlinear patterns that are difficult to identify clinically [12]. Previous studies have attempted to apply ML methods in developing diagnostic models for BA, but most of these cohorts consisted of mixed-age pediatric populations [13–18]. Since age at surgery is a critical determinant of prognosis, delayed diagnosis markedly reduces the effectiveness of the Kasai procedure, thereby limiting the applicability of these models in the early neonatal period. A few studies have explored differentiating BA from other causes of neonatal cholestasis; however, their clinical applicability remains restricted due to small sample sizes, model overfitting, and insufficient external validation, leading to limited generalizability [19]. Transfer learning (TL), which allows knowledge reuse from related tasks, provides a novel strategy to address data scarcity [20]. However, research focusing on TL-based early diagnosis of neonatal BA is still lacking.

Therefore, the present study aimed to develop and validate a TL-based diagnostic model for the early identification of BA in cholestatic neonates by integrating routine clinical and biochemical indicators. This approach seeks to provide an interpretable and clinically applicable tool to facilitate timely surgical intervention and improve patient outcomes.

2. Materials and Methods

2.1. Patient Selection and Data Collection

This retrospective cohort study enrolled neonates evaluated for cholestasis at Beijing Children's Hospital, Capital Medical University, between January 2017 and June 2025. Cholestasis was defined based on established biochemical criteria: (1) direct/conjugated bilirubin > 17.1 μmol/L when total bilirubin (TBIL) ≤ 85.5 μmol/L, or (2) a direct/conjugated bilirubin fraction exceeding 20% of TBIL when TBIL > 85.5 μmol/L. BA was confirmed by intraoperative cholangiography. Non-BA etiologies were determined through a combination of intraoperative cholangiography, liver histopathology, genetic testing, and clinical response to conservative management. Inclusion criteria comprised neonates aged ≤ 30 days with biochemically confirmed cholestasis. Exclusion criteria included incomplete clinical data, inconclusive final diagnosis, or lack of follow-up records at our institution. The study was approved by the Institutional Review Board of Beijing Children's Hospital (2019-k-386). Informed consent was waived due to the retrospective nature of the study and the use of anonymized data. Ultimately, 268 neonates met the criteria, including 86 BA patients and 182 non-BA cholestatic patients.

2.2. Data Collection and Predictor Variables

Demographic characteristics and baseline laboratory parameters obtained at the first visit (prior to any therapeutic intervention) were extracted from the electronic medical record system. The analytical dataset included only the variables listed in Table 1. Demographic and clinical features consisted of age, sex, and the presence of acholic stools (yes/no). Biochemical indicators encompassed 37 parameters across three physiological domains: (1) electrolytes and renal function markers (potassium, sodium, chloride, carbon dioxide, anion gap, osmolality, creatinine, urea, urea/creatinine ratio); (2) hepatic metabolic markers (total protein, albumin, globulin, albumin/globulin ratio, total cholesterol, triglycerides, glucose, calcium, phosphorus, magnesium, uric acid, HDL-C, LDL-C, VLDL-C); and (3) cholestasis-specific biomarkers (total bilirubin, direct bilirubin, indirect bilirubin, direct/indirect bilirubin ratio, total bile acids, alkaline phosphatase, gamma-glutamyl transferase, aspartate aminotransferase, alanine aminotransferase, AST/ALT ratio, prealbumin, creatine kinase, lactate dehydrogenase).

2.3. Machine Learning Algorithm Development and Validation

We employed five supervised machine learning algorithms to develop diagnostic screening models for neonatal BA using the available biochemical and clinical data: Logistic Regression (LR), Decision Tree (DT), Multi-layer Perceptron (MLP), Support Vector Classifier (SVC), and Random Forest (RF). All analyses were performed using Python 3.10 and the scikit-learn library (v1.2.2). The cohort of 268 neonates was stratified by BA/non-BA diagnosis and randomly split into a training set (70%, n = 188) and a validation set (30%, n = 80). Given the relatively high number of biochemical variables, Least Absolute Shrinkage and Selection Operator (LASSO) regression was applied for feature selection to identify the most predictive subset of variables and avoid overfitting. Each model was rigorously tuned via 10-fold cross-validation combined with hyperparameter optimization using the Optuna framework. Model performance was evaluated on the validation cohort using standard metrics: area under the receiver operating characteristic curve (AUC), accuracy, precision, recall, and F1-score. Model interpretability was

achieved using SHapley Additive exPlanations (SHAP) analysis to quantify the contribution of each biomarker.

2.4. Transfer Learning Model Development

To further enhance diagnostic performance, we developed a TL model based on the best-performing traditional model (Random Forest). This model adopted a two-stage strategy: first, a Random Forest model was pre-trained on the entire derivation cohort to obtain feature importance weights; subsequently, an Artificial Neural Network (ANN) with three fully connected layers (node structure: 64-32-1) was constructed. The Rectified Linear Unit (ReLU) activation function was used, and regularization was implemented via Dropout (rate = 0.3). The weights of the input layer were initialized using the feature importance scores derived from the pre-trained RF model to facilitate knowledge transfer. The network was trained using the Adam optimizer (learning rate = 0.001) and binary cross-entropy loss. A sample weighting strategy was applied to mitigate class imbalance. The final model selected for evaluation was the one that demonstrated the best performance on the validation set.

2.5. Clinical Deployment and Graphical User Interface Development

To facilitate the clinical translation and point-of-care application of the optimized TL diagnostic model, we developed a user-friendly graphical user interface (GUI). The GUI was implemented as a lightweight web application using the Flask framework (v2.2.3) in Python 3.10. This interface allows clinicians to input the six key predictor variables (sex, acholic stools, albumin, GGT, total bilirubin, and total bile acids) selected by the LASSO regression. Upon submission, the application returns a real-time prediction of the biliary atresia probability (as a percentage) based on the pre-trained TL model. The source code and a user manual for this GUI are provided in the Supplementary Materials to promote reproducibility and potential external validation.

2.6. Statistical Analysis

Continuous variables conforming to a normal distribution are presented as mean ± standard deviation and were compared using the independent samples t-test. Non-normally distributed variables are presented as median (interquartile range) and were compared using the Mann-Whitney U test. Categorical variables were compared using the Chi-square test or Fisher's exact test, as appropriate. Statistical significance was defined as a two-sided P-value < 0.05. All statistical and machine learning analyses were conducted using Python v3.10.

3. Results

3.1. Baseline Demographic and Biochemical Characteristics

A total of 268 cholestatic neonates (age ≤ 30 days) were ultimately included in this study. The detailed process of patient screening and enrollment is illustrated in **Figure 1**. Briefly, from the initial pool of

infants evaluated for jaundice at Beijing Children's Hospital between 2017 and 2023, 268 met the biochemical criteria for cholestasis and all inclusion/exclusion criteria. This study ultimately included data from 268 cholestatic neonates (age ≤ 30 days), among whom 86 were diagnosed with BA and 182 had non-BA etiologies. The baseline age was comparable between the two groups (BA group median [IQR]: 19.00 [13.00, 26.00] days vs. non-BA group: 19.00 [12.25, 25.00] days; P = 0.653). The proportion of females was significantly higher in the BA group than in the non-BA group (61.63% vs. 34.62%; P < 0.001), and the incidence of acholic stools was also markedly higher (60.49% vs. 3.55%; P < 0.001). Regarding biochemical profiles, BA infants exhibited more severe cholestatic features, with significantly elevated levels of gamma-glutamyl transferase (GGT), total bilirubin (TBIL), direct bilirubin (DBIL), and total bile acids (TBA) (all P < 0.001). Notably, despite cholestasis, BA patients had higher levels of liver synthetic function markers (such as total protein, albumin, and albumin/globulin ratio), along with lower urea and creatinine levels (all P < 0.001). Detailed comparisons for all variables are provided in **Table 1**.

3.2. Univariate and Multivariate Predictors of Neonatal BA

Univariate logistic regression analysis identified multiple variables significantly associated with BA (P < 0.05), as listed in **Table 1**. Among these, the presence of acholic stools (OR = 41.65) and elevated GGT (OR = 1.01) showed particularly strong associations. In the multivariate analysis, four factors were identified as independent predictors of BA: acholic stools (OR = 123.67), elevated GGT (OR = 1.01), elevated TBIL (OR = 1.01), and elevated TBA (OR = 1.01). The results of the univariate and multivariate analyses are summarized in **Table 2**.

3.3. Machine Learning-Based Diagnostic Screening Models for Neonatal BA

As illustrated in **Figure 1**, we developed machine learning-based models for the early detection of BA using data from the 268 neonates. We first assessed the multicollinearity of all candidate variables in **Table 1**, and the results are shown in **Supplementary Figure 1** (no severe multicollinearity, all |r| < 0.9). Feature selection via LASSO regression performed on all variables in **Table 1** identified six optimal predictors for model construction: sex, acholic stools, ALB, GGT, TBIL, and TBA (**Figure 2A**). After 10-fold cross-validation and hyperparameter tuning, all traditional models (LR, DT, MLP, SVC, RF) demonstrated strong performance on the training set (mean AUC range: 0.880 - 0.965; **Table 3** and **Figure 2B**). In the validation set, the RF model achieved the best classification performance (Accuracy = 0.963, AUC = 0.973; **Table 3** and **Figure 2C**), indicating good reproducibility. Calibration curves showed good agreement between predicted and observed probabilities for all traditional models except SVC (**Figure 2D**).

3.4. Diagnostic Performance of the Transfer Learning Model

Building upon the best-performing RF model, we developed a Transfer Learning (TL) model to further improve diagnostic capability. As shown in **Table 3** and **Figure 3A-D**, the TL model demonstrated exceptional diagnostic performance, achieving an AUC of 0.965 in the training set. In the validation set, the TL model attained an AUC of 0.987, with accuracy, precision, recall, and F1-score of 0.938, 0.917,

0.880, and 0.898, respectively. This performance surpassed all traditional models, including its base RF model. Histograms of predicted probabilities showed minimal overlap between BA and non-BA cases, indicating strong discriminative ability (**Figure 3E-F**). The calibration curve indicated excellent agreement between predicted and observed probabilities (**Figure 4A-B**).

3.5. Model Interpretation and Clinical Application via a GUI Platform

We analyzed the feature importance of the five traditional ML models using SHAP, and the results are presented in **Supplementary Figure 2A-E**. SHAP analysis of the TL model consistently identified acholic stools, elevated GGT, and elevated TBIL as the most important predictors in both the training and validation sets (**Figure 4C-D**). Compared to traditional models, the TL model exhibited a more concentrated distribution of feature contributions, suggesting stronger and more reliable dependence on key predictors (**Figure 4E-F and Supplementary Figure 3A-E**). Given its superior and balanced performance, the TL model was deployed as a graphical user interface (GUI) application (**Figure 5A-B**). Clinicians can input the six key variables (sex, acholic stools, ALB, GGT, TBIL, TBA) selected by LASSO to rapidly obtain the predicted probability of BA in a neonate.

4. Discussion

In this retrospective study, we developed and validated a transfer learning TL-based diagnostic model for the early identification of BA among 268 cholestatic neonates. Our results demonstrated that sex, acholic stools, GGT, TBIL, and TBA were the most important independent predictors of BA. Compared with traditional ML algorithms, the TL model exhibited markedly superior performance, achieving an AUC of 0.987 in the validation cohort. This result represents one of the highest diagnostic accuracies reported to date for early BA screening based solely on routine clinical and biochemical parameters.

Acholic stools are a characteristic clinical feature of BA, and previous studies have shown that stool discoloration is closely associated with biliary obstruction, serving as an early warning indicator for BA [21, 22]. In recent years, several countries have implemented stool color card screening programs, which have been demonstrated to substantially improve early identification of BA, facilitate timely KPE within the optimal surgical window, and enhance clinical outcomes [23–25]. In addition, mobile phone–based applications have been developed to assist in identifying acholic stools, offering a practical tool for promoting early detection and timely referral in both community and home settings [26]. However, stool color assessment remains subject to caregiver interpretation and observational variability. Therefore, despite its high specificity, the use of acholic stools in clinical screening should be combined with objective markers to further improve diagnostic accuracy and reproducibility.

For neonates presenting with obstructive jaundice, serological testing is non-invasive and provides valuable information for differential diagnosis, making it a crucial initial step in evaluating the underlying etiology. Among these laboratory indicators, GGT is the most widely used biomarker for the early detection of BA [27, 28]. A meta-analysis including 12 studies reported a pooled sensitivity of 81.5%, specificity of 72.1%, and an AUC of 0.8386 for GGT in diagnosing BA, indicating relatively good

discriminatory performance [29]. Some studies have shown that nomograms incorporating GGT with other serum biomarkers further improve diagnostic accuracy compared with using GGT alone [30]. Similarly, our study confirmed that GGT is an important independent predictor of BA, and its diagnostic performance is enhanced when combined with other clinical or biochemical parameters in the model.

Elevation of TBA is also an important biochemical feature of BA. Increased TBA likely reflects impaired hepatobiliary excretion and indicates the degree of cholestasis. Previous studies have shown that TBA demonstrates moderate sensitivity and specificity in differentiating BA from other causes of neonatal cholestasis, supporting its role as an auxiliary diagnostic marker [31]. Zhao et al. developed a bile acid-based scoring system in which the ratio of glycochenodeoxycholic acid to chenodeoxycholic acid contributed substantially to BA identification [32]. However, bile acid profiling is not routinely implemented worldwide due to its limited accessibility and high cost. In our study, TBA demonstrated notable contribution within the transfer learning model, suggesting its potential utility for early BA risk stratification and its value in informing future diagnostic strategies.

A variety of models and scoring systems have been developed to differentiate BA from other causes of infantile cholestasis [27, 29, 33, 34]. El-Guindi et al. established a 12-point scoring system incorporating clinical, laboratory, ultrasonographic, and histopathological features, achieving an accuracy of 98.83% in the validation cohort. Nevertheless, this approach relies on liver biopsy, an invasive procedure that may limit routine or primary-care application [33]. Kim et al. proposed a Magnetic Resonance Imaging (MRI)-based decision tree model with excellent performance (sensitivity 97.3%, specificity 94.8%, accuracy 96.2%) [35]. However, MRI examinations in infants often require sedation or general anesthesia, which increases procedural risk and limits accessibility. Yan et al. developed a nomogram including gallbladder abnormalities, liver 2D shear-wave elastography, serum GGT, and ALT, which yielded an AUC of 0.99 with a sensitivity of 92% and specificity of 100% [36]. Growing evidence supports the diagnostic value of ultrasonography as a noninvasive adjunctive tool in identifying BA [6, 37]. However, interpretation of key findings such as the triangular cord sign and gallbladder abnormalities is highly operator-dependent, resulting in inter-center variability and limiting broad implementation in early screening.

MMP-7 has emerged as a promising serum biomarker for BA, with many studies reporting AUC values > 0.9, possibly surpassing traditional serum or imaging-based methods [9, 16, 38, 39]. Nonetheless, its diagnostic cutoff values vary substantially across studies, and optimal thresholds differ between neonates and older infants, restricting its immediate clinical utility [9]. In addition, assay standardization remains lacking, and testing costs are high, limiting accessibility in primary hospitals. Future studies are needed to define neonatal-specific cutoff values for MMP-7 and to evaluate its utility in combination with routine biochemical markers to enhance diagnostic performance.

In comparison, the ML-assisted diagnostic model developed in this study is non-invasive, straightforward to apply, and feasible for routine clinical use. All clinical and laboratory variables included in the model are part of standard assessments at hospital admission and can be readily obtained even in primary healthcare settings, supporting its practical applicability and broad dissemination. Although previous

studies have explored ML approaches to distinguish BA from non-BA cholestasis, most included children older than 30 days, limiting their utility for early neonatal diagnosis [13, 34, 40]. In contrast, the present study focused on cholestatic neonates ≤ 30 days of age, aligning with the optimal diagnostic window for BA and enhancing the clinical relevance for early screening and timely referral. Regarding performance, the TL model outperformed traditional ML algorithms, achieving an AUC of 0.987 in the validation cohort. This enhanced performance may be attributed to a two-stage training approach, in which a random forest was first used for feature learning, followed by knowledge transfer to an ANN architecture. This strategy facilitates effective learning under limited sample conditions and helps mitigate the data scarcity inherent to neonatal BA.

Furthermore, we implemented the TL model as a graphical user interface, allowing clinicians to estimate BA risk using only a few input indicators, thereby supporting clinical decision-making and early referral. Nevertheless, as this was a single-center retrospective study, further multi-center prospective validation is necessary to assess the robustness and utility of the model across diverse clinical settings.

Nevertheless, several limitations should be acknowledged. First, this was a single-center retrospective study, which may restrict external generalizability. Second, despite the excellent internal performance, further validation in independent multi-center cohorts is necessary before broader clinical implementation. Third, our model was constructed using only clinical features and standard biochemical parameters. Incorporation of imaging results, genetic information, or more sophisticated biomarkers (MMP-7) may further enhance diagnostic accuracy. Future work should explore prospective validation, integration of multi-omics data, and development of automated pipelines combining imaging and clinical profiles.

5. Conclusion

In summary, our findings suggest that a transfer learning–based model incorporating routine clinical and biochemical markers may help improve early identification of biliary atresia among neonates. Although the model performed well within our cohort, its clinical utility requires further validation. Future multicenter, prospective studies are needed to determine whether this approach can reliably enhance early diagnosis, optimize surgical timing, and potentially improve clinical outcomes.

Abbreviations

Biliary atresia (BA), Kasai portoenterostomy (KPE), Machine learning (ML), Transfer learning (TL),Matrix metalloproteinase-7 (MMP-7), Potassium (K), Sodium (Na), Chloride (Cl), Carbon dioxide (CO_2), Anion gap (AG), Osmolality (OSM), Total protein (TP), Prealbumin (PA), Albumin (ALB), Globulin (GLO), Albumin/globulin ratio (A/G), Urea (Urea), Creatinine (Cr), Urea/creatinine ratio (Urea/Cr), Total cholesterol (Chol), Uric acid (UA), Glucose (GLU), Calcium (Ca), Phosphorus (P), Alkaline phosphatase (ALP), Aspartate aminotransferase (AST), Alanine aminotransferase (ALT), AST/ALT ratio (AST/ALT), Gamma-glutamyl transferase (GGT), Total bilirubin (TBIL), Direct bilirubin (DBIL), Indirect bilirubin (IBIL),

DBIL/IBIL ratio (D/I), Total bile acids (TBA), Magnesium (Mg), Triglycerides (TG), Creatine kinase (CK), Lactate dehydrogenase (LDH), High-density lipoprotein cholesterol (HDL-C), Low-density lipoprotein cholesterol (LDL-C), Very-low-density lipoprotein cholesterol (VLDL-C), Logistic Regression (LR), Decision Tree (DT), Multi-layer Perceptron (MLP), Support Vector Classifier (SVC), Random Forest (RF), Least Absolute Shrinkage and Selection Operator (LASSO), Standard deviation (SD), Area under the receiver operating characteristic curve (AUC), Confidence Interval (CI), Odds ratio (OR), SHapley Additive exPlanations (SHAP), Artificial Neural Network (ANN), Rectified Linear Unit (ReLU), Graphical user interface (GUI), and Magnetic Resonance Imaging (MRI).

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee of Beijing Children's Hospital, Capital Medical University, National Center for Children's Health (Approval No. 2019-k-386). All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. Informed consent to participate was obtained from the parents or legal guardians of all participants under the age of 16.

Consent for publication

Not applicable.

Competing interests

The authors declare no conflicts of interest.

Clinical trial number

Not applicable.

Funding

This work was supported by the Beijing Municipal Natural Science Foundation (Grant No. 7252043 to Jinshi Huang), the Jiangxi (Ganpo) Talent Program-Health Innovation Talent Project (to Jinshi Huang), and the National Natural Science Foundation of China (Grant No. 82400592 to Dingding Wang; Grant No. 82300574 to Dayan Sun).

Author Contribution

Jinshi Huang and Guantong Li designed the study. Dingding Wang, Zonghan Li, and Zhaozhou Liu wrote the main manuscript text. Dayan Sun and Yuyan Jin prepared table 1-3. Zhaozhou Liu, and Yuyan Jin prepared Fig 1-3. All authors reviewed the manuscript.

Acknowledgements

The authors thank all members of the Department of Neonatal Surgery, Beijing Children's Hospital, Capital Medical University, for their contributions to patient care and data collection.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Tam PKH, Wells RG, Tang CSM et al (2024) Biliary atresia. Nat reviews Disease primers 10(1):47. 10.1038/s41572-024-00533-x
2. Fanna M, Masson G, Capito C et al (2019) Management of Biliary Atresia in France 1986 to 2015: Long-term Results. J Pediatr Gastroenterol Nutr 69(4):416–424. 10.1097/mpg.0000000000002446
3. Xie C, Wang P, Wang D et al (2025) Is performing the Kasai portoenterostomy in the neonatal period associated with a better prognosis? A single-center, retrospective cohort study from China. BMC Pediatr 25(1):454. 10.1186/s12887-025-05780-9
4. Davenport M, Makin E, Ong EG et al (2025) The Outcome of a Centralization Program in Biliary Atresia: Twenty Years and Beyond. Ann Surg 281(4):608–614. 10.1097/sla.0000000000006273
5. Davenport M (2025) Updates in Biliary Atresia: Aetiology, Diagnosis and Surgery. Child (Basel Switzerland) 12(1). 10.3390/children12010095
6. Gunda ST, Chambara N, Chen XF et al (2022) *Diagnostic Efficacy of Advanced Ultrasonography Imaging Techniques in Infants with Biliary Atresia (BA): A Systematic Review and Meta-Analysis.* Children (Basel, Switzerland), 9(11). 10.3390/children9111676
7. Hou J, Xiao W, Zhou S et al (2025) Identification of Biliary Atresia in Infantile Cholestasis: Integrating Radiomics With MRCP for Unobservable Extrahepatic Biliary Systems. J Comput Assist Tomogr 49(5):833–840. 10.1097/rct.0000000000001729
8. Wang H, Liu H, Li W et al (2025) Effect of serum MMP-7 on the diagnostic accuracy of biliary atresia: systematic review and meta-analysis. Front Pharmacol 16:1581053. 10.3389/fphar.2025.1581053

9. Li Z, Deng Y, Liao Q (2025) The diagnostic accuracy of MMP-7 for the diagnosis for biliary atresia- a systematic review and meta-analysis. BMC Pediatr 25(1):652. 10.1186/s12887-025-06032-6
10. Hsiao Y-H, Hung W-L, Yang Y-J et al (2025) Non-invasive simple predictors of biliary atresia in cholestatic infants - A preliminary report. Pediatr Neonatol. 10.1016/j.pedneo.2024.11.006
11. Harpavat S, Garcia-Prats JA, Anaya C et al (2020) Diagnostic Yield of Newborn Screening for Biliary Atresia Using Direct or Conjugated Bilirubin Measurements. JAMA 323(12):1141–1150. 10.1001/jama.2020.0837
12. Elsayed H, Singer G, Till T et al *Latest developments of Artificial Intelligence (AI) and Machine learning (ML) models in general pediatric surgery.* European journal of pediatric surgery: official journal of Austrian Association of Pediatric Surgery … = Zeitschrift fur Kinderchirurgie, 2025. 10.1055/a-2689-8280
13. Zhou W, Yang Y, Yu C et al (2021) Ensembled deep learning model outperforms human experts in diagnosing biliary atresia from sonographic gallbladder images. Nat Commun 12(1):1259. 10.1038/s41467-021-21466-z
14. Hsu F-R, Dai S-T, Chou C-M et al (2022) The application of artificial intelligence to support biliary atresia screening by ultrasound images: A study based on deep learning models. PLoS ONE 17(10):e0276278. 10.1371/journal.pone.0276278
15. Sun Y, Dai S, Shen Z et al (2022) Gamma-glutamyl transpeptidase has different efficacy on biliary atresia diagnosis in different hospital patient groups: an application of machine learning approach. Pediatr Surg Int 38(8):1131–1141. 10.1007/s00383-022-05148-5
16. Zhao Y, Wang A, Wang D et al (2024) Development of a diagnostic model for biliary atresia based on MMP7 and serological tests using machine learning. Pediatr Surg Int 40(1):203. 10.1007/s00383-024-05740-x
17. Chen X, Zhao D, Ji H et al (2024) Predictive modeling for early detection of biliary atresia in infants with cholestasis: Insights from a machine learning study. Comput Biol Med 174:108439. 10.1016/j.compbiomed.2024.108439
18. Ma Y, Yang Y, Du Y et al (2025) Development of an artificial intelligence-based multimodal diagnostic system for early detection of biliary atresia. BMC Med 23(1):127. 10.1186/s12916-025-03962-x
19. Wang G, Zhang N, Zhang X et al (2021) Ultrasound characteristics combined with gamma-glutamyl transpeptidase for diagnosis of biliary atresia in infants less than 30 days. Pediatr Surg Int 37(9):1175–1182. 10.1007/s00383-021-04923-0
20. He F, Li G, Zhang Z et al (2025) Transfer learning method for prenatal ultrasound diagnosis of biliary atresia. NPJ Digit Med 8(1):131. 10.1038/s41746-025-01525-1
21. Gopal SH, Zebda R, Mohan A et al (2024) Population-based screening strategies for biliary atresia in the newborn: A systematic review and meta-analysis. PLoS ONE 19(8):e0307837. 10.1371/journal.pone.0307837
22. Gong Z, Lin L, Lu G et al (2023) Development and validation of a model for early diagnosis of biliary atresia. BMC Pediatr 23(1):549. 10.1186/s12887-023-04370-x

23. Zheng J, Ye Y, Wang B et al (2020) Biliary atresia screening in Shenzhen: implementation and achievements. Arch Dis Child 105(8):720–723. 10.1136/archdischild-2019-317787
24. Arshad A, Gardiner J, Ho C et al (2023) Population-based screening methods in biliary atresia: a systematic review and meta-analysis. Arch Dis Child 108(6):468–473. 10.1136/archdischild-2022-324946
25. Hoshino E, Moriwaki K, Morimoto K et al (2023) *Cost-Effectiveness Analysis of Universal Screening for Biliary Atresia in Japan.* The Journal of pediatrics, 253: pp. 101–106.e2. 10.1016/j.jpeds.2022.09.028
26. Angelico R, Liccardo D, Paoletti M et al (2021) A novel mobile phone application for infant stool color recognition: An easy and effective tool to identify acholic stools in newborns. J Med Screen 28(3):230–237. 10.1177/0969141320974413
27. Toh Q, Chen Y, Lee YY et al (2024) Simple biliary atresia score-a validated diagnostic aid for infantile cholestasis. Pediatr Surg Int 40(1):212. 10.1007/s00383-024-05785-y
28. Chen W-C, Lo H-Y, Tsai Y-S et al (2023) The accuracy of magnetic resonance cholangiopancreatography in the diagnosis of biliary atresia in preterm infants with cholestasis. Pediatr Neonatol 64(1):12–18. 10.1016/j.pedneo.2022.07.004
29. Zhang Y, Li T, Wang T et al (2024) Comparison for the diagnostic performance of early diagnostic methods for biliary atresia: a systematic review and network meta-analysis. Pediatr Surg Int 40(1):146. 10.1007/s00383-024-05730-z
30. Dong R, Jiang J, Zhang S et al (2018) Development and Validation of Novel Diagnostic Models for Biliary Atresia in a Large Cohort of Chinese Patients. EBioMedicine 34:223–230. 10.1016/j.ebiom.2018.07.025
31. Han Y-J, Hu S-Q, Zhu J-H et al (2024) Accurate prediction of biliary atresia with an integrated model using MMP-7 levels and bile acids. World J pediatrics: WJP 20(8):822–833. 10.1007/s12519-023-00779-7
32. Zhao D, Zhou K, Chen Y et al (2020) Development and validation of bile acid profile-based scoring system for identification of biliary atresia: a prospective study. BMC Pediatr 20(1):255. 10.1186/s12887-020-02169-8
33. El-Guindi MA-S, Sira MM, Sira AM et al (2014) Design and validation of a diagnostic score for biliary atresia. J Hepatol 61(1):116–123. 10.1016/j.jhep.2014.03.016
34. Zhou W, Lin R, Zheng Y et al (2025) Multimodal model for the diagnosis of biliary atresia based on sonographic images and clinical parameters. NPJ Digit Med 8(1):371. 10.1038/s41746-025-01694-z
35. Kim YH, Kim M-J, Shin HJ et al (2018) MRI-based decision tree model for diagnosis of biliary atresia. Eur Radiol 28(8):3422–3431. 10.1007/s00330-018-5327-0
36. Yan H, Liu J, Jin S et al (2023) A novel prediction tool based on shear wave elastography, gallbladder ultrasound, and serum biomarkers for the early diagnosis of biliary atresia in infants younger than 60 days old. Quant imaging Med Surg 13(1):259–270. 10.21037/qims-22-324

37. Yang Q, Zhang Y, Meng Y et al (2025) The diagnostic performance of ultrasound features for biliary atresia: a systematic review and updated meta-analysis. Pediatr Surg Int 41(1):251. 10.1007/s00383-025-06118-3
38. Pandurangi S, Mourya R, Nalluri S et al (2024) *Diagnostic accuracy of serum matrix metalloproteinase-7 as a biomarker of biliary atresia in a large North American cohort.* Hepatology (Baltimore, Md.), 80(1): pp. 152–162. 10.1097/hep.0000000000000827
39. Jiang J, Dong R, Du M et al (2024) Serum matrix metalloproteinase-7 for discriminating biliary atresia: a diagnostic accuracy and validation study. J translational Med 22(1):636. 10.1186/s12967-024-05442-x
40. Niu Y, Li J, Xu X et al (2024) Deep learning-driven ultrasound-assisted diagnosis: optimizing GallScopeNet for precise identification of biliary atresia. Front Med 11:1445069. 10.3389/fmed.2024.1445069

Tables

Table 1 to 3 are available in the Supplementary Files section.

Figures

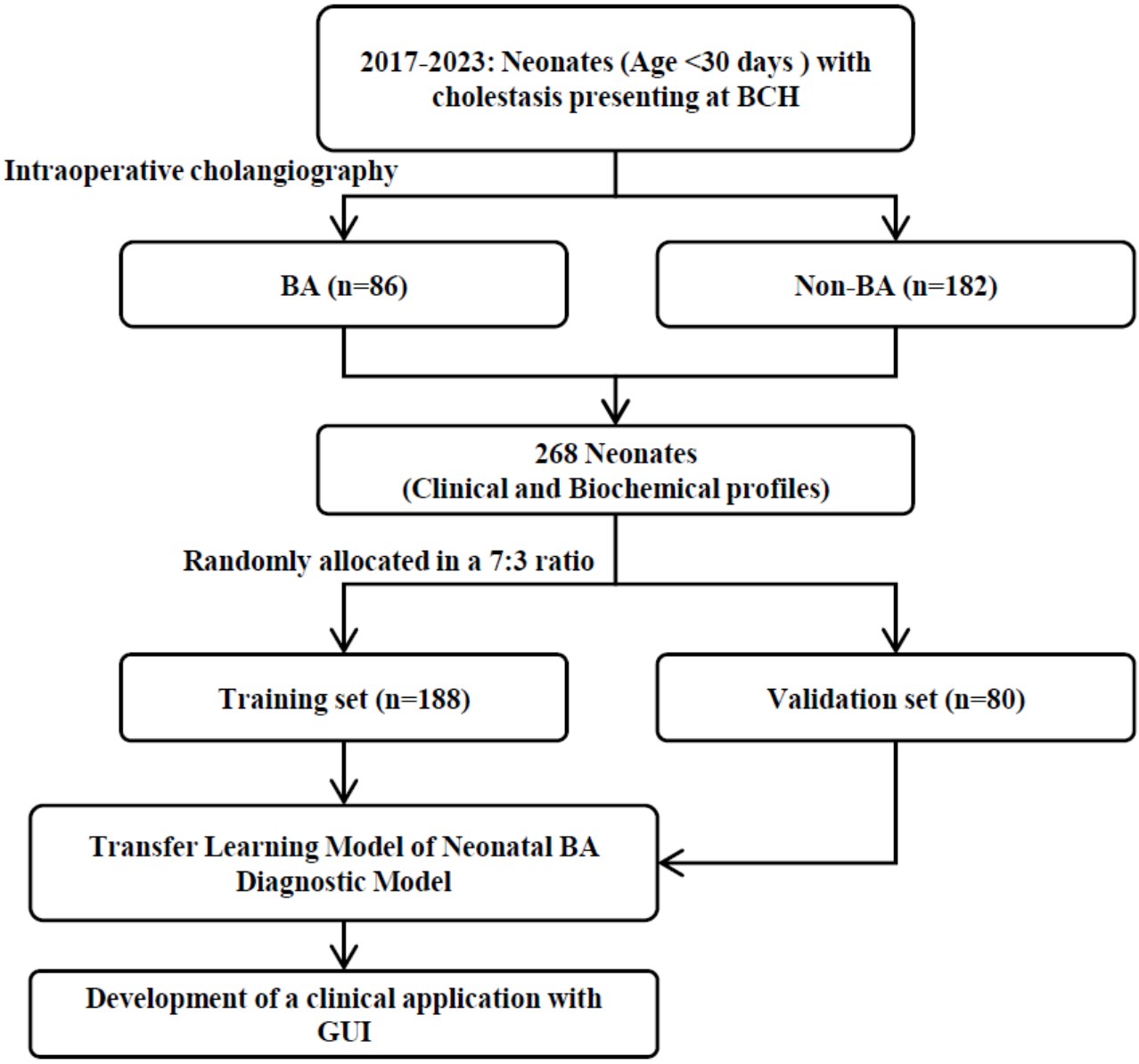


Figure 1

Flowchart of the study cohort.

The enrollment process for 268 jaundiced neonates (age ≤30 days) evaluated at Beijing Children's Hospital from 2017 to 2023. All enrolled infants met biochemical criteria for cholestasis. Biliary atresia (BA) diagnosis (n = 86) was confirmed by intraoperative cholangiography; non-BA cholestasis (n = 182) was determined via multimodal evaluation. The entire cohort was stratified by diagnosis and randomly split into model training and validation sets in a 7:3 ratio. Machine learning and transfer learning techniques were applied to build and validate early diagnostic screening models for neonatal BA. Abbreviation: BA, biliary atresia.

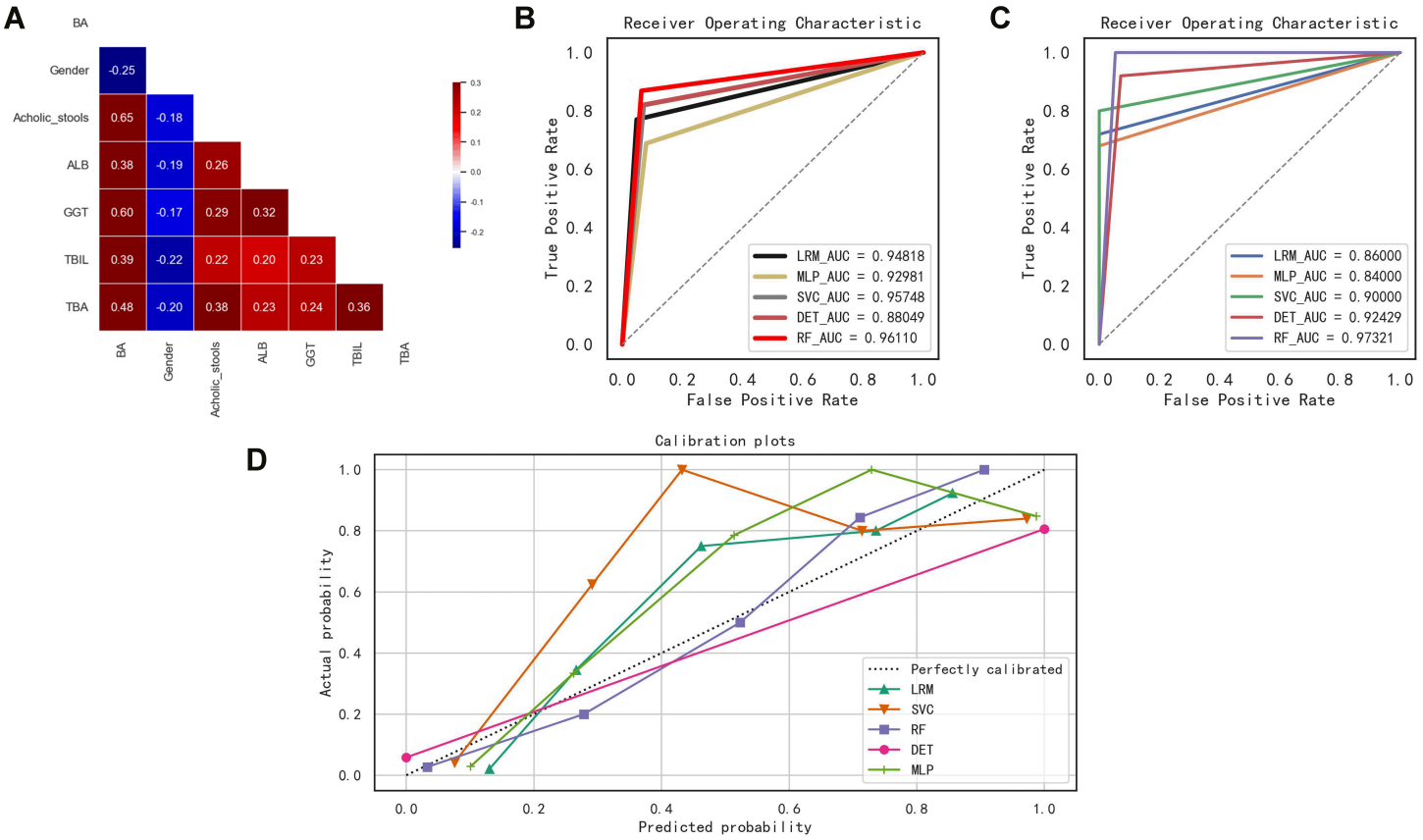


Figure 2

Development and validation of traditional machine learning models.

(A) Feature selection using LASSO regression identified six optimal predictors: Sex, Acholic Stools, Albumin (ALB), Gamma-glutamyl transferase (GGT), Total Bilirubin (TBIL), and Total Bile Acids (TBA). (B) Receiver operating characteristic (ROC) curves of the five traditional machine learning models on the training set (n = 188). (C) ROC curves of these models on the validation set (n = 80), showing the Random Forest (RF) model had the best performance among traditional algorithms. (D) Calibration curves of the five traditional machine learning models, demonstrating the agreement between predicted and observed probabilities. Abbreviations: LASSO, Least Absolute Shrinkage and Selection Operator; AUC, Area Under the Curve; LR, Logistic Regression; DT, Decision Tree; MLP, Multi-layer Perceptron; SVC, Support Vector Classifier.

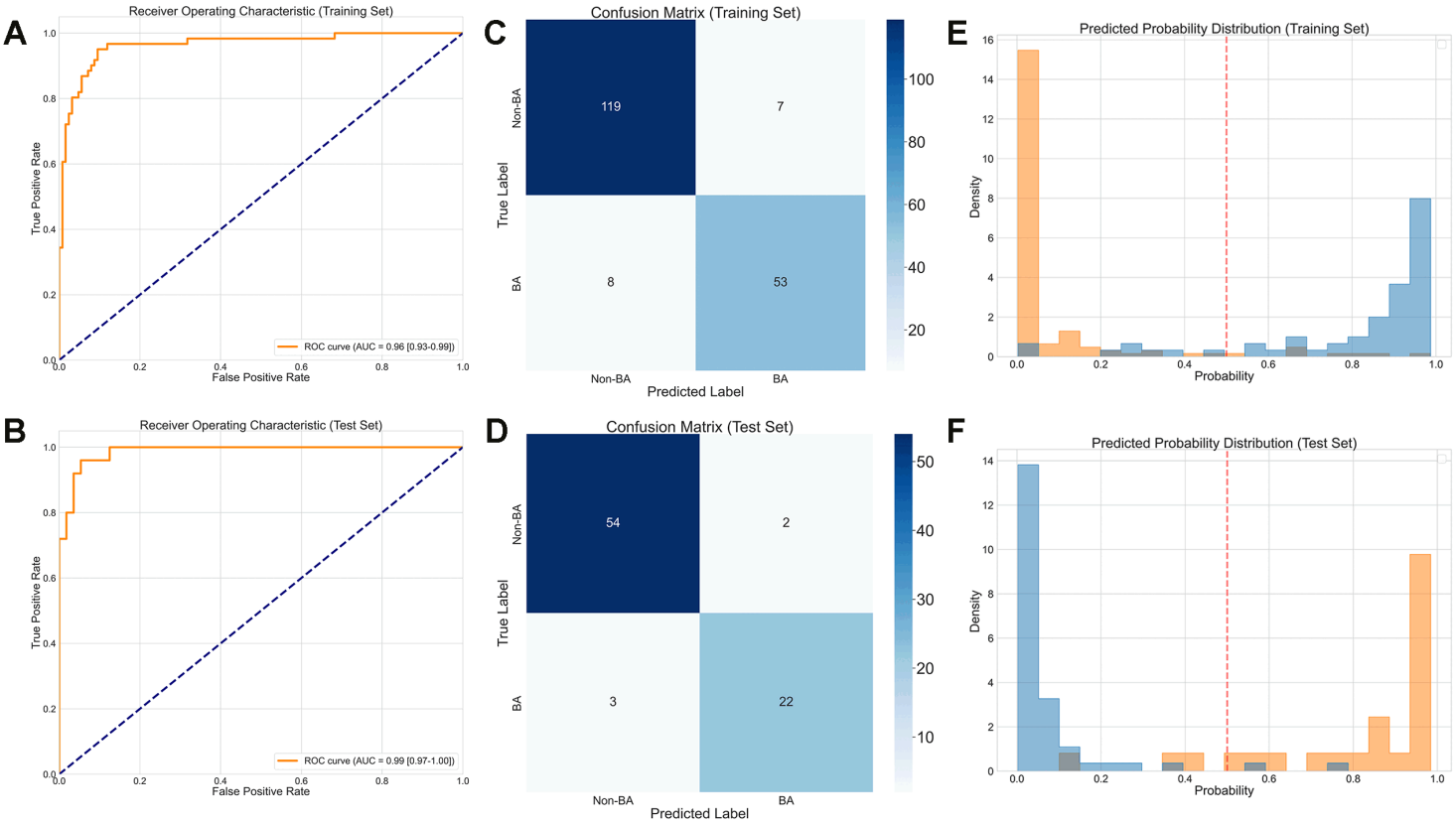


Figure 3

Performance of the transfer learning model built upon Random Forest.

(A) ROC curve of the TL model in the training set, showing performance (AUC=0.965) close to the base RF model (AUC=0.961). (B) ROC curve of the TL model in the validation set, showing superior performance (AUC=0.987) compared to the base RF model (AUC=0.973). (C) Confusion matrix of the TL model in the training set. (D) Confusion matrix of the TL model in the validation set. (E) Histogram of predicted probabilities from the TL model in the training set, showing clear separation between BA and non-BA cases. (F) Histogram of predicted probabilities from the TL model in the validation set, also demonstrating good discrimination.

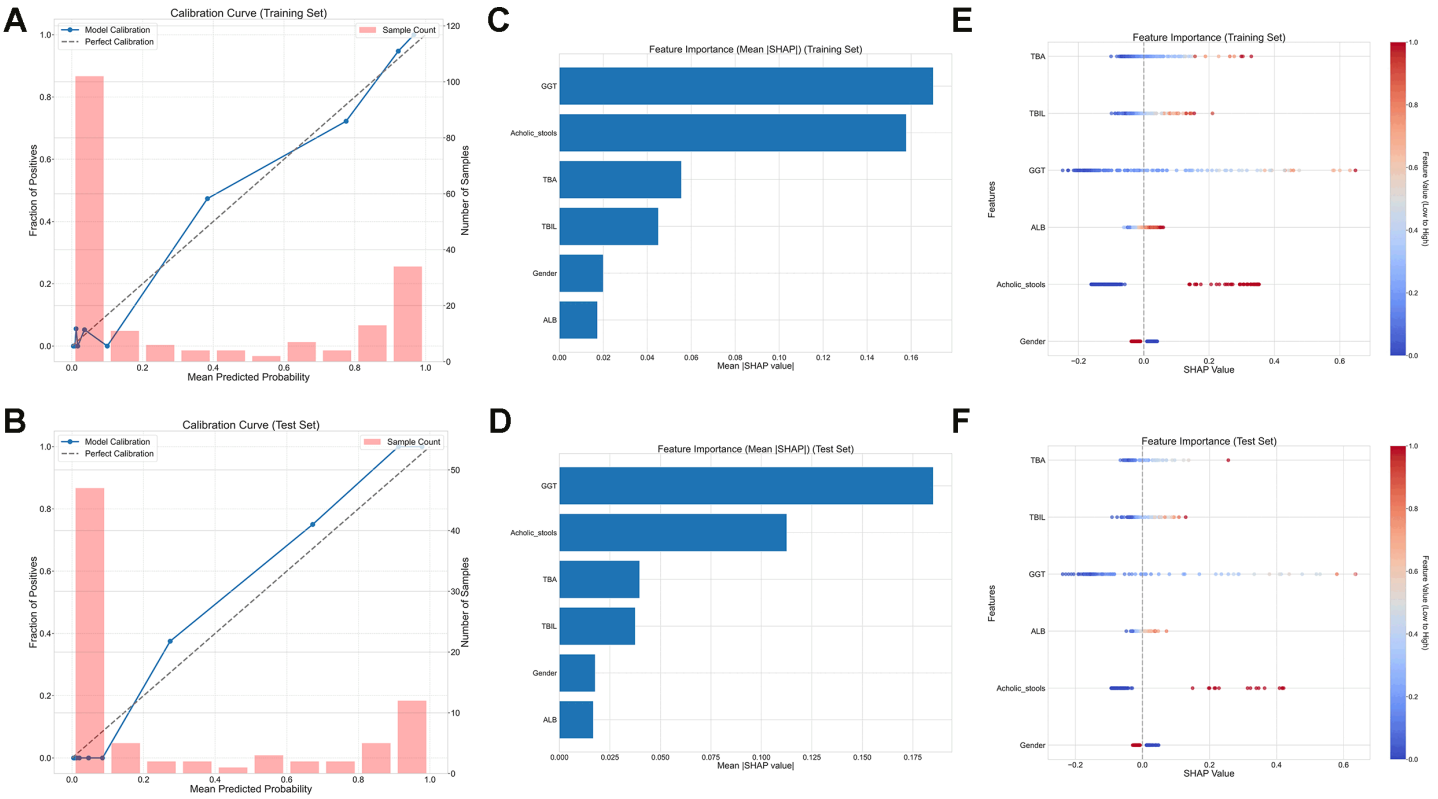


Figure 4

Interpretability analysis of the transfer learning model.

(A-B) Calibration curves of the TL model in the (A) training set and (B) validation set, indicating high agreement between predicted probabilities and observed outcomes. (C-D) Summary bar plots of mean absolute SHAP values for the TL model in the (C) training set and (D) validation set, ranking the relative importance of input features. Elevated GGT, elevated TBA, and acholic stools were consistently identified as the top three most important predictors. (E-F) SHAP dependence plots for the most important feature in the TL model within the (E) training set and (F) validation set, illustrating how the feature value influences the model's output (predicted probability).

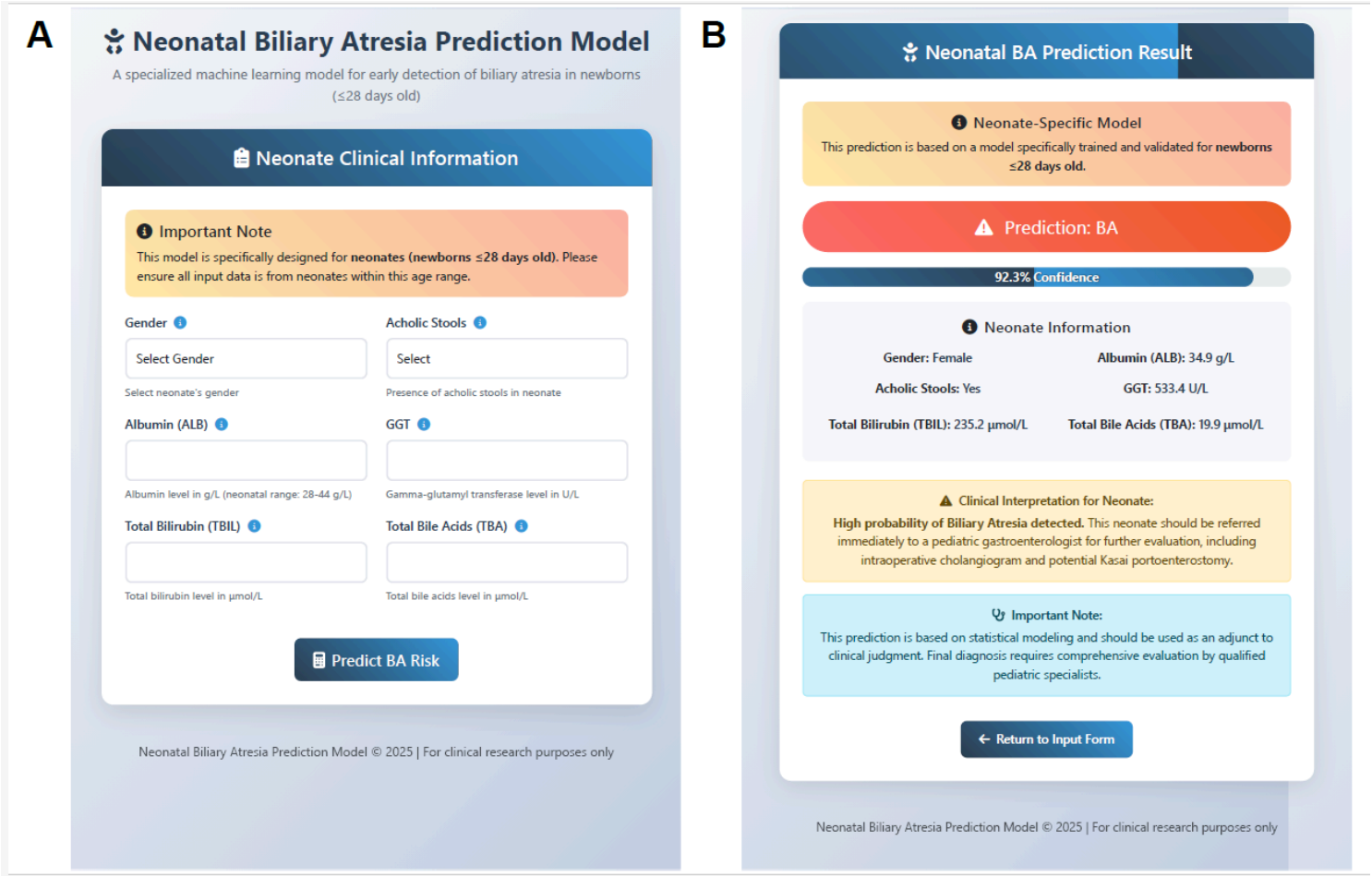


Figure 5

Clinical deployment of the optimized transfer learning model via a graphical user interface.

(A) Input interface for the six key clinical and biochemical parameters used for model prediction. (B) Output interface displaying the diagnostic prediction result (BA probability %). This web-based interface was developed using the Flask framework (Python 3.10). Source code and user manual are available in the Supplementary Materials.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementFigure1.jpg
- SupplementFigure2.tif
- SupplementFigure3.tif
- Table13.docx